Informatica

An International Journal of Computing and Informatics

Special Issue:
Information and Communication Technology at European Universities
Guest Editors:
Viljan Mahnič
Kristel Sarlin

Introduction:

Information and Communication Technology at European Universities

The aim of this issue is twofold: to address different aspects of the use of information and communication technology at European universities, and to promote the activities of EUNIS – European University Information Systems Organisation.

EUNIS was formed to bring together university leaders, higher education information technology leaders, and directors of libraries and media centers from hundreds of European universities and other organizations to explore new directions in the use of information technology in teaching and learning, university administration, information management, and electronic libraries.

One of the major events organized by EUNIS each year is the EUNIS Conference which brings together researchers and practitioners not only from Europe, but also from USA, Canada, and Australia. The 10th jubilee-conference will take place in Slovenia from June, 29th till July 2nd 2004, and this is another reason to devote this special issue to activities of EUNIS.

Professor Matjaž Gams, Executive Associate Editor of Informatica, has kindly given us the opportunity to publish thirteen papers, written by the authors that presented outstanding contributions at the last two EUNIS Conferences. Taking into account the different areas of the use of information and communication technology at universities the papers were divided into five groups:

- E-learning in Higher Education,
- University Information Systems,
- Information Management and Decision Support Systems,
- E-libraries,
- IT Infrastructures.

The first group consists of three papers. The first, *Experiences with Distributed Open Source Courses* by Kirsti Ala-Mutka and Tommi Mikkonen, describes an open source courseware project that started in autumn 2001 by three Finnish universities with the aim of producing open course materials and using these materials for distributing courses to several universities. A distributed course model is

presented that offers universities an easy and inexpensive way to broaden their course selection and distribute knowledge between teaching personnel from different institutions. The second, *Concourse: The design of an online Collaborated Writing Center* by Sjoerd de Vries, describes the development of an online writing community as a study support environment for students in higher education. The proposed center offers services required for professional development of students and for the development and exploitation of knowledge about scientific writing. The last paper in this group, Practice *Related E-learning – the VIP framework* by Mang Li et al., describes a learning framework based on pedagogical principles for constructivist learning. Pedagogical concepts for a representative virtual practical course are discussed, which are supported by three technical building blocks: an e-learning platform, a videoconferencing system, and an interactive simulation environment.

Papers from the second group deal with university information systems. In the first paper, *Information and Communication Technologies and Information Systems Planning in Higher Education*, Jacques Bulchand and Jorge Rodriguez describe a systematic approach to the development of strategic information system and information and communication technology plans in higher education. They propose a methodology that consists of nine steps and involves the whole of the university community. At the end, they show how the proposed methodology has been applied at the University of Las Palmas de Grand Canaria. In the second paper, *The Portal of "GreCO-Universities"*, P.Y. Cunin et al. describe the design and implementation of a common portal of the five universities of Grenoble. Although it is student-centered, the portal aims to provide a wide range of services to anyone interested or involved in higher education. Components included in the first version are described more in detail, viz. the implementation of user directories, the virtual desktop, and information and services related to student life.

The third group consists of three papers. The first, *ICE – a Web-Based Information System to Support Higher Education Policy Decisions* by Peter

Muessig-Trapp, Hans Dicken and Helena Kopp, describes an information system developed to support higher education policy decision-making. The system is in use at the German Ministry of Education and Research, at science and research ministries in ten German states, as well as in some other organizations involved in higher education policy. It has been developed as a Java based, platform and database independent web application using XML technology and the Apache Cocoon based publishing framework, thus providing simple data exchange with third party programs and a wide range of output formats. The second paper, *Analyzing educational process through a chain of data marts* by Viljan Mahnic, describes the development strategy, architecture and logical design of a data warehouse that can be built gradually, exploiting the benefits of the bottom-up, data mart approach. A sequence of data marts connected through common dimension tables is proposed that makes it possible to analyze the educational process as a value chain. The last paper in this group, *A Decision Support System for IST Academic Information* by Elsa Cardoso et al., describes the decision support system for academic information being developed at Instituto Sperior Tecnico, the Engineering School of the Technical University of Lisbon. The paper focuses on the early phases of the system development, viz. the business requirements definition, the dimensional modeling, and the physical design decisions.

The fourth group of papers deals with electronic libraries. The first paper, *Integrating VLE and Library Systems: Opportunities and Challenges* by Clare Uhomoibhi, Alan Masson, and Lyn Norris, describes the potential benefits of virtual learning environments and library systems integration as well as the role of emerging authentication technologies in facilitating the implementation of such integration. It reports on the activities of the 4i Project led by the University of Ulster in collaboration with WebCT, Talis and Athens. In the second paper, *Developing a Quality Culture for Digital Library Programmes* by Brian Kelly, Marieke Guy and Hamish James, approaches for the development of quality assurance procedures for a digital library program are discussed. The authors argue that the adoption of open standards is acknowledged as essential, but in a distributed development environment it can be difficult to ensure that program deliverables actually implement appropriate standards and best practices. The last paper in this group, *Not just a portal: Managing Access in a Complex Information*

Environment by Jean Sykes, John Paschoud, and Christine Cooper, describes work in progress by London School of Economics staff to create managed information and knowledge environment in order to provide access to a wide range of appropriate and permitted content for a broad range of users. The authors argue that portal is only one small part of the project. More difficult tasks are identifying information content for different user types and developing suitable middleware for managing access in an institutional, national, and international context.

The last group consists of two papers dealing with the development of appropriate IT infrastructure as a prerequisite for proper operation of university information systems. The first, *Towards Cross-organisational User Administration* by Mikael Linden, describes the need to find ways to identify network users regardless of which organization they represent. The paper presents related new technologies and activities in the academic world, however, the results can be generalized to cover cross-organizational services in other kinds of institutions as well. The second paper in this group, *Providing Quality of Service in Wide Area Networks* by Ursula Hilgers, Peter Holleczek and Richard Hofmann, deals with the problem of using IP networks for transporting distributed applications. The authors argue that new mechanisms must be provided in network components in order to allow data transport with different transmission characteristics. On the basis of measurements they propose a network architecture that fulfills the quality of service demands of all except the most demanding applications on current IP network technology.

We hope that the selected papers provide a good overview of different aspects of the use of information and communication technology at European universities. We also believe that this special issue will contribute to further promotion of activities of EUNIS with the aim of encouraging the communication and transfer of information between information system developers in higher education institutions in Europe. Once more, we would like to thank Professor Matjaž Gams for his support and the authors of papers for their timely contribution.

Guest Editors:

Viljan Mahnič   **Kristel Sarlin**

Experiences with Distributed Open Source Courses

Kirsti Ala-Mutka and Tommi Mikkonen
Tampere University of Technology, P.O.Box 553, 33101 Tampere, Finland
kirsti.ala-mutka@tut.fi, tommi.mikkonen@tut.fi

The field of information technology extends continuously. In universities, the challenge is to manage increasingly large student groups, while at the same time the teachers need to use a lot of time in developing new IT courses and updating the old courses. Unfortunately, many universities face the lack of teaching resources to deal with the situation. To address these problems, we have started an Open Source Courseware (OSCu) project. The fundamental goal of the project is to increase cooperation between universities in course development. In this project, we produce open course materials, and use these materials for distributing courses to several universities. The distributed course model offers universities an easy and inexpensive way to broaden their course selection and to distribute knowledge between teaching personnel from different universities. This paper introduces the principles of the course material production and the distributed course organization in the OSCu project. We also discuss the issues identified when planning and implementing our first course according these principles during spring 2002.

1 Introduction

The continuous extension of the field of information technology results in two basic problems in university education. First, more and more students need to be educated in IT skills. Therefore, there should be more professors to teach and to mentor the students. Due to the lack of professors, many courses – especially basic ones – are often given by teaching assistants. Second, the scope of the field is extending so rapidly that the teachers need to invest a lot of time for their personal state-of-the-art knowledge. A lot of effort is also needed in order to keep the teaching materials up-to-date with the latest technological advancements.

The two factors above are the basic causes for many practical problems in software engineering education. The courses are difficult to organize due to large class sizes and constantly changing teaching staff. Especially teaching assistants change very often because of changing or finishing their research projects, or leaving the university for a position in the industry. Due to the lack of professors, the leaving assistant may have had a whole course on his/her responsibility. Therefore, the courses must be very independent of the teaching staff. In fact, they must be designed and documented so that the lectures can be given by anyone with enough knowledge and experience of the contents of the course. Another problem is that all universities do not necessarily have any teaching personnel with enough knowledge to build a new course about the latest issues on some emerging expert area.

These problems are difficult to solve with traditional means within any single university. Therefore, we have approached the problem with a scheme that relies on the cooperation between several universities in course development. Open Source Courseware (OSCu) project [21] started autumn 2001 by three Finnish universities; Tampere University of Technology (TUT), Oulu University (OU) and the University of Tampere (UTA). Since then, several new parties have joined the project. In the short term, the idea of the project is to offer new courses and to improve cooperation between universities. In the long run, we aim at an eased exchange of high-quality courses between universities, with an option to train teaching staff via distributed courses. Courses given in the scope of OSCu will be documented thoroughly for easy adoption by the teaching staff in other universities. With the centralized course material organization, we have a possibility to build a bank of always up-to-date course material that collects experiences and latest updates from several course developers. At the same time, we can offer all the universities an easy access to the course materials.

There are also other ongoing projects dealing with similar issues, for instance Candle [4], CUBER [8] and EducaNext [25]. These projects aim at better cooperation and course exchange between universities, too. However, they mainly concentrate on course material whereas we have concentrated on developing a working cooperation scheme for implementing university courses. Therefore, while the course material production and archive are important issues also in the OSCu project, these issues are considered from the viewpoint and practical needs of the distributed course implementation model and face-to-face education. Just after the planning of the OSCu project was launched, also MIT announced its intentions of releasing open courseware of all their courses in MIT OpenCourseWare initiative [17]. Codewitz project [5], on the other hand, develops open interactive teaching materials for programming education.

MIT concentrates on making its own materials available and Codewitz to cooperation in the development work of the separate materials. In the OSCu project, the additional feature is to emphasize the continuous cooperative development of the whole course entities.

This paper, a revised version of [1], is structured as follows. Section 2 describes the idea of considering a course as an individual project entity that should be developed for best possible portability. Section 3 introduces a distributed course model for organizing courses simultaneously in several universities with local support and tutoring for all students. Section 4 presents the guidelines planned for the course material production in order to improve material reusability. Section 5 introduces the chosen technical implementation for delivering course lectures in the distributed model and section 6 provides an insight to a course organized according to the introduced practices during spring 2002. Section 7 concludes the paper with some discussion of the future plans of the project.

2   Course as a project entity

At the TUT Institute of Software Systems, basic software engineering courses have already for years been typically given to several hundreds of students at a time. The constant lack and change of personnel and other resources have forced us to invent efficient ways for implementing and organizing the courses. The strong engineering background of the laboratory personnel has led to a practice where each course can be carried out as a team project, like any conventional engineering project. The analogy is straightforward. A well organized course has exact dates for the beginning and the closing, agreed personnel and teaching material, well-defined tasks for all the personnel, and checkpoints and documentation regarding the progress of the project. This change from the traditional teacher-centered to course-centered thinking gives us the basis for the OSCu project.

2.1   Teacher independent courses

The basic idea is to give each course a status as an entity of its own, not linked to any person. In TUT, this has been supported by assigning each software engineering course a user account on the computer network. Course home directory provides a static archive for all course related materials, structured according to their role in the course instruction. Moreover, courses have their own webpages and E-mail addresses for interacting with students. When teaching personnel changes, the new lecturer (or other course personnel) immediately has an access to all the old course material for him/her to freely use and to develop further.

When a course is given, a team is gathered to implement the course as a project. During the implementation of the course, this team meets weekly to discuss urgent issues and course progress and to ensure that all the crucial issues have

been addressed. These meetings are recorded in the minutes. By the end of the course the collection of these minutes forms a 'course diary' to document the implementation and progress of the whole course project.

2.2   Cooperative course development

Experiences gained at TUT have evidenced that the above scheme enables smooth transition when people responsible for lectures, exercise sessions, and overall bureaucracy change over the years. Therefore, we have used this idea as a guideline when planning course development practices for the OSCu project. By refining the course material development process to produce as university-independent and well-documented material as possible, the course can be adopted as such also to other universities.

The Institute of Software Systems at TUT had some previous experience of using these principles to provide complete courses successfully to other universities. In this project the goal is, however, instead of just delivering a course material packet, to start a continuous cooperative course development tradition between several universities. With active use of a course, its materials get updated continuously and gradually the course combines the knowledge of several experts of the field. In practice, naturally, it is usually only one teacher who originally creates the course, but when he/she releases it freely for others to use and modify, we will get to our goal of cooperatively maintained bank of course materials.

2.3   Dividing courses into subunits

Traditionally, course curricula differ greatly between different universities. Because universities define their study programs independently, they also have different kinds of basic and advanced courses with different interfaces between the courses. Therefore, it may be difficult to include a new course from another university to the study program without overlap with already existing courses. In such a situation, teachers may not be interested in all the issues covered by the course available, and would like to use only part of the course to complement one of their courses.

As a solution, we have defined in each course how they could be divided into smaller units. Each subunit contains lectures, exercises, course projects and other usual course content, but forms a smaller entity than the whole course. With this practice, we hope to increase the ease of taking these course materials into use. Naturally, each subunit contains prerequisites that must be taken into account when taking the study unit into use.

3   Distributed course model

A complete set of course materials is not always enough to help a teacher in preparing to give a new course. The teacher may, for instance, not know enough about the issues covered in the course. Moreover, it is always difficult

to just adopt somebody else's material if one has not seen the original author using it, or otherwise knows how the material was planned to be presented. Documented instructions about the material usage and videos of the lectures do help, but still, the teacher may not feel confident enough for giving the lecture to an audience.

With open course materials, however, it is possible to let universities and the student groups in them to join the course hosted by another university and to see in practice how the course is given. After joining the course a few times as a remote participant, the university has trained competent course staff to teach the course themselves with the provided materials. The university could also decide to stay with the course as a remote partner with the other university leading the course.

3.1 Distributed tutoring work

In many cases, remote courses are organized by simply allowing students from other institutions to take part in the university courses by travelling to the university premises or studying through the possibilities provided on the internet. This increases the workload of the university providing the course, since each of the distant students needs tutoring by the university teachers. Especially when communicating by, e.g., E-mail teacher's workloads are high; answering questions and providing feedback in writing takes inevitably more time than face-to-face communication.

In contrast to these practices, our approach for sharing courses does not rely on tutoring all the participating students only from one point. All the universities have their hands full with the limited resources to deal with their own students. Therefore, it is not possible to take care of the students from other universities as well. Rather, the additional tasks of the course providing university should be minimized, and each university should take part in the course implementation for their own students.

The basis of our distributed course model is presented in the following. Well-documented course materials define the contents of the course. A hosting university takes care of lecturing, planning, and general control of the course. They set the timetable for the course and deliver lectures and materials for all universities. Other participating universities take care of their own practical arrangements, giving exercise sessions, tutoring their students and grading their course assignments and exams locally. From the viewpoint of the hosting university, the course is different from a local one only in that all the student groups and course assistants are not physically located in the same premises.

3.2 Multi-site course management

With the practices presented above, a local course has changed into a multi-site project. This brings us the challenges of synchronizing the course contents in different locations and keeping the whole project from falling apart.

At TUT we had already some experience of organizing courses with this model. We have used similar model for importing a course on software security from another university during years 1998 and 1999. The constant staff meetings useful in the single-university settings are an absolute necessity in a multi-university setting. These meetings are the only way for controlling the whole course project and keeping close contacts between all teaching staff, no matter which university they work at. Also the minutes of each meeting are irreplaceable when there are many issues to be covered and possibly some persons are unable to attend the meeting.

This coordination and synchronization work inevitably increases the workload of the course providing university when the number of course sites and course personnel increases. Although the workload of the providing university does not increase linearly with the number of distant students, every university adds a new need for coordination when joining the course. But since the course would have been given in the hosting university anyway, this is much less extra work that would have been required if they would have organized a separate course or tutoring for a similar number of distant students. The university that provides the course also gains valuable feedback and development suggestions for the course from the students and staff of other universities.

3.3 Student's point of view

Including courses from other universities to a degree at student's home university often demands a lot of bureaucracy between the institutions. Students are required to find suitable courses and negotiate with professors whether they can participate in the course and if it can be accepted to their degree. Many forms, applications etc. need to be filled by students, professors, student office secretaries etc. Still, this is becoming more and more common, because students are better aware of their field and want to study also issues that are not offered at their home university. In addition to the extra bureaucracy, taking courses from other universities brings forward also other issues. Different universities may have different schedules for semesters and exam periods, different means for intracourse communication and practical exercises on the course. A student needs to adapt the whole 'culture' of the university in order to participate in a course at it. Technical support and course tutoring for visiting students, too, may not be organized as well as for the local students, especially when studying over the internet.

The presented OSCu model reduces the problems above, because the students are studying at their home university. When a teacher decides that their university participates in a course as a remote partner, he/she defines and announces the prerequisite courses needed at their university, the degrees it can be included to and which students are allowed to register for the course. For the student the course appears as an additional optional course and is registered as would

any other course given in that university. If, for example, the course exercises require new tools to be used, a local tutor supports the students in these technical issues as well as in any other course-related questions.

As presented, this model is mainly used for distributing a course between student groups at different universities. It is seen as an important part of the lecture that the students get together to a group (with a lecturer or a local tutor). For the students of the remote university, the course appears like a normal university course, except that the lecturer is not physically present at all universities. It is also possible for the students to form study groups of their own and join the lectures in their own group. This is convenient, for example, if there are several university students working in a same company, and the company offers them a chance to study during the workdays. It is common for today's university students to work while studying. By this method they can more easily combine their work and studies and thus reach their degree although working at the same time.

Students can also be offered a possibility to follow the lectures anywhere by themselves, if the lectures are distributed freely over the internet. However, then the idea of a lecture situation is lost. When gathering a group of students together to follow the lecture, they are more concentrated to the course than when sitting alone by a computer. The OSCu project aims mainly at easing the development of traditionally given courses, consisting mostly of contact education and having a fixed timetable. However, there are no reasons why these basic ideas could not be transferred also to courses that are partly or completely given and tutored over the internet.

3.4 Vision of the scheme

With the presented distributed course model we could have, for example, the following development. Professor A in the university X decides to produce a new course on the latest developments in, say, ubiquitous computing. Universities Y and Z both have a considerable number of students who would like to study these issues, but their university does not offer any courses on them. University Z does not even have any staff to develop such a course from a scratch. Thus Y and Z ask to join the course and hire teaching assistants C and D, respectively, to take care of the local organizing and tutoring tasks. A big group of students from university Y works in company N and the company provides them a possibility to follow the course lectures on the company premises. Figure 1 presents the situation in a graphical format.

Next year university Y decides to change the course language from Finnish to English. They produce a new version of the course and begin to offer it to their students. University Z has both English speaking and Finnish speaking students, so they join both the Finnish version of the course from the university X and the English version from the university Y. They can now offer two optional courses for their students with the work resources of two teaching
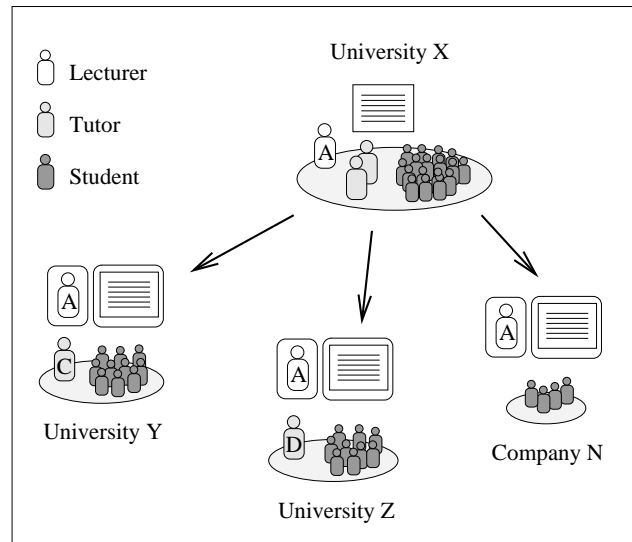
Figure 1: Distributed course organisation.

assistants and without a need for a lecturing professor.

When a couple of years later professor X retires, the teaching assistant D from the university Z has been assisting the course so many years he/she is able to take the responsibility of the whole course. Thus, the university Z begins to host the course and offers it to other universities. The new situation is illustrated in Figure 2.
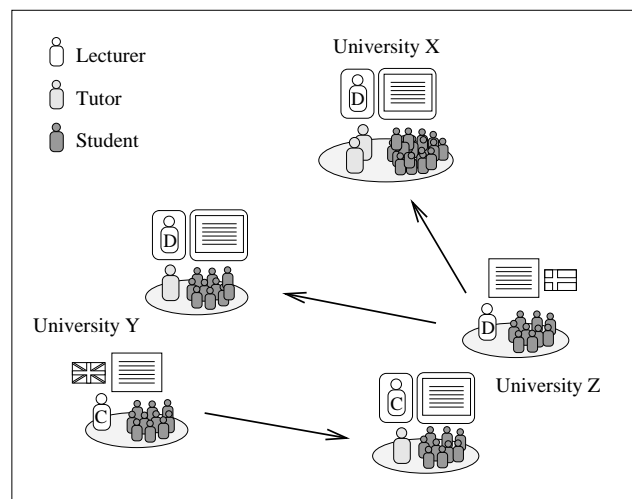
Figure 2: Revisited OSCu course organisation.

4 Guidelines for course material production

OSCu courses are designed to be offered also to other universities and by other teachers than the original providers. Therefore, all the material needs to be structured and pro-

vided with information of the content and teaching methods planned for the course or the content part in question. Since courses will be created by several teachers, we need a set of specific guidelines in order to build a bank of common course materials, where the courses can be easily accessed, used and modified by anyone. The main issues in archiving the materials are introduced in the following.

4.1　Copyrights and material delivery

First, in order to achieve free usability for all teaching material developed in this project, we have made an agreement about the principles of material copyrights. All the course material will be provided under GNU Free Documentation Licence [10]. This assures that the materials will be given freely for anybody to use and modify as long as he/she agrees to the licence, i.e., gives the materials forward after his/her modifications under the same licence. OSCu coordinator maintains the central material archive and the consistence of the course material delivery packages.

Although the materials are offered freely for use, we do want to keep track about the places where these materials are used, and to gather feedback and the latest updates to the materials as well. We also need to control that students cannot automatically load themselves e.g. exam questions and answers from the course they are participating in. This requires authentication and restriction methods to be implemented to the material management and delivery system. At the moment, when the number of courses and participants is small, we manage this with E-mail-based interaction and UNIX file hierarchies. Later, however, when dealing with increasing number of material requests, we plan to implement a WWW-based inquiry system for searching and delivering materials. Until then the materials can be requested by E-mail from the project coordinator.

4.2　Contents and structure of the materials

For understanding how the course materials were planned to be used, each teaching entity should be provided with instructions for the teaching staff. For instance, we have stated for all exercises that the material for an exercise session must consist of at least following items: exercise questions, model solutions and an instruction file containing the goals, main points of the exercise session, and a description of preferred working and tutoring methods. With distributed teaching model these instructions support several teaching assistants and assure students the same level and contents of exercise sessions in all universities. For assuring enough information to the instruction files, we have stated the fields and minimum requirements for their contents.

In addition to stating the required files for each item we have also given instructions about the directory structures for courses. The goal is to have similar file and directory structures in each course so that the materials are easy to maintain and access. Figure 3 gives an example of

the planned file structure for course and also more detailed structure for the course exercises.
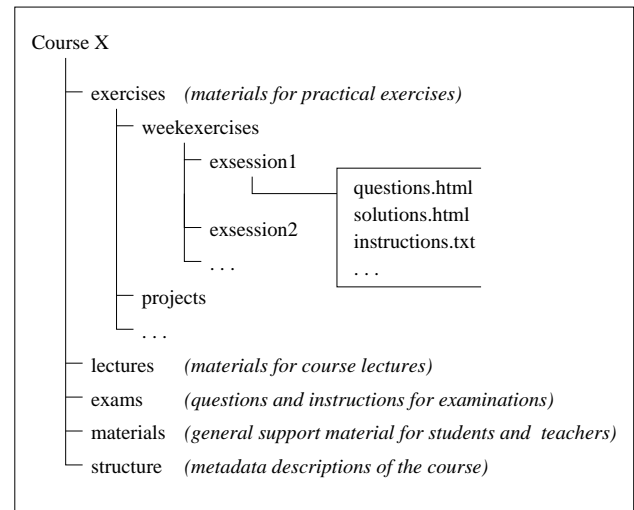
Figure 3: File structure for course materials.

For best possible portability, the materials are instructed to be saved in standard formats, for example as ASCII text or Hypertext Markup Language (HTML) files. No proprietary word processor formats, for example Microsoft Word documents, are allowed, since these formats are not stable and restrict the environment the teachers should work in. We have not, however, found a good solution for a lecture presentation format. The transparencies could be saved and presented as a Portable Document Format (PDF) file, but then they could not be modified by the next user. LaTeX [16] system is a good option for teachers that work in UNIX environment. However, it does restrict the user environment, since it is not commonly used in other platforms. Finally we came to the conclusion, that the lecture transparencies can be saved also as PowerPoint type presentations [18]. In addition to the MS PowerPoint, these presentations can be edited in many environments with OpenOffice tools [20].

4.3　Metadata descriptions

There are many communities, e.g. Ariadne [2], IMS [12], and Dublin Core [9] that develop standards and practices for reusable teaching materials. Their results are combined in the work of the IEEE Learning Technology Standards Committee that has produced a Learning Object Metadata (LOM) Standard draft [11]. There are also both national and European initiatives to support the localization and usage of the standards and recommendations, for instance SCHEMAS project [22] and the Finnish TIEKE organization [24]. Since we aim at a large bank of course materials that are easily available for everyone, we have actively followed the progress of these projects. We see it as essential to be able to attach information about e.g. the contents,

version and keywords to these courses and use LOM as a basis for our metadata descriptions. LOM is meant for describing all teaching material units, starting from very small entities. This, however, sets a heavy workload for the teachers, who at the same time should write and develop the material contents.

At this point, we have come to a consensus of adopting a practice where only major units, such as courses and their major subunits, are provided with formal metadata. Smaller units, e.g. individual exercises, have only textual instruction files for describing their purpose and usage as previously introduced. This provides us sufficient support for reuse at the moment, but we will follow the situation and develop the practices if necessary. With the present implementation, it is possible to later add metadata descriptions also for smaller entities, relying on the present instruction files. However, adding more metadata descriptions definitely requires a good editor for attaching the descriptions to the material files.

An important issue in constructing a data archive is to design the material descriptions for best possible searchability. The LOM specification gives some description fields and value options for the material, for example, user role, resource type, difficulty, and context. It does not, however, define any vocabulary for keywords. This is understandable, since there are no international glossaries that would be actively updated and contained at least reasonably good variety of terms for all possible disciplines. For computer science there are a few glossaries, of which the ACM Computing Classification System [3] is probably the most widely known and used. For example, the CUBER project [8] uses the terms from ACM in their metadata descriptions. With a general vocabulary, however, there is always the problem that it does not contain the latest terms nor the most commonly used. Therefore, we use both terms from the ACM term list and free terms of author's own choice in the material description. This increases the probability of search matches considerably when there are also search terms from everyday life of the field in question. Naturally, the search engine should also have synonym mechanisms.

4.4 Documenting experiences and feedback

Traditionally, when a teacher develops a course and teaches it, he/she gains experience by which he/she can develop the course further. As mentioned before, at TUT this is supported by collecting documentation of the weekly meetings of the course personnel. This documentation provides the teacher next year a possibility to read this course diary and see how the course was implemented and whether any problems were faced. With this kind of collected data, it is much easier to develop the course implementation further. Since in our project the teacher of a certain course is assumed to vary often, it is all the more important to provide practical information about the previous course arrangements and about the possible problems and considerations teachers should take into account.

With distributed courses, the hosting university is required to collect a general course diary and all participating universities are expected to write a report at the end of the course. In this report, all universities are required to document their local arrangements together with conclusions about the course implementation and student feedback from their side. These reports and the general course diary are then saved as a part of the course material for future course teachers.

5 Lecture implementation in the distributed course model

For the students, lectures are an important way of getting introduced to the new ideas and issues of the course. Due to the lack of literature on latest advancements on the field, the lectures often provide information that cannot be obtained anywhere else. Studying lecture transparencies or other documents rarely gives all the same information as when presented by a person with experience on the issue. Learning objectives are usually also supported by practical exercises and course assignments. However, students still need to first get the knowledge and the new ideas in order to be able to apply them in the exercises.

Also with distributed courses, lectures are an important part of the course. In all universities, the students gather together for lecture sessions, and have a possibility to present questions and to discuss course related issues with the lecturer, the tutor or with each other. Our task was to design a technical implementation that would support a near-to normal lecture situation in all universities, no matter where the lecturer was physically located. One of our main requirements was, however, that none of the participating universities would need to make big investments for setting up the course, since this would most likely discourage universities from joining the course.

5.1 Selecting technologies for the lecture delivery

There were some basic requirements for the technical implementation of the lectures. First, we needed to implement two-directional interaction between all the universities. Thus, there should be almost real-time video and audio delivery between all universities. For example, sending a video stream from the hosting university would not have fulfilled the requirement, because of the delay caused by packing and unpacking the stream for presentation. Videoconferencing devices and software make it possible to deliver audio and video of all participants to each other, so we decided to use them.

The next step was to decide the details for the videoconferencing connection. One of the ideas of this project was to develop a model where new participants do not need big financial investments on the technical arrangements. Therefore, we needed a solution for best possible

multipoint videoconferencing connections with moderate costs. An ISDN connection (H.320 standard [13]) would provide a reliable connection with a constant bandwidth, but requires a room with ISDN connection. However, software engineering courses often have several hundreds of students, which would not fit into any videoconferencing room. In fact, we have to be able to organize the lecture in normal class rooms and easily change the place. Thus, we discarded ISDN connections both because of the high costs of such systems and the requirement for a special room.

The biggest Finnish university cities are connected to each other with 1.5 Gbit FUNET connection [7]. The connections to other universities are also good and all universities have a local IP network available at least in some of their class rooms. Thus, the present project participants and most probably also future members have the possibility for an IP-based connection that would not cost any extra. IP-based video conference equipment can be set up with low costs to any room with a network connection. There are several IP-based videoconferencing software products available very economically or even completely free of charge. The drawback of the IP-based equipment is that they are sensitive to problems in network connections and to the load on the network.

There are two IP-based videoconferencing possibilities; Mbone tools [23] and H.323 standard [14]. With Mbone tools, it is possible to form multipoint connections easily. Unfortunately, the tools themselves are difficult to use and the multipoint sessions complicated to set up. We decided to use H.323-based tools for the video conference and use a multipoint control unit (MCU) for forming multipoint connections. This would not add the costs, since some universities already have a MCU and CSC, the Finnish IT center for science [6], offers MCU services for Finnish universities free of charge.

5.2 Videoconferencing arrangements

In a normal lecture situation, the lecturer is situated in front of the audience and students can see the lecturer and the other students in the same room. In a distributed lecture situation all students naturally need to see the lecturer, but it is rather irrelevant for them to see the students from other universities. The situation is meant mainly for lecture presentations, not for discussion between student groups. Student collaboration can be organized during exercises or via internet discussion groups, if considered a good form of education for the course. With large student groups a course wide discussion would be impossible to manage anyway. Thus the questions during lectures are to be presented either to the lecturer or to a local tutor or to the local students. A local tutor or professor is always available in the lecture situation to discuss course-related issues with the students during lecture breaks or exercise sessions. Hence, during lectures we send video only one way, from the hosting university to the other student groups and use the two-way connections only for delivering audio between universities.

The practical lecture arrangement is simple, and can be implemented in any room with internet connections. The lectures are given in the hosting university in an auditorium, equipped with a video camera and two computers. Other universities need two computers and data projectors to participate to the lectures. A video conference connection is established to deliver the picture of the lecturer to all universities and to provide a real-time audio connection between all sites. The video picture contains only the lecturer, lecture transparencies are delivered in a separate connection with the second computer. By this arrangement we do not need a camera expert to mix the picture of the lecture and the transparencies or to plan how to alternate the sending of them in a real-time situation.

This two-connection arrangement adds some technical reliability to the lecture situation. If one of the connections fails, the other can still work and the students are not totally lost during the time it takes to reform the connection. Unfortunately, this is not a rare situation in videoconferencing, especially when implemented with small-scale conference equipment that are not as reliable as heavy videoconferencing systems. The presentation delivery is organized by application sharing over internet via the T.120 standard [15]. This form of connection usually works more reliably than audio/video connections. With application sharing, the remote universities can have the same view of transparencies and other material that is projected from the lecturer's laptop to the audience at the hosting university.

Figure 4 presents the communication connections necessary for the lecture situation. University X delivers the picture and the voice of the lecturer to the universities Y and Z. All universities have a two-way audio connection to each other. MCU takes care of the sharing of the audio and video transmissions to all sites. Application sharing can be carried out without a multipoint connection unit.
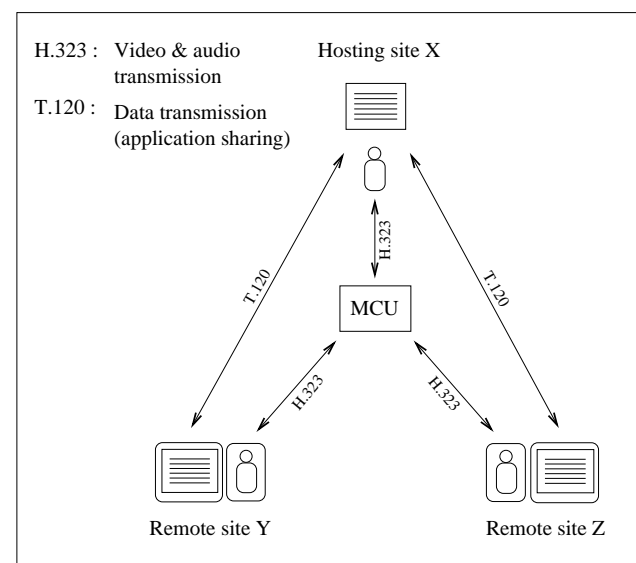

Figure 4: Communication connections between different sites in a lecture situation.

5.3 Following lectures without attending them

IP-based videoconferencing equipment in lower price classes does not provide other features than basic videoconferencing. More advanced systems also offer connections for a video recorder or a possibility to provide video stream of the lecture. Also some MCU's can provide these features. The audio and video broadcast can be transformed to video stream and distributed to internet in real-time. When combined to a WWW-based application sharing, this would provide a mechanism for the students that are unable to be physically present in the lecture session to follow the lecture using their own computer. We do not, however, encourage students to use this as a primary form of following the course lectures. Real-time video stream delivery will be combined to some of the courses later to gather experiences and feedback from students and teachers.

Streaming or taping the lectures to a traditional video provides also a possibility to later edit these lectures into a self-study form with synchronized transparencies. These kinds of lectures can be offered to students that were unable to follow the lecture real-time or otherwise want to review the contents of the lectures. The best way to provide these lectures for students would be a media server, where the students could select easily the parts they are interested in by winding the lecture forwards and backwards. During spring 2002 we stored some of the course lectures in digital form in order to test the equipment and the editing process. Later we plan to save all the streamed lectures and synchronize them with the transparencies for both the students and the teachers to study later.

6 Implementation of a trial course

During spring 2002, we organized the first prototype course according to the presented guidelines. In this instance, the subject was *Programming of mobile devices*, with 4.5 credit units as its size. The course was hosted by Tampere University of Technology, and participated by the universities of Oulu and Tampere. In addition, a group of TUT students working at a local telecommunications company participated in the course by following the lectures on the company premises. This corresponds to the general situation presented in Figure 1, with TUT as the hosting university X and OU, UTA and the company as remote course sites.

The number of participating students ranged between 20 and 80 at the different sites, with the total number of students being around 150. The total number of teachers and assistants involved in organizing the course was 10, with the majority of personnel working for TUT.

6.1 Course material

This was the first time the course was organized. Therefore, the material needed for the course was created and collected more or less single-handedly by the lecturing professor during the autumn 2001. The lecture material was handed out to the students at the beginning of the course, which was considered as a necessity for the course to succeed due to the distribution scheme. After that, all material has been collected, completed with exercise and exam materials, and formatted as described in section 4.

The lecturer developed the course further during the spring. In practice, simultaneous planning of the next and implementation of the current version of the course required two course folders, similar to the basic course folder model presented in Figure 3. One of the folders was used for recording what was been done year 2002, and another included the improvements for next year. The recording folder includes diaries updated each week concerning the current situation at the course, and the feedback coming from the participants. In contrast, new material includes improved lecture handouts and some new ideas on how to organize exercises next year.

6.2 Course personnel organisation

The course lecturer took care of the course material production, lectures and controlling the course progress. He also took care of the coordination of the course and led the weekly meetings of the course personnel. The main task of the local teaching forces was to ensure that practical matters would be taken care of and that the students in all universities would have mentoring with their programming exercises. This worked well and no special problems were encountered.

Local arrangements were also needed for the examination arrangements, because all the participating universities were independent and a professor in one university could not accept credit units for another. Therefore, the course needed professors that took the responsibilities of accepting the course examinations on behalf of his/her university. As a side effect, some personnel involved in the scheme at the different universities were of a higher status than ordinary teaching assistants would have been. In practice, the professors took care of the official tasks and nominated assistants to handle exercise sessions and other practical arrangements. This works well as long as it is clear who has the responsibility and authority over practical course issues. If the tasks are distributed to too many persons, it complicates communication between the universities and the integrity of the course in the remote university may suffer.

As anticipated, the communication between the different organizers was found to be very important. Due to geographical distances, weekly meetings between all the teaching staff would have been be impossible to organize in one physical location. Therefore, these meetings were aided by communications media. Several media were tried out during the course, including E-mail, application sharing and video conferences over the internet, normal telephone, and conference phone calls. These meetings were

mostly focused on the exercises, because they constituted the biggest technology transfer from one university to others. In addition, three face-to-face meetings were agreed, scheduled before, during, and after the course. Moreover, an option was made for an emergency face-to-face meeting if absolutely needed. Fortunately, no emergency meetings were required.

6.3   Technical implementation

The distribution scheme was implemented according to the principles introduced in section 5.2. In our case, the lectures were given in an auditorium which was already equipped with a video camera, a data projector, and an audio system. For the lecture arrangement, we just brought with us to the room two laptop computers and a ViGO video conference device [26]. Oulu University had class rooms with fixed installation of videoconferencing equipment, thus they could use them and did not need to obtain any new equipment for the scheme. University of Tampere joined the video conference first with Netmeeting [19], but later acquired also a ViGO for faster connection and better reliability in the video and audio connections.

One of the computers and connections at each site was used for the video conference. The multipoint connection was formed by using the MCU by CSC. The other computer was used for sharing the presentation application from the lecturer's desktop. This application sharing was implemented with Netmeeting software. At the hosting university, a separate person operated the sending of the lecture and controlled the video camera. This person also monitored the whole conference and all connections with the monitoring tool offered by the CSC. With many conference sites, it would have been impossible for the lecturer himself to take care of all the connections and monitoring while giving a lecture, thus a separate operator was a necessity. The lecturer, however, took care of the application sharing, since those were the same applications that he showed to the students present in the hosting university. Local tutoring assistants at the remote universities took care of the technical arrangements of the lectures. Usually these were the same people who were also tutoring course exercises, thus they got a clear picture of the course entity when they also followed the same lectures as the students.

There were plenty of technical difficulties with the lecture arrangements, especially with the video conference connections. They were only two times in the whole spring semester when the conference connections were set up without problems for any of the participants. On almost half of the lectures, at least one of the participants lost either audio or video connection once or more often. This could usually be quickly solved by disconnecting them from the conference and then calling in again. Unfortunately, this always interrupted the lecture for a few minutes for the university in question. Each university had to restart their devices a couple of times during the spring. Also mi-crophones and their settings gave us some trouble by producing noise that we had difficulties to find the source of.

Some of the problems were caused by the videoconferencing equipment, some by the microphones and loudspeakers, and some were simply due to the load of the network. By selecting ViGO equipment we obtained portable devices to provide video conferences at a moderate price. The total cost of running the course equipment-wise was approximately 5000 EUR, consisting of two ViGO devices and some speaker phones needed for course staff meetings. The laptops and data projectors were already available, so we did not need to invest on them. This solution, however, has proved somewhat unreliable, and we should potentially consider more reliable videoconferencing equipment – even if it is more expensive.

6.4   Special requirements for the lecturer

The distributed lecture delivery was a challenge also for the lecturer. Keeping the audience interested when displayed on a screen in front of a class can be much harder than when physically present. The students from the remote universities hardly ever presented any questions even if they were asked to do so. A few times the participants were activated by presenting them a voting task about issues just covered on the lecture. The local tutors collected the results and delivered them to the hosting university. The lecturer also often used small example applications that brought some variation to the normal presentation. These applications could be shared from his computer desktop with the same application sharing software as the PowerPoint application and thus did not need any special arrangements.

The video conference arrangements restrict the physical appearance and presentation possibilities of the lecturer. Due to the camera, any rapid movements are unacceptable, because they would make the picture worse in the receiving end. Wide movements of the lecturer would also make it difficult for the camera operator to keep the lecturer in the picture. Moving in general was also more difficult because of the fact that the camera was pointed at the lecturer, not to the projected transparencies. Thus, all pointing to the presentation material needed to be done on a computer screen. This can be helped with a computer that has a touchscreen, so that the lecturer can use a pen pointer instead of a mouse. Additional challenges were to keep things focused and organized around precomposed slides. The use of a blackboard was not possible due to the restrictions of image quality over the video connection.

For the lecturer, this means that presentations must be planned in detail, as ad hoc discussions are hard to arrange over the video conference connection. The lecturer should always keep in mind that there is more audience than present in the same physical room. Moreover, even simple things like questions coming from the audience require discipline, as the question must always be repeated in order to ensure that all the participants have correctly received it. The lecturer responsible for this course managed

to pay attention to these issues better and better during the course. It was very difficult for a firsttimer, though, as evidenced by the visiting lecturers. Although they were instructed about the special considerations and requirements, e.g. repeating all questions of the audience, they often forgot these issues and had to be reminded even during the lectures.

Some practical issues were also noticed, that become important when following the lecture by video. First, the camera should be placed in the center behind or among the audience, so that the lecturer looks at the direction of it. This way the remote students watching the video do not feel as if the lecturer was talking to someone else all the time. Our lectures were given in an auditorium with fixed camera installation at the back wall and this forced the students to watch the lecturer from a bit too high an angle. Second, most of the teachers have a tendency to turn to the projected presentation on the wall when explaining issues on the transparency. Although a teacher is instructed not to do so but to give the lecture straight at the audience, he/she still does it every now and then without noticing. Therefore, it is recommendable that in the remote universities the presentation is projected on the same side of the lecturer projection as it is situated in the original class room. Also a wireless microphone, if used, should be kept on the side of the presentation projection if it is not possible to attach it in the middle of the body. These issues can be helped with a touchscreen computer that gives the lecturer an easy way to point the slides using the computer in front of him, and to face the audience all the time.

6.5 Differences between universities

As expected, there were some special considerations that had to be taken into account when arranging a course in three independent universities at the same time. Although in Finland the general teaching semesters are approximately the same, the lecture and examination periods somewhat varied between these participating three universities. The solution was to try to organize lectures only at weeks that belonged to the lecture period in all participating universities. For this reason there were less lecture hours for use than in a normal 4.5 cu course given at the participating universities. However, since this was known in advance, the issue could be taken into account when designing the course and gave us no special problems.

The final exam was organized in the last lecture session. Thus, while organized locally, it was synchronized so that we could use the same exam questions in all the universities. Overall, this was a major deviation from the practices of all participant universities, as exams are usually organized after the lecture period is over. Moreover, we had problems in the sense that the normal time reserved for exams was different at different universities. In the end, we had an exam that lasted 3 hours, which is the time that the lecture session normally took. In addition, an agreement was made that the universities are allowed to have exams

locally whenever found convenient after the first exam. For this purpose, a question pool was created, based on which additional local exams can be easily arranged.

Another minor problem that emerged in an early phase of the course was that the different universities normally used different information systems for group communication between students and personnel: OU used a proprietary information system, UTA had a practice of using a self-developed WWW-based system, and TUT relied on NNTP-based UNIX News system. We solved the problem by creating a newsgroup that had a bridge to World Wide Web via public domain gateway software. This enabled an approach where both News and WWW could be used for accessing the information. Unfortunately, it was found out during the course that the selected software was not too reliable, and some special practices had to be adopted in using the system.

We also had some discussions related to the acceptance of exercises, and mentoring given to students at different universities. In the end, this was left to individual universities to decide. However, the overall guidelines defined by the hosting university were to be followed. We later noticed some differences in the level of accepted course programming projects between the universities, but with better guidelines and more communication between course staff we believe that also these differences can be removed.

On this first course, we wanted to compare student performances and thus course lecturer graded the examinations for all students. The scale of the grades distributed very evenly to all universities. Thus, the difference between remote and local lecturing had not affected to the students' performance in the exam. Actually, the best examination paper was completed by a student from a remote university. Figure 5 shows the distribution of the scores.
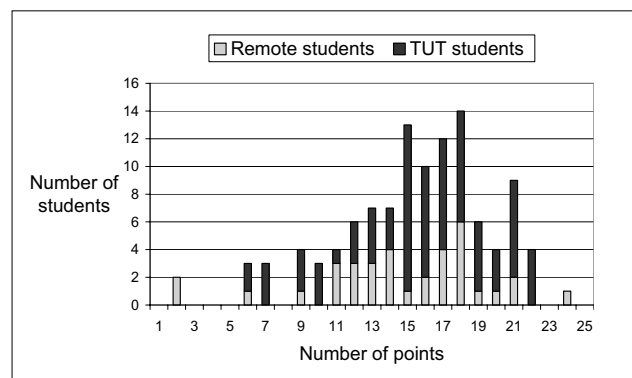

Figure 5: Distribution of examination scores.

6.6 Course feedback

We collected written feedback from course students during the last lecture, and naturally also received oral feedback during the whole spring in exercise sessions and lectures. The main points of the received feedback concentrated on

the contents of the course, not on the technical implementation or organization. This was considered positive. Although we did have several technical difficulties during the lectures, they had not discouraged the students. The TUT students had practically no comments about the technical equipment, delays and strange noises they had been listening to in their lecture sessions. The students from the remote universities had some more comments, but relatively few persons had major complaints regarding technical problems.

Admittingly, the feedback was gathered only from those students who stayed with the course until the end of it. There were also students who dropped out during the course, but the general impression was that this was mostly due to the very laborious first programming project required from the students. There were approximately 170 students in the beginning of the course, 130 students submitted the first course programming project and 106 students attended the final examination. This is a normal dropout percent at TUT on the more advanced non-compulsory courses. On the other hand, there were now participating students that usually would not have a possibility to take the course because of their busy work schedules. For example, it was now possible for a TUT student presently working at a company in Oulu (500 kilometers away from Tampere), to take a TUT course at Oulu University. He attended the course with the students of OU, but the course examination and credits were directly accepted at TUT as a part of his degree.

The feedback from the course personnel was gathered in writing from all the universities in the form of the course reports that were saved with the course material. All course personnel gathered together to a feedback discussion after the course to analyze the course and the implementation practices more closely. The results of this discussion were documented and taken into account when refining the process and methods for the next courses in the project.

In conclusion, the overall feeling about the course was positive. In fact, due to the requests coming from the participating universities, the same course was implemented spring 2003 with a similar organization and process. There are also six universities that have expressed their interests in participating the course during spring 2004. We take this as a credit to the scheme that is used in the project as well as a tribute to all the participating course personnel.

7   Conclusion

In this paper, we have described the basic principles of the Open Source Courseware project. The main goal of the project is to improve cooperation between universities in course material production and course implementation. In this project, we gather a group of interested course developers to use and develop the course materials for each course. In good cooperation the benefits are obvious: when each participating university originally produces a course,

they receive several other courses in return. Moreover, also other experts evaluate, update and improve the course contents.

The presented scheme allows easy transfer of knowledge and well developed courses between universities. This reduces the need to develop new course materials when a university wants to broaden the course selection offered to its students. We also presented a model for multi-site courses. In this model, a course can be distributed to several universities without too much extra burden for the hosting university. This model provides the students a possibility to study courses that are organized and given by another university, but without extra bureaucracy and with good local support and tutoring. In this project, we do not aim at providing time and space independent education, rather at supporting the basic education activities with course development and implementation in universities. However, the principles of developing well-planned and documented course materials provide also a possibility and good basis to create course packages to be studied as distance education over the internet.

During 2002–2003, altogether nine courses have been implemented and six universities have been participating in the project. By the experiences this far, we conclude that the basic ideas seem to work well. Both the students and course personnel have given encouraging feedback. Material production process and course personnel organization and tasks have been supported by developing thorough instructions and also training for participants. Although basic working models have been established, we still need active coordination and organization work for developing the working practices and supporting tools for this project.

OSCu project continues with the support of Finnish Ministry of Education until the end of year 2006. By the year 2006, we wish to have proven the benefits of both the open course material production and distributed course model as a form of university cooperation. When having acknowledged benefits for all participants, this project will continue as a part of normal university course development and implementation practices.

Acknowledgement

References

[1] Ala-Mutka, K. and Mikkonen, T. (2002) Experiences with Distributed Open Source Courses. *Proc. of the 8th International Conference of European University Information Systems "The Changing Universities: The Challenge of New Technologies"*, 19–22 June, University of Porto, Portugal. pp. 26–37.

[2] ARIADNE Foundation. http://www.ariadne-eu.org/ (1.10.2003)

[3] Association for Computing Machinery (ACM). Computing Classification System, http://www.acm.org/class/ (1.10.2003)

[4] Candle project. http://www.candle.eu.org/ (1.10.2003)

[5] Codewitz project. http://www.codewitz.net/ (1.10.2003)

[6] CSC, the Finnish IT center for science, http://tv.funet.fi/videoneuvottelu/index.jsp.en (1.10.2003)

[7] CSC, the Finnish IT center for science, Funet network, http://www.csc.fi/suomi/funet/index.html.en (1.10.2003)

[8] CUBER project. http://www.cuber.net/ (1.10.2003)

[9] Dublin Core Metadata Initiative. http://dublincore.org/ (1.10.2003)

[10] Gnu project. Free Documentation Licence, http://www.gnu.org/licenses/fdl.html (1.10.2003)

[11] IEEE Learning Technology Standards Committee. Final LOM Draft Standard, http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html (1.10.2003)

[12] IMS Global Learning Consortium Inc. http://www.imsglobal.org/ (1.10.2003)

[13] International Telecommunication Union Telecommunication Standardization Sector (1999) H.320, Narrow-band visual telephone systems and terminal equipment.

[14] International Telecommunication Union Telecommunication Standardization Sector (2000) H.323, Packet-based multimedia communications systems.

[15] International Telecommunication Union Telecommunication Standardization Sector (1996) T.120, Data protocols for multimedia conferencing.

[16] LaTeX system. http://www.latex-project.org/ (1.10.2003)

[17] Massachusetts Institute of Technology. MIT OpenCourseWare, http://ocw.mit.edu/ (1.10.2003)

[18] Microsoft PowerPoint. http://www.microsoft.com/office/powerpoint/ (1.10.2003)

[19] Microsoft Windows NetMeeting. http://www.microsoft.com/windows/netmeeting/ (1.10.2003)

[20] OpenOffice.org source project. http://www.openoffice.org/ (1.10.2003)

[21] OSCu project. http://www.cs.tut.fi/˜oscu/english.html (1.10.2003)

[22] SCHEMAS project. http://www.schemas-forum.org/ (1.10.2003)

[23] Svetz, K., Randall, N. and Lepage, Y. (1996) *MBone: Multicasting Tomorrow's Internet*, IDG Books Worldwide, Inc. Available also at http://www.savetz.com/mbone/.

[24] TIEKE, Finnish Information Society Development Centre. http://www.tieke.fi/english.nsf (1.10.2003)

[25] EducaNext community. http://www.educanext.org/ (1.10.2003)

[26] Vcon ViGO. http://www.vcon.com/solutions/videoconferencing/desktop/ViGO/ (1.10.2003)

Concourse: The Design of an Online Collaborated Writing Center

Sjoerd de Vries
Communication Studies
Faculty Behavioral Science
Institutenweg
University of Twente
PO Box 217, NL-7500 AE Enschede
The Netherlands.
E-Mail: sjoerd.devries@utwente.nl
Web Address: www.rinc.nl

The project presented here is Concourse. The project is aiming at the development of an online writing community as a study support environment for students in Higher Education. Concourse is the name of an online collaborated writing center and is intended as a virtual space for online interaction, a place for collaborated writing in a knowledge rich environment. It offers services required for professional development of students and for the development and exploitation of knowledge about scientific writing. Examples of services provided by the system are services for information access, information dissemination, communication, study rooms, research, support, coaching, training and study, and knowledge management. In this paper, we present our design research approach that we applied in the Concourse project. After that, we elaborate on the design guidelines and the online collaborated writing center Concourse based on these guidelines, as a part of this approach.

1 Introduction

Attractive online places for continuing professional development of (junior) professionals, which are legitimate, fully personalized, offering a wide range of knowledge services, flexible, highly interactive, and reliable - that is the focus of our research into the development of successful online knowledge communities. We are investigating the relationships between essential design guidelines for and the development of online knowledge communities.

The project presented here is Concourse (www.concourse.nl). The Concourse project is aiming at the development of an online writing community as a writing support environment for students in Higher Education. The Concourse project is a cooperation between the University of Utrecht (www.uu.nl), namely the Faculty of Arts, in particular German and Spanish Language and Culture Studies, Information Studies, and IVLOS, and Konict BV (www.konict.nl), the project is co-funded by the SURF foundation (www.surf.nl). This community is an example of an online knowledge community, a group of people working online and sharing an interest in the development and exploitation of a particular knowledge domain, in this case academic and business writing [8, 9]. The purpose of this community is to obtain, create, manage and use writing support relevant to the community and for its members to share. Therefore there is a need for an online center as a

meeting place and as a place for collaborated writing in a knowledge rich environment.

The project Concourse started in 1999 and is expected to be finished in February 2003. We are aiming to continue the community as a part of the curricula of the involved studies German and Spanish languages. The writing community is located a didactical university setting, called the 'Workshop'. The Workshop is a one-year course (160 hours of study) for developing students' skills in business and scientific writing in various languages. The Workshop provides support for students' writing in different courses. The support can be by means of online information, drill & practice, or tutorials, but also by peer, or senior students, teachers or experts from outside the University. In the study, participants are mainly university students and teachers. Students can use the Workshop when they are taking courses that require particular writing skills. The Workshop started in the academic Year 2000-2001 in the Faculty of Arts, in the German and Spanish Language and Culture Studies. In 2002-2003 we are introducing and online meeting place for establishing an online community.

Participants are mainly first and second-year students and teachers. The participants are not expected to be specifically interested in the use of ICTs and they do not receive a specific introduction in the use of digital media. It is expected that they possess limited skills in using these media. However, a majority of the students is expected to have access to a computer at home, and a minority also to the Internet. We consider ths students and teachers as (junior) professionals.

As basic communication media e-mail and a WWW-based system Concourse (www.concourse.nl) are used. Concourse is the so-called online collaborated writing center and is intended as a virtual space for online interaction, a place for collaborated writing in a knowledge rich environment. It offers services required for professional development of students and for the development and exploitation of knowledge about scientific writing. Examples of services provided by the system are services for information access, information dissemination, communication, study rooms, research, support, coaching, training and study, and knowledge management.

Not all Workshop activities, such as communication and training, have to take place online. In many situations, if possible, face-to-face communication is preferred over online communication because of its limitations, for example loss of communication cues and loss of perceived social presence. Nonetheless, with further technological developments, such as the growth in bandwidth of data, the quality of online communication is expected to improve further. In addition, it is expected that participants will become more familiar with and experienced in the use of online communication, which might have a positive effect on the quality of communication. Bearing in mind what has been said above, a mixture of face-to-face and online communication is expected to be ideal.

In this paper, we present our design research approach that we apply in the Concourse project. After that, we elaborate on the design guidelines and the online collaborated writing center Concourse based on these guidelines, as a part of this approach.

2   The design research approach

The design research approach is comparable to that used in other disciplines such as Human Factors [2, 5]. We outline the phases and describe outcomes, differences and relationships between them. Our approach includes four research phases (see Figure 1), represented as rectangles. Inputs and outputs are represented by parallelograms and input-output relationships by arrows.
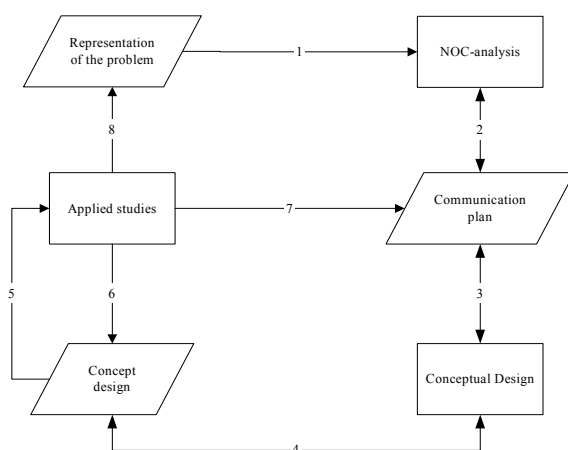
Figure 1: Approach for design research

Analysis. The first phase, Analysis, is the analysis of a 'networked organisational communication' situation. This situation can be considered for example as a communication problem, communication opportunity, or communication hypothesis, for instance as a result an existing communication design process. In the case, this phase addressed questions such as 'how can a Workshop contribute to the quality of student's academic writing'. Expectations included 'active writing, writing in online intensive co-operative groups within an information-rich environment and purposeful writing will have a positive effect on writing quality'. This type of expectation makes a claim about use of the Workshop, but does not specify the design of the Workshop in any way. However, expectations do impose constraints on system design, in this case that active writing, intensive communication and purposeful writing should at least be supported and stimulated if possible. Based on the analysis, decisions can be made regarding the situation that is aimed for. These decisions can be described in a plan, for instance a communication plan or, as we did in our case, 'The global functional structure of Concourse' [10]. In the case, the plan was the input to the design of the Workshop including an online collaborated writing center and constraints that the design must satisfy in order to be successful.

Conceptual Design. The second phase is Conceptual Design, where guidelines are drafted for the communication technologies to be used and a system design based on these guidelines. The 'concept design' is a combination of guidelines and a system design and will be discussed in detail later. In the case, the question addressed in the Conceptual Design phase was 'which guidelines should the design of Concourse adhere to in order allow active writing, intensive co-operation and purposeful writing and stimulate if possible?'. A design guideline is seen as a high level and broadly applicable directing principle. More concrete than guidelines, we also use patterns. A pattern is a named description of a problem and solution for it that can be applied to a variety of contexts [4]. One or more patterns may be derived from a design guideline. In our case, we also distinguished between proof of concept and proven concept. In general, we consider a prototype as a proof of concept if it meets the stated guidelines and if it is acceptable to users. If the community uses the system and if it has positive effects on the quality of scientific writing, then we consider the concept design as a proven concept. A proven concept meets the expectations of users. In the Concourse project, we described one leading guideline and four patterns [10]. The leading design guideline was: 'the centre has to enable a social network of members to develop and efficiently and effectively exploit all relevant information as knowledge assets at the time'. The patterns were 'Knowledge Object model', 'information data profiles', 'knowledge service model', and 'knowledge activating study'. These will be discussed later.

Applied studies. The third research phase is called applied studies. We consider applied studies as studies focussing on a specific concept design in a specific

context of use. These studies investigate a concept design as it is used. The purpose of this type of studies is to collect data to test, documented expectations (from the Analysis phase); second, guidelines from the Conceptual Design phase); or third, effects of interventions that are intended to understand or influence the NOC. Questions that are addressed may include 'how do students study using the Workshop?', 'how do patterns of relations develop using the Workshop?' and 'how does the Workshop develop?'. Such questions are more descriptive of nature. Questions addressed may also include 'what is the effect of collaboration opportunities within the Workshop on study behaviour?', 'what is the effect of an online information-rich environment on the quality of 'writing products'?', 'what is the effect of online coaching on instructors' behaviour?'. These questions focus on for instance efficiency or effectiveness of the use of a concept design and use empirical research, employing quasi-experimental or non-experimental research methods. The aim of these studies is to investigate relations between characteristics of participants, context and media.

The breakdown of communication research into phases that has been presented above is a variant of a standard empirical research process, with specific inputs and outputs of each phase indicated. The role of Conceptual Design was clarified and some indication has been given of how a concept design can be evaluated. Standard social science research methods and techniques [3] can be used in empirical implementation and effect studies of NOC systems.

In this paper, we want to present the Conceptual Design of Concourse in more detail. First, we describe two interfaces of Concourse as the system design based on guidelines -- the concept design -- that are presented later.

3 Concourse: An online collaborated writing center

Concourse has a number of interfaces, depending on the number of user groups and tasks. The basic interface is called the back office. It is a menu driven interface, on the left hand the tree which shows objects hierarchically, and the right screen shows the input and output views of the objects. Below, a number of typical views can be found, for instance, 'who is online', 'what's new', and 'messages'. We have chosen a 'traditional' tree-based interface to lower the learning curve for users and to make them feel comfortable with the program.

The collapsible object tree on the left side shows the wide variety of available objects. These objects can be messages, folders, persons, groups, chats, discussions, tasks, and etcetera. The top of the hierarchy consists of members or groups. We have chosen for the members or groups as the top, because it offers the user a consequent view on the objects. Moreover, members and groups form the professional network, the hearth of an online knowledge community. The tree does not allow for drag-and-drop operations, but is managed by the standard, cut, copy, paste, delete and link to operations. If members

want to share specific information or interact with other members, they can form a new group and do so. Also, they can share for instance a discussion board with members, in that case the board is available in a shared items folder. A third way of making information available for for instance non-members is the opportunity to publish objects to others by means of a front office, of which an example is shown in figure 3. Examples of front offices are a showcase, portfolio, and an online magazine. Another example is a guest book in the public area, or a discussion thread. In this example, the folder 'my groups' contains links to the most important groups for this user.

Figure 2: An impression of the back office interface of Concourse

The menus refer to the basic services. The menu is role driven. Here we give an overview of the main items. The Object menu enables the user to enter new objects and new object types. Entering objects implies filling in the object attributes. The possible choice of entries and values depends on the object types and the context of the center. The entries and values can be modified online and changed. For instance, typical values of the 'folder genre' are: archive, portfolio, and showcase. Values of a discussion board are: brainstorm discussions, announcements, and project proceedings. Each context has a typical set of genre concepts and these can be added by certified users in the role of, for instance, an editorial board. New object types offer the opportunity to add object types without the need of programming. An example is the object 'project'. A project is a group with a specific task given limited resources in a specific period. Members can use the object type group for these purposes and, for instance, add the value project to the entry group genre. However, they can also decide to add the object type project based on the group object type. In that case, a project type has the same functionality as a group object type. This offers opportunities to also add the entry project genres, so that a specific set op projects can be classified. It also contains functions like: cut, copy, paste, modify, copy as link, etcetera. These operations concern the objects in the object tree. The admin menu offers a number of administration options. Often-used items are presented here. 'Manage profile' allows users to choose between different interface

profiles, for instance, simple, basic, and advanced. Personal preferences allow altering the personal member information. 'Manage users/groups' allows users to add users and if needed to create their personal area, to add groups, to modify users and groups, to close and delete. 'Manage keywords' allows users to maintain the fixed and free keywords. 'Associate keywords/object types' permits users to associate members and groups to keywords and object types. This offers the opportunity to distinct between different knowledge areas in one center. If members become a group member, then they are automatically associated to the group keywords and object types. 'Manage roles' allows users to maintain and add user roles, by filling in the rights to menu and object functions. The user guide menu offers the standard support options.

Members and member groups can design tailored interfaces for their 'rooms' in Concourse. For instance, a member is able to make use of a portfolio and can develop a portfolio interface for those who have access to the portfolio information. A member group may like to have a typical classromm interface, or wants to to make use of a 'electronic magazine' interface in order to present course results to the 'outside world'. An example of a classroom interface is shown in Figure 3.
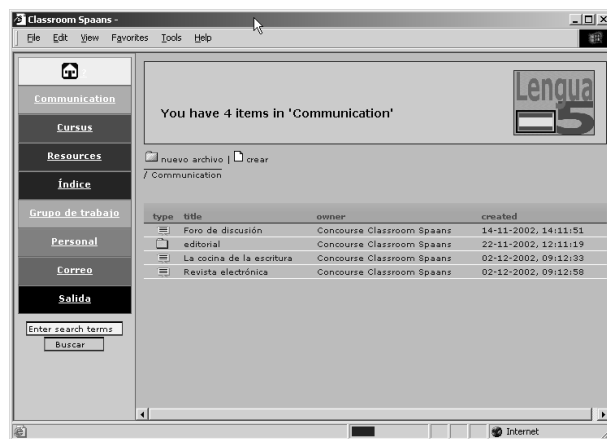

Figure 3: An impression of a member group interface in Concourse

The interface shown here, concerns the same room in Concourse as shown in Figure 1. However, the interface is designed by the user group, a class.

4   Concept design of Concourse

As described earlier, the function of Concourse is a meeting place that enables members to take part in the professional network, to develop them and to develop and exploit knowledge. In this part, we describe one leading design guideline and four related patterns, based on the function of Concourse. The leading design guideline is: the center has to enable a social network of members to develop and efficiently and effectively exploit all relevant information as knowledge assets at the right moment. To meet this leading guideline we

deduced the following four patterns, which are described in Table 1.

Pattern name	Knowledge objects
Solution	Model all relevant information units in digital knowledge assets, where an asset is seen as an object containing information and activities.
Problem it solves	How to model information so that it can be looked upon as knowledge assets?
Pattern name	Information meta data profile
Solution	Describe all information objects by means of appropriate meta data profiles.
Problem it solves	How to model information so that it can be efficiently and effectively developed and exploited as knowledge assets at the right moment?
Pattern name	Knowledge services
Solution	Develop a comprehensive service model that encourages the development of the professional network and enables knowledge development and exploitation.
Problem it solves	How to facilitate a social network to develop and exploit knowledge assets?
Pattern name	Knowledge activating study
Solution	Develop a knowledge activating study as the primary work environment for the professional.
Problem it solves	How to enable members to take part in the professional network, develop them and exploit knowledge?

Table 1: Four patterns for the design of Concourse.

In the next sections, we describe the patterns.

4.1   Knowledge objects

We need to model all relevant information units in digital knowledge assets, where an asset is seen as an object containing information and activities. So, from a software system point of view knowledge assets are seen as knowledge objects. In order to describe this view, we use the concepts and notation of the Unified Modeling Language (UML) as described by Larman [4], among others. As a widely accepted standard, UML is used for modeling object-oriented concepts.

We make a distinction between four super classes of knowledge objects (See Table 2). A class is a description of a set of objects that share the same attributes, operations, methods, relationships and semantics. A super class is a more generic object that encapsulates two or more subtypes.

Super classes	Simple object	Container object
Information object	1) Information object: File, URL, Message, Simple Chat, Simple Discussion board, ...	3) Information view: Folder, Archive, Interactive magazine, Brainstorm discussion, Concluded project, Concluded workshop, ...
Enterprise object	2) Enterprise object: Person, Simple Organization, Simple Group, ...	4) Enterprise space: Project, Workforce, Conference, Workshop, ...

Table 2: The four super classes of knowledge objects

Firstly, we make a distinction between a simple and a container object. A simple object is a singular object,

while a container object is a plural object, containing two or more simple objects. Secondly, we distinguish between an information and an enterprise object. An information object is an object that has no 'development' or 'exploitation' capabilities. An enterprise object is an object that is capable to learn, and develop and exploit knowledge. Based on this distinction it is possible to record objects that offer information and objects that have a disposition to act in a particular way and have value-adding potential [1]. Herein lies the reason we speak of knowledge objects instead of, for instance, only information objects.

The first super class is the information object. Examples are: files, URL's, chats, discussions, tasks, etcetera. Clearly, we also include interactions like chats and discussions in our model. We aim to refer to all information units that can be relevant for the professional development and the development and exploitation of knowledge.

The second is the enterprise object. Examples are: persons, simple organizations, simple groups, etcetera. These can be described partly by competences, and capabilities that refer to their disposition to act in a particular way. One major characteristic is that these objects can be described by their competences like knowledge, skills, attitudes and experiences and that these competences can change during the time.

The third is the information view. Examples are: folders, archives, interactive magazine, brainstorm discussions, portfolio's, showcases, finished projects, finished workshops, etcetera. These are information sources containing two or more objects. It can contain for instance all finished projects, written articles, showcases, or combinations of these. A wide variety of genres of information views is possible. These can be static or dynamic. A static information view is edited and filled manually by, for instance, a project or a person, while a dynamic information view builds up a set of relevant information objects based on a designed query request.

The last super class is the enterprise space. Examples are: projects, workforces, conferences, workshops, and etcetera. These are objects that can make a difference in what is known. For instance, "the online workshop in 'a certain topic' has shown that …". As soon as these objects are finished, there is no development of the object itself anymore. It becomes a container object. That means it can be used as a source of information but is not 'able to learn' anymore.

A question to be answered is: why do we need this distinction between the four super classes. The first reason is that it offers a clear overview of typical knowledge objects. Secondly, each of the super classes has a typical basic meta data profile. Finally, each of the super classes has typical, basic, knowledge services. We come back to the profiles and services in the next section.

In the center, we realized all four super classes and members have opportunities to add additional object types. The capabilities to learn of the enterprise objects such as person and group are expressed in the meta data profiles of these objects. Each is described by, for instance, a competence profile, which is (partly)

dynamically maintained based on the activities in which the person or group is involved. Starting the design of Concourse, implied a description of the main object types, typical for an online knowledge center in a higher education context. The education concerned the master phase of students, so they were highly motivated and could be seen as junior researchers who took part in the development of the knowledge domain. Examples of typical objects are: references, workshops, project groups, and courses.

4.2   Information meta-data profiles

The second pattern refers to the meta-data profiles. Meta-data has several meanings depending on the context of use. Basically, meta-data describes data. It can refer to a more technical description, for instance, the format of the calendar date, but we are interested in meta-data about the knowledge objects. In this context, we see meta-data as a bibliographic description of these objects. Of course, a technical description is also needed, but not relevant in this context.

The purpose of meta-data is to facilitate search, evaluation, acquisition and use of knowledge objects. A number of initiatives have been initiated to define meta-data standards (www.dublincore.org, www.imsproject.org, and http://ltsc.ieee.org/). A major advantage of an internationally accepted standard is to facilitate sharing and exchanging of electronic knowledge objects among or between users and organizations.

Here, we have chosen the approach of the IMS Global Learning Consortium (IMS). IMS is developing and promoting open specifications for facilitating online distributed learning activities such as locating and using educational content, tracking learner progress, reporting learner performance, and exchanging student records between administrative systems. IMS has two key goals:

- defining the technical specifications for interoperability of applications and services in distributed learning, and

- supporting the incorporation of the IMS specifications into products and services worldwide; IMS endeavors to promote the widespread adoption of specifications that will allow distributed learning environments and content from multiple authors to work together (in technical parlance, "interoperate").

IMS is a global consortium with members from educational, commercial, and government organizations. IMS is producing eight specifications that address logically bounded domains, such as learning resource data, enterprises, competencies, accessibility, etcetera. IMS works closely together with the IEEE Learning Technology Standards Committee's (LTSC) LOM (Learning Object Model) Working Group.

As an example, we give an overview of the learning resource meta-data information model, version 1.2.1 final specification (IMS, 2001). This model describes the

information objects, as described in the former section. Other models can be found on the website of the IMS Consortium. These models list the meta-data elements and describe how they are hierarchically organized. Each element is described with eight pieces of information. These are:

1. Name   How the meta-data element should be spelled.

2. Explanation   The definition of the element.

3. Multiplicity   How many elements are allowed and whether their order is significant.

4. Domain   What the element's vocabulary is limited to and other information.

5. Type   Whether the element's value is textual, numerical or a date; and any constraints on its size and format.

6. Extensible   Whether the element is extensible or not.

7. Note   Why the element was included, guidelines for its use, etc.

8. Example   Sample use of element, where appropriate.

The base scheme of the elements is described in Table 3.

Nr	Name	Explanation
1	General	Group's information describing resource as a whole. For instance: title, language, and description.
2	Lifecycle	History and current state of resource. For instance: version, status, and contribute.
3	Metametadata	Features of the description rather than the resource. For instance: contribute, meta data scheme, and language.
4	Technical	Technical features of the resource. For instance: format, size and requirement.
5	Educational	Educational or pedagogic features of the resource. For instance: interactivity type, learning resource type, and semantic density.
6	Rights	Conditions of use of the resource. For instance: cost, copyright and other restrictions, and description.
7	Relation	Features of the resource relationship to other resources. For instance: kind, and resource.
8	Annotation	Comments on the educational use of the resource. For instance: person, date, and description.
9	Classification	Description of a characteristic of the resource by entries in classifications. For instance: purpose, taxonpath, and description.

Table 3: The base scheme of the elements that describe the resources

Each entry is subdivided. From a bibliographic perspective, there are two important entries. These are Educational and Classification. The entry Educational is subdivided in eleven fixed subentries, Table 3 mentions three examples. Classification however offers the opportunity to add typical, context specific, bibliographic subentries. The context here is continuing professional development by formal and informal learning in the course of practice. Context specific entries can be chosen, based on characteristics such as the particular knowledge domain, the professional network, and the learning routes. The result is a situational base scheme.

The information meta-data profiles are crucial for the functioning of the center. We started with a description

of the knowledge domain 'Networked communication', a so-called information space analysis [1]. We worked out a concept map, containing approximately 50 central concepts. The map is based on an analysis of relevant literature and of the related courses. We took the 50 concepts and ordered them hierarchically. These concepts were added as fixed keywords. We derived a 'table of contents' as the hierarchic description. This table was used as the starting point for the development of archives. We also developed four additional meta-data sub-entries related to Educational entries and Classification entries. These are genres, legitimacies, and usages. The genres are specific. The choices are based on a description of the information genres in organizations [6]. In Concourse, administrators have the functionalities to adapt the profiles based on the user experiences.

4.3   Service model: Description of meeting, development and exploitation services

The third pattern refers to the service model, a description of the services that encourage the development of the professional network and enable knowledge development and exploitation. From a software system point of view, these services are software functions. We prefer the concept services, to make clear that we consider the design of Concourse from a user perspective. Normally, services will refer to a number of software functions.

We distinguish three basic service types: information, interaction, and management services. Information services enable professionals to get information and to present information to others. Interaction services enable professionals to communicate with, to cooperate with, and to support others. Management services enable professionals to manage the center. These services types are related to the four super object classes. Basic knowledge objects can be seen as services. An information object like a simple chat or a simple discussion board, can stand for a number of interaction services like: brainstorm discussions, feedback discussions, frequently asked questions lists, who-knows-who discussions, private discussions, meeting, etcetera. For each service, information services like protocols, templates, and/or additional support can be given, but basically the interaction is (a-)synchronic and is based one or more media such as text, audio and video. Each object is subject to management services, such as: create, edit, copy, paste, monitor, review, etcetera.

The choice for services should be based on an analysis of the characteristics of the use context of Concourse, the workshop. So, who are the members, how are their social relations, what culture do they want, what knowledge assets can be identified, and maybe the most important question, what are the valued information processes, etcetera. This analysis leads to a starting set of objects and related services. Tracking research of the development of the professional network and of the knowledge development and exploitation has to ensure

for a continuing adaptation of the service framework of Concourse.

In Concourse, information services are mainly based on search options. Due to the coherent meta-data profiles, it is possible to develop a wide number of typical searches. One example is the overview object. It behaves like a dynamic folder, and when opened performs a stated query, displaying the actual results. Based on this object type, we are able to develop dynamic archives, containing the latest information. Interactions depend on the communication facilities.

4.4   The knowledge activating study

The last pattern we want to describe here concerns the knowledge activating study as the primary work environment for the professional. We use the word 'study' in the sense of 'to spend time in learning' and in the sense of 'a virtual room used for studying'. Furthermore, this room has to activate knowledge. We prefer activating knowledge above, for instance a more often used concept such as knowledge management. The idea is that the primary task of the 'study' is to facilitate the professional in primary knowledge intensive processes; management of information is seen as a typical secondary process. The last characteristic of our interface concept is 'primary work environment'. More concrete, we see it as the desktop of the professional's computer, instead of, for instance, the desktop of an operating system or the intranet of an organization.

We derived the following three guidelines from this concept. The desktop needs to be communication based. The members, groups and projects and the interaction between or among them have to play a central role in the interface. The reason is that the professional network is the hearth of an online knowledge community. The desktop has to support knowledge intensive tasks. The interface has to allow for a far-reaching level of situational support. The reason is that activating knowledge is an active concept; it implies that there is a continuing performance support while the professional is working. Finally, yet importantly, the desktop has to be practical of use. The interface has to enable the professional to work as a practitioner. It has to be usable and practical.

The last pattern is the knowledge activating study. It has to stimulate communication, support knowledge intensive tasks, and it has to be usable and practical. We try to stimulate communication in a number of ways. First, members are objects, so if you are looking for typical information, you will also find members who most likely have that kind of information. Second, it is always possible to identify the 'owner' of information objects and to contact the owner, if the owner is still a member. Third, we make use of pop up 'online member' windows that show the current online members. This window is situated, so if you open this window while you are in a group space, you are able to see the online group members. Fourth, each group and member has the opportunity to add personal information that can be viewed by others. Still, we think that new ideas are needed, and based on member reactions we are still developing ways to improve. Supporting knowledge intensive tasks is crucial. Our starting point is the 'professional development activity model' that is described in the beginning of this article. This model has led to a typical set of information objects and meta data profiles. Based on these objects and related profiles, we are developing blue prints for each of these activities. A blue print can be a continuing workshop dedicated to a specific set of skills or a predefined support search for experiences, method fragments and best practices. Finally, yet importantly, all these opportunities have to be usable and practical. To enhance the usability, we have chosen for a typical 'windows program' interface. For instance, practicality is enhanced by continually seeking for ways to ease the input of information by the members and to ensure to provide them with practical information. Therefore, we developed an 'information design scheme', that allows for a fast scan of the usefulness of information provided. Still, the usability and the practicality are a constant concern. Based on member experiences, we are constantly trying to improve both.

5   Conclusions and discussion

The Concourse project is aiming at the development of an online writing community as a study support environment for students in Higher Education. Therefore there is a need for an online collaborated writing center as the meeting place and place for writing in a knowledge rich environment. We described the approach for the design of this center, and as part of this approach, the conceptual design of Concourse. We described this design in terms of four patterns. These patterns are 'knowledge object model', 'information data profiles', 'knowledge service model', and 'the knowledge activating study'. From a software engineering point of view, the first two refer to the data model, the third to the functionality, and the last one to the interface of the software system. The center Concourse is described as a proof of concept. The design of the center meets the stated guidelines. The question of usability and eventually the question of effectiveness and efficiency can only be answered after the extensive trials. The next step is to realize a proven concept. From october 2002 till februari 2003, we are carrying out a number of trials. In these pilots, social associations and interactions have to be developed, strengthened and intensified. In these pilots, we are interested in the development of these associations and interactions. Based on the results, we are able to support the development of Concourse and comparable online knowledge communities to proven concepts.

References

[1] Boisot, M.H.. Knowledge Assets. Securing competitive advantage in the information economy. Oxford: University Press. (1999)

[2] Cushman, W. & Rosenberg, D. Human factors in product design. Amsterdam: Elsevier. (1991)

[3] Kaplan, A. The conduct of inquiry. Methodology for behavioral science. New Brunswick: Transaction Publishers. (1998)

[4] Larman, C. Applying UML and Patterns. An introduction to Object-Oriented Analysis and Design. London: Prentice Hall. (1997)

[5] Meister, D. Human factors in system design, development, and testing. Mahwah, NJ: Erlbaum. (2002)

[6] Päivärinta, T. The concept of genre within the critical approach to information systems development. Information and organization, **11**, pp. 207-234. (2001).

[7] Vries, S.A. de. Ontwerpen van communicatiemiddelen. (Design of communication means) In: P.J. Schellens, R. Klaassen, & S.A. de Vries (eds). Communicatiekundig ontwerpen.(Communication Design) Assen: van Gorcum. (2000)

[8] Vries, S.A. de. Online Knowledge Communities: Meeting places for continuing professional development. In: T. van Weert & B. Munro. Social, ethical and cognitive issues of Informatics and ICT. Dordrecht: Kluwer Academic Publisher. (2003)

[9] Vries, S.A. de, Roossink, L.L., & Moonen, J.C.M.M. De ontwikkeling van online leercommunities in het studiehuis. NWO 411-211-04. (The development of online learning communities in secondary education) Enschede: Universiteit Twente. (2001)

[10] Vries, S.A. de, Ten Thij, E., & Cromwijk, J. Het gedetailleerd ontwerp van Concourse. Een online expertisecentrum op het gebied van bedrijfsmatig en wetenschappelijk schrijven. (The detailed design of Concourse. An online knowledge center in the field of business and scientific writing.) SURF Concourse project report. Utrecht: University of Utrecht. (2001)

Practice Related e-Learning - The VIP Framework

Mang Li and Claudia Linnhoff-Popien
Department of Informatics
Ludwig-Maximilians-Universtät München (LMU-Munich), Germany
email: {mang.li, linnhoff}@informatik.uni-muenchen.de

Silvia E. Matalik
Institute for Educational Science
Rheinisch-Westfälische Technische Hochschule Aachen (RWTH-Aachen), Germany
email: matalik@lbw.rwth-aachen.de

Tim Seipold, Carsten Pils and Frank Imhoff
Department of Computer Science IV
Rheinisch-Westfälische Technische Hochschule Aachen (RWTH-Aachen), Germany
email: {seipold, pils, imhoff}@i4.informatik.rwth-aachen.de

Practice is an indispensable aspect in learning. In pedagogic, constructivist learning methods, which promote situated learning in authentic context, are widely used. In an inter-disciplinary project, we have been working on a framework (named VIP) for practice-related e-learning. In particular, practical courses in computer science are considered.

The framework is guided by general principles for constructivist learning environment. Pedagogical concepts for a representative virtual practical course have been developed, which are supported by three technical building blocks in the framework: an e-learning platform, a video-conferencing system and an interactive simulation environment. While for the first two blocks available tools can be reused, new concepts must be developed for the interactive simulation environment, whereas well-proven event-driven and real-time simulation approaches are adopted. In addition, to facilitate situated learning in virtual groups, the session concept and a client-server architecture with CORBA as communication infrastructure are applied. Major design and implementation concepts are presented in the paper.

The representative virtual practical course is performed at our universities to evaluate and to improve the concepts and techniques of the framework.

1 Introduction

Practice plays an essential role in learning. In pedagogic, constructivist learning methods are widely used. A constructivist learning approach such as cognitive apprenticeship [2] is mainly intended to help students:

- obtain *practical experience* through explorative work with real-world items.

- improve *social abilities* (e.g. communication, collaboration, presentation etc.) through teamwork.

Different forms of knowledge and their representation are important. Teaching methods, i.e. modelling, coaching, scaffolding and fading, place also an important role. These concepts should be applied when constructing a constructivist learning environment.

In the era of Internet and multimedia technologies, new communication and representation techniques are available. Their application to constructive learning environment for practice-related e-learning is subject of our project, named VIP - "Virtuelles Informatik Praktikum" (engl. virtual practical course in computer science), where the computer science study is considered as an example application domain. [1]

Currently, in such a course, for example, at our department for computer networks, students experiment with different network equipment, such as hubs, switches, protocol analyzers or other management tools, to explore their functions and to better understand their mechanisms. Usually, students work in groups of three or four members. Due to the availability of experimental equipment, only limited number of students can participate within limited time schedule.

Now, with software representations of the equipment, which are duplicated at no cost and hardly cause any need for maintenance, more flexibility could be achieved at less expense than using the hardware itself. Either over local installations or over remote access, students would even be able to work outside the laboratory, where and when they prefer. The flexibility in location and time is an important benefit of e-learning.

To achieve this flexibility, the e-learning environment must cope with the complexity induced by geographical and temporal distribution. It must also cope with the virtual distance of learners and tutors and therefore with their motivation and their conditions of distance communication and collaboration. These are major concerns in our work. Pedagogical concepts, as well as their technical implementation in a virtual learning environment, which are referred to as the VIP framework in this paper, have been developed in our project.

In this paper, major concepts of the VIP framework are presented. In Section 2, we first introduce the pedagogical concepts of the framework. In Section 3, an overview to the supporting building blocks is given, which are an e-learning platform, a video-conferencing system and an interactive simulation environment. The last building block is the core of the framework. Therefore, details on its design and techniques are presented in Section 4. The evaluation plan is introduced in Section 5. Conclusions close the paper.

2 Pedagogical Concepts

The VIP framework is guided by pedagogical principles for constructivist learning.

2.1 Constructivist Learning

Active research on new constructivist learning approaches, besides the traditional instructional approaches, began at the end of the 1980s. In the constructivist epistemology, "learning is an active, constructive and highly situated process" [15]. It emphasizes on authentic learning contexts, self-regulation, as well as social aspects. Cognitive apprenticeship [2], problem-oriented learning [15] and students-centered learning [12] are some representatives in this realm.

Learning through cognitive apprenticeship [2] follows the model of traditional apprenticeship, which had been practiced long before schools appeared. The approach is that, teachers first demonstrate the task, then support students doing the task, and finally, the students are encouraged to do the task independently. In apprenticeship, group learning is essential to stress social interaction and collaborative learning. For students, this kind of learning is a process involving observation and successive approximation. Whereas from the teacher point of view, cognitive apprenticeship is a process of modelling, coaching, scaf-

folding and fading. It supports students to proceed from observation towards independent-doing.

2.2 Constructivist Learning Environments

Based on the constructive apprenticeship approach, Collins, Brown and Newman [2] proposed a framework for the design of learning environments. It describes four dimensions of an ideal learning environment, namely:

- *Content* includes different types of target knowledge for a learning environment, which are basically domain knowledge and strategic knowledge. Domain knowledge denotes explicit conceptual, factual, and procedural knowledge associated with expertise. Strategic knowledge refers to experts' ability to make use of domain knowledge to solve problems.

- *Methods* are teaching methods to help students to learn how to use, to manage and to discover knowledge. Modelling, coaching, scaffolding and fading are core methods of cognitive apprenticeship. Articulation and reflection help students to gain their own problem-solving strategies through observation and reasoning. Exploration is to push students to solve problems on their own.

- *Sequence* is concerned with the sequencing of learning activities to facilitate the development of robust problem-solving skills, e.g. increasing complexity (from easy to complex), increasing diversity (from single to various), and global before local (first big picture, then details).

- *Sociology* is a critical dimension that involves different social abilities of the students. For example, situated learning puts students into situations that reflect the environment, in which they will apply the acquired knowledge in the future. Culture of expert practice is to encourage students to practice problem-solving in a domain.

Many concepts in the framework by Collins, Brown and Newman are also reflected in other works, e.g. in [12] and in [1].

The application of multimedia- and Internet-technologies for virtual learning environments has been explored in a number of recent research papers [10] [7] [8]. Different empirical studies [10] show already strength of multimedia-supported learning. On the other side, some critical questions need to be further discussed, in particular how to guide learners appropriately, so that they do not feel overwhelmed by new technologies, but profit from self-directed learning. From the technical point of view, many facets of the Internet- and multimedia-technologies have not been sufficiently explored, regarding their application in e-learning. Our work concentrates on the utilization of computer-supported interactive simulation with distributed user environment. This is considered within the scope of a

framework that is guided by the concepts explained in the following.

2.3 Concepts Applied to the VIP Framework

The current focus of the VIP framework are practical courses in the computer science study, which are traditionally carried out in labs. In these courses, students learn problem-solving skills through case studies. Practical courses are in nature constructivist learning methods. Therefore, the general principles introduced above apply well to our work, as outline below (for a more comprehensive elaboration please refer to [16] and [17]):

Situated Learning in authentic context This is the core concept to facilitate active learning in quasi real-life situation. In the VIP virtual learning environment, authentic context is constructed by computer-supported simulations and symbolic graphical representations.

Representation of knowledge Domain knowledge and problem-solving strategies are represented in different electronic forms. Text in the book-like form and hyper-media are complementary to each other. In addition, simulation is used as a dynamic form for knowledge representation. Learning content, either domain knowledge or problem-solving strategies, are cognitively prepared in a flexible manner that can be easily adapted and reassembled for various learning contexts.

Self-directed learning and learning in social context
Both learning forms are important for the development of personal abilities. Self-directed learning allows exploration of knowledge at individual needs. Whereas, collaborative learning is essential to practice teamwork. To complement each other, learning content must be supplied in a well-structured form. In addition, for collaborative learning, learning groups are organized. By means of various communication media of the virtual learning environment, e.g. e-mail, chat, shared white-board and video-conferencing, virtual closeness between group members is provided. The VIP framework provides in particular a simulation environment to support distributed virtual groups, in which group members share the same learning context using a graphical user interface that can be deployed on distributed sites.

Virtual tutoring Over the same communication media of the virtual learning environment for students, tele-tutors can provide scaffolding on demand. A tele-tutor, as other human assistants in the learning environment, has only limited availability. Therefore, some knowledge and functions of a human tutor can be virtualized in the learning environment. In a static form, e.g. as Frequently Asked Question (FAQ), this idea is already widely used. More complex and challenging is to implement it in a dynamic and adaptive form, e.g. as part of the simulation environment. This is a topic for future work.

Didactic design of user interface Learning with pleasure achieves more effectiveness. Therefore, not only the functional, but also the visual representation is crucial. Different aspects of the user interface of the VIP learning environment have been examined from the didactic point of view.

These concepts are applied to a representative virtual practical course. It is performed for the first time in the summer semester 2003, collaboratively at the universities RWTH-Aachen and LMU-Munich (more in Section 5). Conceptually, it is structured as a partial-virtual course, with two kinds of alternating phases:

– *Distributed presence phase.* In this phase, students participate from their lecture rooms in Aachen and Munich. The rooms are connected by a video-conferencing system to distribute lectures and presentations.

– *Virtual self-learning phase.* In this phase, students acquire knowledge through exploration, either individually or in virtual groups. This phase is supported by an e-learning platform and an interactive simulation environment.

Operationally, the course is composed of three distributed presence phases, respectively at the beginning, in the middle and at the end of the semester, and two virtual self-learning phases.

3 Overview of the VIP Learning Environment

As introduced in the previous section, the didactic phases of the virtual practical course are supported by three technical parts: an e-learning platform, a video-conferencing system, and an interactive simulation environment. They are the basic functional building blocks of the VIP learning environment.

3.1 E-learning Platform

This part provides a portal to the learning environment, from which students find learning materials in electronic forms for self-directed learning, e.g. scripts, books or tutorials. It provides also organization of virtual groups, and media for student-student and teacher-student communication.

A number of the so-called group-ware products support the desired functions. We chose an open-source product named CVW (Collaborative Virtual Workspace) [3]. It provides communication tools such as chat and white-board.

Therefore, it follows the house-metaphor. Virtual groups are organized in virtual floors and rooms. In particular, it has a built-in tool for small-format video, which can be useful for self-directed group-learning. Figure 1 shows a screen-shot of CVW.
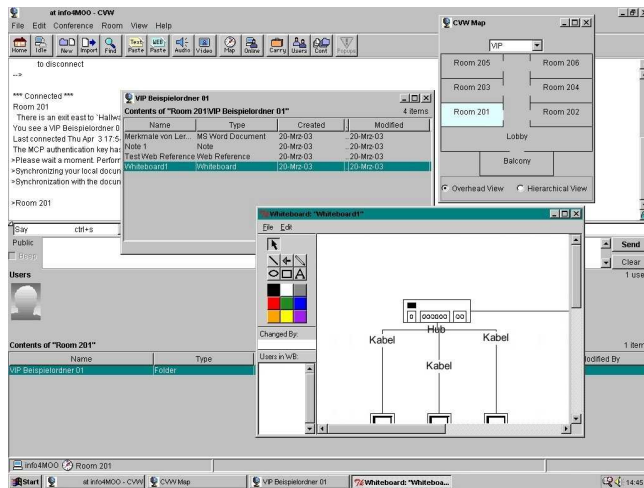

Figure 1: Collaborative Virtual Workspace (CVW)

3.2   Video-Conferencing System

For transmitting the video image in distributed presence phases between lecture rooms, a special video transmission tool - VEX - is used [11]. While there is a large number of (more or less) professional software systems for video conferences or the transfer of video streams under real-time conditions on the market, this number significantly drops when a video quality is required which permits a projection of the transferred pictures on a screen of approximately 4 meters by 3 meters.

Considering video transmission, important characteristics include end-to-end delay, available bandwidth, and packet loss. Extensive research work was done to adapt traffic to variations of these parameters. Two main research directions can be identified. On one hand, rate control mechanisms have been developed that attempt to minimize network congestion by utilizing a predefined bandwidth. On the other hand, error control mechanisms attempt to minimize visual impact of loss.

The H.261 coding algorithm used is a hybrid of inter-picture prediction, transform coding, and motion compensation. The data rate of the coding algorithm may be set between 40 kbps and 2 Mbps. The inter-picture prediction removes temporal redundancy. The transform coding removes the spatial redundancy. Motion vectors are used to help the codex compensate for motion. To remove any further redundancy in the transmitted bit stream, variable length coding is used. H.261 supports two resolutions, Quarter Common Interchange format (QCIF) and Common Interchange format (CIF). We used CIF, as it pro-vides a better quality, especially considering the output scale mentioned before. H.263 encoding is provided as well, deliver-ing superior image quality at the cost of much more computational power necessary for realtime encoding and decoding.

Due to time dependencies of sequentially transmitted video frames, lost packets will cause perception errors at the receiver display. These errors may extend until subsequent packets update the corresponding video frames. That is, the error may be propagated in time. To improve the image quality, an intelligent error correction was implemented. This Feedback Error Control Scheme for block based video communication in packet-switched networks uses a Timestamp Table Scheme, and significantly improves the commonly used Full Intra-frame Refresh patterns. The sender produces a table containing the time stamps of the latest transmitted macro blocks. Receiving a control package the arrived macro blocks which haven't arrived at the receiver can be identified and inserted into the next picture. The advantage of these methods is a substantially more efficient error correction, since no substantial additional expenditure results from transmitting complete pictures in case of packet losses. A comparatively small overhead of this procedure is the additional expenditure for control packets.

3.3   Interactive Simulation Environment

This part supports explorative work through interactive simulations which are "look and feel" of real-world items. It takes care of interactions from distributed users who work in virtual groups. To our best knowledge, there are no such "off-the-shelf" products available on the market, that fulfil the requirements as outlined in Section 4.1. This part is the focus of our work, and its details are elaborated in the following section.

4   An Interactive Simulation Environment

In this section, different facets of the VIP simulation environment are discussed. It starts with a consideration of key aspects according to the pedagogical concepts. Then, the general design concepts are conducted, before the architecture of the simulation environment is introduced. Further, implementation concepts for components of the architecture are explained. Some earlier results have been presented in [13].

4.1   Key Aspects

As discussed in [16], simulation is appropriate to represent knowledge in a dynamic form. For the pedagogical concepts presented in Section 2.3, the following aspects for the VIP simulation environment are conducted:

1. *Adequate representation.* For situated learning, "look and feel" of the physical systems in the desired authentic context must be provided. "Look" is available

over a graphical user interface, whereas it needs not to be necessarily "virtual reality". "Feel" is reflected in the behavior of simulation models for the physical systems. Under the condition of conceptual correctness, simplifications are allowed to ease the development of domain specific simulation models.

2. *Interactivity*. It is an essential aspect to promote "doing" in situated learning. Besides "look and feel", students are able to interact with simulations, in order to learn problem solving strategies in dynamic situations. Issues on interactive simulation are addressed by the consideration of simulation kernel (Section 4.5).

3. *Support for virtual groups*. Collaborative knowledge construction and social skills are practiced in group work. In addition to general communication media provided by group-ware, this property supports domain specific collaboration in a simulated context. For this, the session concept is applied to the management of the simulation environment. It is reflected in the architecture (see Section 4.4).

4. *Flexibility in location and time*. It is a determining benefit of e-learning [20], either for self-directly learning or learning in groups. The location independency is achieved, first of all, using a client-server architecture on CORBA-based communication infrastructure (Section 4.2/4.4), and secondly, by means of the realization of clients with Java-based technologies. In comparison with location independency, time independency is not only the question when clients and server are available. The context, in which a student learns, either alone or together with others in a virtual group, must be able to be created, interrupted, continued and terminated at any desired time. This property is implemented by the management of the simulation environment, where the session concept also applies.

5. *Customization*. It allows the execution of simulations be adapted to individual ability and purpose of a user. Configuration before or during execution is achieved over simulation parameters. Customization in terms of automatic adaptation, e.g. depending on user profiles or behavior, is for future work.

6. *Instruction and explanation*. Scaffolding in a virtual learning environment is a crucial point. Tele-tutoring is e.g. an approach, in which a human tutor answers questions over chat or email [8]. Instruction and explanation as built-in functionalities would be beneficial for self-directed learning. In some approaches, human expertise is directly implemented into the simulation model [6] [19]. Support for this aspect in a more flexible and integrated way is under future consideration.

4.2   General Design

Many different technical and non-technical requirements lead to the design of the simulation environment, including support for distributed interaction to facilitate the explorative group learning process, recording and replay of simulation sessions for analyzing the learning process, flexibility for adding new virtual devices or complete reuse in other fields of teaching, e.g. medical science, by third parties etc. This section gives an overview to some key aspects of the design.

The first issue is the simulation approach, which differs from usual scientific simulations, that are intended to run as fast as possible in order to efficiently generate results. To facilitate interactivity, on one hand, the simulation must be configurable at runtime according to user interaction in difference to the "setup-and-let-run" approach usually found. On the other hand, a real-time coupling of the simulation is required in order to allow the user interactions first hand. Similar requirements may be found in virtual reality research as well, but this project's focus is on simulating only a well-defined abstract subset of the world with scientific precision, not on giving a comprehensive world model.

To support distributed user groups, distribution is a further aspect of the simulation approach. Distributed usage can be supported by distributed simulation. For example, the approach of the High Level Architecture (HLA) [5] proposes federation of distributed simulators of various types (called federates) to achieve interoperability and reuse of simulations. Because federates operate autonomously, the complexity of the synchronization in distributed simulation can be very high. Distributed interactions under real-time conditions may cause problems of time-ordering with complex solutions as e.g. "dead man reconing" known from military applications. To achieve virtual reality with real-time interactivity in a distributed simulation, high requirements (e.g. throughput, delay, error-rate etc.) are put on the underlying networking infrastructure. Such high requirements are not always fulfilled by the networking infrastructure for an e-learning simulation. Last but not least, recording and replaying a distributed simulation is next to impossible due to synchronization issues.

Since distributed user environment does not necessarily imply distributed simulation, we chose an approach using centralized simulation and a distributed interaction environment for clients. Using the model-view-controller pattern we have a single world model - the simulation - on a server at the simulation session site, whereas the view and controller are distributed through client applications among the users' sites.

The simulation environment provides communication facilities for distribution and synchronization of client-server interactions. Locking mechanisms ensure, that only one client obtains write-access to the simulation instance at a time.

The client-server approach implies that every object of the simulation will have to be assembled by two pieces of

code, one on the server side specifying the behavior within the model world and one specifying the look-and-feel presented in the GUI (graphical user interface) at the client. This logical split among the communication line allows use of two different programming languages for server and client side. On the server, it is possible to employ C++ for performance and the availability of simulation core systems, and on the client side the language of choice is Java, as it allows the client to be run on almost any platform, bypassing problems created by the expected heterogeneity of client system.

Communication between the server and the client is carried out using CORBA [18] allowing us to abstract from the underlying networking infrastructure and provides us a direct match between the model objects and the view/controller objects. Therefore, each interaction on the controller can be directly passed to the corresponding model object, and every visible change of the model can be passed to all views immediately. The choice of having a single model does not restrict the framework to be run on a single server, as this single model may be run distributed e.g. on a cluster for performance reasons. The use of CORBA facilitates this, it makes the framework independent from the exact location of the different model objects, which could vary over run-time in such a setting. All synchronization issues arising from the model distribution are of no concern to the framework, as it still sees a single model. Further, distributing the model at the server side takes place in a much more controlled and defined environment, making this far more feasible than distributing the simulation model completely.

4.3 Simulation Session

To manage the distributed usage of a simulation over time, the session concept is used, whereas a simulation session is distinguished from a simulation instance, as defined in the following:

A *simulation instance* is a simulation model in execution:

– It represents the behavior of the physical system to be simulated.

– It can be started, interrupted, continued and terminated over its belonging simulation session.

– A simulation instance is denoted as active in the time after its start and before its termination.

A *simulation session* defines the context for a group of users who are registered with it:

– Its lifetime begins when the first user in the associated group registers with it, and ends when the last user leaves the group.

– It has only one active simulation instance at one time, but may have multiple simulation instances during its lifetime.

While a simulation session is associated with users, a simulation instance is specific to the execution of a simulation model. Therefore, if the same group of users decide for a new run of simulation model, they can remain in the same simulation session. This is comparable with learning under traditional conditions. Students need not to change their partners only because the content of the course has been changed. The key data of each simulation instance can be stored for later reuse or inspection.

4.4 Architecture

In the architecture, the session concept is reflected into the separation of three logical sites, namely, the user site, the manager site and the simulation session site, as illustrated in Figure 2. It is in general a client-server architecture. The simulation session site is the simulation model server, while the manager site is a server for the organization of user groups and their associated simulation sessions.
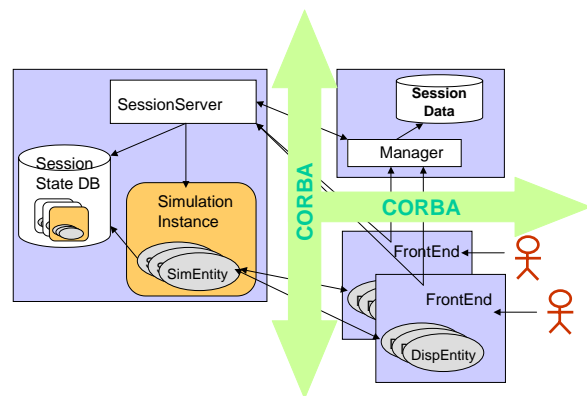
Figure 2: Architecture of VIP simulation environment

User Site It provides a front end with graphical user interface. Over this, a user can register/deregister himself/herself with the Manager to obtain access to a simulation session. After registration, a user observes the progress of the active simulation instance using the graphical user interface. As DispEntities (stands for display entities) are coupled with SimEntities (stands for simulation entities), a user can directly change the configuration of a simulation instance at run-time.

Manager Site Each user group is assigned a simulation session. All simulation sessions are organized at this site. Organizational data, e.g. session identifier, group member, session start/end time, session duration, are stored in a database. An interface for registration and deregistration is provided to users. When a simulation session is required, a simulation session server (see SessionServer in Figure 2) is created. It is terminated when the session ends.

Simulation Session Site A simulation session is controlled by a session server. To manage multiple sim-

ulation instances, the history data of previous simulation instances and the data of the current simulation instance are stored in a session state database, so that earlier simulation progress can be reviewed and reproduced.

The simulation progress is managed centrally at the simulation session site. All interactions from distributed users with a simulation instance, and the display of the current simulation progress propagated to distributed users, are coordinated through CORBA-based interfaces.

4.5 Simulation Kernel

Core of the simulation environment is the simulation kernel. It performs application-independent functions of simulation models. Various concepts and tools resulted from active research in the area of simulation techniques [9] can be adopted to implement the simulation kernel.

4.5.1 Desired Properties

In a simulation, state and time are two important issues. States and state changes over time are used to represent the behavior of a system to be simulated. A simulation kernel provides handling of state and time, in a manner general to domain specific models.

Different notions of time exist, from which the following are adopted:

– *Physical time* refers to time in the physical system. It is the time in the conventional sense.

– *Simulation time* is an abstraction used by the simulation to model the physical time.

– *Wallclock time* is a reference to physical time obtained by reading a hardware clock.

Depending on how the state changes as simulation time advances, simulations can be classified into continuous or discrete simulations. **Event-driven** and time-stepped are two common types of discrete simulations. In an event-driven execution, an event is an abstraction to model some instantaneous action in the simulated system. An instantaneous action can be departure or arrival of a message, or expiration of a timer. In a simulation, each event is assigned a time stamp that indicates the point in simulation time when it occurs. Events usually result in state changes. Simulation time advances from the time stamp of one event to the next.

For applications such as network simulations, which are addressed by the VIP simulation environment, event-driven approach appears to be suitable, because message passing is the primary mechanism used there.

In addition, the simulation kernel must support **real-time execution** to allow interactive work. In a real-time simulation execution, the simulation time advances in synchrony with wallclock time, opposite to common scientific simulation, where the simulation time advances infinitely fast

in the absence of events. To achieve real-time in an event-driven simulation, the advance of simulation time is halted until the wallclock time reaches the time of the next event. By defining the proportion of wallclock time to simulation time, different execution mode, from as-fast-as-possible to real-time, can be realized.

Another important aspect in the consideration of simulation kernel is **scalability**. Scalability is essential to ensure performance for simulation execution, especially under real-time conditions. As shown in Figure 2, the centralized server approach is applied to the VIP simulation environment. On the group/session level, scalability is achieved by the introduction of a SessionServer per simulation session, and its separation from the Manager. On the instance level, to allow execution of complex model, the simulation kernel must provide adequate mechanisms to tackle scalability issues.

4.5.2 RtDaSSF

Desired properties, as outlined above, together with other characteristics such as uniform development interface, open-source implementation, programming language C++, led to the decision for DaSSF [14], an event-driven simulation kernel based on the Scalable Simulation Framework (SSF)[4].

SSF proposes principles for parallel event-driven simulations, with high performance and scalability as its goals. It defines a simple object-oriented programming interface (with language binding for C++ and Java), which is composed of the following basic classes: *entity*, *process*, *event*, *in-channel* and *out-channel*. *Entity* is a container for *process*es that handle *event*s exchanged over *in-channel*s and *out-channel*s. DaSSF is an open-source, C++ implementation of SSF.

DaSSF was not designed for interactive simulations. Therefore, extensions have been made, first of all, to integrate real-time functionality, and secondly, to allow flexible run-time execution control. The extended simulation kernel is called RtDaSSF.

The simulation control interface allows start, suspend, resume and stop of the simulation model. Further, it provides control of the real-time aspects of the simulation run, e.g. controlling the relation between wallclock time and simulation time. Controlling this relation is important for different simulation domains, e.g. continental drift should be simulated at a different time scale than computer networks. Setting this relation appropriately, it is even possible do make the simulation behave in the usually known "run-as-fast-as-possible" way, used when reloading a simulation instance or putting it to a past position. Use of this interface affects all entities in the model at the same time, and it provides also a way to bring the model to a so called safe state.

This is important for the configuration interface, that allows adding and removing entities of the simulation during its runtime as well as changing the object's rela-

tions by changing the way the channels connect the entities. All these changes require the simulation kernel to be brought into the foresaid safe state, basically stopping event processing before making any changes to the model and restarting it afterwards. Putting the simulation into this safe state becomes especially important considering that it may run multi-threaded or even on multiple machines.

All operations committing changes to the model as a whole or single entities provide a time descriptor carrying the current time in all three time scales stating the position at which the event is placed in event queue. This is very important to be able to redo these events later, as they have to be reinserted at the very precise same time slot every time to keep the simulation behavior deterministic.

4.6 Simulation Model

The framework had to make some assumptions about the objects that will be used in a simulation to support this use and provide the additional facilities for distribution etc. As a time-discrete event-driven simulation has been chosen, by inspecting existing simulation frameworks we were able to identify four main concepts (see also Figure 3):

1. An *entity* is a generic abstraction for an object having any behavior within the model. Entities may be added and removed from the simulation any time during its run.

2. *Ports* are used as the endpoints for communication channels that interconnect entities and among which messages are passed between entities within the simulation model. Ports can only exist connected to an entity and may be (dis-)connected during runtime. Ports are typed to prevent connection of incompatible entities.

3. *Parameters* provide a way to influence an entity's behavior during runtime by changing values, e.g. the packet generation rate of a packet generator. Obviously, parameters are immediately connected to an entity.

4. *Variables* provide a way to peek at internal data that an entity collects. Variables are read-only and can not exist without an associated entity.

All of these concepts make no assumptions about implementation specific internal details. An entity may consist of several other sub-entities and processes, but it presents itself as a single object to the framework. Comparing to SSF, one can easily see how this concept maps.

Recording a simulation instance is commonly done by generating a trace file containing all events occurring during the simulation run. This technique is appropriate for later analysis of the simulation, but it does not allow to do an identical rerun of the simulation, as only the events changing the state of the involved entities are recorded, but not the states themselves. Further, the trace files tend to
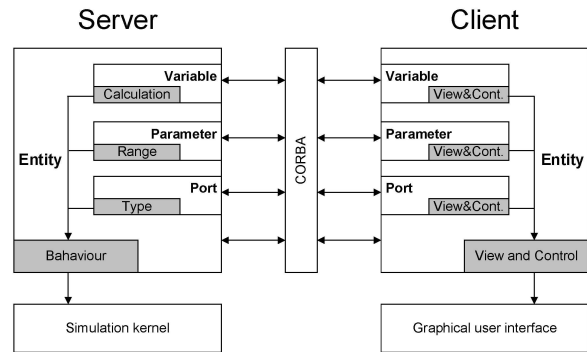

Figure 3: General simulation model

reach prohibitive size to store a significant number as is expected in our application.

This problem is solved by recognizing the fact that the simulation is driven by two different types of events. On one hand, there are internal events (iEvents), that are generated by one entity within the simulation model. And on the other hand, there are external events (eEvents) generated by client interaction. iEvents reassemble the majority of all events within the simulation, as e.g. packet generation proceeds at higher frequencies than any users can make changes to the simulation model. These events are not immediately visible to the user, but some entity is necessary to make them accessible. Therefore, iEvents need not to be distributed for viewing to the clients, reducing the required communication to the clients vastly and thereby making this approach feasible at all.

Given that the state of a simulation instance can be restored identically - including the state of all used pseudo-random number generators and the event queues - all iEvents from that state on will be regenerated automatically by the simulation's deterministic behavior once it is started again. This means, that it is sufficient to save the initial states of the simulation and all following eEvents, in order to be able to replay the complete simulation instance. This causes of course a time penalty when e.g. resuming a stored simulation instance, as the complete simulation has to be rerun from the start, but this problem can be tackled by periodically saving a snapshot of the simulation's state and restarting the simulation later from the last state.

An eEvents created by client interaction needs to be distributed to all other clients, as it changed some visible aspects of the model, and the views have to be updated accordingly. Once a client successfully manipulates an entity, the simulation environment passes the change to the simulation instance, propagates the change in view to all other clients and records the eEvent augmented with its occurring simulation time, so it may be inserted at the right time when loading and replaying a simulation instance. The eEvents are stored together with the information about which entity they affected and which client caused them, so an entity records all external changes done to it during its lifetime. By recording the wallclock time as well, it becomes possible to store and replay the set of simulation runs originat-

ing from the same scenario. This allows a teacher to get a complete trace of the learning process that the students have gone through, allowing him a very profound insight in their advances.

4.7 GUI - View and Control

The GUI provides generic functions for session control, as starting, stopping, saving and leaving a simulation. It provides a workspace on which the view-objects are placed. The base classes for these view-objects contain the framework's functionality of communicating any interaction to the server. This way the model-programmer is shielded from the complexities of CORBA programming and the framework from him making mistakes regarding the used communication protocol.

The framework synchronizes the workspace among the clients, but each client is free to choose which part of the overall scenario he is interested in. Each entity offers the client access to the parameters and variables provided by it. Access to these facets is provided on request only in order to keep the communication requirements between server and client as low as possible.

Manipulating entities or their parameters requires a client to lock the access to that object. Manipulations may be done only while holding the object's lock, providing us the well-defined time order of actions that is necessary for the desired deterministic behavior of the simulation. The access is only granted for a certain amount of time (15 seconds by default), so no user is able to monopolize an object. If the lock is not freed before the end of the granting time, the lock is revoked by the server side of the framework. The client is informed of that fact and his actions are not committed.

Besides the hidden intelligence in the base classes to be used for the view/control objects, the visible part of the client application is important as well, as it is the front-end to the simulation and needs therefore be written with focus on usability. For example, as the desktop workspace is expected to be bigger than the visible screen area, a radar-widget was written and added, that provides an abstracted, shrunken view of the complete workspace together with indication of the currently visible section.

Evaluation and extension of the GUIs visibility will be a major issue during the conducted test practical and - as with every software - never be perfect. A screen-shot of the current GUI is shown by Figure 4.

4.8 Extensibility

In order to provide maximum reuse of the framework and the components used, it is necessary to modularize as much as possible. Especially the framework has to be as generic as it can be and must be clearly separated from the content using it. Therefore, no further restriction than the time-discrete event-driven simulation type is put on the model
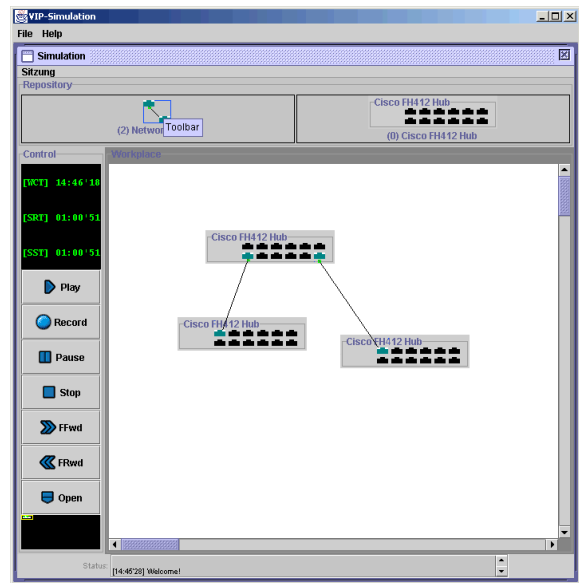
Figure 4: Graphic user interface

developer, as he is given the complete power of the C++ programming language on the server side.

To shield the model programmer from the complexities of the underlying framework, provided base classes have to hide most of the framework's mechanisms - communication, event recording etc. - and to provide an API (Application Programming Interface) for using the simulation kernel with the specific extensions.

The same concepts apply to the client side GUI objects, as the model programmer needs to write them in Java as well, but is granted complete freedom in terms of look and behavior.

The code is loaded dynamically into the framework at runtime, on the server side as system native binary code by a plug-in mechanism, on the client side using the Java class-loading facilities. In this way, it is possible to minimize server downtime and foremost to make recompilation of the server and the client application unnecessary.

Using modularization and delegating much of the responsibility to the model programmer, we are able to make content creation for actual scenarios especially easy. A scenario using any given set of model objects is stored as XML data and loaded when initiating a simulation instance. Therefore, the content creator needs no programming knowledge to create tasks and settings fitting his teaching activities, just an XML-editor. In this way, best possible reuse is guaranteed, as the framework with a given set of model objects can easily be tuned to fit the specific teaching needs at a site without having to dig into all the details. Only the model programmer needs to know the APIs to the core to programm the models.

Mixing models from different sources is theoretically possible, but is an unsure thing to do by now, as the internal data representations of messages exchanged between entities have to be compatible, which is not guaranteed by

the framework. Adding this, and e.g. providing graphical frontend for assisting the content creation process, is work to do.

4.9   VIPNet - A Network Simulation

For the exemplary practical course for telecommunications and networking, a network simulation called VIPNet based on the VIP simulation environment has been developed. An object-oriented VIPNet API (illustrated by Figure 5) eases the implementation of simulation models. The base classes of the API apply fundamental concepts of the ISO-OSI 7-layer model and the Internet TCP/IP model:

–   A *Component* is the basic building block of a simulation model. It is an *entity* according to Section 4.6. It is conceptually a protocol container, which may represent different abstractions in terms of protocol layer. For example, it can be used to simulate a host, a switch, a router, a hub or a cable. For interconnection, a *Component* may have one to more *ComponentPort*s.

–   The in-life of a *Component* is realized by its *Protocol*s, which are in general simulations of network protocols. According to the concept of Service Access Point (SAP), the API provides the base class *SAPPort* for inter-protocol information exchange, from which *SAPProviderPort* and *SAPUserPort* derive. The base class for messages (e.g. Protocol Data Units (PDUs) or frames) is *ProtocolMessage*. In addition, *ProtocolTimer* and *ProtocolSession* are also provided for protocol simulations.
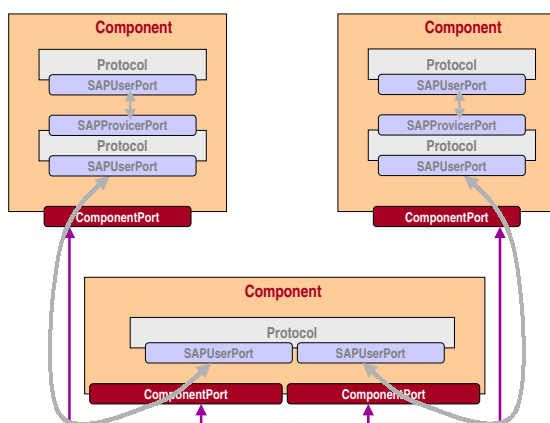

Figure 5: VIPNet API

Based on VIPNet API, simulations for different network protocols (such as IP, TCP, UDP) and components (e.g. layer2-switch, router, web-server) have been developed. An overall functional and performance evaluation of them will be made in the test practical course.

5   Testing and Evaluation

At the time of writing, in the summer semester 2003, a practice course for telecommunications and networking based on the VIP framework is being tested, collaboratively at the universities RWTH-Aachen and LMU-Munich. Goals of the test are to evaluate and to improve the technical realization of the framework, together with its didactic design. For the evaluation, a qualitative approach is used, which is complemented with quantitative questionnaires.

For the didactic design, a formative, i.e. process-accompanying, and a summarized evaluation is in progress. They are essential for the selection and control of didactic approaches to achieve goals, such as to deepen fundamental knowledge in telecommunications and networking, and to apply the knowledge in real-world situation. In addition, the development of key qualifications in terms of communication and cooperation are examined. Students, tele-tutors and developers are interviewed. Accomplishing to the test course, participating observation is made to obtain insight to the communication and cooperation process between students, as well as between students and tutors.

For interviews with tele-tutors and developers, three sets of questionnaires, partially uniform, have been developed. They are applied respectively briefly after the kick-off meeting, in the half- and in the final-presentations. Tele-tutors and developers are interviewed orally over telephone. Students are questioned in written form in the three phases, for which the questionnaires have been made available over the Internet. Data collected by the questionnaires are clustered and quantified to generate assessments to the didactic design of the practical course. An improved design will be used in the regular computer science study at universities.

6   Conclusions

In an inter-disciplinary project, we have been working on a framework named VIP for practice-related e-learning. The framework is based on well-proven principles for constructivist learning environment.

Pedagogical concepts for a representative virtual practical course for computer science study have been developed. Two conceptual phases compose such a course, namely distributed presence phase and virtual self-learning phase. In these two alternating phases, students acquire key qualifications, that include not only domain specific knowledge, but also social abilities through teamwork in virtual groups.

The pedagogical concepts are supported by three technical building blocks in the framework: the group-ware Collaborative Virtual Space (CVW) as e-learning platform, the video-conferencing system VEX developed at RWTH-Aachen in a previous project, and an interactive simulation environment, which is core of our work.

Well-proven approaches from active research on simulation techniques since decades, e.g. event-driven simulation,

real-time simulation and Scalable Simulation Framework (SSF), have been reused and extended. Furthermore, to facilitate constructivist learning in virtual groups, a client-server architecture based on the model-view-controller pattern for the simulation environment has been developed. The architecture implements the session concept to support group work over time. A graphical user interface, which can be deployed on distributed sites, provides users with "look and feel" of authentic learning context over direct access to simulation models executed on the server. In this paper, different facets of the VIP simulation environment are presented.

To test and evaluate the framework, the representative virtual practical course is performed in the summer semester 2003 at our universities. The evaluation results will be used to improve and to further develop the framework.

Acknowledgment

References

[1] J. B. Black, W. Thalheimer, H. Wilder, D. de Soto, and P. Picard. Constructivist Design of Graphic Computer Simulations. *National Convention of the Association for Educational Communications and Technology*, 1994.

[2] A. Collins, J. S. Brown, and S. E. Newman. Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics. In L. B. Resnick, editor, *Knowing, Learning, and Instruction. Essays in the Honour of Robert Glaser*, pages 453–494. Erlbaum, Hillsdale, NJ, USA, 1989.

[3] The MITRE Corporation. Collaborative Virtual Workspace (CVW). http://cvw.sourceforge.net, 2001.

[4] J. H. Cowie. *Scalable Simulation Framework API Reference Manual, Version 1.0*, Mar. 1999.

[5] Defense Modeling and Simulation Organization. The High Level Architecture. Technical report, http://www.dmso.mil/hla, 1999.

[6] P. Dillenbourg. The Design of a Self-Improving Tutor: PROTO-TEG. *Instructional Science*, 18(3):193–216, 1989.

[7] F. Fischer and H. Mandl. Lehren und Lernen mit neuen Medien (Teaching and Learning with New Media). In *Handbuch Bildungsforschung (Handbook of Educational Research)*, pages 623–637. Leske + Budrich, 2002.

[8] F. Fischer, P. Tröndle, and H. Mandl. Using the Internet to Improve University Education: Problem-Oriented Web-Based Learning and the MUNICS Environment. Technical Report 138, Ludwig-Maximilians-University Munich, Institute for Empirical Pedagogy and Pedagogical Psychology, Jul. 2001.

[9] R. M. Fujimoto. *Parallel and Distributed Simulation Systems*. John Wiley & Sons, Inc., New York, 2000.

[10] B. Herzig. Lernföderliche Potenziale von Multimedia: Medienbezogene, lerntheoretische und didaktische Aspekte (Potentials of Multimedia for Learning)). In M. K. W. Schweer, editor, *Aktuelle Aspekte Medien Pädagogischer Forschung (Current Aspects of Media Pedagogical Research)*, pages 149–186. Westdeutscher Verlag, Wiesbaden, Germany, 2001.

[11] F. Imhoff and C. Linnhoff-Popien. Internet-based Teleteaching using the IP-based German Broadband Science Network. In *Proceedings of the Intern. Conference on Intelligent Multimedia and Distance Education*, Fargo, ND, USA, Jun. 2001.

[12] S. M. Land and M. J. Hannafin. Student-Centered Learning Environment. In D. Jonassen and S. Land, editors, *Theoretical Foundations of Learning Environments*, pages 1–23. Lawrence Erlbaum, Mahwah, NJ, USA, 2000.

[13] M. Li and C. Linnhoff-Popien. Interactive Simulation in E-Learning. In *Proceedings of the IASTED International Conference on Applied Modelling and Simulation*, pages 229–234, Cambridge, MA, USA, Nov. 2002.

[14] J. Liu and D. M. Nicol. *Dartmouth Scalable Simulation Framework, User's Manual, Version 3.1*, Aug. 2001.

[15] H. Mandl, C. Gräsel, and F. Fischer. Problem-Oriented Learning: Facilitating the Use of Domain-Specific and Control Strategies through Modeling by an Expert. In W. Perrig and A. Grob, editors, *Control of Human Behaviour, Mental Processes and Awareness*, pages 165–182. Lawrence Erlbaum, Hillsdale, NJ, USA, 2000.

[16] S. E. Matalik. Mediendidaktische Funktionen von Simulationen im virtuellen Informatik-Praktikum - VIP (Media-Didactic Functions of Simulations in VIP). In *Referenzmodelle netzbasierten Lehrens und Lernens. Virtuelle Komponenten der Präsenzlehre (Reference Models for Network-Based Teaching and Learning)*, 2002.

[17] S. E. Matalik. Didaktik-Ansätze für ein virtuelles Informatik-Praktikum (Didactic Approaches to a Virtual Practical Course for Computer Science). In B. Bachmair, P. Diepold, and C. de Witt, editors,

Jahrbuch Medienpädagogik 3. Leverkusen (Annual Report for Media-Didactic). Leske und Budrich, appearing in spring 2003.

[18] Object Management Group (OMG). *Common Object Request Broker Architecture (CORBA/IIOP), Version 3.0.2, formal/02-12-02*, 2002.

[19] E. Shaw, R. Ganeshan, W. L. Johnson, and D. Millar. Building a Case for Agent-Assisted Learning as a Catalyst for Curriculum Reform in Medical Education. In *Proceedings of the International Conference on Artificial Intelligence in Education*, Jul. 1999.

[20] K. Stiller. Möglichkeiten und Grenzen des Medieneinsatzes in Lehr-Lern-Prozessen (Opportunities and Limits of the Application of Media in Teaching-Learning-Process). In M. K. W. Schweer, editor, *Aktuelle Aspekte Medien Pädagogischer Forschung (Current Aspects of Media Pedagogical Research)*, pages 119–148. Westdeutscher Verlag, Wiesbaden, Germany, 2001.

Information and Communication Technologies and Information Systems Planning in Higher Education

Jacques Bulchand and Jorge Rodríguez
Universidad de Las Palmas de Gran Canaria
Departamento de Economía y Dirección de Empresas
Campus de Tafira
35017 Las Palmas de Gran Canaria
jbulchand@dede.ulpgc.es, jrodriguez@dede.ulpgc.es

In this paper, we propose a methodology for the development of strategic information system (IS) and information and communication technology (ICT) plans in higher education. The methodology we propose is composed of nine steps and involves the whole of the university community, not just IS/ICT technicians. These nine phases, derived from strategic planning procedures, are preplanning, external environment assessment, internal evaluation, strategic interest themes identification, mission and vision statements declaration, strategic axes identification, goals and strategies definition, project and specific actions definition and implementation and evaluation. Paper ends showing application of this planning procedure to the Universidad de Las Palmas de Gran Canaria.

1 Why IS/ICT planning is necessary

Lately, we are seeing higher education institutions going though a series of very important changes. On one hand they are subject to growing economical restrictions. On the other, there is a vast necessity of information systems (IS) and information and communication technologies (ICT) in all the environments of university life [2].

Under this perspective, information systems and information and communication technologies (IS/ICT) development requires careful planning. This planning should involve the whole university community. Indeed, they should be given the chance to opine about which are the priorities that must be attended in the short term and which should be the financial resources distribution.

Besides allowing participation, IS/ICT planning lets us to achieve the following goals [1,6,8]:

- Alignment between the IS/ICT strategy and the corporate strategy, as well as providing an appropriate environment for the long term IS/ICT management. This means being able to satisfy present and future information needs.
- Guarantee the necessary resources so that the ICT area is able to face rapidly changing environments. This means being able to satisfy urgent requirements while guaranteeing high quality in resulting IS.
- An efficient and feasible structure for the IS/ICT area. This area should be able to define IS needed by the organization, and how to use efficiently ICT. This use should not be focused exclusively on the internal of the organization but should also acknowledge external trends.

- Improve the communication between management and the IS/ICT technicians. This means generating a sense of corresponsability between them. It also means management getting to know the ICT area and computer specialists getting to know the direction the organization is going to take in the next few years.
- Dealing with a critical and expensive organizational resource. Many times, managers tend to see ICT as a necessary problem instead as a critical function of their business.
- Obtain an organizational knowledge by doing the planning process. This way future development of similar processes will be easier.

A detailed analysis of these six goals allows us to conclude that the most important of them, the one that makes these processes not only necessary but almost essential is the problem that the IS/ICT area means for many managers [6].

Indeed, managers usually consider IS/ICT as a high resource consumer, although they do not really perceive the utility of these investments [8]. What is worse, consumed resources grow every year. Besides, they are not able to talk to the IS/ICT heads and if they do, they very rarely get to an agreement. Due to this, managers do not know what is going on in the IS/ICT area and technicians do not know about business direction and how technology can help business management to achieve the company's objectives [6]. But, above all this, managers have the sensation the area is out of control. It

is not productive and it has become a bottleneck that blocks business' growth and improvement [1].

Causes of this complex situation are several. Very frequently IS departments are former ICT or data processing departments [1]. Due to this, a communication barrier is developed between computer technicians and the rest of the organization. Besides, IS/ICT planning is guided more by technological reasoning than by business needs [8].

Fast changes in ICT is another factor that makes IS/ICT planning processes compulsory. Without them, it is almost impossible to follow ICT changes because every new technology has to be individually faced [8]. Thus it is usual to find organizations in which the ICT responsibles stay anchored to the ICT they know. When this happens, decisions do not take into account all the possibilities.

Last we should consider that IS/ICT management in organizations has usually been carried out with short-term approaches, instead of long-term or strategic approaches.

The situation found in many organizations resumes all this exposition. Chaos in IS/ICT area is such that many directives would prefer to design their systems starting from scratch [8]. Even if it was possible, success would not be guaranteed. Indeed, if something has been done wrong in the past, there is no reason to expect it being done right in the future. Unless, obviously, techniques like strategic planning are used, since these techniques try to learn from mistakes made in the past.

1.1   Barriers to IS/ICT planning

From what has been exposed, it seems obvious that organizations need to carry out formal IS/ICT planning processes. However, it is verifiable that very few do. What is the reason for this behavior? Basically, the existence of a series of barriers to IS/ICT planning processes. We now present these barriers.

The literature distinguishes two main groups who resist to the development of these processes. They happen to be those involved in the planning process, i.e. managers and IS/ICT technicians.

To understand the resistance of IS/ICT technicians, we must consider that almost all the causes previously described (growing expenses, failure to communicate and fast changes in ICT) have been happening since several years. Due to this, organizations have tried to apply in this area every single possible management technique, such as outsourcing, downsizing or total quality management. Unfortunately, results obtained up to date have not really been significant. As a consequence, technicians tend to be skeptical and consider each new initiative and strategy as a trend which is in vogue but which will shortly disappear. What is worse, technicians are convinced managers do not really care about these questions [3].

Managers, on their hand are not favorable either to IS/ICT planning. First they do not really see the impact that IS/ICT have, because they are not able to understand how to obtain competitive advantage through them.

Second, we have a credibility gap in ICT use. Managers see procedures that ICT should be able to solve easily, but they find their organization is not able to automate these procedures. Third, managers do not view information as a business resource to be managed for long-term benefit, as economical and human resources are [8].

This situation shows us why organizations do not normally have IS/ICT planning processes. On one hand managers are not really convinced of their necessity. On the other, if they try to carry them out, they find a frontal opposition from technicians.

From this we obtain two conclusions. First, before starting a planning process, managers must be convinced of the benefits it is going to mean. Second, it is essential to transmit thoroughly to all IS/ICT members the phases that are going to be carried out and their goals. Also implementation and evaluation must be carried out so results can clearly be seen.

1.2   Problems when there is not an IS/ICT plan

Once the main goals of the planning process and the main existing barriers to them have been defined, some authors describe the problems organizations that do not have strategic IS/ICT plans will face [8].

First, there will be a loss of business opportunities and the organization will incur in a competitive disadvantage in respect to competitors, since systems and technologies will not be aligned with business objectives. Due to this they can become a restrictive factor to their development.

Second, there will be a lack of integration between systems and an inefficient management of data. This is duplication of efforts, lack of precision, delays and information not useful for business management.

Third, development priorities will not arise from business needs. Instead, projects to be developed will arrive from available technologies and from the search of an application for them [6]. This means technology will guide the institution instead of the institution deciding what it wants to do with technology. Due to this, projects will change frequently, so productivity of the IS/ICT area will be low, costs will increase and IS quality will be low, too.

These three kinds of problems should help convince managers about the need of carrying out IS/ICT planning processes.

2   Planning methodology

We now propose a methodology to develop information system and information and communication technology plans in higher education institutions. We base our work in the existing methodology for the development of strategic plans applying it to the information system and technology area [2,4,5]. Thus, we propose a methodology composed of nine phases that we now present.

Each of these phases requires a series of techniques and tools. These will be referred to in each of the phases although they will thoroughly described in next section.

2.1　Preparation for planning

This phase can be considered previous to the planning process. Its main goals are three. To define why we are going to carry out the planning, to determine the exact process that is going to be used and to define the involved work teams.

The total execution time should be between four to six months, comprising all the phases. Regarding the human resources meant to develop the plan, we propose five different figures and workgroups:

- The directive sponsor who must be a relevant figure in the organization with decision and convening capacity. In the special case of universities we propose this figure to be developed by the general manager or the vice-rector in charge of the technology area.
- The project manager. This figure is in charge of the direction of the project at all levels. It can be held by a teacher with knowledge in this area, by the technology vice-rector or by an external consultant.
- The responsible of developing the plan, in charge of all operational matters.
- The steering committee, composed by the rector, the general manager, one or more vice-rectors involved with the area, and the IS/ICT head, plus all figures previously mentioned. We consider this group should meet about three to four times while the plan is being developed. Once at the beginning, once or twice during the development and once when the final redaction process is ending.
- The work team composed by the project manager, the top responsible of the project, and some members of the organization's system and technology area.

2.2　External environment assessment

The second phase is external environmental assessment. Its main goal is to detect existing trends and to check their influence on the university being studied. This phase is very important for the planning process, as changes in IS/ICT are very frequent. In this phase, we obtain two of the SWOT matrix components: opportunities and threats.

In higher education environments, this diagnosis is even more complex than in other institutions, since it is expected that universities lead the process of starting up new technologies, not only in research areas, but also in management areas.

The techniques and tools to be used in this phase are other university IS/ICT plan examination, observation and in-depth interviews. Other university IS/ICT plan examination allows us to have a perspective of what other similar organizations are doing. Observation is probably the most useful technique in this phase. A good environmental diagnosis is only possible if we access and process properly all the information that universities have at their disposal.

In-depth interviews to IS/ICT experts, from our own university, from other universities or from outside the university environment, are an extremely valuable resource of information.

2.3　Internal evaluation

Third phase is dedicated to internal evaluation. Its goal is to analyze current and under development IS, user perceptions, infrastructures, current human resources and economical resources dedicated in the last few years. As a result of this phase we will obtain strengths and weakness of the SWOT matrix.

Some characteristics of universities make it especially difficult to accomplish this phase. Expenditure decentralization and department autonomy makes it complex to evaluate the ICT infrastructure and possible shortages. Evaluation of human resources is difficult too, since it is complex to find out technical support needed, especially by teachers. Indeed, this collective is extremely heterogeneous. We have teachers that need help for all operations, while others are sure of having enough knowledge so they demand being able to manage their own ICT. Unfortunately this last situation is very complex, since it is common that institutional norms get usually skipped, in things like server configuration, updating frequency of the antivirus file, etc.

The techniques and tools to be used in this phase are several. Gibson and Nolan's model, value chain model, Rockart's critical success factors model, discussion groups, in-depth interviews, questionnaires, other university IS/ICT plan examination, detailed examination of all the existing documents regarding the IS/ICT area and direct observation.

Gibson and Nolan's model allows us to examine the current position of the institution and to foresee what future should bring. Value chain and critical success factors models may be used to understand the activities carried out in universities and the relative importance of each of them. From them we find out which systems must be developed and their development priority.

Discussion groups let us find out the perception users have from current systems. In-depth interviews should be done to relevant people in the university and they allow us to explore thoroughly the main problems of current systems and to evaluate service provided by the IS/ICT area human resources.

Questionnaire allows us to verify and measure length of IS, ICT and human resource problems. They also allow participation of the whole community. Other university IS/ICT plans help us find out the main problems that these universities are facing, which usually will be very similar to those in our institution.

Documents such as the job post relation and those related to current and under development IS give us a theoretical view of the situation. Last, direct observation of IS/ICT staff at work can help us complement information got from other sources.

2.4   Strategic interest themes identification

Fourth phase is meant to identify strategic interest themes, which can be considered as the union of challenges and trends.

Challenges appear formally expressed in the institution's strategic plan, if it exists, or in its SWOT matrix. Even if the institution's strategic plan exists, we may find that in its development IS/ICT have not been taken into account specifically. When this happens, no challenges will be found in the strategic plan regarding the IS/ICT area, so we will have to use the SWOT matrix too.

Trends, on their hand can be identified through Delphi method and through in-depth interviews. As seen above, detecting trends in the university environment is a very complex task.

Techniques to be used in this phase are four. First the institution's strategic plan and its SWOT matrix help us finding challenges. On the other hand, other university IS/ICT plans and articles by experts in the area help us finding trends. Results must be validated through in-depth interviews to rectoral team members, to department directors and to faculty deans. Also through a questionnaire send to the whole university community and, especially, through a Delphi method, creating a panel whose members are university IS/ICT experts.

2.5   Mission and vision declaration

Fifth phase is dedicated mission and vision statements declaration. Mission is a precise definition that justifies the existence of the organization's IS/ICT area and where functions provided by it to the rest of the organization is found. Vision is where the organization wants to arrive in the long-term. To develop a proper definition of mission and vision, we must find out what university community members expect from the IS/ICT area.

Techniques to be used in this phase are four. First, in-depth interviews to rectoral team members, to department directors and to faculty deans. Second, examination of other university IS/ICT plans, from where we will get examples of mission and vision statements that can be used as guides. Third, observation of tasks usually accomplished by IS/ICT staff, which can help in mission definition. Last, the questionnaire that will be sent to the whole community, allowing us to know what expectations they have from the area.

2.6   Strategic axes identification

Sixth phase consists of strategic axes identification. These axes are a few pillars over which future can be organized. When possible, they should be derived from the institution's strategic plan. Else, in-depth interviews must be used.

If we are under perfect alignment between the corporate and the IS/ICT strategies, the institution's strategic plan will contain the necessary references to the IS/ICT area. Alas, this will not be the usual situation. So we will have to face situations progressively further away from ideal. First, we may find there is a corporate strategic plan, but that it does not include IS/ICT references at all. If this is so, we will have to add to the strategic axes found in the corporate strategic plan one or more specifically referred to the IS/ICT area. Second, they may not be a strategic plan at all. In this case, strategic axes must be defined from scratch.

Considering how important the IS/ICT area is in universities, if there is a strategic plan, it will reference the area without any doubt. If we must start from scratch, they are some axes that always exist in all universities, like teaching and research. The rest of the axes, and even if these two are placed together or separate, depend on the profile of the university.

From what has been exposed, it is obvious that the main technique to be used in this phase is the observation of the institution's strategic plan, if it exists. If it does not, we will have to derive the axes from the in-depth interviews to the to rectoral team members, to department directors and to faculty deans and from other IS/ICT plans that we may be able to access.

2.7   Goals and strategies

In the seventh phase, goals and strategies are proposed for each of the axes defined. Usually they will be between three and five goals for each axis. Goals concerning IS, ICT and information management must be present.

Techniques to be used are in-depth interviews, questionnaires, Delphi method and other university IS/ICT plan examination. In-depth interviews to rectoral team members, to department directors and to faculty deans should help solve problems that have arisen in previous points. Questionnaires sent to the university community will help measure importance of each goal. Delphi method is used to validate goals obtained from other techniques. Last, other university IS/ICT plans should be used, since goals that can be valid in our university may be obtained from them.

2.8   Projects and specific actions definition

In the eighth phase, a series of projects and specific actions are defined for each of the strategies previously outlined. Each of them is assigned a budget, a responsible and one or more indicators that allow us to control them.

In universities, this phase is slightly more complex than in other organizations, due to the structure of the university and the different existing interests. For example, in definition of the budget for each action, it is very difficult to value matters like the resistance that we may find in some of the members of the university when trying to start up homogeneous systems for the whole university.

These specific actions are derived from the questionnaires, interviews and observation of internal documents and from examination of other university IS/ICT plans. All of them will be validated in the Delphi method.

2.9 Implementation and evaluation

Ninth and last phase is called implementation and evaluation. Its goal is to guarantee that proposed actions are put into practice and to evaluate obstacles found.

If implementation and evaluation is important in all organizations, in universities it is even more important. Indeed, organization's size, its complexity and IS/ICT decision making decentralization make the implementation process much more complex. Techniques to be used in this phase are several, but we consider communication to involved parties the most important. This communication can take place through e-mail or the institution's web page.

In the evaluation phase we use the indicators that have been generated for each action. We suggest using the Balanced Scorecard, which allows us to get and integrated vision of all the indicators.

3 Field work

All through the developed proposal, we have pointed several methods and techniques to be used in each of the phases. However, it is not necessary to develop them individually for each phase. Instead, carrying them out just once though the planning process is enough to achieve, simultaneously, all the intended goals.

For example, it is not necessary to carry out a questionnaire for internal evaluation, another one to detect strategic interest themes and another one for goals. Instead, in the same one all required questions can be incorporated. This way, important synergies are achieved and the community is disturbed the minimum possible number of times.

We now present each of the techniques and tools to be used, outlining an approximate proposal of when they should be developed.

3.1 University strategic plan

We have already mentioned that the best situation is when the university has a strategic plan and that it is aligned with the IS/ICT plan. In this case, the strategic plan and the IS/ICT plan will be developed in parallel. Thus the strategic plan will take into account the IS/ICT area, so challenges needed for strategic interest themes and strategic axes will be developed in the institution's plan.

If the strategic plan did not take into account specifically the IS/ICT area when developed, we will probably have to add a strategic axis specifically focused on systems and technologies. Last, if we do not have an institution's strategic plan at all, the strategic interest themes and the axes for the IS/ICT plan will have to be drawn from the external and internal analysis of the institution using other research techniques.

Obviously, the analysis of the institution's strategic must be carried out at the beginning of the planning process, since it will determine the IS/ICT plan development.

3.2 Internal documents

Internal documents are an important source of information for internal evaluation and for mission and vision definition. In a university, two kinds of documents must be analyzed. On one hand, documents referred to the institution as statutes and normatives. These should be taken into account especially if we do not have an institution's strategic plan. On the other hand, documents specifically referred to the IS/ICT area, like the job post relation with their profiles and competencies.

3.3 Direct observation

Direct observation of the work carried out by the IS/ICT staff is useful to complete the results obtained from the examination of internal documents. With it, we also find out which are the tasks really done by the IS/ICT staff above what is officially stated. Obviously, in every organization's area, staffs do tasks apart from those formally assigned. But in the IS/ICT area this situation is extremely frequent, since changes in ICT make it almost impossible to reflect in formal documents changes in the tasks to be carried out.

3.4 Other university IS/ICT plans

Examination of other university IS/ICT plans is one of the most useful information sources in the planning process. This is because environment in all universities is fairly similar. They will be used as part of environmental analysis, internal evaluation, and in trend definition for the strategic interest themes. If we do not have an institution's strategic plan, we will use this examination for the axes definition, too.

Obviously, the most interesting plans will be those concerning universities close to ours from geographical, size, kind of institution and orientation (polytechnic, humanist) points of view. However, experience proves that even very different universities face very similar problems in the IS/ICT area.

In this sense, we consider interesting to use benchmarking, which means identifying the critical processes in our university and selecting universities that carry out the best practices in these areas. Analyzing how this excellence is obtained, organizations learn how to improve their own processes [7].

3.5 Discussion groups

The main goal of discussion groups is to detect the problems associated with IS/ICT use in the university under study. Due to this it is a technique to be used during the internal evaluation phase. We understand discussion groups should not try to find solutions to problems, since in initial phases the main goal is to find out the current situation and not to solve existing problems.

We also advise that ICT technicians do not participate in these groups. Instead, we propose only non-expert users to participate. However, if a solution appears, this will not be a problem. We only must bear in

mind that the discussion group must not become a place to find solutions.

We propose carrying out a minimum of five discussion groups with university community members, trying to respect the criteria of groups being heterogeneous between them, but homogeneous internally. So, two would be dedicated to administrative staff, two to teachers and researchers and one to students.

In the two dedicated to administrative staff, one should be preferably used for middle management and the other one for administrative assistants, excluding the IS/ICT department. If possible, it would be advisable to carry out a greater number of groups (four to six), allowing to segment staff using other variables such as if they do attention to public or not, central or peripheral units, and even considering the number or years in the university. Groups should be carried out during working hours but far from the usual working place, to allow staff to isolate themselves.

For teachers and researchers, we also propose a minimum of two groups, selecting members based on the academic category. We would place in one group full professors. In the other group, associate and assistant professors. If it was possible to celebrate more than two groups, additional divisions could be made based on number of years in the university. Since we also consider that in these groups proposing solutions should be avoided we think not too many members from technical departments should be called in.

As for students, we propose a minimum of one group in which at least one student from each of the academic areas should be present. These students should belong to different study years. If possible, three groups should be carried out. Two for undergraduates and one for postgraduates.

3.6   In-depth interviews

In-depth interviews should be carried out to rectoral team members, department directors and faculty deans. Not all these members must be interviewed, since the number could be considerable. Instead we propose selecting a smaller number through a directed random selection process. The goals of the in-depth interviews are:

- Validate problems detected during internal evaluation.
- Detect perceptions and sensations that relevant university members have about the university's IS/ICT.
- Analyze what these members expect from the university IS/ICT that can be useful for the mission and vision definition.
- Find some solutions to the main problems already detected, identify goals and define challenges that will be part of the strategic interest themes.
- Define strategic axes if an institution's strategic plan does not exist.

In-depth interviews should be carried out once we have ended and analyzed discussion groups. Thus, they will take place during internal evaluation, in its ending part. Interviews should be based on a semistructured

questionnaire of about 20 to 40 questions, and should take between 60 and 90 minutes.

The result of the process will be a document that presents for each question the most usual answer and significant exceptions. From interviews, we will obtain conclusions for the SWOT matrix, for the critical success factors and for the goals and action plans.

3.7   Questionnaire

Questionnaire is destined to all members of the university community, and its main goals are the following:

- Confirm problems that arose during the discussion groups and that were not resolved in the in-depth interviews.
- Allow all university members to take place in the elaboration of the plan, allowing them to give their opinions, which is extremely important for the plan development and for the implementation phase.
- Find out the opinion that the community has about some trends that are happening, making easier strategic interest themes definition.
- Help in defining goals and action plans.

We think it is advisable to carry out the questionnaire after finishing in-depth interviews, once they have been analyzed. Normally, as we prefer the maximum number of participants to answer the questionnaire, it should be possible to answer it for as long as possible. This is, from when analysis of the in-depth interviews finishes until when plan writing starts.

3.8   Delphi method

Last, we propose developing a Delphi method in the ending part of the planning process, when a first outlining of goals, strategies and action plans has already been made. In this case, we think the participants should be experts in university IS/ICT, relevant rectoral team members and some external members, such as experts from local government, nation-wide system experts or university IS/ICT suppliers. We now detail who should be selected in each group.

- Rectoral members. We think some members of the rectoral team should be selected. Rector, general manager and some vice-rectors, basically those involved in institutional planning, research, systems and technologies and those directly involved in some of the actions of the plan (e.g., currently, a plan will probably include a lot of measures related to online training, so vice-rector of this area should be present).
- Administrative staff. Those directly involved in actions suggested by the plan, as the responsible of the training area.
- Researchers and teachers. If in the university there is a group whose main research subject is IS/ICT planning they should be in the expert panel.
- Local and national government. Usually public universities have a close relationship with government, so we suggest including some of their

members, from technical areas and from university relationship area.

- IS/ICT responsibles from other universities. Some IS/ICT managers from universities similar to the one under study, in size means, geographically, kind of university or orientation (polytechnic or humanistic).

- Other institutions related to university. There is an important number of organizations that work almost exclusively in the university area, so their points of view will be interesting. E.g., in Spain we have Universia and the Oficina de Cooperación Universitaria (OCU).

- Suppliers. Those suppliers with a greater presence in universities could also take place in the Delphi method, such as telecommunication operators or hardware and software constructors.

3.9   Analysis tools

In this section we describe some analysis tools that can be used in the IS/ICT planning process in universities. They are the value chain model, the critical success factors and Gibson and Nolan's model.

The value chain model is extremely useful in the university environment. First it helps separate clearly support activities from production ones. We believe this is extremely important, because if it is not done, we risk that administrative support services to become the university's IS/ICT goal. Second, value chain model allows dividing information requirements of a complex organization without loosing a global point of view.

Use of critical success factors model in university environment should make us conclude that basic tasks are those related to teaching, researching and services provided to the community. When applied in IS/ICT planning they allow us to find out which IS must work optimally to achieve those goals. Value chain and critical success factors models are used in internal analysis to help detect areas to which the IS/ICT function should pay special attention.

Last, Gibson and Nolan's model allows us to evaluate the exact point in which we are in the IS/ICT implementation process. In complex and decentralized organizations as universities, institutional evaluation from this point of view can help us know what to expect from the next few years.

4   ULPGC IS/ICT plan

We now describe the process taken to develop the Universidad de Las Palmas de Gran Canaria's (ULPGC) IS/ICT plan. We start characterising this institution, then the process carried out and last the resulting plan.

4.1   Characterization of ULPGC

In the context of Spanish universities, ULPGC places it self in the first by twenty-five in size. It has about 1.600 teachers and researchers, 800 administrative staff and 22.000 students, with 5 campuses geographically separated and a total of 17 faculties and 36 departments placed in 12 different buildings. ULPGC was created in 1989, based on the existing Politechnical University of the Canary Islands. This is why ULPGC has a strong technological component in its members.

Internationally, and due to its privileged geographical position, ULPGC has important relations with Europe, Africa and America. It has exchange programs with 111 universities and cooperates with institutions and universities from Senegal, Mauritania and Morocco in Africa; and with almost the whole of Latin America, but especially with Cuba, Venezuela, Mexico, Brazil and Puerto Rico. Due to these relations, accessibility to computer services from the exterior is fundamental.

Technologically we can quote about 5.000 devices connected to the network, about 50 computer rooms, considering free access ones and teaching ones. About 8.000 mail messages are processed daily.

4.2   Development of the ULPGC IS/ICT plan

We now describe the steps taken to validate this methodology through its application to ULPGC.

The proposed plan was developed in eight months, those between November 2001 and June 2002. This is slightly above the amount recommended in the methodology, but we believe it can be justified because only internal resources were used and because of institution's complexity.

During the eight months, a direct observation of the IS/ICT area of the university was carried out, and during the first four months internal documents were examined.

Five systems and technology plans of higher education institutions from the United States were examined. These five examined plans were obtained from Internet and were those of Berkeley, MIT, University of Arizona, Penn State University and East Tennessee State University.

IS/ICT and institutional strategic plan were developed with a certain overlapping, so some degree of alignment was achieved. However, the institutional plan was developed without really considering strictly IS/ICT so we did not find too many strengths, weakness and trends in the institution's strategic plan that really affected IS/ICT.

During November 2001, five discussion groups with non expert users were carried out. The first two of them were with researchers and teachers, one with full professors and another one with associate and assistant processors. The next two were with administrative staff, one with directors and second level directing staff, and one with administrative assistants. Last one involved students from different areas of the university.

From these discussion groups, a structured list of more than 120 items was created. These items were classified in 10 subjects, each item reflecting a problem detected by users or a wish about how IS/ICT should work.

Using these problems and desires found in the discussion groups, a semistructured questionnaire composed of 31 questions was elaborated. This

questionnaire was used in a series of in-depth interviews that were carried out in December 2001 and January 2002. These in-depth interviews were carried out to 8 rectoral team members, 6 deans and 6 department directors, for a total of 20 interviews.

Last, in April 2002 two simultaneous processes started. On one hand, a questionnaire was sent out to all the members of the University. On the other hand, a Delphi method was developed.

This questionnaire was only filled using electronic media, basically the institution's web page. A total of 544 answers arrived. 234 from teachers and researchers, 143 from administrative staff and 167 from students.

For the Delphi method, an expert panel composed by 22 experts in IS/ICT from university and non-university environments was used. First round took place in April and May 2002 and the second one in June 2002. In each of these rounds, experts where asked about three different items. First, trends in general environment and their possible influence in our university. Second, about the main goals to be carried out. Third, the actions needed to achieve the goals and qualitative and quantitative indicators for each of them.

In the first round, economical and temporal estimations where asked for and in the second round a confirmation of these was required. The Delphi method was implemented only through electronic procedures (i.e., e-mail) reducing considerably expected times for the development of such a method.

Also two of the three proposed tools were used, the value chain model and Gibson and Nolan's model, not having used the critical success factors model.

4.3   ULPGC IS/ICT Plan

To end this paper, we now present the main contents of the Information and Communication System and Technology Plan of the ULPGC (Plan de Sistemas y Tecnologías de la Información y las Comunicaciones 2003-2006, Plan STIC ULPGC).

Following the proposed methodology, the Plan has been structured into four axes. In each of them some goals have been defined. In each goal some action plans are defined, each of which has a priority level (between one and three). For each plan an execution time has also been estimated (short, medium and long term). Last, for each action plan we have a series of concrete actions, with a budget and a timetable for those to be executed in 2003.

4.3.1   SWOT Matrix

From the analysis developed by IS/ICT heads, from the discussion groups and from other university IS/ICT plans, we got the main environment trends that affect the university's IS/ICT. The analysis was considered externally (threats and opportunities) and internally (strengths and weaknesses).

External environmental analysis was structured in three kind of variables: technological, sociocultural and economical. As a result we got 8 opportunities and 2 threats.

Internal analysis was structured in four areas: teaching, researching, services to the university community and technology and available resources. As a result, 8 strengths and 20 weaknesses were found.

4.3.2   Mission and vision

We define the ULPGC mission as "Give University Community a stable, productive and efficient ICT environment, that makes teaching and researching easier, provides services to university community and to society, and helps knowledge management processes".

Vision is defined as "We pretend to place our University in the top-ten in the IS/ICT area in our country, at the same level as reference universities in our context. ULPGC should be innovative in its daily tasks and must promote information and knowledge society in its environment."

4.3.3   Axes

We found four axes in total. Three were derived from the institution's strategic plan, those related to teaching, researching and management and services to the community. Fourth axis is specifically for IS/ICT and has been called "Technology and Available Resources".

4.3.4   Goals

Axis 1, "Teaching" is composed of three goals,
- 1.1. Help the massive incorporation of ICT to teaching activities, changing the learning paradigm.
- 1.2. Place ULPGC in the online teaching and learning market.
- 1.3. Ease student access to ICT.

Axis 2, "Research" has the following two goals,
- 2.1. Help the use of ICT in research processes.
- 2.2. Ease knowledge transference between researchers.

Axis 3, "Management and Services to Community", has four goals,
- 3.1. Give the community only those services that market does not offer or in which ULPGC is able to offer a high added value.
- 3.2. Offer university community electronic personalized services.
- 3.3. Define, increase and spread the services to which the community has a right to access remotely.
- 3.4. Adapt the IS/ICT department to the university community needs.

Last, axis 4, called "Technology and Available Resources", has three goals,
- 4.1. Manage ICT strategically.
- 4.2. Achieve integration of all University databases.
- 4.3. Guarantee human and financial resources needed to face the plan.

Each of these 12 goals has a series of actions. In total, we have proposed up to 60 actions and 36 indicators to evaluate them. The Plan is available for download (in English and Spanish) at http://www.ulpgc.es, in the option "Presentación" → "Planes y Presupuestos".

5   Conclusions

Using a systematic approach to the development of system and technology plans in universities can be a great aid to achieve a successful distribution of limited financial resources in an expansive area such as that related to new technologies.

But with this technique, not only a better resource distribution is achieved. Other important results are also obtained which should not be underestimated. First, the whole university community participates in the process, which makes them feel part of it. Second, as they participate in each of the phases, they get to know the plan. This is also very important, and helps the ninth phase, implementation and evaluation, as one of the main goals of this phase is diffusion of the plan to the university community.

6   References

[1] R. Andreu, J.E. Ricart, J.E., J. Valor. Estrategia y sistemas de información. Barcelona, Spain. McGraw-Hill, Interamericana de España, S.A. (1996).

[2] A.W. Bates. Managing technological change. Strategies for college and university leaders. San Francisco (CA), EE.UU. Jossey-Bass Publishers. (2000).

[3] B. Boar. The Art of Strategic Planning for Information Technology. New York (NY), USA. John Wiley & Sons, Ltd. (2001).

[4] J.M. Bryson. "A strategic planning process for public and non-profit organizations". Long Range Planning, 21(1), pp. 73-81, (1988).

[5] J.M. Bryson. Strategic planning for public and nonprofit organizations (2ª edición revisada). San Francisco (CA), USA. Jossey-Bass. (1995).

[6] A. Cassidy. A practical guide to information systems strategic planning. Boca Raton (FL), USA. St. Lucie Press. (1998).

[7] CIO. Uneasy pieces. Document from Internet: http://www.cio.com/archive/060196_uneasy_4.html. Visit date: June, 18th 2002. (1996).

[8] J. Ward, P. Griffiths. Strategic planning for information systems. Chichester, UK. John Wiley & Sons. (1996).

The Portal of "GreCO-Universités"

Pierre-Yves Cunin, Christine Lacombe, Jean-François Desnos and Christian Lenne
CICG 351 Avenue de la Bibliothèque, BP 53, 38041 Grenoble Cedex 9
Christine.Lacombe@grenet.fr

GreCO is a common project of the five universities of Grenoble [2] aiming at widespread deployment, within themselves, of information and communication technologies for education in every aspect. One major cornerstone is the design and implementation of a common portal able to support the full set of services, dedicated to education, in a way ensuring coherence (with respect to the five universities) flexibility and customizability. The paper first presents briefly the major objectives and components of the GreCO project. Then it focuses on the strategic project "Portal", enlightening goals, organization and technical approaches and solutions. The GreCO (Grenoble Open Campus) project is the result of a study launched in June 1997, and brings together the five universities in the district of Grenoble. Its aim is the harmonious deployment of Information and Communication Technology applied to Education (ICTE) within the universities. It has the clear ambition of making Grenoble into a model for university education, both nationally and internationally, by 2004.

1 Introduction and background: the GreCO project

1.1 Structure and tasks

The GreCO project [2] is subdivided into eight "**strategic**" projects and many "**operational**" projects, in order to reach the following objectives:

- in the medium term: to promote teaching initiatives in the universities by setting up operational projects (100 such operational projects have been implemented so far);
- in the longer term: to integrate ICTE into every area of university life, with the help of unifying strategic projects.

A management team is responsible for the coordination and oversight of all aspects of the project.

The eight strategic projects prepare the universities for the profound changes that ICTE brings into further education, and on the horizontal plane take care of the coordination and implementation of transferable methods. The eight projects are as follows:

1. To heighten the ICTE-awareness of all teaching staff and other concerned personnel.
2. To provide more thorough training to teaching staff when necessary.
3. To define pedagogical models in order to structure the courses offered down to the level of a single module of lessons.
4. To design follow-up modules for students who may live far from the campus.
5. To define a production and distribution policy.
6. To define and implement an adequate hardware and software infrastructure.
7. To define quality standards for all services offered to users.
8. To implement a single reception area - the **portal** - able to support the full set of services.

1.2 Objectives for year 2003

The GreCO project is a very large project universities (of the Grenoble district) have decided to undertake together. Such a project has many strategic common objectives. In order to be pragmatic and to phase the work to be done "reasonable" objectives have been chosen for year 2003. The most important ones are:

- Access for 12-14000 students (25% of the total) to all or some of the services.
- At least 50% of the teaching staff made aware of the integration of ICTE in teaching.
- 30 courses redesigned (ECTS, modularization), to enable part-time and distance learning.
- 100 complete or partial training courses using these methods (operational projects).
- Implementation of a plan to widen the scope of quality certification (e.g. ISO).
- Elaboration of a regional, national and international communication policy.

2 Objectives of the portal project (strategic project 8)

Higher education establishments must obtain or create new ways of passing information, producing knowledge, organizing and communicating between students, teachers, administrative staff, technical personnel and the wider socio-economic environment.

We are currently witnessing an upsurge in online information and services, developed in a chaotic manner, with no real consideration of users' needs. In contrast to this multitude of partial solutions, the portal is defined as an integrated environment providing a common vision for all the universities, while allowing each one to customize its services.

2.1   General objectives.

The portal project aims to provide a range of ICTE-based services to anyone interested or involved in education, in particular:

- to offer exhaustive information on the courses available and the enrolment procedure;
- to enable enrolment without requiring the students physical presence;
- to have access to, or to produce, all information concerning the training process;
- to present the various educational activities in an efficient and organized way;
- to enable communication and collaboration (groupware) between all users.

This group of services will be accessed in an integrated and coherent manner from a single entry point via a web browser. The offer will be generic, customizable, open, secure and modifiable.

2.2   Expected benefits

The benefits expected from this project are as follows:

- improved student education: learning, organization, communication, job-seeking, etc.;
- more efficient administrative procedures;
- higher quality of information;
- more transparent universities, with a more positive image;
- a first step towards the information system of tomorrow's university.

In the preliminary phase of the portal project (during the academic year 200-2001), the desired services were identified and an exhaustive needs analysis produced [1]. The second phase, which began in October 2001, aims to achieve the following technical goals:

- elaboration of a large-scale (LDAP type) directory that will enable any person in the universities to be identified and to have secure access to the portal, according to his or her roles;
- implementation of high-priority services in all of the universities, concerning courses offered, enrolment, course organization, student life, access to documentation, etc.;
- creation of pilot (moke-up) portals in order to test various technical options and to assess the desired usability and accessibility .

3   Technical presentation of the portal project

3.1   The target users

The project concerns all those people who are involved in university life (80,000 people), but is centered on the student. It also addresses the outside world. Portal users fall into the following categories:

- students (in undergraduate or postgraduate education, from abroad, etc.),
- teaching staff,
- administrative and technical staff,
- members of the industrial world who are closely linked with higher education.

3.2   Sub-projects

Within the portal project, the work has been divided into sub-projects, each dealing with a particular aspect of university life. Some sub-projects are purely technical ones, others are more functional ones (service oriented).

A project team made up of a supervisory board (project manager's role) and the manager of each sub-project has been set up for this phase of the overall project.

All the work is carried out on an inter-university basis with the relevant university departments (joint and single departments).

Figure 1 shows the organisation of the project.

Architecture sub-project: Integration. Launch of the portal, platform for the integration of services and components, Single Sign On (SSO) and authentification mechanisms.

Directory sub-project. Design and implementation of one or more portal user directories. This sub-project constitutes the technical foundation for the development of any large-scale portal.

Architecture sub-project: uses. Design of portal architecture, from the point of view of the user : Design of screens, navigation, scenarios, graphical design of the MMI…

Information and orientation sub-project. Design and implementation of services in the areas of courses offered, work placements, job seeking, etc.

Student life sub-project. Identification of useful information for students (accommodation, grants, arts, health, etc.). This sub-project helps to improve students' living standards.

Administration of studies sub-project. Survey of existing procedures, definition of desired services (enrolment, notes, calendar, timetables, etc.), implementation and evaluation of the portal administrative services.

Special needs sub-project. Identification and consideration of users with special needs: the disabled [3], high-level sportspeople, foreign students, students suffering from long-term illness, etc.
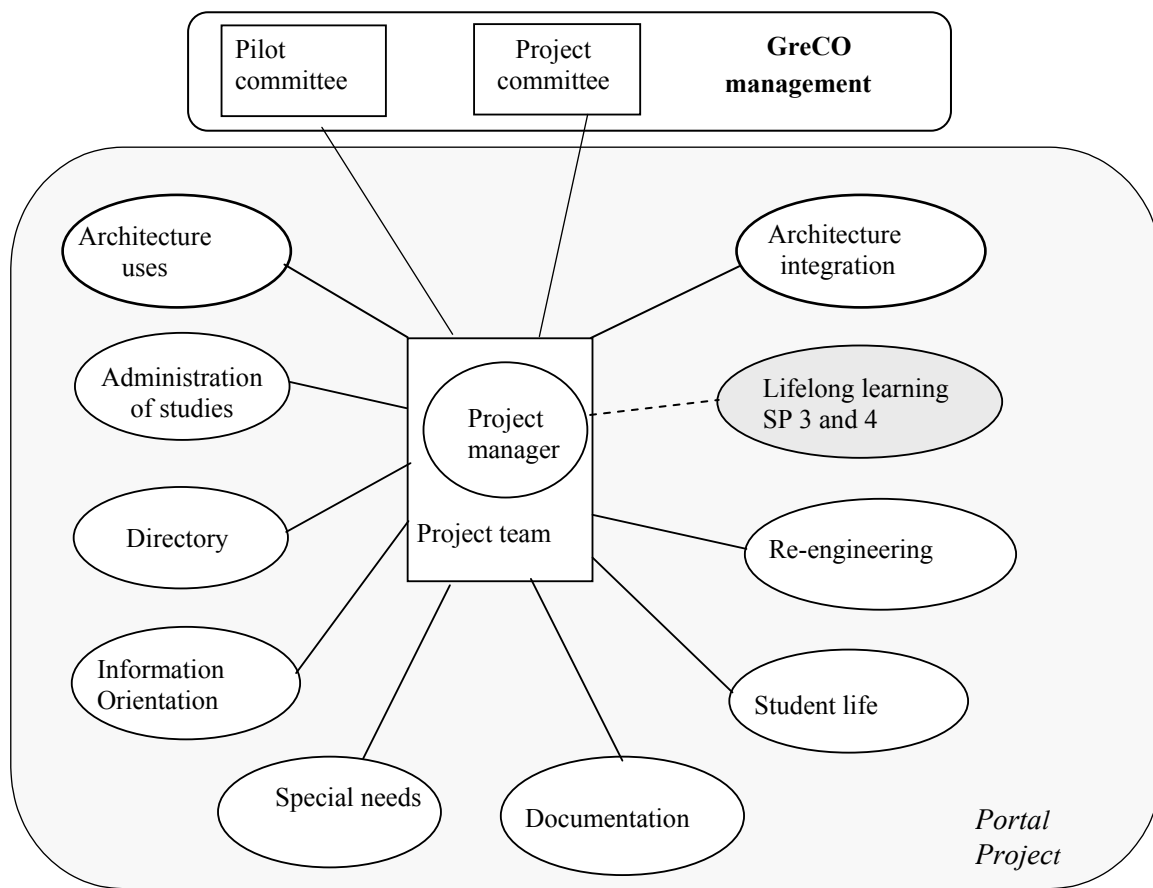
Figure 1 : Project structure diagram

Documentation sub-project. Inventory of existing computerized library services, transparent interconnection of the departments concerned with the various documentary resources (loans, catalogues, etc.).

Lifelong learning sub-project. This sub-project is concerned with the production and access of education resources (lessons, distance supervision, etc.). In GreCO this task forms part of strategic projects 3 and 4 (see section 1). The portal should enable a movement towards suitable educational platforms such as Learning Space, WebCT, TopClass, or simply an electronic notice board for storing and/or sharing educational files.

Re-engineering sub-project. Technical problems due to existing computer systems (interoperability of inherited equipment), coherence of databases developed in the universities, integration of users' favourite tools, etc. The legacy system can be preserved while retaining the flexibility to choose future technologies.

3.3   Implemented mechanisms

In this section, we describe the main mechanisms which must be provided in the portal to implement services. They constitute the basic tools supporting services.

- Mechanisms to integrate well known software, that is applications we have full control on their development or on their source code.

- Mechanisms to integrate black box web tools and applications. Further development is impossible because of the unavailability of source code. However, such an application provides a web interface and is accessible from a navigator (Web compliant)

- Mechanisms to integrate black box applications where the source code is unavailable and which are not accessible from a web interface.

- Mechanism for secured transfer with FTP protocol and button-operated to transfer or deposit forms, transfer, documents, tools etc.

- Information profiling management : this mechanisms allows the user to subscribe to his favourite interests and to receive associated "flash" information. Information is distributed to the right students at the right time.

- Access rights management according to user profile (classes, groups to which the user belongs determine rights and roles).

- Content production: the application enables the user to author new content and publish it, for example, to subscribers. We provide an editor, import facilities with encapsulation and a docflow tool to support the document production.

- Single Sign On: User can sign-on once to the portal and receive personalized Web pages providing access to the information, people, and applications they need and they are allowed to.
- Process support: Workflow tools for defining modeling and automating (enacting) the various processes, e.g. administrative procedures.
- Access to information in databases: a large part of the universities information systems is based on classical database systems. A high level request language is provided to access and interoperate with those databases.
- Connectors: we need connectors between the main directory (LDAP based) and the various legacy applications used by the universities, e.g. national scholarshipmanagent applications, accounting and financial applications including staff salary.

To summarise, these mechanisms should provide :

- A single access to all resources associated with the portal
- Personalized interaction with the portal services
- Integration with applications and workflow systems

4   The first version

This section provides an overview of components included in the first version. To begin with, the implementation of user directories constitutes the technical foundation for the development of any large-scale portal.

Furthermore, portal users fall into the following categories: students, teaching staff, administrative and technical staff…. Students, however, represent the main target of the portal. Therefore, this version aims to provide the components students are more interested in: a virtual desktop and information and services related to student life.

4.1   The directory sub-project

This sub-project, named AGALAN, deals with the design and implementation of one or more portal user directories [7].

Furthermore, the portal is designed to integrate with existing and future systems and applications. This extensibility must provide data synchronization and role-based single sign-on processes. Single sign-on means that after a user logs into the portal, that user can log in to any integrated (external) system (according to his or her role) simply by clicking a link – and without entering another user name and password.

SSO relies on AGALAN which must implement an adequate data structure in an efficient and organized way. Thereby, AGALAN provides data structure giving information concerning all resources of the GreCO Universities portal.

The figure 2 presents the overall structure of the sub-project which expected results are as follows :

- Development of a joint directory (LDAP type) and associated tools, that will enable any person in the universities to be identified and to have secure access to the portal, according to his or her role;
- General instructions for the elaboration of a large-scale implementation
- Shared development tools for specific needs of a university according to the diversity of the establishments involved.
- Integration of this common directory in the information systems of each university. This implies the elaboration of connectors with different existing services. A connector is a piece of code designed to interpret the URL query strings sent by the LDAP server, communicate with the external system code or directly with the system's databases, and respond to the LDAP server using the appropriate syntax.
- This set of connectors is the main resource of the portal for authenticated access to its various services. Each connector is developed only once, for all the establishments.

4.2   The virtual desktop

One of the main components of this portal is the virtual desktop. This is not a component fully specific to the academic world despite some needs expressed by the potential users interviewed during the initial phase (collecting needs).

After user identification and authentification, the virtual desktop becomes the user's homepage. This homepage can be profiled according to his needs and priorities or habits in accordance with his roles. In other words, users only need to go to one website to access a whole range of information and services which they can customize to their own unique needs.

It includes community applications listed below such as webmail, chat, forum, shared folders, calendar, targeted announcements, etc.

Shared folders

Groups can take advantage of the community tools to interact more effectively and efficiently.

Folders may be shared between training staff and students, administration and students, between teachers or between students. This application enables communication and collaboration between all users.

For example, training staff and students are interested in sharing course documents, students homework. Files and graphic uploads are made available to only those members included in the group.

Files, such as educational documents and presentations can be rendered to training staff belonging to a group. For example, a Word document could be made available only to faculty members teaching Computer Science somewhere in the portal.

Similarly administration of studies may share documents with a relevant student group.
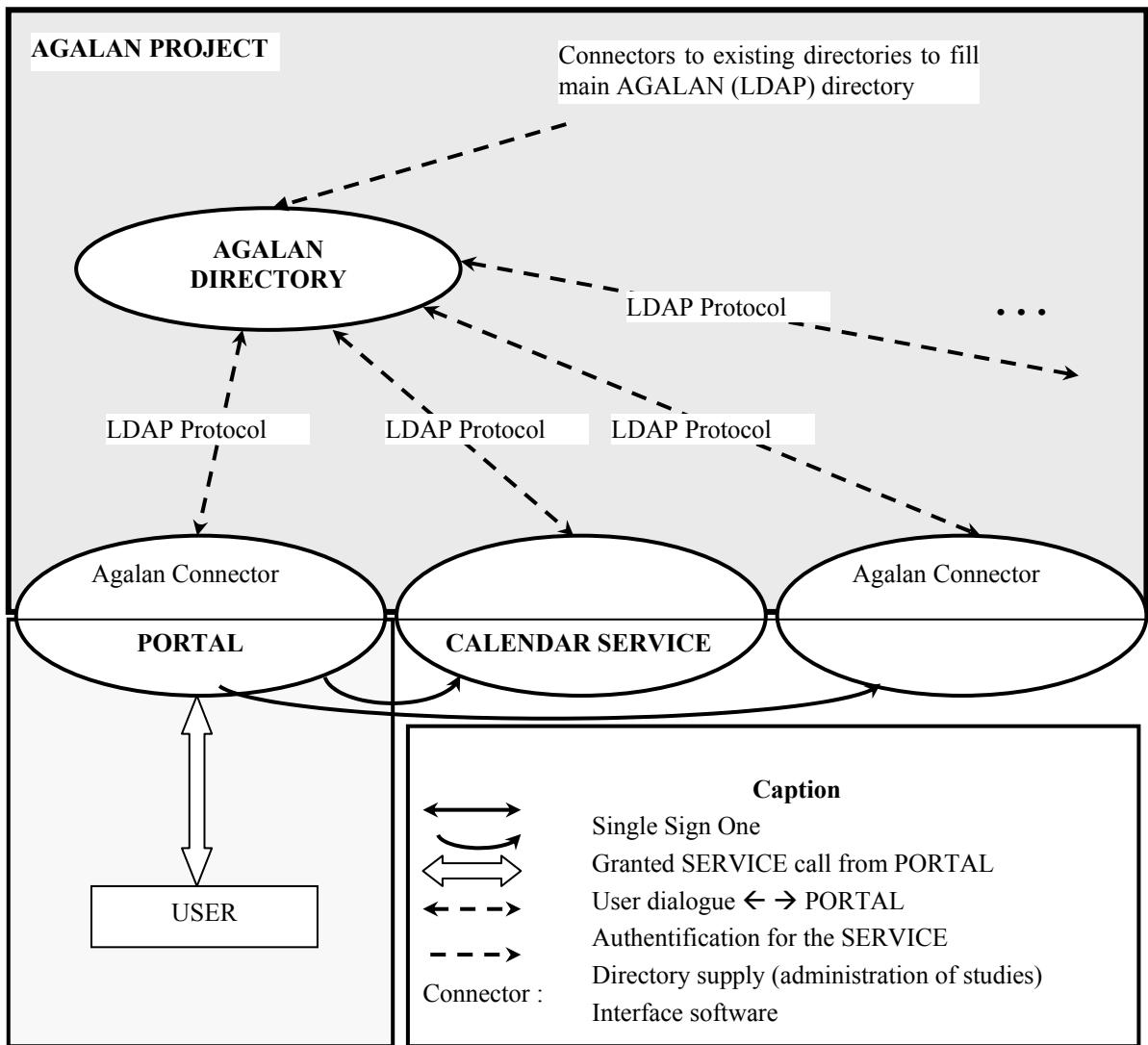
Figure 2 : Structure of the AGALAN sub-project

All of this implies the existence of a group concept. In the AGALAN sub-project we find two categories of groups:

- groups connected to administrative organisation of training.
- groups connected to a short-lived educational organization (for example, 6 students who work on the same project during 5 weeks)

At the virtual desktop level, the system provides drag and drop functionality or even more sophisticated services to deposit or get information into or from these shared folders which are implemented and managed on a set of servers.

Email

Email is an integrated on line service. The arrival of a new mail in his/her mail box is notified to the user. SSO mechanism enables message delivery from the mail server to the desktop. The mail server is one component of the user profile. Furthermore, we use a WebMail toensure the security unlike mail programs used from a desktop computer (Eudora, Outlook).

Targeted announcements

The portal provides a targeted announcement utility that can be used to send messages to groups of users. Messages can be sent to a user's email address or configured to appear on the user's default page depending on the subscription made by the user. Messages and events announcements can be targeted and sent according to attributes such as group membership.

Calendar

We include integrated individual and group calendar applications. Users can create and manage a personal Web-based calendar. They are also integrated with the administration database to populate the individual's calendar with course data and events. Users also have access to class calendars that are created automatically by the system.

Group calendar applications are supplied with educational activities data such as training schedules or

room distribution. The final calendar may be the superposition of both individual and group calendars.

Chat, forums

We find included a chat application to create a dedicated secure room. The chat room acts as a forum where the teacher or leader can conduct synchronous, on line discussions with students and other participants. In a chat room, participants type their comments into a portion of a window that is visible only to them. Participants post their comments to a message window, which is visible to every-one accessing the room. Each comment is identified with the author's name. We envisage a system allowing the user to send comments to selected participants without involving the entire chat room (management of sub-groups).

The communication will be implemented using available standard tools and provided protocols for broadcasting.

Distribution lists

In addition to mail, calendar, chat and message, the virtual desktop provides distribution lists. For example, an email list allows the training staff to send messages to an entire class or to select members of a class without having to manually create and maintain the list.

Office tools

Different office tools will be included to provide services, such as text editing, slide presentation, electronic sheet, drawing and imaging, etc.

Page hosting

We need content management tools to enable groups and individuals users to establish personal home pages and to maintain Web information without learning programming skills.

4.3   Student life application

4.3.1   The objectives

The objective is to provide a suitable reply to the questions (prospective and current) students may ask. A student will find helpful tips here about the various formalities to be completed, daily life, lodging and leisure activities. This application has also to ensure a common presentation of information and services to students.

When a prospective student wants to choose a university , information and services provided in such areas may be as important as the course topics or training quality.

Student living conditions and good integration in social life are very important to succeed in one's studies. For example :

- foreign students must be aware of the necessary formalities both in Grenoble and before leaving their home country (resident permit, insurance etc.)

- for accommodation, students may have the choice between rooms in a university, private flat or private lodgings. They need to know the prices, geographical location and level of comfort.
- for leisure activities, we can help the student visit the cultural attractions of Grenoble and surrounding area (e.g., details of films or theatre schedule) and make life on the campus more pleasant.

4.3.2   Information and set of services

Data is stored in different data bases. Data include text, images, links or data from external applications. In other words, we want to 'feed' off data that sit in external databases. The portal aggregates and summarizes information content for users. Information is associated to an author name, to an updating procedure, and timelessness management. We need docflow mechanisms to manage all those contents and their life-cycle.

To achieve that goal the system relies on basic mechanisms already mentioned, such as, for example:
- updating data according to the user' s rights.
- allowing an administrator platform to create new headings , to grant rights to users etc.

General information

The user finds information about administrative formalities, health, sports, leisure, culture. The system provides the author in charge of the information update with personalized access and an any easy to use graphical interface suitable for occasional users.

Clubs and societies

The users has access to a list of clubs and societies. A specific database is created. The administrator of the database can read, and update associated information. Access is granted to clubs allowing each one to update its data.

Current events

There are three user levels. The administrator creates section headings. He grants rights to a writer.

The writer cannot create a section but can add and update information in a given section. The final user can sort events by theme, location, date, price (free/not free), by name, etc. It is possible to combine all those various criteria. The calendar regulator is informed when an event is added or modified.

Forums

The application provides thematic forums. They are created by the administrator. Nevertheless, the administrator can give certain people access to utilities that will allow them to execute some tasks as long as they act in accordance with the rules expressed in a "good usage" chart.

Classified advertisements

Data is acquired using an online form, with criteria such as theme, advertisement shelf-life, etc.

Access may be granted to "ordinary" users in order to publish advertisements or to create forums after agreement with a moderator.

Usually, for all the above-mentioned applications, although the software allows the system administrators to perform every task, some of these tasks can be delegated to other groups of people.

5   Project outcome

This work enabled us to respond to the call for tenders for the digital campus 2002. Four out of the sixteen submitted projects were selected, including our own, known as ENCORA. This project has been developed in partnership with Sun Microsystems and Everteam (a software house specialized in documentation and knowledge engineering).

For several years now the government has been actively supporting French higher education establishments in their drive to take on board the advantages of information and communication technology. This policy is part of the common aim of the EU countries to build a European higher education network. Through a series of four-year contracts, the education and research ministers have strongly supported the projects developed by each university.

The 'French digital campus' tender complements the above projects. It supports inter-university projects which are open to international partners and the world of industry. This principle of cooperation is necessary in order to share skills, guarantee a high standard of quality, pool results and obtain recognition both nationally and internationally.

The principal objective is for students and staff of higher education establishments to have access via a digital working environment to services and content, some of which are already available while others have yet to be developed. For example, a digital working environment can allow the user, through computer networks (principally the Internet):

- To have a digital desktop personalised according to his/her profile and activities: e-mail, videoconferencing, diary, address book, document storage, tools to produce text or multimedia documents, shared working environments, etc.;
- To personalise the appearance of the interface;
- To use personalised search tools;
- To have access to or produce information relevant to the teaching process (access to teaching and documentary resources, exam results, lecture notes, etc.);
- To have full access to information concerning the type of training offered and the inscription procedures;
- To enrol without having to be physically present in the administrative offices;
- To present the various elements of the course in an organised and efficient way.
- To have online access to relevant services;

- To perform cooperative work with other users (students, teachers, ...))
- To have online access to information concerning 'student life': culture, sport, leisure, career guidance, job-hunting, etc.

The target users are all students, including both regular and distance learners, teachers and other staff.

ENCORA will be a step within the comprehensive digital workspace project, which aims to define and implement the basic infrastructure of a digital working environment.

Furthermore, the implementation of the project is described in the article (or survey) "Organisation et Architecture de l'ENT ENCORA" [8] presented at JRES conferences.

At the moment a piece of software is being developed which will present the organisation of the teaching element of the digital working environment. It will be structured around the three-stage progression: bachelor's degree, master's degree, doctorate. It is being developed within the framework of 'Universities Developing Digital Services for Students', as outlined in the call for tenders. It will also respect the various European directives, as defined within 'Building a European Environment for Higher Education', in order to give a European scope to the universities' projects.

The three-stage model present in each university provides the basic structure for all teaching in every subject area represented within the establishment.

For pedagogical reasons, the Digital Workspace should reflect this structure and should therefore be organised accordingly.

6   Conclusion

The portal project is intended to give adaptable results, and thereby to function as a pilot project at national level (a national call for proposals has been issued on that domain). With this in mind, the diversity of the five establishments involved (a science university, a group of engineering schools, a multi-disciplinary university, a university of social sciences and an arts and humanities university) and the number of students are important assets. To provide the means to develop the portal is to ensure prominence on the national - and international - landscape.

References

[1] Portail étudiants des universités de Grenoble Analyse des besoins, Projet GreCO PS8, Septembre 2001.

[2] http://greco.grenet.fr

[3] Web Content Accessibility Guidelines 2.0 J. White, W. Chisholm, G. Vanderheiden, 24 August 2001, http://www.w3.org/TR

[4]   Schéma Directeur des Espaces numériques de Travail, Ministère de la jeunesse, de l'Education nationale et de la Recherche, 17 Mars 2003

[5] Campus Numériques : Appel à Projet 2002 ; Objectifs et Modalités, Ministère de l'Education nationale et de la Recherche

[6] Campus Numériques : Appel à Projet 2002 ; volet 2, Ministère de l'Education nationale et de la Recherche

[7] Projet AGALAN
     http://www.agalan.org

[8] Organisation et Architecture de l'ENT ENCORA Christian LENNE, Dominique MERLE, Stéphane PICHEVIN JRES 2003, Lille, 17 – 21 Novembre 2003 forthcoming

ICE – a Web-Based Information System to Support Higher Education Policy Decisions

Peter Müßig-Trapp, Hans Dicken and Helena Kopp
HIS Hochschul-Informations-System
Hannover, Germany
muessig@his.de, dicken@his.de, kopp@his.de

ICE stands for Information, Controlling, Entscheidung (in English: Information, Controlling, Decision) and is an information system developed to support higher education policy decision-making which has proven itself in practice. The system is currently in use at the German Ministry of Education and Research, at science and research ministries in ten German federal states and at other organisations active in the field of higher education policy (eg, the German Science Council and the German Academic Exchange Service)

The debate on protecting personal data which arose in the 1980s led to increasing sensitivity in the population regarding data protection questions and to stricter data protection rules. Both resulted in fewer possibilities of statistical data analysis being allowed: In many fields, statistical analyses can only make use of so-called aggregated datasets. With its ICE information system, the HIS now provides a solution which is capable of extracting a maximum of information from fundamentally limited aggregated datasets. At the same time - subject to their appropriate availability - the system also allows the analysis of individual case data records.

The following outlines the main system features. ICE

- *is a web application, is Java based, platform independent (for example, backend runnable under Windows, Linux, SUN OS), and database independent (eg, Oracle, Informix, MySQL),*
- *offers very high data import and data analysis flexibility,*
- *uses XML technology and the Apache Cocoon based ICE Publishing Framework, which means that it provides a wide range of output formats (XML, HTML, XHTML, Excel, Gnumeric, PDF, etc), simple data exchange with third party programs, and is capable of handling future technologies.*

has further developments planned in the direction of data expansion, internationalisation, and towards mobile computing (functional expansion of the existing cellular version into a handheld version, ICE-Mobil)

1   History

At the start of the 1990s, ICE evolved as a commissioned project on behalf of the German education and science ministry[1], which later merged with the research and technology ministry[2] to form the BMBF. At the time, the ministry used the Macintosh operating system, which explains why the ICE system was originally developed on the proprietary HyperCard basis. So the application was only available under Mac-OS. The mid 1990s saw desktop environments migrate towards Microsoft Windows. This called for a new development. This need was used as an occasion to completely re-engineer the system architecture and the technical framework. An Intranet system was developed based a modern multi-tier-architecture, with a database at the backend and Java

as its central development platform. This made the system accessible with an Internet browser, regardless of platform. A Content Management System was added as an extra component to visualise the structures and procedures within the participating ministry departments and served to support general document exchange.

The year 1997 saw the system adopted by the German Science Council albeit with a specifically defined data basis: System architecture and software components were very largely adopted direct; only the data basis (the backend module) was changed. Analogue extensions were undertaken in subsequent years for the science and education sections of the relevant ministries of the German federal states [1] (protected Internet access) and for the DAAD [2].

Changing and expanding user requirements resulted in the system being continually carried forward. For example, the presentation and export options were greatly enhanced and extended for the BMBF. XML-

---

1   Bundeministerium für Bildung und Wissenschaft (BMBW)
2   Bundesministerium für Forschung und Technologie (BMFT)

based modules were developed to allow the ministry to automatically generate the layout for its annual book of tables "Basic and Structural Data"[3] direct from the ICE database. The needs of the federal states called for user authentication and a system of rights to provide finely-structured access control to the modules in the system and especially to the user-generated analyses.

ICE can also be understood as a response to the restrictions introduced on the analysis of statistical information at the end of the 1980s. While access to official micro data had been managed quite liberally for science, education and administration up until the end of the 1970s, extremely restrictive access rules were introduced against the background of the data protection debate which set in at the beginning of the 1980s. At the time, an occasionally controversial political and legislative debate on the planned census and on the protection of personal data raged in Germany. In the wake of great public sensitivity, the existing legislative provisions were interpreted restrictively, while further hurdles were established by introducing additional rules. This is why administrators analysing statistical data cannot, as a rule, access individual case data; they can only access so-called aggregated datasets.

Besides the individual university administrations, only the offices of statistics are allowed to analyse official individual case data in the higher education sector. However, these offices regularly publish their standard analyses (aggregated sets). Composition and structure of these aggregated datasets were defined in time-consuming negotiations with advisory bodies and, as a rule, have now been set for many years. The offices of statistics are able to meet the information needs that can be satisfied by these standard aggregated datasets relatively quickly; however, any structured analysis requests which differ from these can - depending on the workload and technical specifications involved - only be made available, if at all, in the form of special analyses (as a rule, assembler programming) after quite a time delay.

The range of analysis options allowed by standard aggregated datasets are comparatively limited and, generally, allow only precisely *those* analyses to be made which had been the intention of the previously-defined respective aggregated dataset. It is absolutely impossible to completely predict the information needs of politics for even just a few months, let alone over years. This is why it was obvious that long-term planned standard aggregated datasets and the answers they contained to questions that had been anticipated years before would only be able to partly satisfy the higher education policy questions which arise now. This is why there was and still always is a wish for making individual case data records available for statistical analyses.

Recently, a development has been observed in Germany which may be viewed - depending on viewpoint and subject area - either as a relaxation of the rigorous approach to data protection questions or as a crisis in the data protection field. A wide range of

differing causes underlie this development, of which overregulation by rules, decrees and laws (data protection has become a domain of the lawyers) is certainly an important one. The increasing willingness of far-sighted players in the field of data protection (eg, the Lower Saxony Data Protection Officer) to allow the use of statistical analyses from individual case data records for academic and administrative purposes, as long as the individual's right to informational self-determination has not been impacted, is a consequence of this development.

In the face of this background, a major new development was started on behalf of the Lower Saxony Ministry of Science and Education[4]; this project is currently in its implementation phase: An ICE with deeply-structured data for individual federal states which not only allows the regularly used standard aggregated datasets to be processed but now also allows very broad aggregated records and also individual case data records to be processed.

Therefore, ICE provides the possibility to process standard aggregated data as well as individual case data. A combined analysis of both kinds of data is also comfortably possible. This is important, because huge portions of higher education statistical data still will only be available as aggregated data in the future.

2   Goals

ICE was and continues to be developed with the following five central goals in mind: Firstly, to make available a system with which the maximum of possible analysis options can be achieved from essentially restricted aggregated data records. Secondly, and at the same time, to develop a data warehouse, that is a system with which the widest possible range of different data stocks (with various depths of structure, quality levels, presentation options, etc) can be integrated and flexibly analysed. Thirdly, to make the system platform independent in order to avoid - as has already happened in the past - a change of a customer's operating system or central database resulting in large sections of the system having to be redeveloped. Fourthly, to make the system easy to operate with the minimum possible need for computing skills or detailed data knowledge. As far as possible the necessary know-how should be contained in the system itself, so that the user - as far as possible free from technical and formal considerations - can concentrate on formulating data content requests. This means that the target group for the system not only extends to experts well-versed in computing, statistics and data content, such as specialists from the ministries of science and education, but also to executive-level decision-makers. Fifthly and finally - and this goal was only added recently - to provide a system which also allows the quick analysis and interpretation of comprehensive individual case data as well as combinations of individual case data and aggregated data records. The background to this is formed by the cautiously growing willingness on the part of German

---

3   Grund- und Strukturdaten

4   Niedersächsisches Ministerium für Wissenschaft und Kultur

data protection officers to allow the analysis and interpretation of anonymous individual record data.

These principle goals have not yet been fully achieved in all areas. However, the system is being continuously developed. Focuses and priorities are set on the basis of the needs of the customers who are financing the development. The current state of development and so the performance spectrum of the system will be outlined below.

3   System features

- *Web application.* Access to an ICE installation is enabled via a network using a Java capable web browser (such as the open source browser Mozilla, Netscape Navigator or Microsoft Internet Explorer). In principle, this means that the system can be accessed from any computer on the Internet or Intranet which is registered with the ICE server. Access to the system (or to parts of the system) can be restricted to authenticated users as required. Where necessary, the system (or parts of the system, eg, collections of standard tables) can be set up to allow access from the Internet.

- *Platform independence.* On the server-side, the system can be installed under Microsoft Windows as well as under Unix (Solaris) and Linux. The client end (user) only needs a Java capable web browser. Browsers are available free of charge for all the commonly-used platforms (such as MS Windows, MacOS, Linux). The system is also independent in terms of the relational database management system that is chosen: The system has also been installed under Oracle and Informix. We are currently testing the use of open source databases (MySQL, PostgreSQL).

- *Flexible data import.* Data with any structure and depth of structure can be imported. The system can also be expanded to include new topics. Since very recently, it has been possible not only to analyse and interpret aggregated data but also comprehensive individual case record data with good performance. Similarly, the combined analysis of aggregated and individual case record data is possible.

- *Flexible data analysis.* The very flexible import of data stocks is mirrored by an equally flexible range of analysis and interpretation modules. Using the so-called "flexible table generator", data stocks available in the system can be used to output any extract in tables. Analyses using information from various data stocks can also be easily requested: It is no problem combining information from several data stocks in a single results table.

- *Flexible data export.* Results tables produced with the flexible table generator can be stored in HTML and MS Excel format. This makes it possible to process the tables using third party programs, to pass

on statistical information to other interested persons, eg, by e-mail, and to create information and data collections on the web. The data export options were extended substantially in 2002. The ICE Publishing Framework provides a tool which makes additional output formats available. On the one hand, this contains an XML interface which can be used for data exchange and as a universal interface to third party programs (eg, other databases, spreadsheets, graphics programmes, geographical information systems, and so on). On the other hand, this format is suitable both for web-based presentations as well as for high-quality print-outs. The user can influence format and appearance of the PDF output in many different ways.

- *Data harmonisation with an integrated key*: All the data contained in the system are encoded with a uniform ICE key. The project team centrally updates and hosts the key. This ensures that - as far as meaningful for the content - various stocks, also possibly from various sources, can be analysed together. Equivalency rules are defined where necessary to make possible comparison of variously encoded data, which do have like content however, (example: subject groups in the staff statistics ←→ subject groups in the student statistics). The system can also recognise key internal hierarchies and places knowledge at the disposal of the user for carrying out sorting functions, for example. (The system "knows" that the University of Hannover, for example, belongs to the state of Lower Saxony and to the higher education institution type "university").

- *ICE standard tables (with integrated automatic updating).* All results table produced using flexible table generation can be stored as so-called ICE standard tables in the standard table collections. These table collections can also be made accessible to third parties on the Intranet or Internet and can be searched both by a hierarchical directory structure as well as by a keyword search. The integrated automatic update is a particularly useful feature of the ICE standard tables: A once generated table which has been stored as a standard table can automatically be updated at the touch of a button with data imported into the system at a later point. The user can choose from various update options (eg, time series addition, time series shift, substitution of the whole table with the latest available data). Using the ICE Publishing Framework it is possible to request these standard tables in various formats. At present, the following formats are available: HTML, XML, MS-Excel, Gnumeric, PDF. The XML-based technology (Apache Cocoon) we used means that the creation of corresponding XSLT style sheets allows other output formats to be made available with little effort. Just as unproblematic are the user-defined modifications to the output format (colours, document size, fonts, etc) which can be produced. This is relatively easy

because all output layouts are produced on demand and on-the-fly and the relevant information is extracted from a database; it is not necessary to fall back on prefabricated files.

- *Reliable and reasonably-priced data updates.* The import of new data stocks and updates of existing stocks is done by the HIS-ICE group. Central system maintenance, data processing and data administration assure a high degree of data quality and reduce costs by producing synergies (eg, keys developed for one customer can be modified for another customer).

- *Hotline.* The ICE group provides advice and support on all questions relating to the information system (both by telephone and by e-mail). This service includes technical questions (network problems, security settings) and operating questions on the software (browser, ICE application) just as it does subject-related/statistical questions (regarding data and their analysis, keys, and other aspects).

- *Development carried forward all the time.* The development of the ICE is being carried forward all the time. Upgrades and improvements produced for a particular customer will be made available to other customers a short while later.

4   Data

In general, the system can process practically all kinds of data; the only condition lies in the development of an appropriate key for the classification and integration of that data. Data encoding with a uniform ICE key means that different data stocks, possibly also from various sources with differing levels of data quality, can be analysed together. Each data stock is registered with the help of the ICE key as far as data source and data quality are concerned and can be requested for analysis and interpretation. At present, the system largely contains data from the Federal Office of Statistics[5]. However, data from other suppliers - eg, German Science Council, BMBF, KMK[6], BLK[7], Bundesanstalt für Arbeit[8], HIS, EUROSTAT, OECD, UNESCO and a number of others - can be integrated into the system and analysed. Appropriate data were, for example, integrated into the system for the implementation of the BMBF's Basic and Structural Data publication. The "data quality" feature provides information on the credibility of data (final or preliminary data from the official statistics, random samples, forecasts, etc). This facilitates an extremely high degree of flexibility as regards the provision and

processing of statistical data from various sources/systems. It satisfies both the need for reliable and generally comparable data from the official statistics as well as for the very latest corresponding data from other sources.

The system can import data of any structure and depth of structure. The system can also be extended flexibly as far as topics are concerned. The content focus lies specially on higher education statistical data: University staff, student and examination/degree data, holders of higher education entrance qualifications, university funding statistics. At present, the system is being expanded to cover a wider range of educational statistics. Besides higher education statistics, the system is also capable of analysing and interpreting data on kindergartens, schools and vocational training, continuing education, resources for education, science and research, educational assistance, etc (as implemented in the ICE used by the BMBF).

To ensure that the user does not have to concentrate too much examining the various statistics and data in detail, intelligence has been implemented into the system itself at many points. For example, if the user wishes to generate a table which contains both student numbers as well as the case numbers on university staff, each arranged by subject groups, then the user should know that the subject group classification in the official statistics differs for both cases. Equivalency rules are maintained in the system for such comparison of variously keyed but - in terms of content - comparable data. They make it possible - completely transparently for the user - to assign the corresponding data from differing statistics and so facilitate the integrated analysis of different data.

Besides equivalent relationships between ICE keys, the system also has so-called implication rules which reflect a large part of the key contexts. For example, implications determine to which federal state or which type of higher education institution an individual university belongs. This knowledge is placed at the user's disposal in certain cases, eg, for carrying out sorting functions.

5   The system from the user's perspective

The ICE user accesses the various functions of the ICE system from its start page. Two methods are available for searching for data in the ICE. On the one hand, the available, updated and maintained *ICE standard table collections* can be used. ICE standard tables offer frequently needed statistical data in easy to use and quickly accessible formats. On the other hand, data can be quickly requested from the database with the help of the central module known as *table generation*. In addition, the start page contains tools to manage the tables, to update the standard tables as well as lists of the available data stocks. All these functions are protected against unauthorised access by an authentication system. The following will briefly outline the table generation process and its management.

---

5   Statistisches Bundesamt
6   KMK = Kultusministerkonferenz (Standing Conference of the Ministers of Education and Cultural Affairs of the Länder of the Federal Republic of Germany)
7   BLK = Bund-Länder Kommission für Bildungsplanung und Forschungs-förderung (Federal-State Commission for Educational Planning and Research Promotion)
8   Federal Labour Office

5.1 Table generation

In principle, it is possible to differentiate between two kinds of table generation: *Single data stock* and *multi data stock table generation*. Single data stock table generation (Table Generation I) can only be used to produce tables whose data stem from *one* predefined virtual single data stock within the ICE. A single data stock is a given basic stock within the system of matching data (eg, a time series on students and study entrants, categorised by several features). The particular advantage of single data stock table generation lies in the fact that only those data can be combined with each other which also "fit together". So, it is not possible with this tool, for example, to combine staff data (staff, staffing positions) with data on teaching demand (students, study entrants). So this rules out the possibility of making subject-related mistakes, because such data are by definition not contained in a single data stock. This safety leads to less flexibility than the multi data stock table generation (Table Generation II) allows, because the latter allows several data stocks to be combined with each other in a table presentation. Single data stock table generation represents the standard form of system usage, because it requires no in-depth specialist knowledge on the subject area to be reflected by the data.

The *Table Generation* window has a menu bar (with the following index tabs: "Keywords", "ICE Data Stock", "Data Stock Selection", "Table Definition" and "ICE Key"). These index tabs are sequentially processed from left to right up to the table definition tab to produce a table. Essentially, a table is produced in three steps.

Users:
- select suitable keywords from a list of keywords to circumscribe the desired topic area,
- select a data stock from a number of suggestions, and
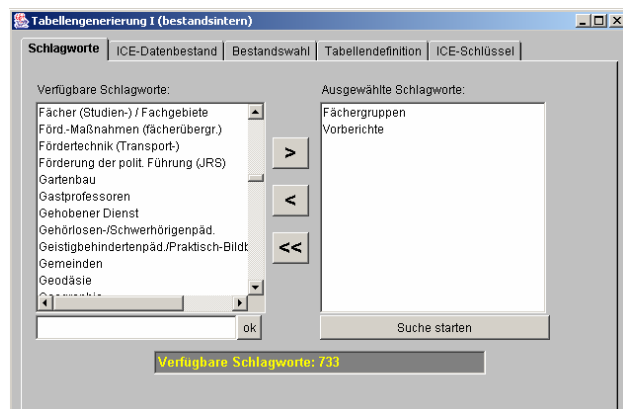- define the table layout and table content.

Figure 1: Window for selecting keywords

Keywords reflect the whole stock of features and categories in the ICE key. There are keywords which reflect a feature within the ICE key (eg, *sex*) and keywords which relate to a form of the ICE key (eg, the form *male* of the feature *sex*). Once the keywords have been selected, the system lists all the data stocks which

contain information on the given keywords under the menu item *ICE Data Stock* including a short description. In the following window ("Data Stock Selection") the selected data stock is once again described in detail along with the available forms. It is easily possible to change the data stock selection.
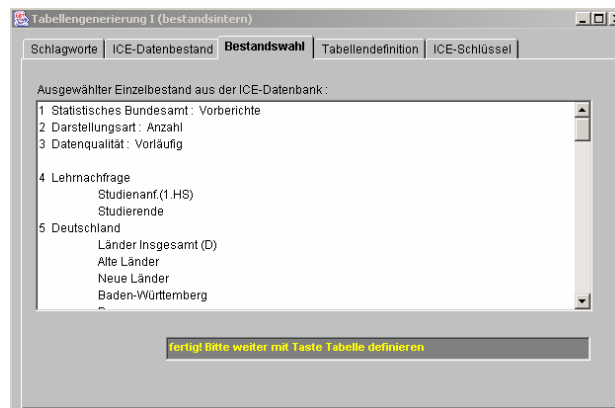

Figure 2: Window for selecting a data stock with detailed data stock description

To *define the table* the user determines in the next window (index tab "Table Definition") which features along with which values are to be assigned to columns and which to rows.
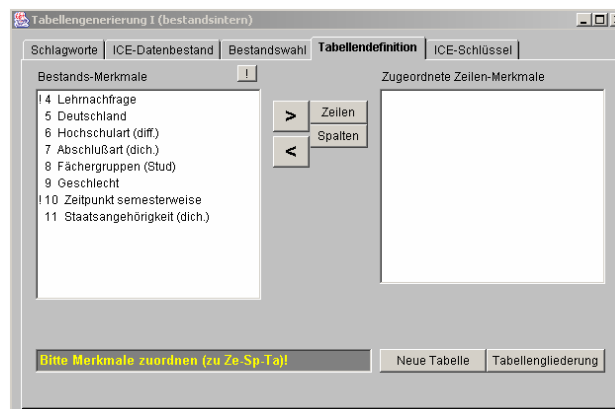

Figure 3: Window for defining the table

It is possible but not necessary to use all the features offered. If a feature is not selected (eg, sex) then the tables will contain values for the respective "Total" ("male" *and* "female"). Some features have no meaningful "Total" (eg, timepoints "1998", "1999", "2000"; …). In such cases, the user must make a selection. The order of the selected features corresponds with that in the later table. It can subsequently be modified by shifting, and it is also possible to remove features and, when so desired, to reselect them in a different order. In the case of extensive features, the system offers a sorting function; this allows university towns, for example, to be sorted by federal states and/or types of higher education institutions. The user can check

the selected values at any time via the table structure function.

With the help of the *Table Generation II* function it is possible to produce tables whose data stem from various single data stocks. For users, this ICE function is like generating a series of tables under version I. A table produced with the help of a multi data stock generation is made up of a series of table sections which have each been individually defined. The structure of such a table corresponds with that shown in Figure 4 below.

Generation of this table begins with "Table Section 1/1". This step automatically labels Row 1 and Column 1 accordingly. The second step then adds "Table Section 1/2" (with given Row label 1) or "Table Section 2/1" (with given Column label 1). All other table sections are then added in the same way (Table Sections 2/2 and 2/3 follow automatically and do not therefore need to be defined). Any number of further extensions of the table are allowed, which means that columns and/or rows can be added and extended at any time. Since several table sections are merged to form the overall table with in some cases identical labels, it is particularly important when using this version of the ICE to think through the table structure especially thoroughly beforehand.

The table can be generated once the table structure has been described the desired data are requested from the database and the table assembled in accordance with the table definition. Beforehand, a window indicates the number of requested values and enables the user to correct the table definition if the data should turn out to be too extensive.

	Column Heading1	Column Heading 2	Column Heading 3
Row Heading 1	Table Section 1/1	Table Section ½	Table Section 1/3
Row Heading 2	Table Section 2/1	Table Section 2/2	Table Section 2/3

Figure 4: Table structure after multi data stock table generation with several table sections

The multi data stock generation of tables allows subsequent *data calculations to be carried out*, eg, percentage calculation, indexing, quota formation and difference formation. Finally, the user can output the request result in various formats (eg, Excel, HTML), or also store it on the sever as a so-called standard table.

5.2   ICE standard tables

Each table produced with the Table Generation II method can be stored as a standard table. This means that each user can compile personal standard table collections. To store the request results as a standard table the user only has to click the appropriate button in the generation window.

Figure 5: ICE table with typical calculation

Besides this, there are also the *standard table collections* which ICE makes available, maintains and administers, and which can be viewed under Directory on the start page. The structure of the topic areas allows quick access to the desired tables. The output of these tables is possible in various formats: *HTML, XML, Excel or PDF* and opens up a wide range of further processing and flexible data export options. Moreover, the central ICE collections as well as the collections compiled by users can be searched using a *keyword search* option.

ICE standard tables have a special format which allows them to be *updated* at the "touch of a button". This activates a special tool which automatically updates and extends the table. At present, three update types have been implemented which are described in further detail in Section 8 of this paper.

5.3   Standard table administration

This tool can be used to compile, rename or delete collections. The option of allocating reading and writing rights makes it possible to completely block collections for third parties or to allow them to read or to collaboratively process collections. Standard table administration thus makes it possible (if not blocked by the "rights allocation") to access all available standard table collections. This means that ICE standard tables can be used as a basis for individual table collections and can be further developed.

A number of functions are also available for individual tables. They can be deleted or copied into other collections, the table title can be changed or headers and footers added. Metadata from individual tables, such as topic area, data source or "rights allocation" can be read or changed using this tool, such as the type of updating.

6   System architecture

ICE was originally developed on HyperCard basis which meant that it was only accessible for computers running the MacOS operating system. Files were stored in the file system as were the metadata which described the corresponding file contents. Data were imported into the system using special programs which converted these

from the supplied formats into the uniform ICE format and correspondingly adapted and complemented the metadata. The import routines were largely written in the programming language Fortran and implemented using Unix-Shell-Scripts.
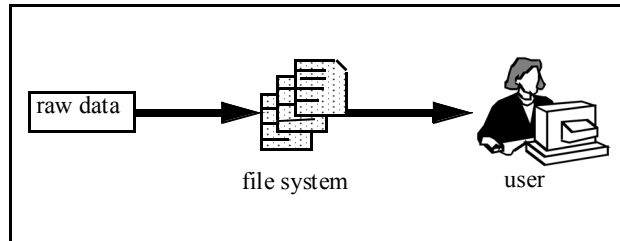

Figure 6: HyperCard-based architecture

Following the restructuring of ICE into a web-based architecture, the import mechanisms were largely adopted as a first step, because the specifically-developed routines contain numerous correctness and plausibility tests on the raw data. The new architecture now takes the data which have been checked for plausibility and transferred into the file system and imports them with the help of new developments into a relational database, where further tests (in most cases technical) are carried out on the data.
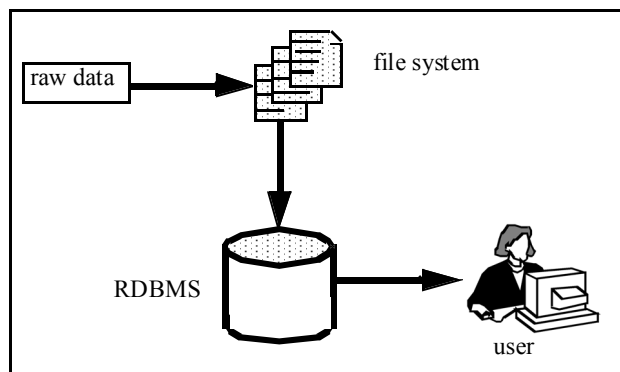

Figure 7: Use of relational databases

The system was designed as a Multi-Tier-System, whereby an application server located between the database backend and the application level (Java Applets) converted abstract requests received by the system into requests which the database could understand, transferred these to the database, processed the results and transferred them back to the user. Apache was used as a HTTP server, communication between the applets and the application server was originally handled by low level socket connections which proved to be comparatively simple and stable.

Recently, this was changed over to HTTP in connection with Java Servlets. The Tomcat Server developed by the Apache Group has proven reliable as a Servlet Engine. The changeover to HTTP communications proved advantageous, especially in the context of the more recent XML developments for the conversion of request results into various file formats

(eg, PDF) with the help of publishing frameworks (see Section 7). The changeover was also needed after the state ministries received a system in the form of ICE which had to be accessible on the Internet, ie, beyond the bounds of agency or company-wide Intranets. The firewalls for the ministry networks, which were generally restrictively configured for security reasons, did not allow communication via sockets. This would have required additional ports to be opened which, as a rule, would not have been permitted or, if so, then only after a great deal of persuasion.
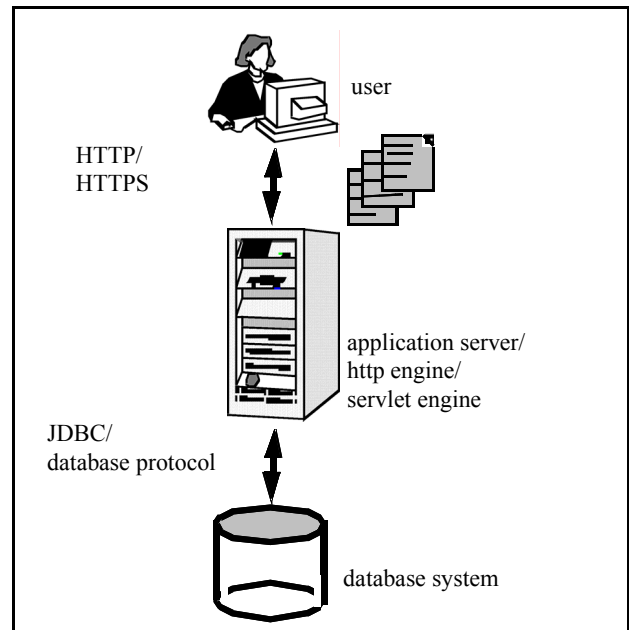
Figure 8: Multi-tier-architecture of ICE

The decision in favour of Java as the development platform was made as early as in 1996. One goal in connection with the multi-tier decision consisted of shifting as many application tasks as possible from the client to the application server in order to, in particular, achieve better system scalability. The programming language Java from Sun Microsystems had hardly been tested in practice at the time; looking back, however, we can say that the decision to go for this environment was right, and not only in terms of platform independence. Indeed, Java has meanwhile developed into one of the most important programming environments [3].

However, the use of applet technology failed to relieve the workload of the client resources ("thin client" principle) to the desired extent, even though, instead of the more modern swing technology we chose an even leaner implementation with the help of the AWT (Abstract Window Toolkit), also because no more advanced technology was available at the time of the implementation of the client components. The presentation of request results in table format was realised with the help of the JKit/Grid class library from Objectshare. At the time of the relevant developments, this proved powerful enough to also depict interleaved tables, for example.

More recent considerations go in the direction of using two different requirement levels:

Thin Clients, especially for ICE Internet applications: In such environments it is only possible to expect users to meet low hardware requirements, because these are hard to "check", since the users are more or less unknown. In such cases, Java Applets will be replaced by HTML/XML solutions, used on the server-side by Servlets, Java Server Pages (JSP) or Extensible Server Pages (XSP). This means that no Java code is run on the user computer and, consequently, there is no need for pre-installations on user computers.

Modern Applet technologies using swing components, especially for Intranet applications. The circle of users is manageable in such environments and this is why certain hardware requirements and pre-installations can be expected of users (eg, a modern Java run-time environment). This makes it possible to run very powerful libraries for the presentation of the applications on the clients.

Communication between the database level and the application level is managed by JDBC-API (Java Database Connectivity). [4] This makes it possible to incorporate SQL commands into the Java program components which are understandable for relational databases and to transmit these to the database; it also makes it possible to receive and process request results. A major advantage of this solution was seen in the fact that (insofar as standards are adhered to in the development) the database systems of various providers can be exchanged for each other practically without problem. Indeed, ICE is meanwhile used in production operations both with Oracle databases as well as with Informix and MySQL. [5] The interface between the SQL sections and the DBMS systems is made by so-called JDBC drivers which are available for all common relational database systems. Besides the direct translation of SQL into database language, it is also possible to address proprietary database constructs, such as network components from Oracle or procedures stored in the database. However, we have only made use of these options when it was absolutely certain that porting to other database management system could be completely ruled out.

New questions arose in connection with the making available of ICE for users in various state agencies via the Internet. Such issues had not arisen to this extent in controlled Intranet environments. This applies, in particular, to data security/data protection and system availability. According to German law (from a data protection perspective), data within ICE do not require any particular protection, for they are not personal data. Nevertheless, economic aspects, aspects of state competition and other aspects would suggest the use of technical arrangements to prevent uncontrolled access to the data. This is why a group and role-based authentication and rights management system was designed and implemented which allows the finely-structured control both of access to the system and its various modules as well, and in particular, access to the

analyses and interpretations generated by users (ICE standard tables).

For web-based applications, the system was made highly available by introducing multiple replication and further measures, and its performance was raised by using a simply extendable server cluster. Replication takes place both at database level as well as at application server level. A scheduler serves to distribute the requests, depending on the current load of the individual server knots on the computer with the least load. For reasons of cost, this solution made use of the open source database MySQL. Together with self-developed applications and further open components (Apache, Tomcat, etc.) this meant that is was possible to achieve quite an economic solution.

7   ICE Publishing Framework

The ICE Publishing Framework is a system component designed for the dynamic generation of (table) outputs in various formats. It is run *after* users have defined the logical structure of their desired analysis and have extracted the data from the database. At present, users can choose to output the tables in one of the following formats: HTML, XML, MS-Excel, Gnumeric, PDF. Since the PDF format is especially suitable for the production of print templates, several standard output formats are offered (eg, DIN A4 landscape, DIN A4 portrait). The architecture of the Publishing Framework, described in more detail below, means that further output formats (eg, addition of a logo, change of font, table types, paper formats, etc.) can be made available as necessary with very little effort. In addition, a layout tool was developed to let the user influence the PDF appearance of each table in a wide range of different ways. Among the aspects which users can influence are the font, font size, line spacing, column width, page breaks, etc.

The ICE Publishing Framework is based on the XML application platform Cocoon developed by the Apache Group. Cocoon is free software, which means: The source code is available (open source) and may be changed and extended. Licence fees neither need to be paid today nor in the future. Cocoon was chosen because it incorporates three key principles of modern software architecture:

1. *The strict separation of data, logic and layout*: Such separation makes work sharing easier in the development, maintenance and extension of the software. This separation differentiates Cocoon from other technologies such as Active Server Pages (ASP), as well as Java Server Pages (JSP).

2. *Consistent use of XML:* XML is an internationally-recognised standard used and supported both in the open source sector as well as in commercial fields, and has meanwhile asserted itself across a broad front. The main areas of application for the mega
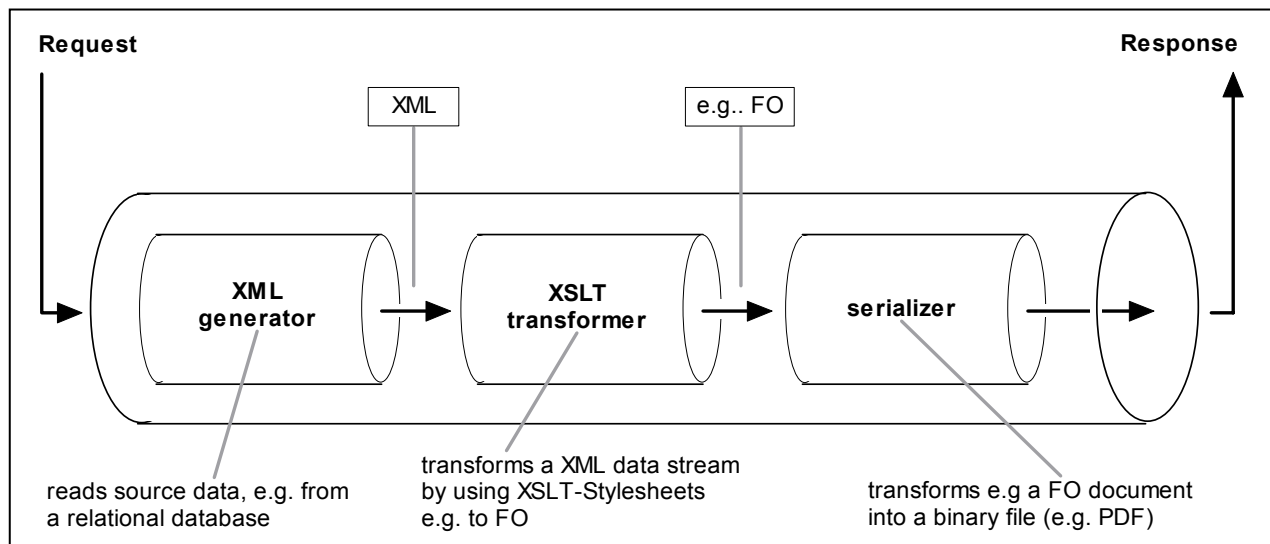
Figure 9: Cocoon XML-Pipeline

language are integration tasks as well as data exchange between applications and companies. The standardisation process for the key components has been completed (XML, XSLT and XML schema have successfully completed the W3C standardisation process and have the status of a "recommendation").

3. *Component integration*: All necessary XML components for the creation of a Publishing Framework are already contained in Cocoon, since it is possible to make use of the large collection of Apache projects. In ICE it is possible therefore to concentrate on the development of the application as such. For example, Cocoon uses Apache Xerces as an XML-Parser, Apache Xalan as an XSLT-Processor and Apache FOP for PDF generation. The architecture is based on the Apache project Avalon, a Java approach to component-oriented software development.

This has the following advantages for ICE:

- The Publishing Framework can easily be adapted to bring it into line with the existing infrastructure. Only very limited changes are necessary to the existing ICE.
- Very flexible, easily expandable system. Besides the presently available formats HTML, XML, MS-Excel, Gnumeric, PDF, it is possible with very little effort to create further formats (eg, WML for display on handhelds or mobiles/cell phones). For example, the generation of vector graphics or maps from an XML base file is also conceivable.
- Use of generally available standard software from the Apache Group (producers of easily the most frequently used web server worldwide).
- Use of international standards

Figure 10 illustrates the production process in the ICE Publishing Framework. The user request via a web browser, such as Netscape Navigator, Microsoft Internet Explorer or Opera browser ("Client"), marks the starting point. On a usual webpage, the user selects the XSP page (eXtensible Server Pages, called "XSP page: table list" in Figure 10) to be depicted from a list by using the table title for a table and the desired format. The one and the same XSP page is always requested via the eXtensible Server Pages. Although, possibly, thousands of standard tables are available in the system, they are requested via a single document ("ICE-Tabellen.xsp"). Upon receipt of the request, this XSP page, the table identification and the desired format are transmitted as parameters. XSP is an advanced development of the Java Server Pages (JSP). An XSP page is an XML document with dynamic content. This dynamism is achieved by means of directives defined in tags. In ICE, the XSP page calls up a program logic which, using the parameters communicated to the page, extracts the data required to present the tables (table title, remarks, footnotes, table values, etc) from the ICE standard tables database. Using the ESQL-TagLib it is easy to send appropriate SQL commands. The ESQL-TagLib is a further component of Cocoon. And so by accessing the database, the system gradually creates a well-formed XML document from an initially empty XSP page. The generation of native spreadsheet formats represents a particular challenge, especially the only rudimentarily documented Excel format. The realisation made use of the POI Library. POI (Poor Obfuscation Implementation) is meanwhile a firm component of the Apache Jakarta Project. In the form of the HSSF Library (Horrible Spreadsheet Format - programmers seem to feel a certain degree of dislike towards the Excel format which they reimplemented) the POI provides an implementation of the Excel file format which makes the writing of simple tables easy. POI is fully integrated into Cocoon via the HSSFSerializer.

Figure 10: ICE Publishing Framework
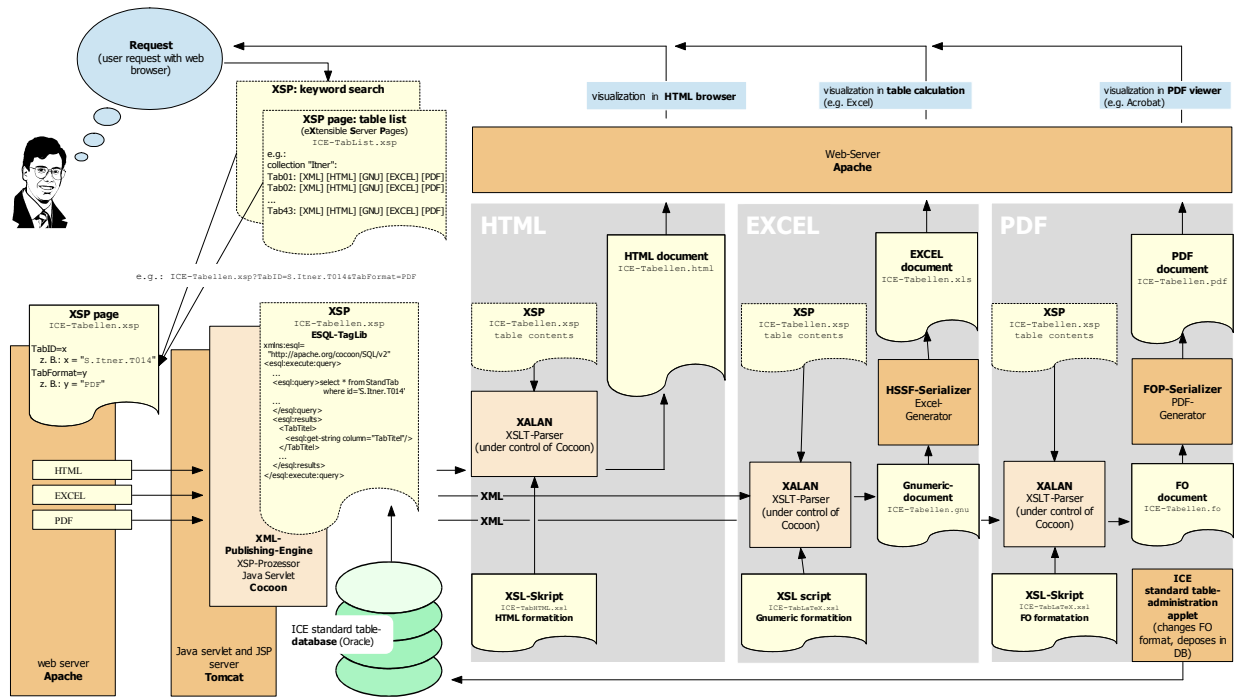
It makes it possible to output XML documents which correspond as Excel files with the Gnumeric Spreadsheet Format in Cocoon. This is why, initially, the generation of Gnumeric-XML-files first needed to be implemented. And so ICE offers a further output format which, against the background of the emerging broad rejection of proprietary formats, will probably play an even greater role in the future as an open standard. At present, this format can be read above all by the ever more popular office packages OpenOffice.org (an efficient open source package) as well as its commercial counterpart Staroffice from SUN.

8 Self-updating standard tables

Tables produced with ICE tools can be stored locally on the client computer by converting them into HTML or MS Excel format or as so-called standard tables on the server. The advantage of storing a results table as a standard table lies in the fact that is can be updated or extended and overwritten with the latest data at a later date. The changes that are to be made when more current timepoints are available are defined when the standard table is compiled. The type of update of each individual standard table depends on the table structure, according to whether the time-based assignment of data is made up of only one timepoint (so-called cross-sectional table) or of a time series sequence. Users can choose from three *update types* at any time:

1. *Timepoint related update:* In this type of update, all the table data are replaced with the latest data available in the system (for example, the table contains the data for 2001 and these are replaced with data for 2002).

2. *Time series addition*: Depending on the table structure, more recent data are added to rows or columns, ie, the table grows with the addition of more recent timepoints.

3. *Time series shift:* Depending on the table structure, more recent data are added to rows or columns while simultaneously the corresponding number of older timepoints are chronologically left out, meaning that the table size remains unchanged.

For each of these three update types it is additionally necessary to define whether the years, dates of the next winter semester or of the following semester (possibly also of the summer semester) are to be used. It is possible at any time to change the update time of a standard table (eg, after reaching a certain number of timepoints, the time series addition option can be replaced with the time series shift option), so as to influence the table appearance over the course of time.

Technical background:

A results table which is subsequently to be stored on the server as a standard table is produced from a so-called virtual data stock which the user accesses by entering keywords. Preselection in the form of virtual data stocks was chosen because the data contained in the database can, in principle, be combined in any which way. When generating the standard table, the system remembers the virtual data stock from which the data were extracted.
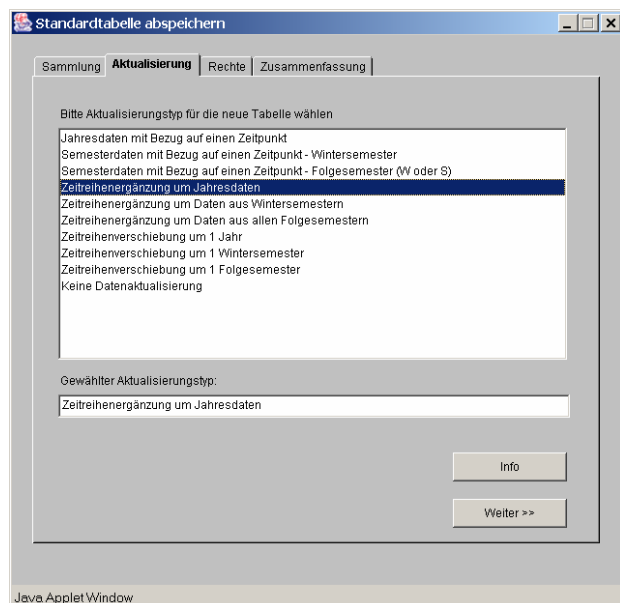
Figure 11: Selecting an update type when storing a standard table

As far as the timepoint is concerned, each virtual data stock is registered and requestable for analysis and interpretation with the help of an ICE key. When importing new data into the system, the time information for the virtual data stock also changes. When the update of standard tables is triggered by pressing the appropriate button, the first step involves comparison of the latest timepoint in the results table with the original data stock. If the system finds that the virtual data stock meanwhile has more up-to-date time information, then the structure of the standard table is analysed and changed in accordance with the chosen update type. In the case of time series additions, the table structure is amended at the relevant places by adding the new timepoints. In the case of time series shifts and the updating of cross-sectional tables (as a special case of time series shift), the table structure remains unchanged. At one end, new timepoints are added while at the other end older timepoints are removed to make space. The data for the new timepoints are subsequently extracted from the database, while any old data which the updated table will continue to show are simply adopted. This avoids having to fetch table sections which had already been found at an earlier stage.

The advantage of automatically updating standard tables not only lies in the fact the table structure is automatically brought into line with the new timepoints. In addition, when a table is updated the new table sections are subjected to the calculations which had been defined for the old table. This modification of the updated table (percentage calculation, indexing, difference formation) also occurs automatically, therefore.

Further planned update types:

The currently-defined update types offer users the opportunity to maintain standard analyses over time and,

as necessary, to change them easily. In the future, we intend to add further update types which make the updating of analysis tables even more flexible or will facilitate the updating of special table structures.

Time series with various qualities of data: This update type will allow a table which is made up of several virtual data stocks and which uses data of varying degrees of data quality to be intelligently updated. Final data are already available for the older data in the time series, while for more recent timepoints only provisional data are available and for the most recent timepoint only flash reports. When updating such a table, the system checks whether the time series can be supplemented with new timepoints and, at the same time, whether data of poorer quality can be replaced with data of better quality.

User-defined time series amendments: In the case of tables which have several timepoints, we intend to give the user more influence over the table timepoints so that it is possible to choose from all the available timepoints in the virtual data stock when updating a table.

9   Outlook

9.1   Data

The utility of the system depends essentially on the data which it makes available. It would seem to be a law of nature of statistical data analysis that it is precisely those data or differentiations which are only incompletely available which are just now needed to answer current questions. This is why it is our constant endeavour to increase the data stocks and to achieve greater depth of structure. If, in the future, it becomes possible to use individual case data records more than has been allowed in the past, then this problem will probably also solve itself from that side.

9.2   Internationalisation

HIS is currently thinking about extending the system by adding the option of outputting one and the same data stocks in various languages. Specifically, this involves enabling the publication of the German Ministry of Education and Research's Basic and Structural Data on the education system in several languages as a web application. [6]

A second consideration goes in the direction of using the system for European comparison. To do this, we would need the appropriate data, of course, and the above-mentioned multilingualism would have to be implemented.

9.3   ICE-Mobil

In the future, we want ICE not only to be available to users from a stationary workplace computer (or from an Internet café), but also to be available as a mobile system.

Access to elementary data via cell phone or WAP interface has already been implemented. A further interface is planned (Thin Client) with which access

would also be possible via handheld computers (Palm, Pocket-PC, Sharp-Zaurus).

References

[1] ICE der Wissenschaftsressorts der Länderministerien (ICE system for the science and education departments in the state ministries) http://iceland.his.de

[2] Wissenschaft Weltoffen website operated by the DAAD und HIS (ICE Data Basis) http://www.wissenschaft-weltoffen.org

[3] Birk, Lothar; Dicken, Hans; Kopp, Helena: Der goldene Schnitt – Statistische Datenauswertung mit Java. In: Java Spektrum 4/2000

[4] Dicken, Hans: JDBC – Internet-Datenbankenan-bindung mit Java; Internat. Thomson Publ., Bonn u. a. 1997

[5] Dicken, Hans; Hipper, Gunther; Müßig-Trapp, Peter: Datenbanken unter Linux; Thomson Publ., Bonn u. a. 2000

[6] Grund- und Strukturdaten des BMBF (Basic and Structural Data) http://www.bmbf.de/pub/GuS2002_ges_dt.pdf

Analyzing Educational Process Through a Chain of Data Marts

Viljan Mahnic
University of Ljubljana
Faculty of Computer and Information Science
Trzaska 25, SI-1000 Ljubljana, Slovenia
viljan.mahnic@fri.uni-lj.si

We describe the development strategy, architecture, and logical design of a data warehouse that can be built gradually, exploiting the benefits of the bottom-up, data mart approach. Connections between individual data marts are planned in advance with the aim of building a sequence of data marts that makes it possible to analyze the educational process as a value chain. Queries can be made across different subject areas (viz. enrolment applications, enrolment, examination, and degree records) in order to obtain a snapshot or a slice of the entire value chain that shows how far a subset of students has moved from the enrolment application to their final degree.

1 Introduction

In the early nineties, Bill Inmon introduced the concept of a data warehouse as a subject oriented, integrated, nonvolatile, time variant collection of data in support of management's decisions [5]. To create a data warehouse, data are extracted from different source systems, and then transformed, integrated, and loaded on an appropriate data store. Since then, data warehousing has grown to become one of the most important areas in the information systems field [3]. The benefits of data warehousing are numerous and some organizations are receiving significant returns [12].

The concepts of data warehousing have attracted substantial attention within the EUNIS community. At the EUNIS 1997 Conference, D. Stevenson [9] presented a data warehouse development project from users' and management's perspective, while at EUNIS 1999 M. Bajec et al. [1] proposed to build a data warehouse in order to analyze enrolment applications. At EUNIS 2001, two French initiatives were presented: J-F. Desnos [2] described a comprehensive data warehouse project for French universities, while Flory et al. [4] presented the design and implementation of a data warehouse for research administration. Additionally, the importance of data quality was described in order to ensure successful data warehouse implementation [8].

The aim of our paper is to describe the development strategy, architecture, and logical design of a data warehouse that should provide a unified and integrated source of data for various analyses of educational process at the University of Ljubljana. The main feature of this data warehouse is that it can be built stepwise as a chain of data marts that use common dimension tables, thus providing a suitable architecture for drill-across applications.

2 Development strategy and data warehouse architecture

Even though data warehouses are widespread, there is no common agreement about the best development methodology to use. While Bill Inmon (who is recognized as "the father of data warehousing") recommends a top-down, enterprise data warehouse approach, Ralph Kimball [6, 7] recommends the bottom-up, data mart approach. Using a top-down approach, a global enterprise data warehouse is built first that serves as a basis for implementation of individual data marts. On the other hand, in a bottom-up approach, individual data marts are developed first and later interconnected through common dimensions into a comprehensive data warehouse.

Considering our specific situation (viz. limited budget and the need for tangible results as soon as possible) as well as positive experience reported in the literature [10] we decided to adopt the bottom-up, data mart approach. This approach provides usable data faster, at a lower cost, and with less financial risk. However, in the long term this approach is successful only if all connections between individual data marts are well planned in advance. Therefore, special attention was devoted to logical data warehouse design in order to develop consistent data definitions and define an appropriate structure of the dimension tables that interconnect different data marts.

Additionally, since data warehouses (in general) play an important role in understanding value chains (e.g., by connecting trading partners along the demand or supply chain), we designed our data warehouse with the aim of representing the educational process as a value chain consisting of the following steps:

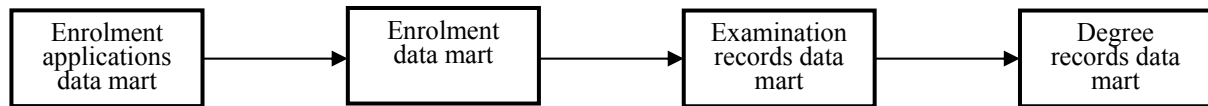| Enrolment applications data mart | → | Enrolment data mart | → | Examination records data mart | → | Degree records data mart |

Figure 1: The sequence of the data marts used to analyze the educational process as a value chain

- enrolment application
- first enrolment
- examination
- next enrolment
- examination
- ... (several enrolment and examination steps repeat here)
- degree

In our design, the aforementioned chain is modeled as a sequence of four data marts shown in Figure 1. The first data mart corresponds to enrolment applications, the second one contains enrolment data, the third one corresponds to examination records, and the last one deals with alumni data.

3 Logical design of data marts

Data marts are designed using dimensional modeling introduced by Kimball [6]. Each data mart is represented by a star schema data model (also called star join schema or dimensional model) that is made up of a fact table in the center of the star and several dimension tables as the points of the star. The fact table contains measurable facts that are recorded for each transaction (viz. enrolment application, enrolment, examination, and degree taken, respectively), while dimension tables describe entities (viz. students, study programs, teachers, courses, etc.) that are involved into these transactions.

Each star schema data model can be implemented individually and later integrated with other star schemas through common dimensions as described in Section 4. Figures 2 and 3 represent two sample star schema data models of our data warehouse. Figure 2 describes the logical design of the enrolment data mart that will be implemented first, while Figure 3 represents data in the examination records data mart.

3.1   Enrolment data mart

The fact table
We modelled each enrolment as an event at the intersection of seven dimensions: student, time (viz. academic year), department, study program, year of study, study mode, and type of enrolment (see Figure 2). Such a fact table represents a robust set of many-to-many relationships among these seven dimensions; however, it has only one measurable fact: the fee paid for studies.[1] This means that applications will perform mostly counts.

-----
[1] In Slovenia only part-time students pay fees for their studies.

Nevertheless, this table can be queried to answer any number of interesting questions, such as:

- How many students enrolled at each department or study program?

- What is the structure of enrolled students (considering secondary school, profession, secondary school grade, study mode, and/or type of enrolment)?

- Is the number of students (at a particular department or study program) increasing or decreasing?

- What is the progress rate of a particular generation of students?

The student dimension
The student dimension table contains data on students that are enrolled at the university. Each row corresponds to one student and contains his or her personal data. Secondary school, secondary school grade, and profession attributes enable the construction of correlations between students' progress and secondary school, secondary school grades, and secondary school profile of students. On the other hand, zip, county, and region attributes represent a geographic hierarchy that is useful in performing analyses of novice students in connection with the enrolment applications data mart.

The time dimension
Considering the enrolment data mart alone, an explicit time dimension table is not necessary because it is enough to keep the academic year of each enrolment as a degenerate dimension within the fact table. However, since the time dimension is common to all data marts a more elaborate version of time dimension table is necessary that is represented in Figure 3.

The study program dimension
At the University of Ljubljana study programs usually consist of several elective modules which can be further divided into submodules (e.g., in the final year of studies). Therefore, the study program dimension in fact describes individual submodules and defines a useful study program-module-submodule hierarchy that enables the generation of reports with different levels of detail (viz. drill-up and drill-down queries).

The year of study dimension
This is a degenerate dimension since the year of study attribute in the fact table is the only attribute of this dimension. It can be used as a grouping key for pulling together all the students enrolled in the same year of studies.
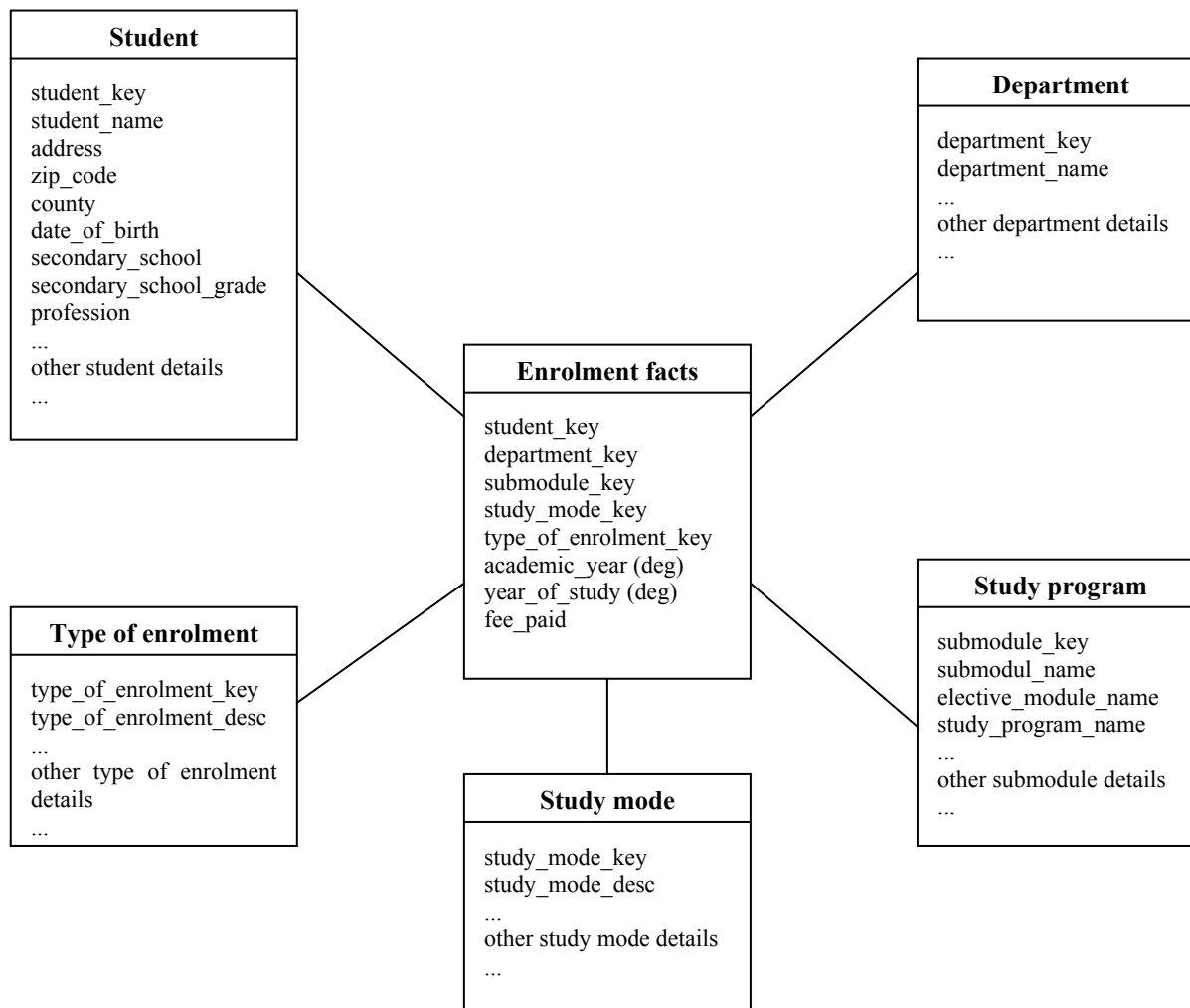
Student

student_key
student_name
address
zip_code
county
date_of_birth
secondary_school
secondary_school_grade
profession
...
other student details
...

Department

department_key
department_name
...
other department details
...

Enrolment facts

student_key
department_key
submodule_key
study_mode_key
type_of_enrolment_key
academic_year (deg)
year_of_study (deg)
fee_paid

Type of enrolment

type_of_enrolment_key
type_of_enrolment_desc
...
other type of enrolment
details
...

Study program

submodule_key
submodul_name
elective_module_name
study_program_name
...
other submodule details
...

Study mode

study_mode_key
study_mode_desc
...
other study mode details
...

Figure 2: The star schema data model representing the logical design of the enrolment data mart.

The department dimension
The department dimension table describes each member institution. At present, the University of Ljubljana consists of 26 member institutions (22 faculties, 3 academies, and 1 high school).

The study mode dimension
The study mode dimension describes every possible manner of studies (e.g. full-time, part-time, etc.).

The type of enrolment dimension
The type of enrolment dimension describes all possible enrolment types (e.g. first enrolment, repeated enrolment, etc.).

3.2   Examination records data mart

The fact table
Each record of the fact table in Figure 3 corresponds to one examination (viz. the grain of the fact table) and is uniquely defined by a compound key consisting of keys of all dimension tables. There are two measurable facts

that can be taken at the intersection of all the dimensions: grade and sequential number of examination attempt.

Dimension tables
Logical design of the examination records data mart comprises six dimension tables through which the data in the fact table can be analyzed: the student dimension, the course dimension, the teacher dimension, the department dimension, the time dimension, and the study program dimension.

The student dimension is the same as in the enrolment data mart. Secondary school, secondary school grade, and profession attributes can again be used in the construction of correlations between examination results and secondary school, secondary school grades, and secondary school profile of students.

Given the fact that the key of the time dimension is simply a date, the time dimension table could be omitted, but we found it useful because additional attributes (such as semester, academic year, and day number overall) enable to slice data by semesters and academic years, as well as to perform simple arithmetic between days across year and month boundaries (e.g., to compute time elapsed from first enrolment till graduation).
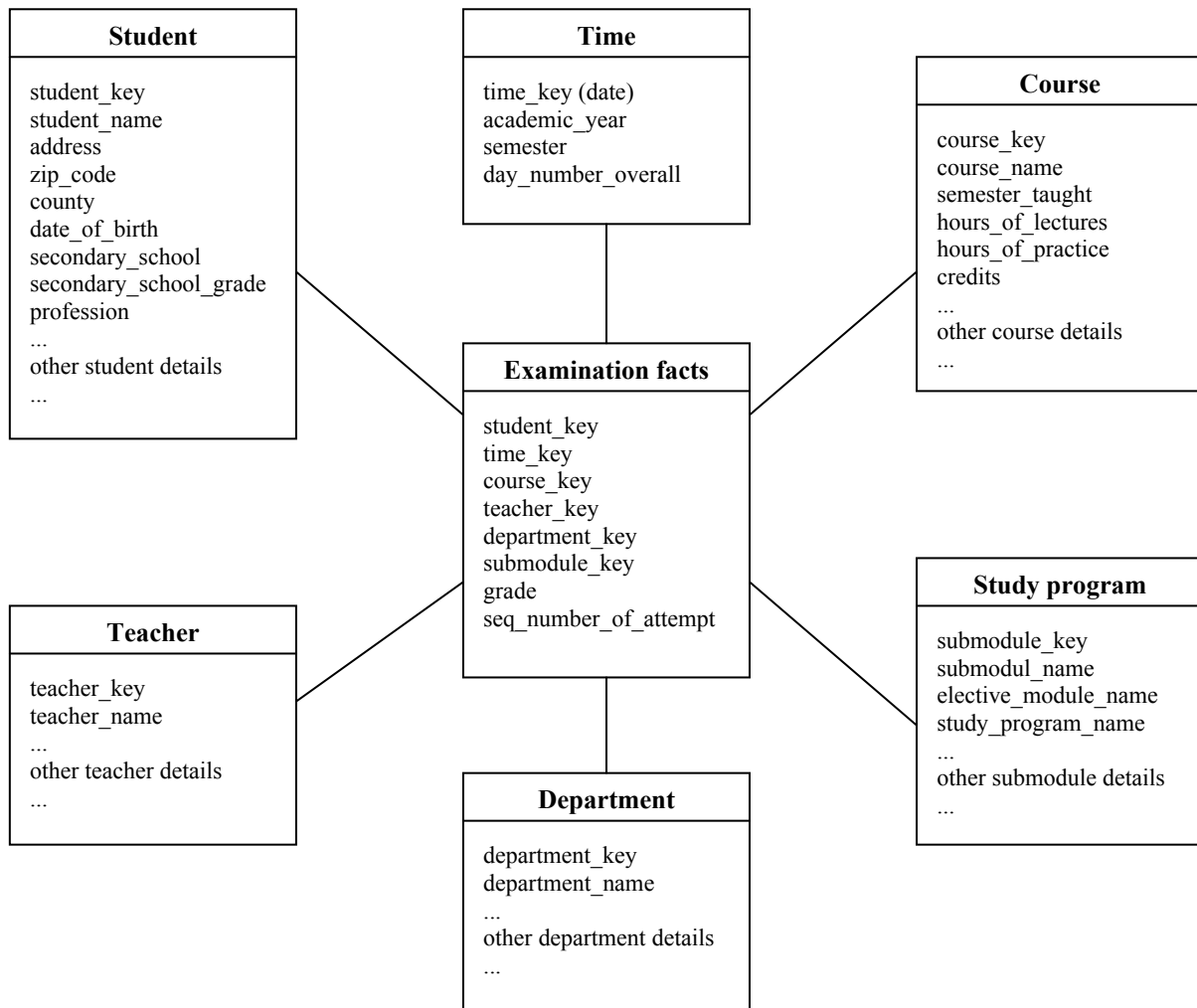
Figure 3: The star schema data model representing the logical design of the examination records data mart.

The course dimension describes every course, and the teacher dimension describes every teacher. The department dimension and the study program dimension are the same as in the enrolment data mart. Using the examination records data mart various analyses of examination results are possible, e.g.:

- What is the average grade of a specified subset of students?[2]

- How many students passed an exam in a given time period at each department or study program?

- What is the average grade and number of examination attempts at given course and/or teacher?

- How do average grades compare among different study programs and/or faculties?

etc.

4 Connecting data marts through common dimensions and drill-across applications

Some dimension tables (e.g., the student dimension, the study program dimension, the time dimension, etc.) are common to all data marts in our data warehouse. These tables can act as "glue" that connects the data marts together and allows meaningful queries to be made across different subject areas (viz. enrolment applications, enrolment records, examination records, and degree records). Using data warehousing terminology, these queries are often called drill-across applications.

In order to support drill-across applications, all constraints on dimension attributes must evaluate to exactly the same set of dimensional entities from one data mart in the value chain to the next data mart in the value chain. For example, a constraint on student dimension at any point in the value chain must mean exactly the same subset of students at all points in the chain. The easiest way to achieve this requirement is to

---

[2] The subset of students can be specified using attributes in the student dimension table as a source of constraints.
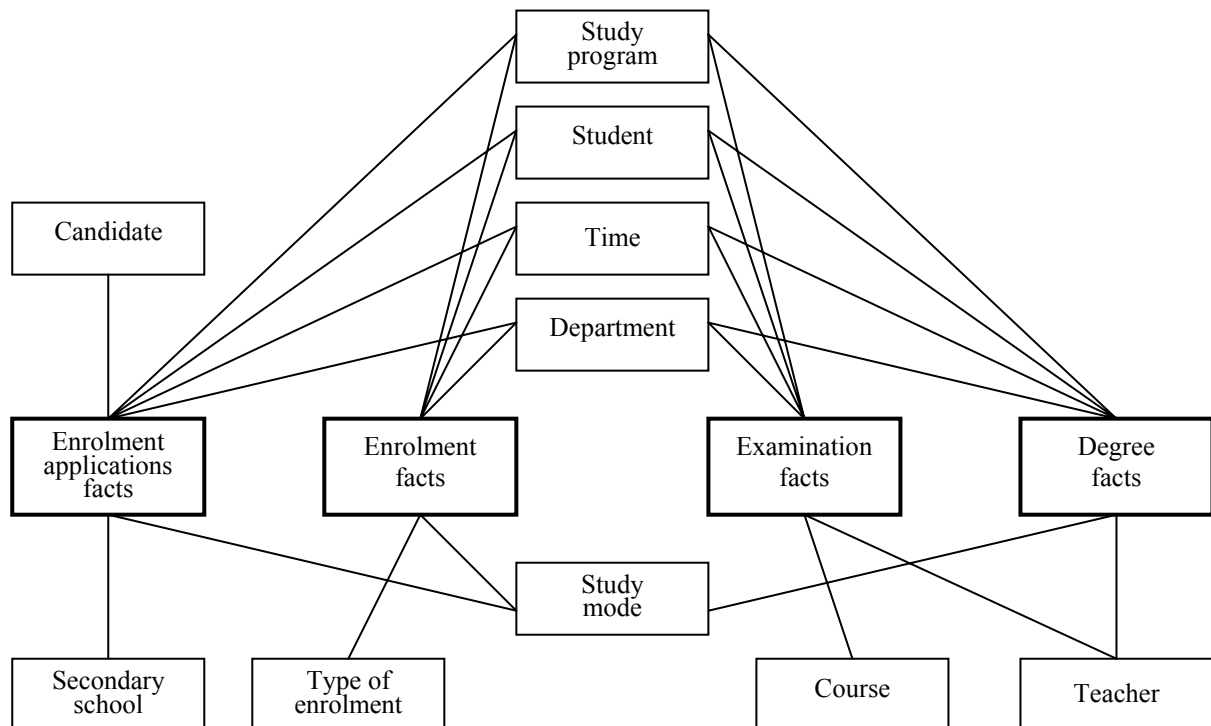
Figure 4: The chain of data marts connected through common dimension tables that are physically implemented only once.

physically implement all common tables only once as shown in Figure 4.[3]

Therefore, it is extremely important to plan the structure of common dimension tables in advance not only to satisfy the needs of individual data marts, but also to provide the necessary connections for drill-across applications. Given the fact that the University of Ljubljana is extremely decentralized and each member institution maintains its own data about students, teachers, and courses, substantial effort was necessary to integrate and cleanse these data in order to build common dimension tables.

Beside dimension tables that have already been described in previous section, there are two additional dimensions shown in Figure 4:

- The candidate dimension table corresponds to all candidates for enrolment. The accepted candidates become part of the student dimension after first enrolment.
- The secondary school dimension describes every secondary school in Slovenia.

---

[3] There are some special situations when the aforementioned requirement can be achieved without implementing the common dimension only once, e.g. in the case of dimensions with reduced detail and derived dimensions that support aggregates. The interested reader can find more information in [6, pp. 84-85].

5 A sample drill-across report

Using our data warehouse we can imagine that students "move" sequentially through the value chain, and a drill-across report can show a snapshot or a slice of the entire value chain that shows how far a subset of students has moved from the enrolment application to their final degree. For example, using the enrolment applications data mart a subset of candidates that applied for enrolment in a given academic year can be defined. In combination with the enrolment data mart, only those candidates that actually enrolled (viz. became students) can be isolated. For this subset of students, their examination records can be analyzed using the examination records data mart and the number of students who finished their studies can be determined from degree records data mart.

Table 1 represents a sample report obtained by drilling across. Suppose we have several generations of candidates that applied for enrolment at a given department in five consecutive years and we want to track their progress towards the graduation. Given the fact that the department and time dimensions are exactly the same for all data marts, only those data that belong to the department and academic years in question are processed in each data mart. Similarly, the common student dimension assures that the same subset of students is taken into account in the whole value chain. Using the enrolment applications data mart the exact number of applicants and the number of approved

Department XXX

Academic Year	Applicants	Approved	Enrolled	Average Grade	Exams Passed	Graduated	Length of Studies
1992/93	237	162	150	8.16	29.3	77	6.15
1993/94	199	165	143	8.02	27.1	71	6.32
1994/95	182	158	144	7.57	24.3	64	6.87
1995/96	176	151	145	7.03	19.5	56	7.02
1996/97	183	154	150	7.43	22.4	57	6.21

Table 1: A sample drill-across report

Department XXX

Academic Year	Study Program	Applicants	Approved	Enrolled	Average Grade	Exams Passed	Graduated	Length of Studies
1992/93	AAA	125	101	95	8.21	30.1	52	6.08
	BBB	112	61	55	8.07	27.9	25	6.32
Total 1992/93		237	162	150	8.16	29.3	77	6.15
1993/94	AAA	121	110	96	8.12	28.1	52	6.12
	BBB	78	61	47	7.81	25.0	19	6.89
Total 1993/94		199	165	143	8.02	27.1	71	6.32
1994/95	AAA	119	108	94	7.78	27.1	47	6.65
	BBB	63	50	50	7.18	19.0	17	7.47
Total 1994/95		182	158	144	7.57	24.3	64	6.87

etc.

Table 2: Report refinement by drilling down

applications can be determined. The enrolment data mart enables the computation of the actual number of enrolled students, while the examination records data mart provides the average grade[4] and average number of exams passed for these students. Finally, the degree records data mart is used to compute the number of graduates in each generation as well as the average length of their studies (in years).

By drilling down we can still refine our query in order to obtain more detailed information about each generation of students. By simply adding the study program name as a new row header the same analysis can be obtained for each study program separately (see Table 2). We can further subdivide the subset of students in each generation by using other attributes from the study program hierarchy (viz. elective module name and submodule name) as well as by choosing row headers from dimensional attributes of other dimensions (e.g., secondary school, secondary school grade, secondary school profession).

6 Conclusions

We described the design of a data warehouse that can be implemented gradually as a chain of data marts connected through common dimension tables. Each data mart was represented using the star schema data model and a special attention was devoted to the definition of common dimensions in order to enable drill-across applications. The proposed design describes the educational process as a value chain and makes it possible to analyze how far a subset of students has moved from the enrolment application to their final degree.

---

[4] In Slovenia the following grades are used: 1 to 5 – insufficient; 6 – sufficient; 7 – good; 8, 9 – very good; 10 – excellent.

References

[1] Bajec, M., Rupnik, R., Krisper, M. Using Data Warehouses in University Information Systems, in K. Sarlin (ed.) EUNIS 99 – Information Technology Shaping European Universities, pp. 115-121.

[2] Desnos, J-F. A Data Warehouse for French Universities, Informatica, Vol. 25, No. 2, July 2001, pp. 177-181.

[3] Eckerson, W.W. Evolution of Data Warehousing: The Trend toward Analytical Applications, Boston, MA: The Patricia Seybold Group (April 28, 1999), pp. 1-8.

[4] Flory, A. et al. Design and Implementation of a Data Warehouse for Research Administration Universities, in J. Knop and P. Schirmbacher (eds.) EUNIS 2001 – The Changing Universities, The Role of Technology, Berlin, March 2001, pp. 164-167.

[5] Inmon, B. Building the Data Warehouse, QED Publishing Group, 1992.

[6] Kimball, R. The Data Warehouse Toolkit, John Wiley & Sons, 1996.

[7] Kimball, R. et al. The Data Warehouse Lifecycle Toolkit, John Wiley & Sons, 1998.

[8] Mahnic, V. Rozanc, I. Data Quality: A Prerequisite for Successful Data Warehouse Implementation, Informatica, Vol. 25, No. 2, July 2001, pp. 183-188.

[9] Stevenson, D. Data Warehouse and Executive Information Systems – Ignoring the Hype, in J-F. Desnos and Y. Epelboin (eds.) European Cooperation in Higher Education Information Systems, Grenoble, France, September 1997, pp. 202-207.

[10] Watson, J.H. et al. Sherwin-Williams' Data Mart Strategy: Creating Intelligence across the Supply Chain, Communications of the Association for Information Systems, Vol. 5, Article 9, May 2001.

[11] Watson, J.H. Recent Developments in Data Warehousing, Communications of the Association for Information Systems, Vol. 8, 2001, pp. 1-25.

[12] Watson, J.H. et al. The Benefits of Data Warehousing: Why Some Organizations Realize Exceptional Payoffs, Information & Management, Vol. 39, 2002, pp. 491-502.

A Decision Support System for IST Academic Information

Elsa Cardoso, Helena Galhardas and Rito Silva
Instituto Superior Técnico, INESC-ID, Rua Alves Redol, nº 9, 1000-029 Lisboa, Portugal
Elsa.Cardoso@inesc-id.pt, www.esw.inesc-id.pt/~eac
hig@inesc-id.pt, http://gsi.inesc-id.pt/~hig
Rito.Silva@inesc-id.pt, www.esw.inesc-id.pt/~ars

Maria José Trigueiros
Instituto Superior de Ciências do Trabalho e da Empresa
Avenida das Forças Armadas 1649-026 Lisboa, Portugal
mjtrig@iscte.pt

This article describes the Decision Support System (DSS) for Academic Information being developed at Instituto Superior Técnico, the Engineering School of the Technical University of Lisbon. In Portuguese, this project has been given the acronym SADIA (Sistema de Apoio à Decisão da Informação Académica). This paper focuses on the early phases of the DSS development process, i.e., the business requirements definition and the dimensional modelling. First, we show how the business requirements of the School drive the definition of the DSS dimensional model. Second, we detail the logical dimensional model for a selected business process, the IST Student Admission process. Third, the corresponding physical design decisions are reported. The results obtained from the three phases were successfully validated by business users.

1 Introduction

Instituto Superior Técnico (IST) is the Engineering School of the Technical University of Lisbon, and one of the biggest Higher Education Schools in Portugal since the first decade of the 20th century. Currently, IST offers twenty-two 5-year undergraduate degrees to a population of more than 10,000 students.

Two years ago, IST has started the FENIX project [2] aiming at the integration of academic management information. New Information Systems are being developed and others restructured. As part of this global strategy, the School's Board of Directors has also decided to implement an Academic Information Decision Support System known as the SADIA system (in Portuguese SADIA stands for *Sistema de Apoio à Decisão para a Informação Académica*) [8]. The first prototype of the system is focused only on the School Pedagogic Assessment. Later, the SADIA system will feed a higher-level DSS system, which is being implemented by the Technical University of Lisbon to support the Dean's management decisions.

This paper describes the SADIA system in terms of the development process and results in dimensional modelling of the underlying Data Marts. Of note, the implementation of the system is already in progress.

1.1 Decision Support Systems: Basic Definitions

The SADIA system has been developed according to the **Business Dimensional Lifecycle** proposed by Ralph Kimball [1]. Figure 1 illustrates the basic elements of a Data Warehousing project, based on a four level architecture: (1) Operational Data Source Systems; (2) Staging Area; (3) Presentation Servers; (4) End-user Data Access.
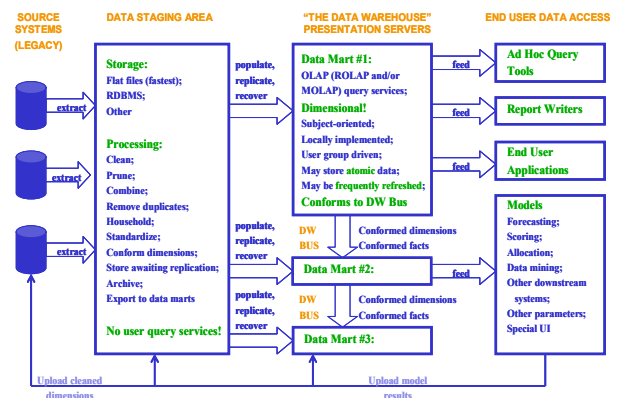
Figure 1: Basic Elements of a Data Warehouse [1]

Source systems, also called operational or legacy systems (in mainframe environments), support the operational nature of the business, recording and managing transactions. The main priorities of these systems are transactional performance, uptime and availability. They usually do not record historical data and are not suitable for the generation of management reports. Traditionally, source systems are developed in organizations to support certain business areas or departments.

The **Staging Area** encompasses a storage area and a set of processes (which may be implemented in one or more machines) that prepare source data for use in the Data Warehouse (DW). Some of these processes are: (1) Data cleaning; (2) Data reformatting; (3) Data transformation; (4) Semantic validation; (5) Store awaiting replication; and (6) Replication to presentation servers. At this level no kind of queries or presentation services should be allowed.

The **Presentation Server** is the target physical machine that stores the DW data for direct querying by end users, report writers and other applications. At this level, data is presented and stored in a dimensional framework. Presentation servers may be implemented using relational databases (with star schemas) or non-relational OLAP (On-line Analytic Processing) databases (with multi-dimensional cubes).

A **Data Mart** is, according to Kimball, a logical subset of the complete Data Warehouse, i.e., a restriction of the DW to a single business process or to a group of related business processes designed for a specific group of users. A Data Mart is usually sponsored and built by a single part of the business.

Each Data Mart is represented by a dimensional model, supported by a set of fact tables. The apparent inconsistency does not really exist. At a high-level analysis, when subject business areas and candidate dimensions are identified, data marts may be restricted to a single star schema (meaning one dimensional model). However, the physical design of the Data Mart may impose the implementation of several star schemas, because different information aggregation levels may be needed or by performance reasons associated with data sparseness.

Kimball considers the Data Warehouse made up of the **union** of all its Data Marts. This statement is valid to Kimball as long as the Data Warehouse Bus Architecture rules are respected. That is, within a Data Warehouse all Data Marts must be built from **conformed dimensions** and **conformed facts** [1], otherwise Data Marts may soon become isolated and obsolete systems.

1.2 Case Study: the SADIA System

The Academic Information Decision Support System of Instituto Superior Técnico (SADIA) [8] is part of the School's global strategic integration plan of academic management. The IST's Computer Centre (CIIST) is the entity in charge of the FENIX project [2]: the new integrated academic management Information System of IST. The purpose of the FENIX project is to create a new Information System able to successfully respond to the current needs of all participants in the tuition process (i.e., teachers, students and administrative services). The system performance should enable gains of both time and effort, and the system ought to be designed in a modular way, in order to be easily expanded.

The FENIX project has two main tracks: the operational and the decision support tracks. The operational track supports the execution of the School's functional business processes. The decision support track

(i.e., the SADIA system) provides present and historical information organized in terms of key performance indicators to support management decisions. One primary goal of the SADIA system, required by the Board of Directors, is the automatic generation of tables including the statistics required by the external processes for accreditation and assessment of undergraduate degrees.

The SADIA system has been designed according to Kimball methodology for developing Data Warehouses entitled *Business Dimensional Lifecycle*.

1.3 Outline of the paper

This paper is organized in six sections. Section 2 presents the *Business Dimensional Lifecycle*, which is the DW development process adopted for the SADIA project. Sections 3 to 6 describe the execution of some activities of the Business Dimensional Lifecycle Process in the SADIA project applied to a single business area. The selected business area is the IST Student Admission process. Section 3 briefly describes the Project Planning and the Management phase. The Business Requirements Definition, described in Section 4, encompasses the Business Modelling activity performed for the SADIA project. This section presents a detailed description of the selected business process and a few examples of user analysis queries, regarding the IST Admission process. The Dimensional modelling activity is described in Section 5. This section presents the step-by-step development of the logical dimensional model of the IST Student Admission Data Mart. Physical Design decisions are reported in Section 6. Finally, we conclude and summarize future work.

2 Development Process

Figure 2 presents the **Business Dimensional Lifecycle**. Kimball compares this methodology to a conductor's score [1], as a way of assuring that every piece of the project is joined correctly at the right moment. Well-succeeded DW implementations depend on the integration of numerous tasks, components and tools. To implement a successful DW, it is imperative to gain skills in all project areas. It is not enough to design the best dimensional model or to buy the most expensive technology of the market... it is necessary to coordinate the multiple features of a DW project.

As seen in Figure 2, the methodology starts with the Project Planning phase. Project Management is active through the entire life of the project. The Business Requirements Definition is the central activity, as business requirements drive the whole Data Warehouse project.

After establishing the project foundations (i.e., the business requirements), three parallel tracks should be followed:

- **Data Track**: with three data activities: (1) the Dimensional Modelling; (2) the Physical Design; and (3) the Data Staging Design and Development.

- **Technology Track**: encloses the activities: (1) the Technical Architecture Design; and (2) the Product Selection and Installation (e.g., SGBD, Data Staging tool and Data Access tool).

- **Application Track**: with two activities concerning the End-user Application design and implementation.
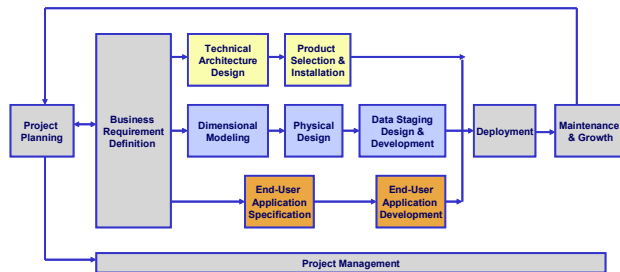

Figure 2: Business Dimensional Lifecycle [1]

Reaching the Deployment phase corresponds to only 25% of total project effort. Another 25% will be spent on system tests and additional 25% should be dedicated to iterative correction and validation procedures to assure the quality of the DW data contents. The remaining 25% will be spent on Maintenance and Growth activities.

2.1 Project Planning and Management Phases

Project Planning and Management are naturally quite similar to homonymous phases in traditional software development processes. As usual, the main activities consist in defining, planning and managing all project tasks.

Kimball starts Project Planning with a test, called the *Readiness Litmus Test* [1], to evaluate the organization's receptiveness to a DW. The existence of one (or more) strong business management sponsor in charge of the project is the most critical factor when assessing the readiness to a DW, for a number of reasons. Since DW projects tend be expensive with rising maintenance and growth costs, a strong sponsor and a solid financial return are indispensable conditions to sustain a long term economic power. The DW may be used to respond to critical business requirements. Some business motivations that may be the driving force for an organization to adopt a Data Warehousing strategic plan are [1]: (1) a highly competitive and ever-changing market; (2) an internal crisis; (3) a strategic vision of a potential marketplace opportunity; and (4) integration problems inherent to acquisition strategies. The success of a DW also depends on a joint effort and shared responsibilities, between the business management and the IS team. Most successful DWs are built by organizations in which fact-based decision making is encouraged and rewarded. On the contrary, in an organization where managers base their decisions on their ideas/opinions ("gut-feelings"), pushing information and analysis to a secondary place, the readiness for a DW is highly questionable. Finally, technical feasibility

issues should be analysed, in particular regarding: (1) data availability; (2) ease of development and deployment; and (3) resource availability and team experience.

Project Scope
The definition of the project scope should be based on business requirements and not on time constraints. Kimball suggests the following five guidelines to determine the preliminary scope during the Project Planning phase:

- The scope should be defined as a joint effort of IS and Management team representatives;

- The initial scope of a DW project should have an impact in the organization (i.e., there must be some added-value to the business, by addressing a well-defined business requirement), while being feasible. Data Warehousing should start with small initiatives, since it is meant to have an iterative development process;

- The project should be initially focused only on one business process, supported by data from a single source system. The Data Staging development effort is estimated to grow exponentially with each additional major source system.

- The initial number of end users should be limited (e.g., to 25 users);

- The success criteria of the project should be identified as soon as the scope is defined.

Project Management
Managing a DW project implies the following activities: (1) to maintain the Project Plan and Project Documentation; (2) to manage the scope; and (3) to elaborate a Communication Plan to manage user expectations.

Maintaining the Project Documentation is mandatory in Data Warehousing, due to the unending nature of this kind of projects. Nowadays, since it is quite difficult to maintain unchangeable development teams, the existence of detailed documentation will ease the integration of newcomers in the project.

Change is inevitable in DW projects! This is not due to poorly defined business requirements but to the impossibility to anticipate all user needs and the variations in the marketplace that may impact on the business. Today's market is a turbulent one, where mergers and acquisitions of enterprises are common events. Kimball considers a change any issue resolution that impacts in the project schedule, budget or scope [1]. Changes should be formally documented and broadcasted to the users, in order to readjust their expectations.

2.2 Business Requirements Definition

Business requirements should have an impact on every DW project area. The starting point of the Requirements

Business Definition phase is the set of business users. The team responsible for the Requirements Process should start by talking to business users, to understand their jobs, goals and challenges and particularly *"how do they take their decisions"*. In parallel, Data Audit interviews should start with the "data *gurus"* of the organization, those crucial elements in the Informatics Department with a deep knowledge of data. The single purpose of Data Audit interviews should be the systematic exploration of the data source systems. The reality of the organizations' data should be identified, namely whether there is data to support the analysis of the user's requests, and start assessing the quality of data.

The requirements capture methodology is based on interviews and facilitating sessions. Kimball provides different sets of key-questions, tailored to the interviewer's profile. These questions were found extremely useful when applied to the interviews of the SADIA project. The selection of the interviewers is crucial. One should start by interviewing business users horizontally across the organization, embracing other groups beyond the "target" group to be addressed by the DW. The objective is to gather a global vision of the organization's common vocabulary, ensuring the data integration process over time, and therefore avoiding the development of stovepipes. These interviews should also address a vertical representation of the target business area. The Executive Business Management staff detains the high-level strategy and an overall vision of the business. However, it is imperative to go down the hierarchy to the Middle Managers, who are the key players for the translation of the high-level strategies into the real business tactics. Middle Managers also have a realistic perspective of the company's strategy concerning information or knowledge management. At the base of this hierarchy, Business Analysts from the target business area possess a detailed and practical know-how concerning the use of the organization's information.

The interviews phase is the right moment to start defining the terminology of the project. The exact definition of this terminology will have a huge impact on the grain and dimensionality of the data model. Issues of vocabulary standardization typically emerge as we conduct interviews across departments in the organization. Vocabulary inconsistencies should not be resolved during interviews. The best approach is to host a facilitating session among the decision makers of different departments.

Business requirements establish the foundations of the project that enable to execute the three parallel tracks: Data Track, Technology Track and Application Track.

2.3   Data Track

Dimensional Modelling

Dimensional Modelling is a logical design technique, used in Data Warehouses, which is an alternative to Entity-Relationship (ER) modelling. A dimensional model contains the same information as an ER model but stores data in a symmetric "star-like" structure optimized

for different design goals [1,5]: user understanding, query performance and resilience to change. A dimensional model is composed of one table with a multipart key – the **fact table** – and a set of smaller tables – the **dimensional tables**. Each dimensional table is connected to the fact table by a single-part primary key that corresponds to one of the multipart key components. Fact tables contain the business measurements or facts, whereas dimensional tables contain many textual attributes that will be used as constraints in DW queries.

There are three types of measures or facts: (1) fully additive; (2) semi-additive; and (3) non-additive. Fully additive facts are the most useful since they can be meaningfully summarized across any dimension. Semi-additive measures can only be summarized across some dimensions (for example, levels such as inventory quantities or account balances that cannot be summarized over time). Non-additive measures (e.g. ratios) cannot be summarized at all. The solution is to break the non-additive measure into its fully additive components and store them in the fact table.

The Dimensional Modelling activity starts with the definition of the *DW Bus Architecture matrix* [1] that displays the key business processes versus the candidate analysis dimensions. This top-down perspective matrix must be conciliated with a bottom-up data source analysis in order to adjust the information needs to the reality of available data. Then, the 4-step method [1] for the design of individual fact tables should be applied to each feasible business process or Data Mart selected for implementation. The 4-step method to design a fact table encloses the following steps:

Step 1: *Choosing the Process.* This step consists of identifying the Data Mart's subject area.

Step 2: *Choosing the Data Mart Grain.* The grain is the level of detail at which each row in a Fact Table is recorded. Choosing the grain means identifying exactly what a single Fact Table record represents.

Step 3: *Identifying and Conforming the Dimensions.* Any dimension that takes a single value in the presence of the grain is a good candidate to be selected for the Data Mart. Conforming dimensions imposes the requirement that the same dimension (e.g., the Customer or Product Dimension) in two different Data Marts be defined precisely in the same way.

Step 4: *Choosing the Facts or Measures.* The selection of facts should be as rich as possible within the context of the declared grain. All facts must be expressed at the level previously defined for the grain. In addition, the facts should be as additive as possible.
In [6,7] Kimball enhances the 4-step method. The new proposed steps for logical dimensional modelling comprise the following activities:

Step 5: *Storing Pre-calculations in the Fact Table.* Some values can be obtained from existing facts (e.g., the net price of a product can be derived from the product price minus the allowances and discounts [7]). Nevertheless, it may be interesting to pre-calculate and explicitly

materialize some of them in the fact table. Although redundant information is stored, this decision avoids mistakes (for example, the user applies the wrong formula) and improves performance.

Step 6: *Enriching the Dimension Tables*. The grain decision in step 2 also determines the grain of each dimension table. The goal in this step is to be as comprehensive and wordy as possible. At least 50 textual attributes should be identified for important dimensions, such as Customer or Product. Dwelling on data sourcing or data quality details should be avoided since these problems will be solved in the Data Staging Design and Development phase.

Step 7: *Choosing the Duration of the Database*. The business area determines the amount of historical information to store in the Data Mart, i.e., the duration of the fact table. For example, insurance companies usually store seven or more year-old facts. Special attention must be taken in this case since older data records tend to be more difficult to interpret.

Step 8: *The Need to Track Slowly Changing Dimensions*. During the DW lifecycle changes in data occur, for example a customer address change. There are three mechanisms to handle and integrate these changes into the DW schema: Type 1, 2 and 3. Type 1 is only used to correct errors since it overwrites the dimension record with the new values. Type 2 creates a new record to store the new value, and is the most used mechanism. Type 3 creates an "old" attribute in the dimension record to store the old value. This mechanism is used less frequently, for "soft changes" situations, where both the old and the new values must be supplied.

Physical Design
This phase deals with the physical database design issues that must be defined in order to support the logical design. Some modelling activities addressed in this phase are: (1) Definition of Aggregates; and (2) Definition of Indexing and Partitioning Strategies.

Aggregates are fact tables with some level of summarization, derived from the most granular or elementary fact table of the Data Mart. Aggregates are used in DW to improve query performance.

Data Staging Design and Development
This activity is typically the most underestimated within DW projects. Yet, it is a critical task consuming a large amount of project resources. Data Staging Design involves three main steps: (1) **E**xtraction; (2) **T**ransformation; and (3) **L**oad, known as the **ETL process**.

The Extraction process extracts data from the source systems. Data must then be transformed, using Data Cleaning techniques/algorithms to overcome data quality problems. The final process loads "cleaned" data into the DW dimensional framework. Two ETL processes must be defined. The first one handles the initial migration of data from source systems, populating the DW for the first time. The second one concerns the periodical and incremental loads of data.

2.4 Technology and Application Tracks
The Technology Track encompasses two activities: (1) Technical Design Architecture; and (2) Product Selection and Installation. The design of the DW technical architecture is essentially driven by business requirements, the current technological environment and the technical directions planned for the organization DSS strategy.

The third parallel track, the Application Track, is composed of two activities: (1) End-User Application Specification; and (2) End-User Application Development. The first activity deals with the definition of a set of standard end-user applications, like report templates and required formulas used in calculations. The second activity involves the metadata tool configuration and the development of specified reports.

3 Project Planning and Management
The SADIA project has one strong sponsor: the President of the CIIST. Since the Board of Directors also supports the project, the partnership between IS and Business has been successful. The business motivation that drives the project is the ability to improve the teaching quality provided by the Faculty, since the number of students (applicants) is decreasing and Faculties struggle to attract the most brilliant students.

The Board of Directors is convinced that SADIA will allow them to base their decisions on factual data and, more importantly, that the system will allow them to start creating a culture of decision-making based on factual data.

4 Business Requirements Definition
The SADIA development team interviewed the key business users, i.e., the School's Pedagogical Council (several members of the Executive Commission including the President) and the School's Office for Studies and Planning (GEP, standing for *Gabinete de Estudos e Planeamento* in Portuguese). In parallel, data "*gurus*" were also interviewed to identify the reality of the School's data. The following deliverables were produced [8]: (1) the **Requirements Specification Document** that includes a detailed analysis of GEP Studies, a summary of all interviews and the analysis of two source systems – the data model reengineering of the current students enrolment application (managed by CIIST) and the SIGLA database (managed by GEP); and (2) the **Organizational Model of the School**, enclosing the static organization of IST and the business processes modelling.

The application of Kimball's methodological proposals to the SADIA project was found extremely useful. In particular, the interview questionnaires [1] for business users (managers and analysts) and for Data Auditing produced excellent results for capturing business requirements. However, this methodology lacks

a notation to detail business requirements. Due to the SADIA team experience in Software Engineering, UML (Unified Modelling Language) [3] was selected as the language to model business processes. The identified SADIA business processes are represented as UML business use cases, illustrated in Figures 3 and 4.
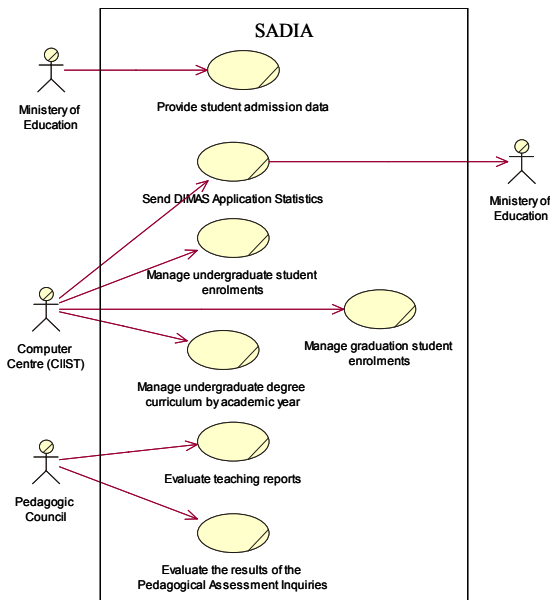

Figure 3: SADIA Business Processes

Clearly, this represents a major system that cannot be implemented all at once. We have selected the following business processes or major subject areas to be supported by the first release of the SADIA system:

- Elaborate IST Student Admission Study

- Elaborate Student Performance Study

- Elaborate Undergraduate Degree Self-assessment Study

These studies have well-defined and documented business requirements. The scope of the first release of the SADIA system will have an impact on the Pedagogic Assessment of the School.

Some statistics of the Student Admission and Student Performance studies are also required by the Undergraduate Degree Self-assessment study. The SADIA system is focused on the following subset of business processes required by the Undergraduate Degree Self-assessment study: (1) IST Student Admission; (2) Undergraduate Degree Performance Evaluation; (3) Course Performance Evaluation; and (4) Student Performance Evaluation.

For each of the selected processes the corresponding business indicators were identified. From this point forward, the IST Student Admission process is used to illustrate all activities already performed in the SADIA project. Business indicators were identified through the analysis of existing School documentation and user interviews. Section 4.1 presents the description of the

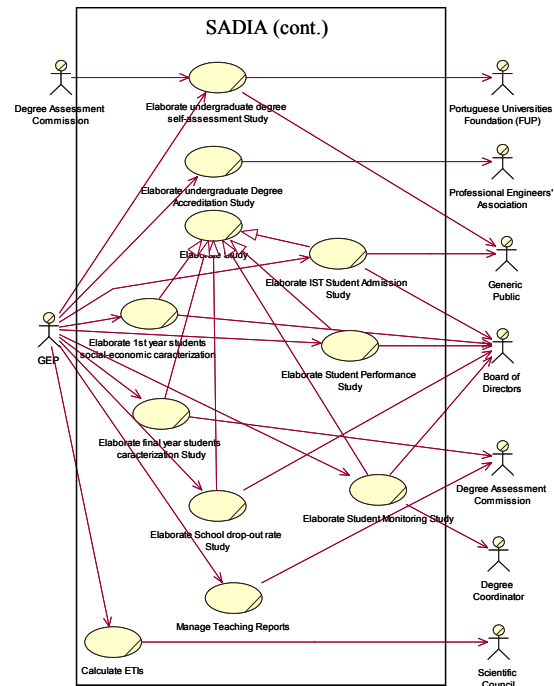Student Admission process and a few user queries are exemplified in Section 4.2.

Figure 4: SADIA Business Processes (cont.)

4.1   The IST Student Admission Process

The IST Student Admission process for each academic year may be analysed from three perspectives: (1) the overall admission process of the School; (2) the student admission process for each undergraduate degree of the School; and (3) the admission process for a particular applicant.

Overall Admission Process of the School

The global business indicators that characterize the admission process of the School for an academic year are displayed in Table 1. The Admission process also includes the analysis of the admission average classification of all admitted students for an academic year, represented in Table 2.

The admission process corresponds to the fulfilment of the number of vacancies by applicants, depending on their application classification and on their application option. Applicants are ordered according to their ranking, which is currently defined as 50% of the High School final classification plus 50% of the classification obtained in the specific set of admission tests required by the degree.

Student Admission Process for each Undergraduate Degree

Table 3 presents an analysis map illustrating the Student Admission Business Indicators calculated for each undergraduate degree for an academic year. In order to assess the quality of the admitted students, some admission classifications were also calculated (see Table

Global Business Indicators	97/98	98/99
Number of vacancies	1250	1300
Total number of applicants	8457	7184
Occupation rate	100%	100%
% admissions from the General Admission Contingent	93%	95%

Table 1: Global business indicators for the IST Student Admission Process

Admission Grades	97/98	98/99
Average of application classification	77,5%	81,4%
Average classification of Maths admission test	78,6%	82,8%
Average classification of Physics admission test	68,9%	81,8%
Average classification of Chemistry admission test	82,9%	80,3%
Average classification Geology admission test	76,3%	74,4%
Average classification of Biology admission test	90,2%	91,3%
Average classification of Drawing and Geometry admission test	---	93,6%
Average High School final classification	15,9	16,1

Table 2: Average admission grades of all students admitted to IST

4). The values represented in Tables 3 and 4 refer to the Student Admission process to the Electrical Engineering undergraduate degree (LEEC, in Portuguese) in 1997/98 and 1998/1999.

The geographic origin of students is also an important issue regarding the characterization of the admitted students. The business indicator underlying this analysis is the Number of Applicants per Geographic District.

Student Admission Business Indicators	97/98	98/99
Number of vacancies	250	250
Number of applicants	982	961
Number of first option applicants	401	395
Ratio applicants/vacancies	3,9	3,8
Ratio first option applicants/vacancies	1,6	1,6
Occupation rate	100%	100%
% admitted from the General Admission Contingent	92,8%	94%

Table 3: Student admission analysis map for each Undergraduate Degree

Admission Grades	97/98	98/99
Minimum application classification	69,3%	74,4%
Average of application classification	77,3%	82,7%
Average classification of Maths admission test	79,0%	85,0%
Average classification of Physics admission test	69,5%	83,7%
Average High School final classification	16,1	16,2

Table 4: Minimum and average admission grades of the admitted students in the LEEC degree

Admission Process for a particular Applicant

The performance of an individual student in an Admission Contest is measured in terms of his/her application classification, the order of entrance of the applicant, the High School final classification, and the classifications obtained in the specific admission tests he/she performed.

4.2　User Queries

Another result of the interviews performed during the Business Requirements Definition Phase is a set of typical user analysis queries. Consider the following three questions related to the Analysis of the Admission for an Undergraduate Degree in a particular academic year (see [8] for the complete set of user queries):

Q1: Number of admitted students per Admission Contingent?

Q2: Number of students that did not registered?

Q3: Distribution of admitted students by application option?

5　Dimensional Modelling

The Dimensional Modelling activity started with the definition of the DW Bus Architecture matrix [1], represented in Table 5. This matrix displays the key business processes versus the candidate analysis dimensions. Then, we followed the enhanced 4-step method previously described in Section 2.3.

Business Processes	TIME	STUDENT	ADMISSION	GEOGRAPHY	COURSE	DEGREE	DEPARTMENT
IST Student Admission	X	X	X	X		X	X
Undergraduate Degree Performance Evaluation	X				X	X	X
Course Performance Evaluation	X				X	X	X
Student Performance Evaluation	X	X		X		X	X

Table 5: DW Bus Architecture Matrix

Step 1: Choosing the Process

The admission process of potential students is modelled as a specific Data Mart, with one accumulating snapshot fact table, as suggested in [4]. This kind of fact tables is suitable for short-lived processes, such as the admission pipeline, with well-defined beginning and ending dates and a set of standard milestones.

In the admission process to an undergraduate degree, the potential students (our clients!) progress through a set of admission milestones. The process is similar to a funnel, where many candidates enter the pipeline, but few reach the final milestone. GEP considered only the following milestones: application, admission and registration. In the future, additional milestones considered relevant may be added to this model.

Step 2: Choosing the Data Mart Grain

The grain of the accumulating snapshot to track the applicant's lifecycle is one row per potential student [4]. This granularity represents the lowest level of detail

captured when the student enters the pipeline. The candidate's state is updated in the fact table row with the information collected while he/she progresses through the application, admission and registration milestones.

Step 3: Identifying and Conforming the Dimensions

The candidate dimensions defined for the IST Admission Data Mart are: Time (including the academic year hierarchy), Student (encompassing the candidates' attributes), Admission (including admission types and contingents), Geography, Degree and Department.

The Geography dimension was isolated from the Student dimension to enable more intuitive geography-driven analysis. Assuming that the natural evolution of the SADIA system will be the implementation of a Data Mart to support the decisions of the Scientific Council, a conformed Geography dimension will be the most helpful.

Step 4: Choosing the Facts or Measures

As stated in Section 2.3, there are three types of measures or facts: (1) fully additive; (2) semi-additive; and (3) non-additive. Apart from this classification, in the SADIA project we have identified elementary aggregated and derived measures or facts [10]. **Elementary facts** designate the fundamental or key business measures. **Aggregated facts** are measures with some level of aggregation. **Derived facts** are those measures calculated from the values of other elementary or aggregated measures.

The elementary measures considered for the IST Student Admission fact table are presented in the star schema model of Figure 5. Dates (e.g., the Application Date) are treated as *role-playing dimensions*, using surrogate keys to overcome the inevitable unknown dates when we first load the row. The measures *applicant*, *admitted* and *registered* are factless (i.e., with values 0 or 1). This choice represents a design optimization to facilitate the frequent counting of the number of students in each of the process milestones, that is, the number of applicants, admitted and registered students.
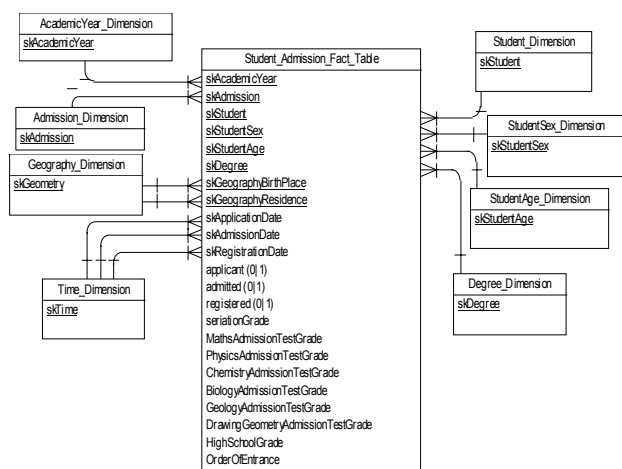

Figure 5: IST Student Admission Star Schema

Step 5: Storing Pre-calculations in the Fact Table

In order to identify the set of pre-calculations to store in the fact table, we gathered all derived and aggregated facts. In practice, we have anticipated the beginning of one step of the Physical Design, i.e., the definition of the Aggregation Strategy. This modelling step will be described in Section 6.

Step 6: Enriching the Dimension Tables

In this step, dimension tables are enriched with all the attributes identified for analysis constraints.

Time Dimension and Academic Year Dimension

The Time Dimension is based on the traditional Time Dimension proposed by Kimball [1]. Initially, the academic year was considered a parallel hierarchy inside the Time Dimension similar to the fiscal year hierarchy, as represented in Figure 6. However, since the majority of the user queries intended to analyse data for a particular academic year and semester, a new dimension for the Academic Year was created. Another argument for this decision is related to the navigation in aggregated models (that will be presented in the following sections). These models are meaningful only in the context of one Semester/Academic Year and not for a particular day (i.e., the grain of the Time Dimension).

Figure 6 represents the detailed diagram produced for the Time Dimension. This kind of diagram highlights dimension hierarchies and the expected cardinality of each hierarchy level. In the diagram, the relationships between attributes specify the possible navigation or *drill* paths. Each record in the Time Dimension has a surrogate key generated sequentially from a fixed reference date (e.g., the 1st of January of 1920).
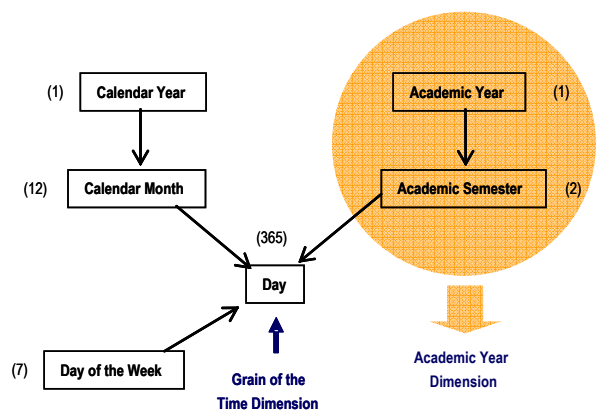

Figure 6: Detailed diagram of the Time Dimension

Degree Dimension

Currently, IST offers twenty-two 5-year-long undergraduate degrees, some of them with branches or profiles of specialization. The partition into degree years is common to all undergraduate degrees. Figure 7 represents the detailed diagram of the Degree Dimension. The set of attributes considered for this dimension is listed in Table 6.
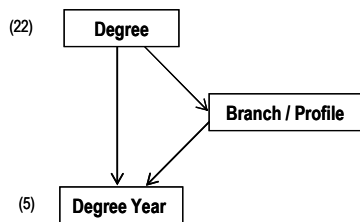
Figure 7: Detailed diagram of the Degree Dimension

Dimension Attributes	Example Data
Degree Name	Undergraduate Degree in Electrical Engineering and Computer Systems
Degree Acronym	LEEC
Degree year	2
Branch/Profile	Control and Robotics
Branch/Profile acronym	CR
Degree Coordinator	Professor XZ
Maths admission test	Required
Physics admission test	Required
Chemistry admission test	Not required
Geology admission test	Not required
Biology admission test	Not required
Drawing and Geometry admission test	Not required
Coordinating Department #1	3
Coordinating Department #2	1
Coordinating Department #3	1
Coordinating Department #4	1
Coordinating Department #5	1

Table 6: Attributes of the Degree Dimension

Department Dimension

The Department Dimension exhibits one hierarchy partitions departments into sections. This dimension still needs to be conformed, but we need to model the Scientific Council business processes to accomplish this task. In this version of the SADIA system the attributes considered for the Department Dimension are: (1) Department name; (2) Department acronym; (3) Head of the Department; (4) Department Section name; (5) Department Section acronym; and (6) Department Section Coordinator.

Student Dimension

The Student Dimension corresponds to the traditional DW Customer Dimension. Therefore, some Customer Relationship Management (CRM) techniques can be applied to the design of this dimension. The basic idea of CRM is to determine an integrated and unique view of each customer, enabling the establishment of long-lasting and profitable business relations [4]. CRM objectives are not confined to maintain the business relations with the most profitable customers of the organization. It is also important to turn current non-profitable customers into profitable ones. From the point of view of a University it is extremely important to develop "attraction mechanisms" to capture the best students in the market, since the number of applicants is declining. By defining an integrated and unique view of the student (as a customer or client!) the University may develop mechanisms to maintain profitable relations with the

students. Moreover, it becomes possible to identify early on situations of risk, e.g. a student with low classifications, and act before the probable drop-out occurs.

The first CRM modelling mechanism used in the Student Dimension concerns the parsing of names and addresses. The way operational systems deal with customers' names and addresses is usually too simplistic to be useful in Data Warehousing [4]. The most common design with generic columns for names (name-1 to name-3) and addresses (address-1 to address-6) is useless for analyzing and segmenting the behaviour of customers. Thus, instead of using generic fields, names and addresses should be divided into their elementary components. Attributes should be standardized, e.g., "St" should be replaced by "Street". It is also necessary to verify the correctness of information, for instance, checking if the local or district of a postal code is correct. This is a complex task since Portuguese names and addresses are usually strongly unstructured. In the current version of the SADIA system we did not tokenize and standardize the address attributes.

Table 7 presents the attributes of the Student Dimension. The more detailed is the description of our *customers*¸ the more robust the Student Dimension will be and consequently more interesting analysis become possible.

Some attributes of the Student Dimension were modelled as **mini-dimensions**, namely gender and age. The creation of a new Student-Gender Dimension enhances the performance of the user queries, since most of them impose constraints on the student gender. The Student-Age Dimension is a design optimisation that enables the specification of several age ranks, used in standard user reports, without over sizing the Student Dimension.

Geography Dimension

The Geography Dimension has been enriched with the Portuguese Administrative Geographic Divisions (NUTS that stands for *Nomenclatura das Unidades Territoriais para fins Estatísticos*, in Portuguese) standardized by the National Statistic Institute [9]. These attributes will enable more intuitive geographic-related analysis, such as: "*What is the influence of the origin district of the students in the application classification?*" or "*From which regions in Portugal do most of the IST students come from?*"

Steps 7 and 8 are currently being defined.

6 Physical Design

As explained in Section 2.3, Aggregates are fact tables with some level of summarization, derived from the Data Mart most granular fact table. The elementary fact table, presented in Figure 5, corresponds to the student admission fact table. Each aggregated fact table should be designed in a different physical table, since facts of different grains should not be mixed in the same table [4].

Dimension Attributes	Example Data
Salutation	Sra.
Informal name	Elsa
Formal name	Sra. Elsa Cardoso
First and middle names	Elsa Alexandra Cabral da Rocha
Surname	Cardoso
Title	Engineer
Address	Rua XPTO de Baixo, Lote 31, 2º Esq
Local	Sassoeiros
Postal code (4 digits)	2775
Postal code (3 digits)	785
Complete postal code	2775-785
Postal code local	Carcavelos
Home phone	211111111
Mobile phone	911111111
Email	Elsa.cardoso@inesc-id.pt
Alternative email	Elsa.cardoso@iscte.pt
Web site	www.esw.inesc-id.pt/~eac
Date of birth	29-12-1970
Year of birth	1970
Name of the father	Manuel Cardoso
Name of the mother	Maria Cândida Rocha
Nationality	Portuguesa
Country of birth	Moçambique
Civil state	Single
Identification document number	11111111
Identification document type	Bilhete de Identidade
Identification document emission locality	Lisboa
Identification document expiration date	02-08-2006
Fiscal number	222222
Application Option	1

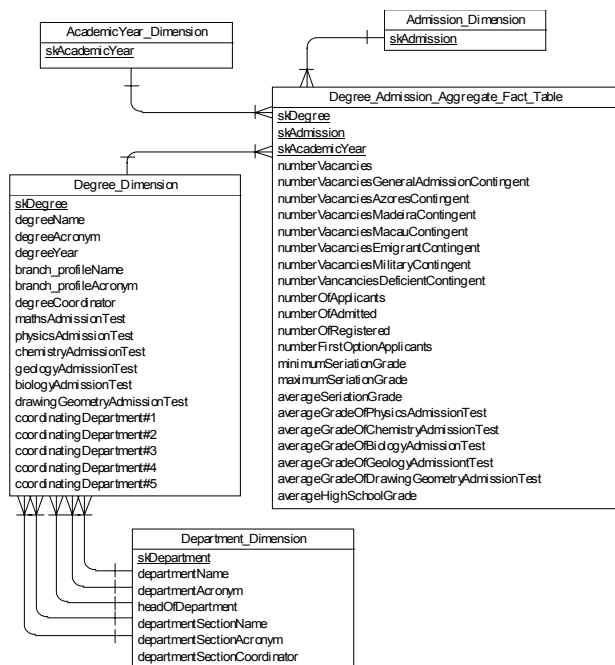Table 7: Attributes of the Student Dimension

Figure 8: Aggregated Model for the Admission Process for a Undergraduate Degree

Table 8 presents the set of derived and aggregated facts identified for the Admission process. The users were questioned about the real usage pattern of these facts. The facts were then parameterised in terms of usage frequency and required computation effort. This classification (from 1 to 5) was the basis for determining whether each fact should be stored.

Based on this study, we have created a new aggregated fact table for the Admission Process for an Undergraduate Degree. The grain of this fact table, presented in Figure 8, is one record for each undergraduate degree offered by IST per academic year.

STUDENT ADMISSION PROCESS	TIME	ACADEMIC YEAR	STUDENT	STUDENT-GENDER	STUDENT-AGE	GEOGRAPHY	DEGREE	DEPARTMENT	COURSE	ADMISSION	USAGE FREQUENCY	COMPUTATION EFFORT	INCLUDE?
No of candidates		X					X	X		X	1	2	Y
No of admitted students		X					X	X		X	1	2	Y
No of registered students		X					X	X		X	1	2	Y
No of first option candidates		X					X	X		X	1	2	Y
Minimum application classification		X					X	X		X	1	2	Y
Maximum application classification		X					X	X		X	1	2	Y
Average application classification		X					X	X		X	1	3	Y
Average classification of Maths Admission Test		X					X	X		X	1	3	Y
Average classification of Physics Admission Test		X					X	X		X	1	3	Y
Average classification of Chemistry Admission Test		X					X	X		X	1	3	Y
Average classification of Geology Admission Test		X					X	X		X	1	3	Y
Average classification of Biology Admission Test		X					X	X		X	1	3	Y
Average classification of Drawing and Geometry Admission Test		X					X	X		X	1	3	Y
Average High School final classification		X					X	X		X	1	3	Y
Occupation rate = (No of admitted students/ No of vacancies)		X					X	X		X	1	1	N
Ratio No of candidates/ No of vacancies		X					X	X		X	1	1	N
Ratio No of first option candidates/ No of vacancies		X					X	X		X	1	1	N
No of unregistered students= (No of admitted – No of registered)		X					X	X		X	1	1	N

Table 8: Aggregated and Derived Facts

Validating the Dimensional Model

Although the methodology does not preclude a validation phase, it is fundamental to check if the logical and physical models completely respond to user analysis queries. The validation process of the dimensional model has been performed, query-by-query, using validity matrixes built with a Microsoft-Excel spreadsheet.

7   Conclusions

This paper presented the SADIA system [8], which is the Academic Information Decision Support System of Instituto Superior Técnico. We have defined the business requirements and reported on the dimensional modelling achievements concerning the first release of the DSS.

Apart from validating the dimensional model, the key business users have also approved the design decisions. We also concluded that users completely understood star schema models, since they easily realized the business advantages of decision support systems.

Current ongoing work encompasses the Technology and Application tracks. A business intelligence tool will be used to develop the end-user application. Future work includes the deployment of the first SADIA prototype.

Acknowledgement

The authors would like to thank José Barateiro for his valuable comments to this paper.

References

[1] Kimball, R., and Reeves, L., and Thornthwaite, W. The Data Warehouse Lifecycle Toolkit. John Wiley & Sons Inc., New York, 1998.

[2] Silva, A. Rito. The FENIX Project. Technical Report. IST, 2002.

[3] Booch, G., Rumbaugh, J. and Jacobson, I. The Unified Modeling Language User Guide. Addison-Wesley, 1999.

[4] Kimball, R. and Ross, M. The Data Warehouse Toolkit – the Complete Guide to Dimensional Modeling, 2ª Ed. John Wiley & Sons Inc., 2002.

[5] Kimball, R. A Dimensional Modelling Manifesto. DBMS online, Data Warehouse Architect, 1997.

[6] Kimball, R. Letting the Users Sleep, Part 1. DBMS online, Data Warehouse Architect, December 1996.

[7] Kimball, R. Letting the Users Sleep, Part 2. DBMS online, Data Warehouse Architect, January 1997.

[8] Cardoso, E. Sistema de Apoio à Decisão para a Informação Académica do Instituto Superior Técnico. Master Thesis, 2003 (in Portuguese).

[9] INE, Instituto Nacional de Estatística. Nomenclaturas – Divisão Administrativa, 2002 (in Portuguese).
http://www.ine.pt/prodserv/nomenclaturas/refter/div_admin.htm

[10]   Soares, J. Soluções de Data Warehousing – Fundamentos Teóricos, Metodologias e Práticas de Implementação. Master Thesis, 2002 (in Portuguese).

Integrating VLE and Library Systems: Opportunities and Challenges

Clare Uhomoibhi
Department of Information Services,
University of Ulster, Shore Road,
Newtownabbey BT37 0QB, United Kingdom
email: c.uhomoibhi@ulster.ac.uk

Alan Masson
Institute of Lifelong Learning,
University of Ulster, Shore Road,
Newtownabbey BT37 0QB, United Kingdom
email: aj.masson@ulster.ac.uk

Lyn Norris
EduServ,
Queen Anne House, 11 Charlotte Street,
Bath BA1 2NE, United Kingdom
email: lyn.norris@eduserv.org.uk

This paper describes the potential benefits of VLE - Library system integration to learners, libraries and content providers and examines the role of emerging authentication technologies in facilitating the practical realisation of such integration. It reports on the activities of the 4i Project (Interoperable Institutional, Integrated Implementation) led by the University of Ulster in collaboration with WebCT, Talis and Athens. This project is funded by the Joint Information Systems Committee (JISC) under the Linking Digital Libraries and Virtual Learning Environments (DiVLE) programme which aims to explore the technical, pedagogical and organisational issues of linking digital library systems and virtual learning environments.

1 Introduction

The University of Ulster is the largest higher education establishment on the island of Ireland, with over 21,000 students. It has four physical campuses spread over a distance of 100 miles and a virtual campus, Campus One, which was launched in 2001.

Since 1999 the University of Ulster has sought to take a strategic approach to the development and implementation of e-learning [14]. To support this work, it has developed an institutional e-learning infrastructure comprising a consolidated server system, new video conferencing facilities and the procurement of an institutional VLE (WebCT). At the same time, the University established the Institute of Lifelong Learning, with responsibility for promoting e-learning across the institution. This strategic approach to the development and support of e-learning has allowed the University to rapidly roll out fully online post graduate masters courses (17 at time of writing) and to support an increasing number of traditionally taught programmes and modules. As a matter of policy, every University of Ulster module is currently provided with a student populated WebCT course area containing, as a minimum service, a calendar and a set of dynamic library links as described later in this paper.

This institutional approach to e-learning brought together a number of relevant stakeholders, in particular library, IT infrastructure and pedagogic support staff.

At an early stage of this initiative, it was noted that the information known about a user in an online course (specifically person and course data) could be used to direct contextual links to library resources and services based on the users' expected needs for that course. This observation initiated the current VLE – Library integration initiative.

1.1 UU Library perspective

On considering the role of the library in the digital age, Pinfield [15] noted "The library is first and foremost a service. It's primary mission is to support the learning and teaching and research activity of its parent institution by providing access to information resources. Subject Librarians can help to ensure that the service is directed at existing user needs and also be instrumental in developing and implementing new services that proactively address changing needs. This applies in the new electronic library environment just as it always has done in the traditional library".

Pinfield's objectives, and the need to ensure library services do not become marginalised by user preference for more convenient generic search engines [13],

provided the University of Ulster with a framework for the development of its VLE – Library integration activities. In particular, the need to ensure that any such system integration went beyond the access and retrieval of documents and resources was identified as being of utmost importance and that effective integration should provide learners with access to the fullest range of library services.

The rapid development of flexible learning pathways for full- and part-time students along with introduction of programmes specifically designed for open and distance learning has produced a student body with very different needs and expectations than would have been the case just a few years ago [6] [11]. The establishment of appropriate mechanisms to support institutional e-learning brought together a number of key stakeholders, in particular Library, IT infrastructure and pedagogic support staff. The dynamics of a multidisciplinary team have been found to promote the implementation of a holistic solution to the challenges and opportunities presented when integrating VLE - Library systems supporting the work of Johnston [10] who noted "integrated access to learning materials and the information resources which support them, cannot be achieved without new collaborations and levels of co-operation between information managers, teachers and VLE suppliers".

The University of Ulster Library supports teaching, learning and research by providing access to the scholarly materials required by university staff and students, together with the services needed to support their best use. To ensure equity of service to students, including the growing number of part-time and distance learners, the Library has sought to leverage user and course data residing in its VLE (WebCT) to introduce a range of new services and resources to provide seamless contextual links for VLE users, direct from the course menu bar.

1.2   Library Resources

In the area of resource management, the Library has made significant moves towards subscribing to full text electronic journals and online reference material in support of the increasing number of students who are physically remote from the university campus. In some cases this led to the cancellation of paper-based journal subscriptions however many resources are currently available from the Library in both paper and electronic form offering users a degree of flexibility in accessing required learning material.

In addition to an extensive collection of shelf stock comprising various media, the Library offers access to an array of approximately 270 different electronic information services. These include an extensive range of online databases and retrieval services that encompass approximately 4000 full text electronic journals.

1.3   Library Services

The Library has increased the range of self-service options available to students via online access to their borrower account in the Library Management System (Talis). In addition to checking their account status, renewing loans and reserving items online users can also request Inter Library Loans via the same online interface. Successful requests generate an email notification to the student granting access to the material requested in PDF format.

Added value services offered by the Library include an option through which users can view material recently acquired on behalf of a particular School hence creating an opportunity for further customisation of users' view of the Library. New approaches to document delivery include subscription to the HERON (Higher Education Resources ON-demand) a national service for copyright clearance, digitisation and delivery of selected journal articles and book chapters. [7] The Library has also undertaken a digitisation exercise to provide online access to past examination papers. Such initiatives increase flexibility in terms of provision of and access to learning material.

1.4   Library Support

A range of training initiatives, along with the provision of web-based user guides and support material, direct students to recommended resource discovery tools and furnish them with the knowledge and skills required for independent study. It is of strategic importance that this traditional role of the Library is maintained and adapted as required as librarians find themselves working in a Library without walls supporting users in a Virtual Learning Environment (VLE).

This paper describes how the Library, in collaboration with colleagues throughout the university, addressed the challenge of integrating access to relevant learning resources, services and support for students in our virtual campus – CampusOne - whilst continuing to provide for the needs of traditional on-campus students.

2   Background to the University of Ulster VLE - Library integration model

2.1   Overview

During the rollout of its institutional VLE and its integration with library resources and services, the University of Ulster Library has worked closely with the Institute of Lifelong Learning to ensure that students are not channelled directly from the VLE to a variety of abstract online resources, bypassing the Library.

The promotion of effective liaison between librarians and academics regarding the provision of comprehensive online resource lists for all courses pointing to appropriate tools and resources will ensure that courses are adequately supported and equity of service is maintained.

In 2001, the Library introduced a sophisticated web-based resource management facility (TalisList) to provide a controlled gateway to a range of appropriate

local and remote resources tailored to students course-specific requirements. Work was later undertaken to tightly integrate TalisList with WebCT supporting distance learning by providing coherent access to resources. These resources may comprise the various elements of a hybrid library [3] with links to online bibliographic databases, full-text electronic journals, specific journal articles or book extracts in PDF format, e-books, networked CDROMs, audio/video clips, useful internal and external websites along with direct links to the Library catalogue to display details of a specific item of shelf stock such as books or other non-book media.

2.2 Talislist resource management system

Since its introduction to University of Ulster staff in September 2001, the TalisList system was promoted to academics as a 'resource' rather than 'reading' list management system. The introduction of online module resource lists, and the institutional rollout of WebCT, presented the Library with an opportunity to entice academics away from the rather one-dimensional idea of a reading list (often a static list of library shelf stock) and encourage them to embrace the idea of a dynamic resource list, rich in both content and functionality. In doing so the Library was able to raise awareness of the range of electronic information resources available for use. Given the associated cost of acquiring such resources it is important that they are used to best advantage by all students, not just those studying at a distance.

TalisList forms an integral part of the university's e-learning infrastructure and enables the Library to provide a contextualised gateway to the plethora of learning resources available to students. These resources are hand-picked by academics in consultation with Librarians bringing added value to the online learning experience whilst providing valuable support to traditional on-campus students. This system is centrally hosted by the Library in collaboration with academics who provide the content in a structured format. Notes of guidance on the use of particular resources from academics and recommendations from Librarians are incorporated within the resource lists thereby providing the students in an online environment the same degree of customisation they would encounter in a traditional 'bricks and mortar' learning environment.

2.3 VLE - Library integration model overview

At an early stage in the establishment of an institutional e-learning infrastructure, it was noted that the information known about a user in an online course (specifically person and course data) could be used to direct contextual links to library resources and services based on the users expected needs for that course. In order to achieve this, it was recognised that a means of extracting this user profile data and passing it to the library would need to be developed and that an

appropriately granular view of library resources and services would be required. It was also recognised that an integrated approach to authentication would facilitate seamless user access to protected resources such as databases and online journals. In early 2002, the University of Ulster approached its Library systems and VLE providers (Talis and WebCT) and the Athens Access Management Service to initiate an institutional VLE – Library integration pilot study. This work successfully demonstrated the ability to seamlessly integrate course content in an institutional VLE (WebCT) with relevant resource lists in a resource management system (TalisList) and provide onward access to students' required learning resources using Athens Devolved Authentication.

The sharing of person and module data between systems also offered potential to grant WebCT users pre-authenticated access to selected self-service options within the Library Management System (Talis).

2.4 The 4i Project

This concept has since been further developed and is now being mainstreamed across the University of Ulster with the support of funding from the Joint Information Systems Committee. The JISC-funded '4i Project' (Interoperable Institutional, Integrated Implementation) continued the University's collaboration with WebCT, Talis and Athens in this exciting field. The project was funded under the Linking Digital Libraries and Virtual Learning Environments (DiVLE) programme which aimed to explore the technical, pedagogical and organisational issues of linking digital library systems and Virtual Learning Environments.

One of the key objectives of the 4i project was to explore the hypothesis that integration of VLE and Library authentication processes can simplify user education, increase usage of electronic resources, reduce helpdesk queries and streamline library business processes. An integral part of the VLE - library system integration work of the University of Ulster was the integration its WebCT and Athens authentication processes.

The development of Athens Devolved Authentication enabled users accredited locally by the University of Ulster to acquire the necessary authentication credentials to seamlessly access a range of Athens protected resources from a WebCT course or from a subsequently visited library page (i.e. a resource list in Talislist).

The integration of these processes both simplifies the user experience and allows consolidation of helpdesk activities. It provides remote users and distance learners, as well as campus-based users, with a simple and effective means of utilising a range of online services without having to be issued with and challenged for differing sets of credentials.

3   Access Management

Providing appropriate access to resources is key to the business of a Library and authentication as always been an issue, with the humble borrower card as the basic form of 'offline' authentication that is still in use in academic libraries. However Library staff and users are confronted with a growing range of resources to search and gain access to as today's library offers access to resources beyond the confines of its shelves.

A dramatic rise in the number of university students studying at a distance and increasing demands for flexible access to resources as well as courses, combined with the proliferation of electronic resources being offered by suppliers in response to this emerging new market, presents a number of challenges and opportunities for academic libraries. New approaches to resource management and discovery have been adopted. Rather than simply displaying a Library borrower card to borrow a book, in the online learning environment we need a system to verify who an off-campus user is and what rights they have before granting access to the appropriate library resources.

The provision of a transparent, scalable access management process is a key objective for any HE Library. Such a service must be scalable and flexible if it is to meet the institutions' current and future needs. As Lynch [12] noted "In a world of networked information resources, access management needs to be a basic part of the infrastructure, and must not become a barrier to institutional decisions to change or add resource providers."

Students often have to acquire a new set of 'information skills' to enable effective use of the various electronic information services available to them, regardless of remembering which set of access credentials are required to gain access to each resource. There are inevitable language and time zone constraints with regard to providing distance learners with adequate helpdesk support. As a result innovative methods of user education and support such as the 'Follow the Sun 24 hour helpdesk initiative' [16] have evolved during recent years.

Early experience of providing electronic access to Library resources was a world away from the current situation. Access was often IP-based on campus with a generic institutional password available for those students who wanted to, and were able to, access networked Library resources from home. The majority of required resources were paper based and off-campus access to the relatively small collection of electronic resources was not a huge concern as most students visited the campus frequently.

Within a couple of years, the growth in electronic resources increased dramatically and libraries had to quickly adjust to meet the changing demands associated with providing appropriate access to learning materials. There was increased administration of usernames and passwords from a Library perspective and students were often subject to information overload as there was no standard authentication mechanism in place.   As J.

Edwards [5] noted in 1997, "One of the stumbling blocks to promoting the use of electronic journals is the potential plethora of interfaces and delivery mechanisms with which the user may be required to become familiar with."

3.1   Athens Access Management

Athens is a large scale centralised Access Management System in common use, particularly in the UK. It is a major example of a co-operative approach to authentication in the academic Library community, requiring major planning and commitment. Resources who adopt Athens authentication are typically academic research (content) providers. For a full list see http://www.athensams.net.

Athens has the current JISC contract for an Access Management Service for use in UK Higher and Further Education. Athens has a similar contract with the NHS (National Health Service) Information Authority for centralised registration, account management and authorisation facilities, for access to the same set of online academic research material.
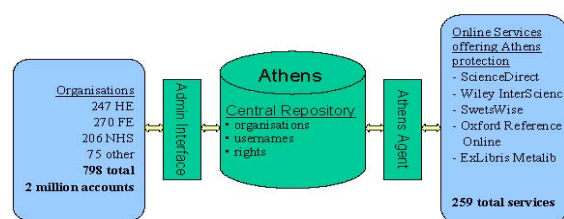

Figure 1: Schematic of Athens Access Management

In its simplest terms, Athens is a central database that contains

- Details of institutions that use Athens, with links to their preferred database of credentials for access to Athens-protected services. These can be held centrally in Athens, or delegated to the institution itself, using a local authentication service. The latter is referred to as Athens Devolved Authentication or AthensDA.
- Usernames and passwords for institutions that wish to use its central database for its usernames and passwords. Athens offers extensive web-based account management facilities for these usernames and passwords. This is delegated account management.
- Authorisation rights for access to the online services. Athens does not sell services, it simply holds the rights information. When an Athens-protected service is sold by the content owner or their agent to an institution using Athens, Athens is simply informed by the accredited agent and the institution is given access to the service.

At the time of writing, there are over 250 Athens protected web services and there has also been a shift from the use of generic institutional passwords to individual user credentials i.e. Athens personal user accounts.

Increased availability of Athens-protected electronic resources in recent years has benefited both users and HE libraries as one set of credentials granted access to an ever-increasing range of resources. For example, major services such as SwetsWise and ScienceDirect use Athens authentication providing a standard method of off-campus access to approximately 4000 full text journals for University of Ulster students. However the administration and secure distribution of Athens usernames and passwords is still a time-consuming and costly task for libraries. The plethora of available online resources requires extensive user education programmes and is often bewildering for learners, resulting in high numbers of helpdesk queries [4].

Secure distribution of Athens credentials can be problematic, especially when dealing with distance learning students. Perhaps due to their isolation from general Library procedures this group of students can also experience particular difficulty in understanding when and how to use their Athens username and password.

A survey conducted amongst University of Ulster Library support staff indicated that the majority of helpdesk queries from distance learners were associated with obtaining and appropriate use of their Athens credentials. This supports the finding that the complexities of authentication and authorisation can inhibit user awareness and utilisation of an academic Library's collection of electronic resources [2], even when a standard mechanism such as Athens personal accounts is in use.

3.2 Athens Devolved Authentication (AthensDA)

One of the key aspects of Athens is that it maintains a central repository of all its users and hence a separate namespace with different Athens usernames. This poses a series of problems to administrators at participating institutions, relating to unnecessary duplication of user record data such as data freshness, correctness and data sharing.

AthensDA [1] solves some of these problems by integrating with an institution's existing user data and authentication frameworks. This extension will allow members of participating institutions to gain remote single sign-on access to Athens authenticated resources using their own institutional credentials transparently with a level of security acceptable to Data Service Providers (DSPs), users and institutions.

AthensDA works by referring a user to some software running in their institutional domain which will authenticate the user and then alias that user an Athens entity which has the correct authorisation permissions for that user. The user is still authenticated by DSPs, using the existing Athens infrastructure.

4 UU VLE - Library integration model

The use of a common data schema across legacy systems is one of the fundamental requirements for successful VLE-Library integration on an institutional scale. The University of Ulster uses the same categorisation schemas (module code and student number) in its Student Records, VLE and Library systems, with the population of each system being generated from a common data source. Initial integration work utilised these two key data elements to provide contextual linkages from WebCT to resource lists (in TalisList) and self-service options (in Talis). This shaping of Library services to the context in which the student is presently operating is achieved generating a cookie containing the student's registration number and module code from WebCT. Each link read and acted on the cookie independently, extracting the module and person data as required.

Successful implementation of this mechanism to pass person and module data between VLE and Library systems led to the application and enhancement of this technology to ease and integrate student access to other Library services and resources that reside outside the two core Library systems (Talis and TalisList), for example past examination papers and HERON documents.

This process was enhanced by the integration of the WebCT and AthensDA login processes to facilitate pre-authentication of VLE users to Athens protected resources. Such use of the authentication functionality of an institutional system (i.e. the VLE) augmented with local verification of Library privileges from an institutional directory service provided an access management model that could enable such Athens pre-authentication to be realised.

The development of such a hybrid AthensDA implementation utilising VLE authentication and Directory Service authorisation was subsequently supported by Athens, who developed it as Local Authentication Assertion (LAA) mode. In the simplest terms, this means that the user only needs one username, for WebCT.

The final aspect of the integration model is the routing of all such linkages through a single 'Library gateway' rather than having embedded "hard coded" links to specific Library targets. This gateway is locally known as the Library Service Point (LSP).

During the VLE – Library integration process, person and module information passed from WebCT is manipulated by the LSP to generate person-, module-, school- and faculty-specific links to various Library resources and services. These contextual linkages between the institutional VLE and Library are automatically embedded into the global navigation bar of WebCT. This approach, where the point of integration (LSP) sits outside the actual systems concerned, maintains system independence in the integration between VLE and Library systems.
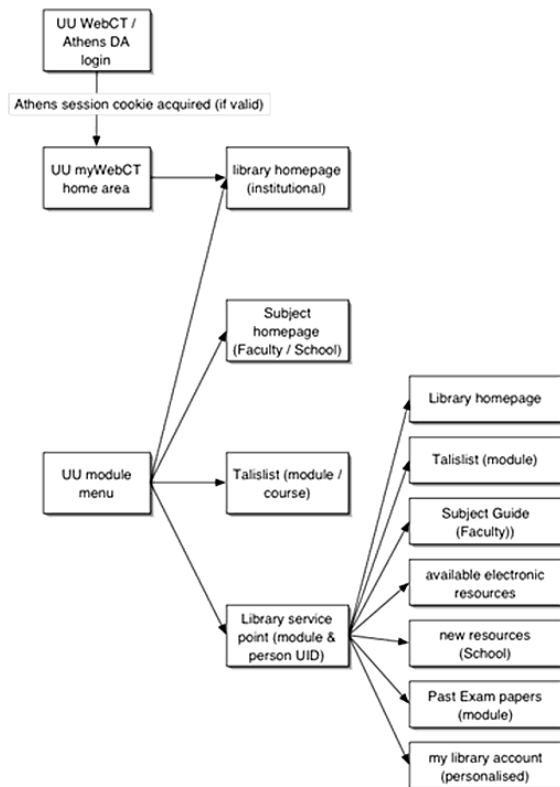
Figure 2: UU Library Service Point Schematic

The Library Service Point (LSP) currently offers students dynamic links to:

- Main Library Website and Catalogue (generic)
- My Module Resource List (module-specific)
- Library Information and Services for My Faculty (faculty-specific)
- My Electronic Information Services (Athens) (person-specific)
- New additions to the Library in My Subject (school-specific)
- Past exam papers for this module (module-specific)
- My Library Account (person-specific)

Figure 3: Typical UU module in WebCT with LSP links in Navigation Bar

All live modules in WebCT have the following links permanently visible in the navigation bar:

- Library (link to Main Library Website and Catalogue)
- Library services (links to LSP)
- Module resources (link to My Module Reading List in TalisList)
- Subject guide (link to Library Information & Services for My Faculty in Library website).

4.1 Benefits of the VLE – Library Gateway (LSP)

The implementation of the LSP as 'middleware' picking up requests from the VLE and routing them onward to the appropriate Library service or resource, depending on the users current requirements, offers a number of benefits to the student as well as those working on maintaining the respective VLE and Library systems.

Learner benefits

Students benefit from seamless access to a variety of customised Library services and resources they would previously have had to use their own search and navigation skills to locate, and individual authentication credentials to access.

The provision of contextual links to specific module resource lists and subject related discovery gateways support both student learning and research needs.

VLE benefits

Requests to a fixed reference point rather than direct addressing to the back-end Library services from the VLE avoids hard coding of links. Any changes to the URL of the Library service/resource requested need only be implemented at the LSP and not in WebCT. The LSP offers flexibility in terms of either providing a direct link or indirect links via a menu of service and support offerings, which can be amended without maintenance issues on the VLE side. This reduces the workload of the instructional designer as they no longer have to worry about incorporating appropriate Library links within their online courses.

References to learning resources are not stored within the VLE, rather they are centrally managed by the Library in collaboration with academics using the TalisList resource management system. This further reduces associated maintenance tasks such as ensuring currency of links and relevance of material that is managed by the Library.

Users can simply locate and access relevant Library resources and services directly from online course pages. This will promote student utilisation of Library resources and research tools.

Library benefits

The LSP is advantageous to the Library as it enables Library services to have a single authoritative and stable source of information. This can then be used to authorise trusted access to other Library services such as past

examination papers or HERON documents on the basis that the user has been authenticated by another UU service (in this case WebCT). The need for a separate system of authentication is effectively eliminated.

The provision of a consistent ergonomic user interface within ALL online module/course areas simplifies user access, reducing user education and helpdesk workloads.

4.2   Evaluation

The University formally enabled its VLE – Library integration in February 2003. Formal evaluation of the integration is ongoing at the time of writing of this paper. Initial investigations indicate:

1. Institutional pre-authentication of UU WebCT users to AthensDA protected resources using the LAA mode was achieved. All WebCT sessions from February 2003 onwards have acquired, where appropriate, AthensDA credentials. In addition, every WebCT user received, upon successful login, a message notifying them of their Athens status and the range of services available to them. During the months of Feb – April, the average number of daily page requests to the WebCT server averaged in excess of 140,000.

2. Automated, contextual integration with a range of Library systems and services leveraging VLE provided user and course information can be implemented on an institutional basis. All UU modules are provided with a WebCT module area which includes the LSP facilitated Library links described earlier. These WebCT module areas are populated from the student record system, providing students with a contextual gateway to the Library for all their modules of study.

3. The VLE – Library gateway was successfully used by learners accessing the Library from off campus. In percentage terms, 55% of LSP activity emanated from outside the .ulster.ac.uk domain. Utilisation of the service by distance and off campus learners can be evidenced by 53% of LSP activity emanating from outside the .uk domain. Detailed analysis of system and server logs is in progress at the time of writing of this paper.

4. Managed access to protected resources can be provided to distance learners without the need for Library specific credentials. Indeed a potentially complex user education programme to describe how distance learners can gain access to such resources has be replaced with a personalised awareness and information service, supported with contextual links on the UU WebCT homepage, WebCT course pages, Library hosted module resource lists and other Library pages. The success of this approach was immediately apparent as Athens related queries from distance learners largely ceased on the introduction

of the Athens LAA process using UU's WebCT system. Residual Athens related queries from distance learning students related to access to resources not yet supported by AthensDA. The opportunity for these queries to re-occur will reduce as more DSPs support AthensDA. It is expected that all DSPs who use Athens will be AthensDA compliant by September 2003.

5. The availability of module resource information to Library systems can be used to inform Library stock management and user education activities.

4.3   Challenges of implementation

In developing and deploying a VLE – Library integration service on an institutional basis a number of challenges were identified at an early stage.

Proliferation of course related roles within e-learning
The proliferation of para-teaching roles within e-learning courses, each requiring access to the VLE, provides a significant access management challenge, as the exact nature of their relationship with the University can, from a Library perspective, be unclear. Some of these users will be eligible under the licence conditions for access to Library resources, others will not. The use of explicit staff categories (such as associate and visitor) to populate both VLE and Library systems can provide some inference to the users Library permissions. However, the users exact status of Library access will be determined by the personal terms that he or she has agreed with the institution. This mapping of personal and group membership credentials with specific institutional (including Library) permissions is at the heart of emerging role based authorisation projects such as Shibboleth. To date, the University of Ulster is only providing Athens access to WebCT users with an existing Library account.

Library business processes
The heart of this VLE – Library integration is the provision of module resource lists. With over 5000 live modules, the creation and maintenance of such large numbers of lists poses a significant challenge. Initial investigations have focussed on the defining consistent list vocabularies and schema and examining the issues surrounding the harvesting and input of large numbers of resource items into Talislist.

The provision of devolved resource input to academics, coupled with librarian approval options provide an opportunity for Talislist to be used as a means of promoting greater academic – librarian collaboration. An output of the 4i project will be an investigation into the issues surrounding the large scale input and management of module resource lists.

Library service that does not 'spoon feed' the learner
The Module Resource Lists hosted in Talislist provide learners with an organised list of resources (paper- based and electronic) selected by academic staff. The level of

detail within each list is appropriate to the requirements of the teaching staff for the module.

The Subject Guide pages are managed by subject librarians and provide a more generic overview of the range of available databases, subject gateways and resources available on a subject by subject basis. These pages provide learners with a useful starting point when embarking on research activities.

The Library Service Point provides learners with a means of obtaining deep links to a range of services; some personal, some subject related and some informative of Library activities.

This combination of complimentary services, supported by academics and librarians respectively, provides a simple interface to a range of Library services and resources appropriate to the needs of the subject, academic and learner.

Interoperability

The current method of WebCT - Library integration developed to date by the University of Ulster lacks a number of key interoperable attributes:

1. If exported to another institution, the course will continue to point to the UU LSP.
2. The integration is realised because both the VLE and the Library use the module code within their systems. Many institutions may wish to utilise other course data (subject, campus etc.) to realise this integration.

An output of the 4i project is the development of explicit use case scenarios to describe how institutions may wish to integrate their VLE and Library systems.

The described integration can readily be adapted to work using a Web Services model. It is the intention of the University of Ulster to develop this approach when the relevant Web Services and interoperability standards have become more fully defined and have broad sector acceptance.

5   Emerging metadata standards and access management technologies

5.1   IMS

IMS Global Learning Consortium, Inc. (IMS) develops and promotes open specifications to facilitate online distributed learning activities such as locating and using educational content, tracking learner progress, reporting learner performance, and exchanging student records between administrative systems. IMS came into existence in 1997 as a project within the National Learning Infrastructure Initiative of EDUCAUSE.

A number of IMS schemas are of relevance to the integration of Library services with online courses hosted within a VLE. In addition to the Content Packaging schema, which provides a means of specifically identifying resources that are utilised in support of a given course, IMS released in March 2003 a Digital Repositories Interoperability (DRI) specification [8]. This specification seeks to "define a specific set of functions and protocols that enable diverse e-learning components to communicate with each other. These functions and protocols draw on XML technologies such as SOAP (Simple Object Access Protocol) and XQuery, and established technologies such as Z39.50, developed by the Library community. The specification acknowledges a wide range of content formats and is applicable internationally to both learning object repositories, as well as to other traditional content sources such as libraries and museum collections."

The development of these schemas is ongoing, with the need for effective interoperability between VLE's and institutional Libraries a key objective. The availability of appropriate metadata standards will facilitate system integration across multiple vendor combinations and allow institutions to readily implement effective VLE – Library integration services that are sustainable beyond their own context.

5.2   Shibboleth

Shibboleth [9] is an emerging Web authorization architecture and software, and is one of the key developments in the field of authentication. It is an Internet2 / MACE (Middleware Architecture Committee for Education) project, whose objectives are "to define the architecture and message protocols for the secure management of authorisation information that can be used in access control decision making" along with practical technologies, and open source implementations. The architecture emphasises federated administration, access control based on attributes rather than identity, and active management of privacy to provide a scalable and extensible framework for inter-institutional authorization.

The architecture of Shibboleth defines roles for the institution and for the resource provider. The institution is responsible for

- A local authentication service providing federated or devolved authentication
- An Attribute Authority providing approved attributes of individuals to external services
- A Handle Server to communicate with Resource Providers. (Provided by Shibboleth)

The Resource Provider is responsible for

- Directing the user to his institutional authentication system and receiving in turn a 'handle' or key to the user's attributes
- Making authorisation decisions based on the user's attributes

Shibboleth deliverables include an architecture definition, a set of message passing specifications, a set of sample code and a reference implementation.

Athens are committed to integration with Shibboleth, so that institutions registered with Athens can access Shibboleth protected resources, and conversely that

institutions who adopt Shibboleth can access Athens protected resources.

The availability of this Athens service to the UK HE sector will provide institutions a robust approach to implementing access management for their users and will facilitate the development of consortia and inter-institutional learning and research initiatives with a user authentication process that is both simple for the user and auditable for DSPs.

6 Conclusions

The benefits of an integrated, interoperable, institutional approach to VLE and Library system implementation, typified by the 4i project, are multi-faceted.

6.1 Streamlining of Library business processes

The implementation of AthensDA addresses the root cause of a large proportion of Athens-related helpdesk queries by firstly removing the need for students to remember an Athens username and password and secondly relieving Library staff of administrative tasks associated with maintaining the external central Athens repository in line with institutional student records and human resource systems.

With full implementation of AthensDA there is no longer a requirement to synchronise internal legacy systems with the central repository of accounts hosted by Athens. Rather the institution maintains its own directory service (e.g. LDAP) which is used by the Athens as a trusted data source for the verification of legitimate members of the university. This reduces the administrative burden on the Library of generating, maintaining and distributing large volumes of Athens personal user accounts.

The use of a resource management system results in a single authoritative source of information on the Library resources required by various courses. Such information can be used to inform an academic Library's resource management strategy. For instance, it is possible to determine how many courses make use of a particular resource. For paper-based resources this enables additional copies to be ordered as required and for electronic resources usage can be traced back to Schools and Faculties to enable costs to be attributed accordingly.

6.2 Simplified user education

Integration of authentication mechanisms, using the LAA mode of AthensDA, eliminates the need for students in an institutional VLE to enter an additional username and password in order to access a large proportion of the Library's electronic resources. This development redefines the role of Library support staff to some degree. The focus shifts from one of user education ("You need to use this set of credentials to access this set of resources") to raising user awareness of what they can access (i.e. timely notification of what resources they can access). This has led to reduced helpdesk enquiries and increased usage of a Library's electronic information services.

6.3 Reduction in helpdesk queries

Prior to the implementation of the VLE-Library gateway (LSP) and AthensDA at the University of Ulster, Library support staff reported a large number of helpdesk enquiries relating to Athens personal accounts. The majority of such queries were from off-campus distance learning students and survey results indicate that these were the most problematic and time-consuming for support staff to resolve.

The end users' relative isolation from Library procedures and practices coupled with the inherent complexities managing Athens credentials for a large user population compounded these problems. Queries relating to verification of usernames and passwords and the appropriate use of these credentials were relatively straight-forward to resolve, providing the time-zone and language differences could be overcome. However helpdesk staff could not always offer an immediate solution to certain queries, such as those relating to currency of data held in the central repository of Athens accounts. These had to be forwarded to systems administration staff for further investigation and where therefore more time-consuming which has a negative impact on the user experience.

6.4 Increased usage of electronic resources

The introduction of a resource list management system (TalisList) to provide coherent access to support material was a catalyst for improved liaison between Library and academic staff. This led to increased awareness amongst academic staff of the full range of electronic information services available to support teaching and learning, which in turn promoted student use of these resources.

Statistics show a general upward trend in the usage of electronic resources, especially those that have recently become Athens-protected, for example, ScienceDirect which offers access to a range of full text electronic journals. There has been a three fold increase in the usage of this particular resource by University of Ulster students during the past year. Statistics show a 93% increase in usage since the institutional launch of the LSP and AthensDA.

References

[1] EduServ, 'Athens Devolved Authentication' (2003) [http://www.athensams.net/development/devolved_a uthentication url checked 08/10/2003]

[2] D. Cohen. 'Course-Management Software: Where's the Library?', EDUCAUSE Review, Volume 37, No. 3, pp 12-13 (2002). [http://www.educause.edu /ir/library/pdf/erm0239.pdf checked 08/10/2003]

[3] M. Corcoran. 'The Hybrid Library: Revolution or Evolution?'. Presentation made at HEAnet User

Group for Libraries AGM. (2003). [http://lirgroup.heanet.ie url checked 08/10/2003]

[4] C. Edwards. 'Change and Uncertainty in Academic Libraries', Ariadne, Volume 11 (1997) [http://www.ariadne.ac.uk/issue11/main url checked 08/10/2003]

[5] J. Edwards. 'Electronic journals: Problem or Panacea?', Ariadne, Volume 10 (1997) [http://www.ariadne.ac.uk/issue10/journals/intro.html url checked 08/10/2003]

[6] A. Gibson, R. Newton, D. Dixon. 'Supporting open and distance learners', Library Review, Volume 48, No. 5, 219-231, (2000)

[7] E. McCubbin 'HERON - opportunities for Further Education in copyright clearance and digitised course readings', CofHE Bulletin, Autumn 2002.

[8] [8] IMS Global Learning Consortium, 'IMS Digital Repositories Specification' (2003) [http://www.imsproject.org/digitalrepositories url checked 08/10/2003]

[9] Internet2, 'Shibboleth Project' (2003) [http://shibboleth.internet2.edu/ url checked 08/10/2003]

[10] P. Johnston. 'After the big bang: forces of change and e-Learning'. Ariadne. Volume 27. (2001) [http://www.ariadne.ac.uk/issue27/johnston/ url checked 08/10/2003]

[11] M.M. Kazmer. 'Distance education students speak to the library: here's how you can help even

more', The Electronic Library, Volume 20, No. 5, pp. 395-400 (2003)

[12] C. Lynch. 'Access Management for Networked Information Resources', CAUSE/EFFECT Journal, Volume 21, No. 4. (1998) [http://www.educause.edu/ir/library/html/cem9842.html url checked 08/10/2003]

[13] N. McLean. 'Library Services for a Managed Learning Environment'. Macquarie University conference paper. (2000) [http://www.lib.mq.edu.au/conference/mclean/managed url checked 08/10/2003]

[14] A. Newman. 'Meeting the Mission: E-Learning Implementation Stories for Twelve Postsecondary Institutions', Eduventures Case Study Report (2001). [http://www.eduventures.com/research/industry_research_resources/mission.cfm url checked 08/10/2003]

[15] S. Pinfield. 'The Changing Role of Subject Librarians in Academic Libraries', Journal of Librarianship and Information Science, Volume 33 (1) (2001).

[16] J. Sykes. "London School of Economics: Follow the Sun- the three continents helpdesk". UCISA Good Practice Award submission. (2002) [http://www.ucisa.ac.uk/groups/tlig/conf/userserv02/lse.doc url checked 08/10/2003]

Developing a Quality Culture for Digital Library Programmes

Brian Kelly and Marieke Guy
UKOLN, University of Bath,
BATH, UK
E-Mail: {B.Kelly,M.Guy}@ukoln.ac.uk

Hamish James
AHDS, Kings College
London, UK
E-Mail: Hamish.James@ahds.ac.uk

In this presentation the authors describe approaches for the development of quality assurance (QA) procedures for a digital library programme. The authors argue that QA procedures are needed in order to ensure that deliverables from digital library programmes will be interoperable and give be easily deployed and repurposed. The adoption of open standards is acknowledged as essential in digital library programmes but in a distributed development environment it can be difficult to ensure that programme deliverables actually implement appropriate standards and best practices. The authors describe the approaches to the development of a quality culture based on encouragement of use of QA by project holders in one digital library programme which is funded by the JISC in the UK.

1 Background

The World-Wide Web is accepted as the key delivery platform for digital library services. The Web promises universal access to resources and provides flexibility, including platform- and application-independence, though use of open standards. In practice however, it can be difficult to achieve this goal. Proprietary formats are appealing and, as we learnt during the "browser wars", software vendors can promise open standards while deploying proprietary extensions which can result in services which fail to be interoperable. Developers can be unsure as to which standards are applicable to their area of work: there is a danger that simple standards, such as HTML, are used when richer standards, such as XML, could provide greater interoperability.

The JISC (Joint Information Systems Committee) has funded a QA Focus post which aims to ensure that projects make use of QA (quality assurance) procedures which will help ensure interoperability through use of appropriate standards and best practices.

A summary of the work of QA Focus is provided in this paper. The paper describes the background to IT development in the UK's Higher Education community, the role of standards and the approaches taken by QA Focus. The paper concludes by outlining future work for QA Focus and the potential for use of similar approaches by other digital library programmes.

2 IT Development Culture

The UK's Higher Education community has a culture which is supportive of open standards in its IT development programmes. Within the eLib Programme, for example, the eLib Standards Guidelines [1] defined the standards funded projects were expected to implement.

Although the Standards Guidelines document was available shortly after the start of the programme, compliance was not enforced. There was recognition of the dangers of enforcing standards too rigidly in those early days of the Web: if the programme had started a few years earlier use of Gopher could well have been chosen as the standard delivery mechanism! In addition the UK Higher Education community had previously attempted to standardise on Coloured Books networking protocols, which subsequently failed to be adopted widely and were eventually superceded by Internet protocols.

The eLib programme encouraged a certain amount of diversity: this approach of letting a "thousand flowers bloom" was probably appropriate for the mid-1990s, before it was clear that the Web would be the killer application which, with hindsight, we recognise that it is. This approach also reflected the culture of software development in a HE environment, in which strict management practices aren't the norm and there has been a tendency to allow software developers a fair amount of freedom.

Nowadays, however, there is increased recognition of the need to have a more managed approach to development. The Web is now recognised as the killer application. Project deliverables, which are often Web-based, can no longer be treated as self-contained services – there is a need for them to interoperate. Also stricter compliance with standards will be needed: Web browsers have been tolerant of errors in HTML resources, but this

will be different in a world in which "Web Services" technologies will be reliant on well-structured resources for machine processing. Finally, JISC has moved on from a research and experimental approach and is now funding programmes in which project deliverables are normally expected to be deployed in a service environment.

3   The JISC Information Environment

The JISC's Information Environment (IE, formerly DNER) [2] seeks to provide seamless access to scholarly resources which are distributed across a range of providers, including centrally-funded JISC services, commercial providers and the institutions themselves. The *Standards and Guidelines To Build A National Resource document* [3] was written to define the standards which form the basis for the IE. The standards document is supported by an IE Architecture [4] which describes the technical architecture of the IE.

The JISC has funded a number of programmes in order to develop the IE, including 5/99 [5] which was followed by the FAIR [6] and X4L [7] programmes.

4   The QA Focus Post

JISC has recognised that there is a need for the JISC-funded programmes to be supported by a post which ensures that projects comply with standards and best practices. The QA Focus post has been funded for two years (from 1 January 2002) to support the JISC 5/99 programme. Initially the post was provided by UKOLN (University of Bath) and ILRT (University of Bristol), but, following a decision to refocus on other areas, in January 2003 ILRT were replaced by AHDS (Arts and Humanities Data Service).

5   Approaches To QA

QA Focus aims to provide a support service to 5/99 projects: the emphasis is on advice and support, based on close links with the projects, rather than a policing role. An important deliverable will be the development of a self-assessment toolkit which can be used by the projects themselves for validation of the project deliverables.

QA Focus is addressing a range of technical areas which include digitisation, Web (including accessibility), digitisation, metadata, software development and service deployment.

The areas of work which are being carried out by QA Focus include:

- Providing advice on standards and best practices.
- Carrying out surveys across projects, looking at compliance with standards and best practices.
- Commissioning case studies which provide examples of best practices.
- Providing documentation on best practices, approaches to compliance checking, etc.
- Developing a Self-Assessment Toolkit

Although QA Focus places an emphasis on its role in supporting projects in developing their own QA procedures, in cases of severe interoperability problems QA Focus will be expected to make contact with the project concerned and seek to ensure that concerns are addressed. If this does not result in a satisfactory solution, the issue will be passed on to the JISC.

6   QA Focus Work To Date

6.1   Links With Projects

A number of workshop sessions have been held with a selection of the projects. The first two workshops aimed to obtain feedback from the projects on (a) the Standards document, (b) implementation experiences and (c) deployment of project deliverables into a service environment.

The workshops provided valuable feedback which has helped to identify key areas which need to be addressed. Useful information was obtained about the Standards document including a lack of awareness of the standards document in some cases, concerns over the change control of the document (since new standards may be developed and other standards may fail to gain acceptance) and some uncertainties as to the appropriateness of some of the standards and deployment difficulties in other cases, especially projects which were reliant on third party software development of existing systems which cannot easily be modified. The feedback on implementation experiences raised several predictable issues, including the poor support for Web standards in many widely-used browsers. The lack of a technical support infrastructure was highlighted by several projects, mainly those based in academic departments or in smaller institutions.

6.2   Surveys

A meeting of 5/99 projects was held at the University of Nottingham, on 30 October - 1 November 1 2002. Prior to the meeting QA Focus carried out a survey of various aspects of 5/99 project Web sites. The survey findings [8] were made available and formed the basis for discussions at the QA Focus workshop sessions.

The surveys made use of a number of freely available tools, all of which had a Web-interface. This meant methodology was open and tools could be used by projects themselves without the need to install software locally. The survey findings were published openly. This allowed examples of best practices to be seen, trends to be monitored and areas which projects found difficult to implement to be identified.

The surveys were complemented by a number of brief advisory documents. In addition a number of case studies have been commissioned which allows the projects themselves to describe their approaches to compliance with standards and best practices and any difficulties they have experienced and lessons they have learnt.

Survey	Tool	Information
HTML Compliance	W3C's HTML validator	Does the home page comply with HTML standards? What DTDs are used?
CSS Compliance	W3C's CSS validator	Does the home page use CSS? Does the CSS comply with standards?
Accessibility	Bobby	Does the home page comply with W3C WAI guidelines?
404 page	Manual observation	Does the 404 page provide navigational facilities and support?
Internet Archive	Manual observation	Is the Web site available in the Internet Archive?
PDA Access	AvantGo	Can the Web site be accessed by a PDA?
XHTML Conversion	W3C's Tidy tool	Can the Web site be converted to XHTML without loss of functionality?
WML Conversion	Google WAP conversion service	Can the Web site be converted to WML without loss of functionality?
HTTP Headers	Dundee's HTTP analysis tool	Are correct HTTP headers sent? What Web server environment is used?
Metadata	W3C's Tidy and RDF validator & UKOLN's DC-dot tools	Is Dublin Core metadata used? Does it comply with standards?

Table 1: Initial QA Focus Surveys

The surveys aimed to establish how well project Web sites complied with standards and best practices. The surveys addressed several areas related to Web technologies including compliance with HTML and CSS standards and compliance with W3C's Web Accessibility Initiatives (WAI) guidelines for project entry points. The HTTP Headers were analysed and details of the Web server platform recorded (together with details of invalid HTTP headers).

As well as testing compliance with well-defined standards the survey also used a number of tools which helped to see if the Web sites allowed repurposing. This included checking availability of project Web sites in the Internet Archive, using the AvantGo service to test access to project Web sites on a PDA and converting the Web site to WML and viewing in the Opera browser which provides a WAP emulator.

The survey also used a simple usability test by reporting on the approach taken to the Web site 404 error

page: whether the 404 error page was branded, provided helpful information and appropriate links, etc.

Metadata embedded in project Web site entry points was tested and any Dublin Core metadata found was validated using a Dublin Core validation tool developed in UKOLN. In addition the Dublin Core metadata was converted to RDF format and then visualised allowing an alternative display of the metadata to be viewed.

6.3 Limitations of Methodology

There is a danger that the publication of the findings can be perceived as threatening to projects. Where the findings indicate lack of compliance with standards or failure to implement best practices projects may point out particular features of their project which the surveys fail to acknowledge, limitations of the tools used the timing of the survey and the available resources.

There is an element of truth in such concerns. The projects are addressing a diverse set of areas, including digitising content, enhancing existing services and software development. The project Web sites will also have a diverse set of objectives, including providing communications with project partners, providing information about the project and providing access to project deliverables. The projects will have different levels of funding, start and completion dates and technical expertise.

Despite these reservations it is felt that significant benefits can be gained from the QA Focus approach. The openness seeks to facilitate dialogue with projects and sharing of best practices. The approach also takes what can be perceived as a dry standards document and places it more centrally in the activities of the projects. It also helps to provide feedback on the standards; if a particular standard has not been adopted this may indicate that the standard is too esoteric or a lack of tools or expertise. Such considerations can be fed back to the authors of the Standards document.

6.4 Documentation

An important role of QA Focus is to ensure that appropriate documentation is provided for the projects. The approach that has been taken is to produce short advisory documents which address specific problems. This approach has the advantage that documents can be written more quickly and can be easily updated.

A summary of the documents published to date is given in Table 2. The documents can be accessed at [9].

Document	Area
Checking Compliance With HTML and CSS Standards	Summarises a number of approaches for checking that HTML resources comply with HTML and CSS standards

Use Of Automated Tools For Testing Web Site Accessibility	Describes tools such as Bobby and summarises the implications of common problem areas
Use Of Proprietary Formats On Web Sites	Provides suggestions for techniques when using common proprietary formats
404 Error Pages On Web Sites	Describes ways of providing user-friendly 404 error pages
Accessing Your Web Site On A PDA	Describes an approach for making a Web site available on a PDA
Approaches To Link Checking	Describes approaches for link-checking, including links to CSS & JavaScript files
Search Facilities For Your Web Site	Describes different approaches for providing search facilities on project Web sites
Enhancing Web Site Navigation Using The LINK Element	Provides advice on use of the HTML `<link>` element to provide enhanced Web site navigation
Image QA In The Digitisation Workflow	Provides advice on QA for images
What Are Open Standards?	Gives an explanation of open standards
Mothballing Your Web Site	Provides advice on "mothballing" a Web site, when funding ceases
How To Evaluate A Web Site's Accessibility Level	Describes approaches for checking Web accessibility

Table 2: QA Focus Advisory Documents

The advisory documents are complemented by case studies which are normally written by the project developers themselves. The case studies provide a solution to the common request of "Can you tell me exactly what approaches I should be using?". It is not possible to provide answers to this question as there are many projects, addressing a range of areas and with their own background and culture. It is also not desirable to impose a particular solution from the centre. The case studies allow projects to describe the solution which they adopted, the approaches they took, any problems of difficulties they experienced and lessons learnt.

The case studies which have been published to date include:

- *Managing And Using Metadata In An E- Journal*
- *Standards and Accessibility Compliance in the FAILTE Project Web Site*
- *Managing a Distributed Development Project: The Subject Portals Project*

- *Creating Accessible Learning And Teaching Resources: The e-MapScholar Experience*
- *Standards for e-learning: The e-MapScholar Experience*
- *Gathering Usage Statistics And Performance Indicators: The NMAP Experience*
- *Using SVG In The ARTWORLD Project*
- *Crafts Study Centre Digitisation Project - and Why 'Born Digital'*
- *Image Digitisation Strategy and Technique: Crafts Study Centre Digitisation Project*
- *Standards and Accessibility Compliance for the DEMOS Project Web Site*
- *Implementing a Communications Infrastructure*
- *Usability Testing for the Non-Visual Access to the Digital Library (NoVA) Project*

Access to these documents is available at [10].

7 Next Steps

Once the QA Focus work in the Web area has been finalised work will move on to a number of other areas including digitisation, multimedia, metadata, software development and deployment into service.

The initial work carried out by QA Focus made use of automated tools to monitor compliance with standards and best practices. In the areas listed above there will be a need to address the use of manual QA processes as well as use of automated tools. For example, the use of correct syntax for storing metadata can be checking using software, but ensuring that textual information is correct cannot be done using only automated processes.

As well as providing advice and support for the projects, QA Focus will also provide advice to JISC on best practices for the termination of the programme and for setting up new programmes. This will include development of FAQs (along the lines of those which have been developed by UKOLN to supports its role in providing the Technical Advisory Service for the nof-digitise programme [11]).

8 Digitisation

Digitisation is the first stage in the creation of a resource, and it represents the link between the analogue and digital worlds. The consequences of poor quality digitisation will flow through the entire project, reducing the value of all later work.

QA for digitisation is therefore very important, but, with the exception of the digitisation of bitmap images [12] there is relatively little advice and support that is accessible to the non-specialist. QA Focus will provide QA guidance for image digitisation, but will also deal with other types of material including text, audio and moving images. We will also link the process of capturing data to the next step, organising the data once it is in digital form, by providing QA for databases and XML applications in particular.

8.1 Workflow

Digitisation typically has some of the qualities of a production line with analogue originals being retrieved, digitised and returned while digital files are created, edited and stored. Rigorous procedures can make sure that this process goes smoothly ensuring that originals are not missed, or mislaid, and the status of digital files (particularly what post-processing has occurred to them and which original or originals they relate to) is tracked. This type of quality assurance for the digitisation workflow is well established for images, and we aim to provide analogous advice for projects digitising other types of material, including checklists and model procedures to follow.

Ensuring consistent quality, and keeping records that demonstrate this, is a vital part of digitisation that indirectly affects interoperability by ensuring, that however the final resource is accessed, users can make informed use of it.

Structured metadata provides a useful mechanism for recording aspects of the digitisation process. QA Focus will review relevant existing and emerging standards. We will also investigate tools for the semi-automatic or automatic creation of technical metadata about digitised material.

8.2 Fitness for Purpose

Before any material is digitised, projects need to define their requirements for the digitised material. QA Focus advocates that projects' take active responsibility for these decisions and avoid allowing the capabilities of available technology set these decisions.

A key part of QA for digitisation is the development of objective, measurable criteria for judging if the digitised material is 'fit for purpose'. Determining what is fit for purpose involves consideration of the acceptable level of accuracy in digitisation in relation to the intended purpose of the digitised material. For example, a low resolution image may be suitable for a Web page, but a product also available on CD-ROM could include higher resolution images. Very similar situations occur with the digitisation of audio and moving images, but we will also address less obviously similar situations, such as rules for the standardisation of place names or the transliteration of text during transcription.

8.3 Rights

Digital files are easily copied and distributed, so it is important for projects to ensure that they have obtained any necessary rights to use the originals. Projects may also want to protect their own rights in the digitised material.

Intellectual Property law is a complex area and QA Focus will not be able to provide definitive answers, but we hope to produce a series of case studies that demonstrate how a project can best minimise the risk of running afoul of copyright infringement. We will liaise with JISC's Legal Information Service [13] which has expertise in this area.

9 Metadata

Metadata has a key role to play in ensuring the projects deliverables can be interoperable. However unless QA procedures are deployed which ensure that the metadata content is correct, the metadata is represented in an appropriate format, complies with appropriate standards and can be processed unambigously we are likely to encounter difficulties in service deployment.

While resource discovery metadata is central to interoperability, we will also investigate requirements for workflow, technical and rights metadata that support the digitisation process and deployment into service.

We are currently planning focus group sessions in which we will obtain feedback from groups with experience in metadata activities. This should provide us with examples of the type of approaches which can be recommended in order to ensure that metadata is interoperable.

Approaches we are currently considering include:

- Checking syntax, encoding, etc. for metadata embedded in HTML and XML resources. This may include documenting the methodology employed in the survey of Dublin Core metadata embedded in project home pages [14] and employing use of XSLT [15].
- Ensuring that the metadata deployed is appropriate for the purposes for which it will be used.
- Ensuring projects have appropriate cataloguing rules for their metadata and processes in place for implementing the rules and monitoring compliance.
- Ensuring that metadata can interoperate with third party services.
- Using techniques for checking metadata such as use of spell-checkers, checking against lists of controlled vocabularies, etc.

The QA procedures will be applied to metadata which is used in various ways including metadata embedded in HTML and XML resources, OAI metadata, educational metadata, RSS newsfeeds, etc.

A case study which describes the use of metadata in an e-journal, including details of the metadata elements used, the purpose of the metadata, the architecture for managing the metadata and the limitations of the approach has been published [16].

10 Software Development

QA is crucial in the development of quality software. It is fundamental to the entire software development process from the initial systems analysis and agreement on standards through to problem handling and testing and software deployment. Once established, QA processes form a thread through the software development lifecycle and help developers focus on possible problem areas and their prevention.

10.1 Development

Before the onset of a software development project the project team should produce a detailed set of specifications that document what exactly the software will do. Questions need to be asked about the purpose of the software and whether this purpose reflects the requirements of the user.

QA Focus will be providing case studies and briefing papers on these areas. Consideration of one possible design process for recording specific software development requirements, Unified Modelling Language (UML), is given in a case study provided by the Subject Portals Project [17].

10.2 Documentation

QA Focus will be providing advice on standards for software documentation, both public and internal. Having clear documentation is especially important in a digital library programme in which short term contracts and high staff turnover are the norm [18]. In the long term good documentation can improve usability, reduce support costs, improve reliability and increase ease of maintainance. Throughout a project's lifetime information should be recorded on the software environment a package has been developed in, language systems used and the libraries accessed.

Project teams will need to agree on standards used when writing software code. This should be done prior to development. QA Focus have produced a briefing paper which provides advice on how projects do this [19].

In the later stages of development work user documentation may be required. Writing documentation is a useful process that can show up bugs which have been missed in testing. Ideally the documentation writers are a different team of people from the developers and provide a different perspective on the software.

10.3 Testing

A software product should only be released after it has gone through a proper process of development, testing and bug fixing. Testing looks at areas such as performance, stability and error handling by setting up test scenarios under controlled conditions and assessing the results.

Before commencing testing it is useful to have a test plan which gives the scope of testing, details on the testing environment (hardware/software) and the test tools to be used. Testers will also have to decide on answers to specific questions for each test case such as what is being tested? How are results documented? How are fixes implemented? How are problems tracked? QA Focus will be looking mainly at automated testing which allows testers reuse code and scripts and standardise the testing process. We will also be considering the documentation that is useful for this type of testing such as logs, bug tracking reports, weekly status report and test scripts. We recognise that there are limits to testing, no programme can be tested completely. However the key is to test for what is important. We will be providing

documentation on testing methodologies which projects should consider using.

As part of the testing procedure it is desirable to provide a range of inputs to the software, in order to ensure that the software can handle unusual input data correctly. It will also be necessary to check the outputs of the software. This is particularly important if the software outputs should comply with an open standard. It will be necessary not only to ensure that the output template complies with standards, but also that data included in the output template complies with standards (for example special characters such as '&' will need to be escaped if the output format is HTML).

11 Deployment Into Service

The final area QA Focus will be looking at is the deployment of project deliverables in a service environment. It is unlikely that project will migrate into a service directly – the intention is that many of the project deliverables will be transferred to a JISC service who will be responsible for deploying the deliverables into a service environment. In addition to the deployment into a service environment for use by end users project resources may also need to be preserved. This is another area in which we will provide appropriate advice. Other work will address the issues involved in deploying software deliverables, digitised resources, Web sites, etc. into a service environment.

There are a number of scenarios for the deployment of projects deliverables: the deliverables may be hosted by a national service, within an institution or on the user's desktop.

It may be necessary to consider any special requirements for the user's desktop PC. For example will the service require a minimum browser version, will it require use of browser plugin technologies, are there any security issues (e.g. use of JavaScript), could institutional firewalls prevent use of the service, etc.

Inevitably there are resource implications for the deployment of project deliverables into a service environment: consideration needs to be given of the time taken for deployment and possible impact on other services (such as security, performance and compatibility issues). As well as these technical and resource issues there will be human aspects, including the potential resistance to change or reluctance to make use of work carried out by others.

An interesting approach which sought to provide a simple syndication tool has been carried out by the RDN. The RDN-include tool provides access to subject gateways and allows the institution to control the look-and-feel of the gateway. However, as this tool is implemented as a CGI script it requires System Administration privileges in order to be deployed. It was felt that System Administrators may be reluctant to deploy the tool, due to concerns over potential security problems. In order to address such concerns RDNi-Lite was devolved, which provides similar functionality but, as it is implemented using JavaScript, can be used by an HTML author: no special System Administration

privileges are required. This example illustrates an approach which acknowledges potential deployment difficulties and provides an alternative solution. Further information on this approach is available [20].

An important aspect of this work will be to ensure that projects describe the development environment at an early stage, in order to ensure that services are aware of potential difficulties in deploying deliverables in a service environment. One could envisage, for example, a project which made use of innovative technologies, open source tools, etc. which the service had no expertise in. This could potentially make service deployment a costly exercise, even if open standards and open source products are used.

In addition to considerations of the deployment technologies, there is also a need to address the licence conditions of digitised resources. Again it would be possible to envisage a scenario in which large numbers of resources were digitised, some with licenses which permitted use by all and some which limited use to the project's organisation. In this scenario it is essential that the right's metadata allows the resources which can be used freely is made available to the service and that the production service can be deployed without making use of resources with licence restrictions.

12 Preservation Of Project Results

Even if a project has a clear idea of its final service deployment environment, there may be additional requirements during the project's development. Within the context of the JISC 5/99 programme there is now an expectation that learning objects funded by the programme will be stored in a learning object repository. The Jorum+ project [21] has been set up to provide repositories of the learning objects.

There is also discussion of the need to provide a records management service to ensure that project documentation, such as project reports, are not lost after the end of the programme.

In both of these areas QA Focus is well-positioned to advise JISC and the projects on appropriate strategies, based on its work in advising on technical interoperability.

13 The QA Focus Toolkit

An important QA Focus deliverable will be a QA Self Assessment Toolkit which will allow projects to check their project QA procedures for themselves.

A pilot version of the toolkit is currently being tested. The pilot covers the QA requirements when mothballing a project Web site and other project deliverables once the project has finished and funding ceases [22].

The toolkit consists of a number of checklists with pointers to appropriate advice or examples of best practice. The toolkit is illustrated below.
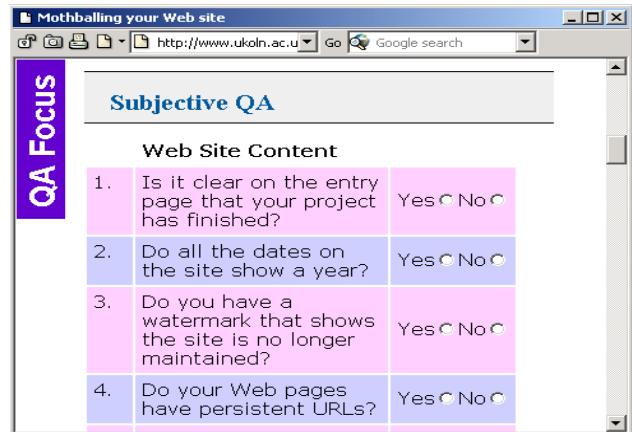

Figure 1: Toolkit For Mothballing Web Sites

The toolkit aims to document the importance of standards in a readable manner, which can be understood by project managers as well as technical developers. The toolkit will make use of case studies which have been commissioned and appropriate advisory documents. Most importantly the toolkit will provide a checklist and, in a number of cases, a set of tools which will allow projects to assess project deliverables for themselves.

The structure of the toolkit is illustrated below.

Figure 2: QA Self-Assessment Toolkit Structure

In the area of standards compliance for Web resources software tools can be used to check for compliance with standards. An article on "*Interfaces To Web Testing Tools*" describes the use of "bookmarklets" and a server-based interface to testing tools [23].

In a number of areas the use of software tools will be documented. The documentation will include a summary of the limitations of the tools, and ways in which the tools can be used for large-scale deliverables. This may include testing of significant deliverables, sampling techniques, etc.

14 Applying QA To QA Focus Web Site

We are using the methodologies described in this paper for in-house QA for the QA Focus Web site. This is being done in order to ensure that the Web site fulfils its role, to test our own procedures and guidelines and to gain experience of potential difficulties.

The approach used is to provide a series of policy documents [24]. The policies follow a standard template, which describes the area covered, the reason for the policy, approaches to checking compliance, allowable exceptions and audit trails, as illustrated below.

Policy On Standards For QA Focus Web Site

Area: Web

Policy: The Web site will be based on XHTML 1.0.

Justification: Compliance with appropriate standards should ensure that access to Web resources is maximised and that resources can be repurposed using tools such as XSLT.

Responsibilities: The QA Focus project manager is responsible for this policy. The Web editor is responsible for ensuring that appropriate procedures are deployed.

Exceptions: Resources which are derived automatically from other formats (such as MS PowerPoint) need not comply with standards. In cases where compliance with this policy is felt to be difficult to implement the policy may be broken. However in such cases the project manager must give agreement and the reasons for the decision must be documented.

Compliance measures: When new resources are added to the Web site or existing resources update the `,validate` tool will be used to check compliance. A batch compliance audit will be carried out monthly.

Audit trail: Reports from the monthly audit will be published on the Web site. The QA Focus Blog will be used to link to the audit.

Further information: Links to appropriate QA Focus documents.

Figure 3: QA Policy For QA Focus Web Site

15 Applying QA Methodology In Other Contexts

Although the approach to QA described in this paper is meant to be developmental, it is likely that projects will, to some extent, feel obligated to deploy the methodologies described. Use of the methodology from projects which are not funded under the JISC 5/99 programme will help to establish the effectiveness of the approach and should provide valuable feedback.

A presentation on the QA Focus work was given to staff from the Centre For Digital Library Research (CDLR) based at the University of Strathclyde in April 2003 [25]. Shortly afterwards CDLR staff felt sufficiently motivated to investigate the potential of the

methodology for two digital library projects: a digitisation project funded by the NOF-digitise programme which is currently under development and a regional digital library project which has been completed with no funding available for additional work.

The following conclusions were drawn:

"CDLR staff attempted to follow QA Focus guidelines retrospectively and to implement appropriate recommendations. This exercise showed that the extent of compliance with guidelines could be categorised into four areas: (1) areas of full compliance, where the project had already made decisions in accordance with QA guidelines; (2) areas in which compliance could be achieved with little extra work or with minor changes to workflow procedures; (3) areas in which QA guidelines were considered desirable but impracticable or too expensive and (4) areas where QA guidelines were not considered appropriate for the project.

The conclusion from the project managers involved was that consideration of the QA guidelines improved the value, flexibility and accessibility of the digital library deliverables, provided they were interpreted as guidelines and not rules. Rather than the QA process imposing additional constraints, the exercise validated decisions that had been made to vary from recommended standards, provided the issues had been considered and the decisions documented. What had been seen as a potentially burdensome exercise was regarded in retrospect as beneficial for the user service, for accessibility, interoperability, future flexibility and even for content management. It was felt that there are a number of areas in which simple developments to scripts or use of tools can provide a significant development to interoperability." [26].

16 The Open Standards Philosophy

The JISC promotes the use of open standards in its development programmes. However feedback from projects indicates that there is not necessarily a clear understanding of what is meant by open standards.

QA Focus has produced a briefing document which seeks to clarify the term 'open standards' [27]. However there is still an unresolved issue as to the role that proprietary standards have in development programmes and the processes needed to evaluate open and proprietary standards and perhaps, in certain circumstances, chose a proprietary standards rather than an open one due to issues such as resources implications, maturity of standard, etc.

On reflection it would appear that an approach based on a simply advocating use of open standards is not necessarily desirable. It is felt that there are several factors which need to be addressed, including:

- Ownership of the standard (owned by an open standards body or by a company).
- In cases of proprietary standards, whether there is a community process for development of the standard.

- In cases of proprietary standards, whether the standard has been published openly or reverse-engineered.
- Whether viewing tools are available, available for free, available as open source and available on multiple platforms.
- Whether authoring tools are available, available for free, available as open source and available on multiple platforms.
- The fitness for purpose of the standard.
- Resource implications in use of the standard.
- Complexity of the standard.
- Interoperability of the standard.
- Organisational culture of the project's organisation.

It is felt that use of a matrix approach when choosing the standards for use in a development programme is well suited to the developmental culture prevalent in many digital library programmes and is preferable to a strict requirement that only open standards may be used.

The approach will, of course, require documentation outlining the decisions made and justification of deviation from use of accepted open standards and best practices.

17 Team Working Within QA Focus

QA Focus is provided by UKOLN and the AHDS, which are located in Bath and London respectively. In order to support working by a distributed team and minimise unnecessary travel team members make use of a number of collaborative tools, including My.Yahoo as a shared repository of resources. YahooGroups for managing the team mailing list and the MSN instant messenger to provide real time communications. We are also making use of a 'Blog' to provide news on QA Focus activities.

This approach appears to be working well. In order to share the experiences with other projects and to highlight potential problems (e.g. reliance on an unfunded third party) a case study has been produced [28].

18 What Next For QA Focus?

Although QA Focus funded is due to finish on 31st December 2003 we will be seeking additional funding to continue our work. We feel that QA for JISC's development programmes will be an ongoing activity, and, indeed, will grow in importance as "Web Service" technologies are developed which will require more rigourous compliance with standards.

We would hope to maintain the resources on the QA Focus Web site and produce new ones in appropriate areas. Additional activities we could engage in could include the deployment, development or purchase of testing tools and services. One possibility would be hosting a JISC compliance service, along the lines of the UK Government's eGIF Compliance Service [29].

As well as providing advice to projects, QA Focus will also advise JISC on approaches to future programmes. We will be well-placed to provide advice

prior to the start of project work, which will help to ensure that best practices are deployed from the start. We will recommend that, in addition to providing training on project management when new programmes begin that training is provided on best practices for ensuring that project deliverables are interoperable in a broad sense. We will also advise on contractual issues, including advice on the persistency of Web sites once project funding has finished. Advice will also be provided for evaluators of project proposals to ensure that consideration is given to issues such as QA procedures as well as technical feasibility.

19 Conclusions

The paper has described the work of the QA Focus project which supports JISC development activities by providing advice and support for projects in ensuring that project deliverables will be widely accessible, are interoperable and can be deployed into a service environment with the minimum of effort.

JISC will not be alone in giving a higher profile to quality assurance and compliance with standards and best practices for its development programmes. Within the UK two examples of standards-based programmes should be mentioned: (1) the e-government interoperability framework (e-GIF) defines the "internet and World Wide Web standards for all government systems" [30]; (2) the New Opportunities Fund's NOF-digitise programme provides funding to digitise cultural heritage resources [11].

We will be exploring the possibilities of shared approaches to QA with these bodies. The author welcomes feedback from those involved in similar activities in the international digital library community.

References

[1] eLib Standards Guidelines, <http://www.ukoln.ac.uk/services/elib/papers/other/standards/>

[2] Welcome To The DNER, <http://www.jisc.ac.uk/index.cfm?name=dner_home>

[3] Standards and Guidelines to Build a National Resource, <http://www.jisc.ac.uk/index.cfm?name=projman_standards>

[4] JISC Information Environment Architecture, <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/>

[5] Learning and Teaching (5/99) Programme, <http://www.jisc.ac.uk/index.cfm?name=programme_learning_teaching>

[6] FAIR Programme, <http://www.jisc.ac.uk/index.cfm?name=programme_fair>

[7] X4L Programme, <http://www.jisc.ac.uk/index.cfm?name=programme_x4l>

[8] QA Focus Surveys Of Project Web Sites, <http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/>

[9] QA Focus Briefing Documents, <http://www.ukoln.ac.uk/qa-focus/documents/briefings/>

[10] QA Focus Case Studies, <http://www.ukoln.ac.uk/qa-focus/documents/case-studies/>

[11] NOF-digitise Technical Advisory Service, <http://www.ukoln.ac.uk/nof/support/>

[12] TASI, <http://www.tasi.ac.uk/>

[13] Legal Information Service, <http://www.jisc.ac.uk/legal/>

[14] Metadata Analysis Of JISC 5/99 Project Entry Points, <http://www.ukoln.ac.uk/qa-focus/surveys/web-10-2002/dc-metadata/>

[15] Approaches To Validation Of Dublin Core Metadata Embedded In (X)HTML Documents, WWW 2003 Poster, P. Johnston, B. Kelly and A. Powell, <http://www.ukoln.ac.uk/web-focus/papers/www2003/>

[16] Managing And Using Metadata In An E- Journal, <http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-01/>

[17] Managing a Distributed Development Project, <http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-03/>

[18] Resolving the Human Issues in LIS Projects, <http://www.rslp.ac.uk/circs/2002/LIS.htm>

[19] Guidelines for Writing Software Code, <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-13/>

[20] RDN-Include: Re-branding Remote Resources, WWW 10 Poster, B. Kelly, P. Cliff and A. Powell, <http://www.rdn.ac.uk/publications/www10/>

[21] Jorum+, <http://www.jorum.ac.uk/>

[22] QA Self-Assessment Toolkit: Mothballing, <http://www.ukoln.ac.uk/qa-focus/toolkit/mothballing/>

[23] Interfaces To Web Testing Tools, Ariadne issue 34, Jan 2003, <http://www.ariadne.ac.uk/issue34/web-focus/>

[24] Inhouse QA, <http://www.ukoln.ac.uk/qa-focus/qa/>

[25] QA For Digital Library Projects, <http://www.ukoln.ac.uk/qa-focus/presentations/strathclyde-04-2003/>

[26] Deployment Of Quality Assurance Procedures For Digital Library Programmes, B. Kelly, A. Dawson and A. Williamson, paper submitted to IADIS 2003 conference, <http://www.ukoln.ac.uk/web-focus/papers/iadis-2003/>

[27] What Are Open Standards?, <http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-11/>

[28] Implementing a Communications Infrastructure, <http://www.ukoln.ac.uk/qa-focus/documents/case-studies/case-study-12/>

[29] e-GIF Compliance Assessment Service, <http://www.egifcompliance.org/>

[30] e-GIF, UK Gov Talk <http://www.govtalk.gov.uk/interoperability/egif.asp>

Not Just a Portal: Managing Access in a Complex Information Environment

Jean Sykes, John Paschoud and Christine Cooper
London School of Economics & Political Science, United Kingdom
Corresponding author: <j.paschoud@lse.ac.uk>
http://www.lse.ac.uk

Work is in progress by LSE staff to create a Managed Information and Knowledge Environment, simple and managed access to a wide range of appropriate and permitted content for a broad range of users. The portal is only one small part. More difficult tasks are identifying information content (internal and external) for different user types, and developing suitable middleware for managing access, in an institutional, national and international context. Two projects will be highlighted: SECURe, working on the access management middleware, and UK Computing Plus, offering electronic information access to certain library visitors.

1 Introduction

Everyone is talking about portals. Software and systems suppliers are creating portals as part of their product portfolios, and universities across the world are becoming aware of the need to have an institutional portal.

Why the need? Because researchers, students and administrators are using more and more electronic access to content across a wide spectrum of information types:

- institutional and third party
- commercial and open source
- stored locally and stored remotely (anywhere in the world)
- quality assured and not
- primary and secondary
- in digest form and full text

Most of this content is available via the web, but there are also many information sources – especially those stored within an institution for internal purposes only – which are still paper-based, or hidden in personal/group databases, or filed in such applications as Microsoft Exchange public folders. Increasingly users want all the information relevant to their work to be available to them electronically on a 24 x 7 basis and from wherever they happen to be. They do not want to have to log in and out of different systems with different logon procedures and passwords; they need to navigate between their sources. They also want to eliminate duplication of effort (for example re-keying the same search in different databases) and they want irrelevant information to be filtered out for them. But what is irrelevant to a user today may become relevant for them tomorrow, and so they want flexibility and inclusivity as well as refinement and exclusivity. This is a tall order

for information providers to fulfil and is a major challenge for us all.

2 The vision for LSE

At the London School of Economics and Political Science (LSE) we are keen to provide our users with a portal to support their specific needs. This means creating different "views" to different users, and it means allowing users to access discrete parts of the information environment if they wish to as well, since always entering through a particular view may not be the quickest and most efficient way in for certain tasks. Unlike many of our colleagues in other universities, therefore, we do not entirely subscribe to the single portal vision.

Moreover, our definition of a portal is that it is just a gateway to the information, just a part of the whole which we are trying to create. The portal is the front-end, the user interface, and of course we do need a new generation of such applications to cope with the complexity of the new need for several user views. But ultimately the portal is the easy bit. Far more difficult are the other two major areas involved in building this information environment for our users:

- the information content at the back end
- the middleware which facilitates dialogue between the portal at the front and the information content at the back

So at the LSE we are not calling our project a portal. We are calling it a Managed Information and Knowledge Environment, or MIKE. The goal of MIKE is to help LSE teachers, researchers, students, administrators, alumni, and visitors to cope better with the spiralling problem of information overload in their daily work. To do this we will:

- present them with a personalised desktop accessible both on and off campus, meeting their priorities and adapting to their changing needs
- link together the wide range of relevant information and administrative processes which are currently dispersed around the institution into a coherent and easily-searchable whole
- provide seamless access across a growing variety of licensed third party information content
- offer well-managed and well-packaged information to enhance the reputation, recruitment, and marketing of the LSE
- encourage content creators within the institution to make their material available online and integrate it into the information environment for easy access for authorised users

We believe that at LSE we are already well placed to create a successful MIKE because we have considerable experience and expertise across all our information services: Library, IT Services, Management Information Services (MIS), Website Services, and Centre for Learning Technology. And there are a number of building blocks already in place across these services which we can bring together now and enhance as a good seedbed for the MIKE. These include:

- LSEForYou, created by MIS and winner of a number of national and international prizes. This is a portal approach to providing administrative services so that students and staff can look up classroom bookings, fee payment status, salary payment status, and many more pieces of information from the corporate database [http://www.lse.ac.uk/lseforyou/]
- Electronic Library, one of the first services on the web to bring together a subject search approach to a large number of electronic sources, including access to over 5000 licensed full text e-journals and a wide range of free websites validated by subject librarians [http://library-2.lse.ac.uk/EL/info/]
- Online learning as part of a growing number of courses. The Centre for Learning Technology, using WebCT, is helping academic staff to embed online elements to most taught courses in the LSE, and through this and their work with the Library on electronic course packs they have a wealth of experience in the way learners use and interact with e-learning materials [http://teaching.lse.ac.uk/tech/]

A number of externally-funded projects can also be regarded as potential building blocks for MIKE, including:

- ANGEL, a Library project which has developed an understanding of the different roles that users have and methods of mediating access to information resources based on those roles. ANGEL has developed a Resource Manager software product which can be used to manage collections of electronic resources, and indeed the Electronic Library makes us of this product [http://www.angel.ac.uk/]
- DELIVER, a joint project between the Library and the Centre for Learning Technology which aims to develop integration tools to enable Virtual Learning Environment-based resources (in the case of the LSE this is WebCT) and Library-based resources to be accessed easily from each other [http://www.angel.ac.uk/DELIVER/]
- SECURe, a collaborative project led by the Library but also involving staff from IT Services and Management Information Services, which is working towards an integrated system for controlling access to resources without multiple username and password challenges [http://www.angel.ac.uk/secure/]

Work for MIKE has started in two broad areas at the moment but still has a long way to go:

1. Content. A Content Group is identifying the large range of information resources required by members of the LSE community by holding focus group meetings with students, administrators, academics, and alumni. The next tasks will be to classify and prioritise the identified needs and design a metadata/content management structure for searching them. We will investigate existing products and schemes such as the proposed RSLP Collection Description Schema [1] and such proprietary products as Metalib [http://www.aleph.co.il/MetaLib/] and iPort [http://oclcpica.org/?id=106&ln=uk]. The information content will remain in separate but linked repositories.

2. Technical architecture. The IT staff involved in MIKE will be responsible for selecting the portal software and for defining the ways in which different components of the system interact. They will also create the access management layer to connect the user at the desktop to the relevant content in one or more of the collection level registries, building on the SECURe Project. This will be achieved through single sign-on authentication and a sophisticated authorisation system which will deliver relevant resources only to users entitled to access them.

3   The portal

Most portal implementations in universities (including the current LSEForYou system) can be described by a generalised two-layer model, in which the portal or presentation layer provides the end-user interface, and is configured to access a number of content sources or collections of resources (Figure 1). This works well and is cost-effective to maintain when the number of content layer sources is relatively small, their structure and interfaces do not change frequently, and they are under local control. Most of the content to which access is enabled by LSEForYou resides in locally-managed Oracle databases. Specialised user interfaces, rather than the general-purpose portal, are used for more complex transactions such as those required by administrative
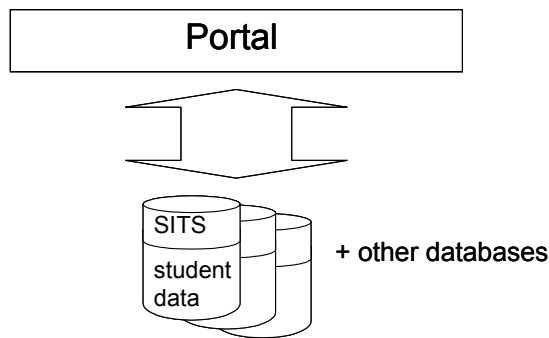
Portal

+ other databases

SITS
student
data

Figure 1: current LSEForYou architecture

staff. For example, the SITS system is used at LSE for management of student record data, but students can access a restricted view of the data via LSEForYou.

Problems arise when this model is extended to include larger numbers of content collections, under more diverse and autonomous management, and these problems may not be evident until attempts are made (as at LSE) to extend the scope of the portal to encompass content sources managed by two or more distinct administrative domains of the university. Such problems for portal implementers are often 'caused' by the university library, which complicates a previously simple task by revealing the existence of hundreds (possibly thousands) of 'new' electronic resources, all dearly beloved by some important users, and all configured, managed and licensed-from a vast array of external suppliers who may, or may not, be willing to negotiate about the way in which their resources are accessed. The university is unlikely to have much control or influence over changes to these resources and the interfaces they expose, which will 'break' links from the portal. If system administration functions for the portal are the responsibility of, say, the IT support service, and a staff member in the library is responsible for liaison with a resource supplier, there will be additional barriers to awareness of an impending change of this sort, and the contacts and information needed to respond to the change.

Academic library consortia already have considerable experience in attempting to integrate collections of such autonomous content sources. For example, the InforM25 virtual union catalogue system [http://www.m25lib.ac.uk/] provides cross-searching of a large number of library catalogues that all nominally conform to the well-defined Z39.50 standard [http://www.loc.gov/z3950/agency/]. But diversity between implementations of the standard by the owners of different content stores (or the suppliers of their catalogue systems) renders the addition of each new source a far from trivial task. Furthermore, changes (such as routine software and configuration upgrades) made by those owners, which they may not even suspect will be evident at the portal, are often the cause of unexpected remedial work which must be done to re-make a broken connection.

The existing information environment of a university could be depicted (with a great deal of simplification) in

a way that models typical organisational domains of the institution. Thus, one or more administrative user interfaces are provided to administrative data (such as student, staff and course/curriculum records); a virtual learning environment such as WebCT [http://www.webct.com/] or Blackboard [http://www.blackboard.com/] may be used as a 'learning portal' to access course materials; and a library management system will be present for access to full-text electronic library materials and metadata (i.e. catalogue and stock management records) describing traditional printed library material. Indeed, most of the information systems in use can be described in a generalised way as a content store, specialised to suit the structure of the data involved; and a presentation layer or user interface, specialised to suit the kinds of transactions between the system and different types of users.

Usually it is necessary to duplicate some information across each of these discrete systems. For example, administrative systems will hold the primary or definitive records of all students and staff. These must be exported, imported and kept updated in a virtual learning environment to control access to course material. Similarly, student and staff records must be duplicated to a library system (often with the addition of records for other library users, who are not otherwise members of the university) as patron records, to manage lending and other functions. The university library often introduces additional problems in this area too, because there may be large numbers of authorised 'external' users of the library who are not otherwise known as 'members' of the university.

The principal reasons for a portal to need the authenticated identity of a user are to enable personalisation of portal content and functions, and to ensure access authorisation - i.e. that users are permitted access only to information for which they have rights, for reasons of confidentiality, security and economy.

Thus the two key areas on which a MIKE must focus are metadata about resources, and metadata about users. As well as enabling access with fewer discontinuities across diverse information resources, a MIKE must also impose access management.

4   The middleware: Collection Level Registries

The architecture for MIKE introduces a third, middleware layer between purpose-specific content management systems and presentation-layer user interfaces (Figure 2). The middleware components are a number of Collection Level Registries (CLRs) that provide a level of indirection between the (or, any number of) user interfaces or portals, and the content stores. This is similar in some ways to ideas being developed in the MIT Open Knowledge Initiative [http://web.mit.edu/oki/].

Separating the interfaces to different content sources from the portal itself brings a number of benefits:

1. Each administrative domain can continue to manage autonomously the content for which it is

responsible, maintaining metadata about each content collection in its own CLR.

2. More than one portal can be economically implemented and maintained. This may be trivially useful where 'live' and test versions of the portal must co-exist; but it may also be productive to experiment with several radically different portals, intended for use by different groups of users such as students, researchers or alumni.

3. The architecture is scaleable to include very large numbers of new content collections; if necessary, by adding CLRs to balance the loading.

Each CLR must, of course conform to agreed standards for machine-to-machine interfaces with the portal(s) or presentation layer for which it provides services.

Although a content collection can be described in the CLR using an accepted schema such as the RLSP Collection Level Description, further metadata is needed for the architecture to operate. Our understanding of what is required has been developed through work on a number of national projects such as HeadLine [http://www.headline.ac.uk/], which implemented such a schema as a relational database, accessed by presentation-layer applications using SQL. The HeadLine Resource Datamodel [2,3] included all the attributes that would be recognised as a bibliographic description, but added metadata to describe how, technically, a resource collection could be searched and content items retrieved and how end-user access to the collection was controlled, anticipating in some ways the devolved authentication schemes that are now being implemented by projects such as AthensDA [http://www.athensams.net/development/devolved_authe ntication/] and Shibboleth [http://shibboleth.internet2. edu/].

5 The middleware: Content management

The ANGEL Project [http://www.angel.ac.uk/] developed this concept further, implementing a Resource Manager (which is available as Open Source software) that encapsulated the relational database within an XML-based messaging protocol for requests and responses. This enables details of the resource collections to be accessed by portals (or any presentation-layer application) in other commonly-used formats, such as the Open Archives Initiative Protocol for Metadata Harvesting [4] and Z39.50 [5]. The ANGEL Resource Manager is designed in such a way that new access protocols can be added relatively easily. For example, RDF Site Summary (RSS)[http://web.resource.org/rss/ 1.0/] is commonly used by many portal products to enable 'news feeds'; but RSS is not inherently tied to this purpose, and is well-suited for rendering brief, linked descriptions of resources, rather like the results of a Web search-engine. So a RSS feed was implemented in the Resource Manager to make library resources easily accessible via portals.

We do not take any particular ideological stance on the selection of appropriate software components to implement elements of the MIKE; neither towards a completely in-house (or Open Source) development; nor in favour of total 'outsourcing' using proprietary commercially supported products. We will inevitably need to make use of many existing 'legacy' systems in the early stages, and the architecture will need to accommodate and include many systems that are required in the future. Instead, the approach we take is to define standards for inter-operation and the interfaces between major components. In the short-term this will incur penalties of lower runtime efficiency (than the alternative of a more tightly-integrated single proprietary solution); but we believe these will be soon outweighed by the greater flexibility and scalability of a high-level architecture over which we have greater control; and that runtime performance, where it poses problems, can be easily improved with more hardware and network capacity. This is relatively very much cheaper than dealing with the maintenance and periodic 'big bang'
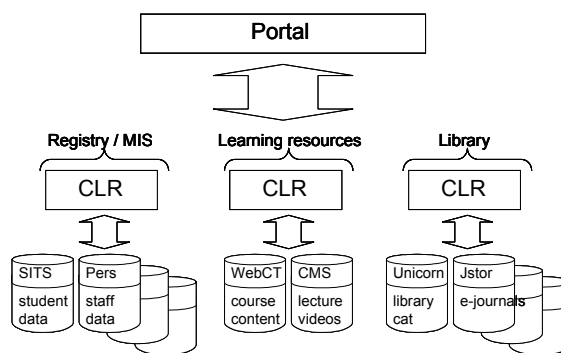

Figure 2: MIKE architecture for federated content management

replacement of a monolithic solution.

As an example of how this 'mixed economy' approach might be implemented, commercial products such as MetaLib/SFX or iPort could be used to implement one (or more) of the CLR components. MetaLib/SFX could co-exist, in the MIKE architecture, with an Open Source product such as the ANGEL Resource Manager, both using the published OpenURL standard [6] to provide client applications with a 'deep link' to items such as individual e-journal articles, in managed content collections that may be in-house repositories or (more likely) the remote holdings of a publisher, accessed under license.

6 Access management: the SECURe Project

In this field as in others, LSE has been extremely active in making and developing relationships with other Higher Education institutions, standards bodies and research programmes; in the UK, Europe, USA, Australasia and other parts of the world. It is clear that no university can or should attempt to be a self-reliant or

self-contained information environment, and the design goals of MIKE are that it should operate (in most respects) regardless of the physical boundaries or network topology of our campus, and of the 'regulatory boundary' defining resources over which we have administrative control. MIKE should also address the problems of authorised LSE users who want access from anywhere (in the world), at any time.

However, there are many boundaries that it must enforce - ensuring that only authorised users have access, and enforcing the restrictions properly required by considerations of intellectual property, privacy, security and plain economics [7].

The access management functions of MIKE are supported by the SECURe (Secure Environment for Certificated Use of Resources) Project. SECURe will implement and test a full-scale solution, portable and scaleable to other Further & Higher Education institutional environments, to the key problem of achieving fine-grain authorisation of access by all users in an institution to the complete spectrum of relevant information resources. The range of resources must include those under a high degree of institutional control, with relatively high security requirements (such as sensitive staff and student personal data), through to such out-of-domain resources (over which the institution has no direct control) as subscription e-journals to which the institutional library negotiates licensed access for users.

In particular, SECURe will use the existing and emerging technologies of Public Key Infrastructure (PKI) with X.509 digital certificates [http://www.ietf.org/html.charters/pkix-charter.html] and Shibboleth [http://shibboleth.internet2.edu/] to implement the necessary authentication infrastructure to support this, on as large a scale as project and institutional resources allow. To meet the practical needs of MIKE, this will involve all staff and student members of LSE (approximately 10,000 individuals). This phase of SECURe will focus on documenting and solving the management problems raised by institutional scale deployment of PKI and Certificate Authority services, including appropriate strategies for certificate issue, revocation, and linking the authorisation of certificates with existing registration processes that establish real-world identities of students, staff and other users of campus IT services and managed information resources.

7   The need for interim solutions

The vision for LSE described earlier in this paper will be realised at some time in the future and whilst some elements are already implemented, others are clearly some way off yet. But the problems and issues that the MIKE seeks to address are real problems that are presenting themselves within many universities today. The answers cannot wait until some unknown date when all will become possible; practical, pragmatic solutions that can be deployed now are needed to fill the gap.

It would be foolish, however, to disregard the vision and to implement solutions today that do not fit our model of the future. These stop-gap systems and services

must take all they can from the MIKE architecture and must only be creative where necessary. It is also preferable that, as the vision is realised, these systems and services will evolve into the MIKE architecture rather than being short term systems in their own right. This has a significant impact on the design and development process. It is more acceptable to deploy a system where some aspects are less than ideal if it is known that those elements will be replaced in six or twelve months time.

There are, of course, many others but the British Library of Political and Economic Science at LSE provides a good example of the problems that universities face today. This major research library serves a much wider community than just the staff and students at LSE. It houses many very important collections which are used by scholars from across the world. It is also a valuable research resource which is used by students, commercial users and members of the general public. In the past, it was necessary to make the journey but once there, all the resources were available in physical printed form. It may have been an enormous task, but collecting these resources, classifying them and making them available to those who wished to read them was at least possible.

The age of electronic information has brought about significant changes though. The cost of purchasing and handling printed media has led a massive drive towards the provision of electronic only versions, particularly of scholarly journals. Increasingly, important academic work is never published in the traditional manner. This clearly has an impact on the professor from China who has made the trip to LSE, as he can physically access the printed material but he now needs a computer to be able to see the whole collection.

A brief application of the MIKE vision and architecture to this problem will demonstrate that in the future this will be a trivial matter. The professor from China will identify himself as such and will be challenged for his authentication token which will then be verified with his institution or some other competent body. The portal will welcome him as a visitor to the LSE and will tailor his view to the resources he is most likely to use. Access to electronic resources will be then be mediated by the owners of those resources and allowed either of the basis of his own identity or through being a visitor to the LSE. If all the information required is held in electronic form, it will be possible to be authenticated as a virtual visitor to the LSE and our professor need not even make the trip.

However, today there are many challenges which must be understood and overcome before any large scale, international scheme which provides this type of facility will be implemented. LSE, like most institutions, requires users to identify themselves and to be authenticated before they can use a networked computer, but how can this process be managed for users who are not members of our institution? The suppliers of electronic information are also anxious to protect their revenue streams and will need to change the way that they control authentication and authorisation if they are

to meet the demands for ubiquitous access to their products. These resources are provided under a myriad of different licenses, each with its own terms and conditions as to what classes of user are permitted, which makes it very difficult to determine who is entitled access to what. The success of the Athens service in the UK has shown that some are willing to entrust a third party with the information to verify authentication; many however are not.

8 Visitor access: the UK Computing Plus Project

The difficulty for visitors to our libraries threatens to become an acute problem for the UK Libraries Plus scheme which has been of great benefit to part time students and distance learners within the UK higher education community for some years. Under the scheme, eligible students are able to register to use the library facilities at up to three other member institutions with students generally choosing those nearest where they live or work. Once registered, users have access to printed reference material and also limited borrowing rights but do not have access to resources that were available in an electronic format only. The extension of the scheme to cover access to IT facilities and to electronic resources would further enhance its value. In order to explore these challenges, six pilot projects were commissioned, each with a different approach to tackling these issues. One of these pilot projects, which are collectively known as UK Computing Plus, is at LSE.

In considering what model should be adopted for UK Computing Plus and in the absence of a common, national scheme for authentication and authorisation, the LSE philosophy has been to make each institution responsible for the administration and access management of its own students. It was also felt that participants in the scheme should be able to access the same resources as a "standard" student and should be neither advantaged nor disadvantaged by physically being in another institution. These two factors have influenced the model developed.

The joint goals are achieved through the creation of a virtual extension to the 'home' institution. As far as possible, the visiting student is in the same position as if they were sat at a workstation in their own institution and so has access to those resources which are managed by (or, to which access is licensed by) their 'home' institution.

There are a number of different technologies available which could be used to implement this service, all of which require the home institution to provide an appropriate server and the receiving institution to provide client software. Many institutions already had such an infrastructure in place, based on either web proxy servers or thin client technology (such as Citrix) in order to support staff and students who work from home. The receiving institution will need client configuration information for each user and some work is required to develop a simple but secure user interface for the pilot scheme.

The operation of the system is very simple. When a UK Libraries Plus student registers to use the library at LSE, they are given the UK Computing Plus username and password. The same username and password are used by all users under the scheme. When a user logs onto an LSE computer with this username and password a program is started which requires them to select their home institution from a menu. The security policy is set so that should they ever exit this program, the session is ended and they are logged out. Having selected their institution, one or both of the following will happen:

1. A web browser window is opened which is pre-configured to route all traffic through their home institution web proxy server. This proxy server should require them to authenticate with their local (i.e. home institution) username and password.

2. A Windows Terminal Server session is opened connected to their home institution's terminal server. Again this should require them to authenticate with their local username and password.

Note that, other than the generic username and password to access the menu program, all username and password challenges are for their home institution credentials. Once a national scheme for authentication is in place, the need to connect to a server at the home institution will disappear as the user's identity credentials will be verified through the national scheme, (although in many instances users will still want to be connected to their home server).

It should also be noted that the user does not have to know anything about how they connect to their home institution. A database has been gathered of the configuration details required for these connections. For this project this is held at LSE and collecting this information has, in fact, proved one of the most difficult parts of the whole project. In the MIKE architecture, this database becomes an item in a Collection Level Registry, (CLR), initially held at LSE but it could move to a national CLR should that prove more appropriate.

One drawback of the project, as implemented at the moment, is that the control of access to resources is still based solely on the basis of the user's home institution. Whilst this is still of benefit to the distance learners who may live too far away to have physical access to a computer on their "home" campus, as well as those who wish to work with a combination of printed and electronic sources, for many the value is limited. The real value will come only when global standards have been implemented for the mediation of access to electronic resources.

The UK Computing Plus initiative is specifically for part-time students and distance learners and so the LSE pilot service is focused on this particular group. However, as stated earlier, the LSE Library's role as a national and international research library brings in many other types of visitor and the model developed for the pilot is suitable for use by any group of visitors where there is an equivalent concept to the 'home' institution. Access to the system has already been extended to other visitors with borrowing rights from UK institutions and further extensions are planned. If a national PKI for the

UK population were ever implemented it might even be possible to extend the scheme to the general public.

Acknowledgements

References

[1] Powell, Andy et al. "RSLP Collection Description", D-Lib, volume 6 number 9, DOI: 10.1045/september2000-powell

[2] Graham, Stephen. "The HeadLine Resource Data Model", VINE, issue 117, pp 13-17.

[3] Paschoud, John. "The filling in the PIE - HeadLine's Resource Data Model", Ariadne, issue 27, [http://www.ariadne.ac.uk /issue27/paschoud/]

[4] Open Archives Initiative. "Protocol for Metadata Harvesting", [http://www.openarchives.org/OAI/2.0/openarchives protocol.htm]

[5] Library of Congress. "Z39.50 Maintenance Agency Page", [http://www.loc.gov/z3950/agency/]

[6] National Information Standards Organization. "The OpenURL Framework for Context-Sensitive Services", [http://www.niso.org/committees/ committee_ax. html]

[7] Paschoud, John. "All Users are Not Created Equal! - How to decide Who Gets What from your Hybrid Library", Internet Librarian International (1999), [http://www.headline.ac.uk/public/diss/index.html#9 0330JP]

Towards Cross-Organisational User Administration

Mikael Linden
CSC, the Finnish IT Center for Science
mikael.linden@csc.fi

The increase of personal services on the web and the co-operation between organisations have made it necessary to find ways to identify network users regardless of which organisation they are representing. New middleware technologies for user authentication and authorisation are being developed and deployed. This paper outlines the problem of cross-organisational user administration and presents related new technologies and activities in the academic world. Although the paper uses higher education as an example, the results can be generalised to cover cross-organisational services in other kinds of institutions as well.

1   Introduction

User administration is considered to mean keeping track of the information system users and their privileges. User administration covers both technology in use and administrative processes deployed in the organisation.

The concepts of identification, authentication and authorisation are relevant for user administration. Their interrelation is clarified in Figure 1.
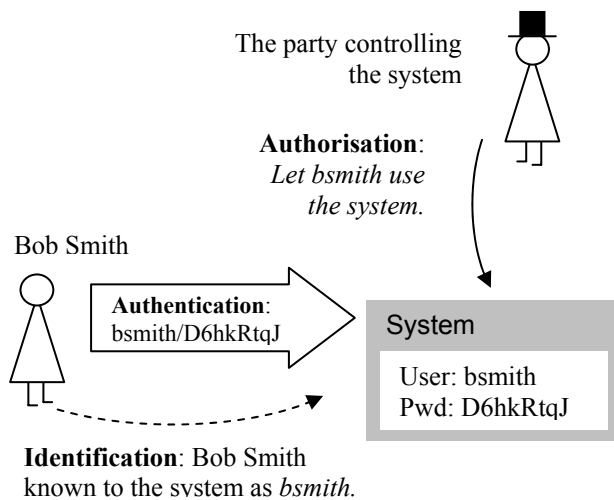
Figure 1. Concepts of identification, authentication and authorisation.

In a distributed environment, the identity of an object is represented by an identifier. In daily life various kinds of identifiers are used in distinguishing between people. Names are the most common identifiers, but they are not very useful as two persons may have the same name. In information systems, the traditional unique user identifier is the username (*bsmith*). Social security numbers are also widely used, although they are not assigned by the organisation in question but by the government. In universities other common identifiers for people are student numbers and employee numbers. A more detailed description of identifiers in universities has been done in Internet2 [8].

Identification and authentication of users are two interrelated concepts. When a service authenticates a user, it obtains assurance about her identity. A common way to authenticate a human user is to ask her to enter a password. The use of passwords is considered as weak authentication, as passwords can be guessed, sniffed, shoulder-surfed or just found on a piece of paper under the keyboard. There are also stronger ways to authenticate a user, such as one-time-passwords and public key infrastructure (PKI).

Authorisation means deciding, who is allowed to access a system (such as a web service) and which operations she is allowed to do in it. Authorisation is done by the party controlling the system. It can be done on individual level by maintaining a list of the identifiers for the authorised users. However, in most cases authorisation is based on the role of the user. An example of role-based access control (RBAC) is that only students are allowed to enroll in an exam of a university course, and only staff is allowed to check, who have enrolled in an exam.

Traditional view of networking has been a set of services, which have been interconnected by the network. Middleware is a layer of abstraction between the applications and the network infrastructure. It covers technologies like remote procedure call (RPC), quality of service, distributed computing (such as Grid) and so on.

One aspect of middleware is administering users and their privileges on services in the network (Figure 2). On one hand it covers identification and authentication of the network users, on the other hand also the mediation of their roles and other attributes that are necessary for deducing what the users are authorised to do in the network.
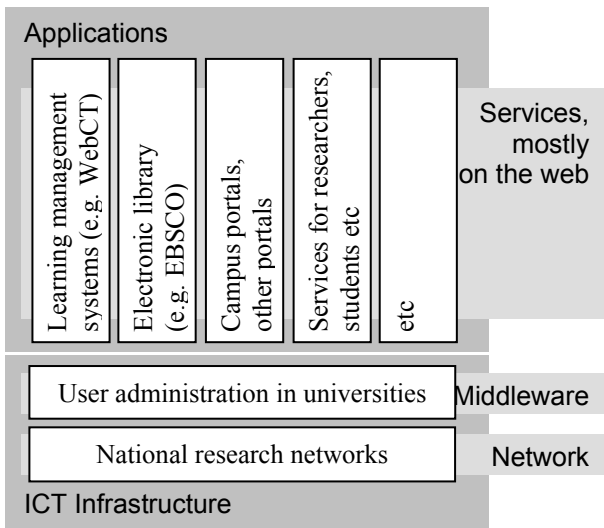
Figure 2. User administration is a middleware component between the applications and the network.

Formerly the network services were provided mostly by the university in which the user was studying or working. As the co-operation of universities increases, the user may not necessarily belong to the same organisation that provides the accessed service. The user, for example, can be a researcher that is accessing a national research portal or a student studying a distant course provided by another university. Most of these services need to be aware of the identity and/or the role of the user in her home organisation.

2   Scope of the User Identity

In an organisation the scope of the user identity can be threefold. The identity can be scoped for only one specific service, or the user may have the same indentity in all the services in the organisation. It is even possible to use the same identity in services outside the organisation. The scope of the identity and its relation to user authentication is sketched in Figure 3.
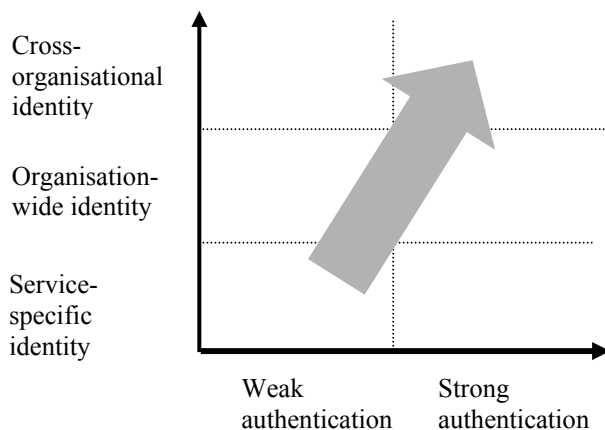
Figure 3. Scope of the user identity and the required reliability of user authentication. [6]

2.1   From Service-Specific Identities to Organisation-wide Identities

In a university an average user is usually authorised to use several information systems, for example workstations and servers in Unix and Windows environments, web based services such as university portals, learning management systems (LMS), dial-up services etc. If the user has different identities in each service, she probably has to remember several username/password pairs in her daily life (lower left corner of Figure 3). In one service Bob Smith is known as *bsmith* and in another as *bobsm*, and the passwords in the services are different unless Bob has synchronised them by himself. If the user administration of the information systems relies on service-specific identities, introducing a new personal service on the network means giving a new username/password pair to the users.

As services in the network proliferate, administration of user identities causes a significant amount of work both for the organisation and the user herself. Replacing the service specific identities by one organisation-wide user identity for each user in the organisation (the middle row of the figure) reduces overlapping work and is also comfortable from the usability perspective. The user has one single username/password that is used in all the services that she is authorised to use in the university.

In other words, organisation-wide identity separates the administration of identity from the administration of authorisation; granting access to a new service does not anymore mean issuing a new service-specific identity to the user. Instead it means authorising an existing user with a known username to use the new service.

The use of an organisation-wide identity is also motivated by information security. When a user leaves the organisation, for example when a student graduates, her user accounts in all the services in the university should be inactivated. In contrast to opening an account, the user is not usually motivated to actively take care that her user accounts are closed as she leaves. Closing all the service specific user accounts is a considerable task unless the user has one organisation-wide identity.

Once organisation-wide identities are introduced, it makes sense to use some time in integrating the user administration to other databases in the organisation, such as to the student registry and the payroll system. If the number of users entering and leaving the organisation is large, manual work and latency can be reduced if the accounts are automatically closed right after the person has left the organisation.

Integration of user account databases and other databases introduces the concept of a metadirectory, which is a directory bringing together the strategic directories that the organisation has. A property of a metadirectory is that once a piece of information is changed in some database, the related changes are mediated to all the other relevant databases in the organisation. For example, as a student graduates, the event is propagated from the student registry to the user administration to close her user accounts, to the library to

close her patron files there, to the alumni database to introduce a new alumnus and so on.

There are different technologies available for user administration in a university. Relational databases, which have been connected to student and employee databases, are used widely. When a new student is added to the student registry, a new user account is automatically created in the user administration database. Directories based on Light-weight Directory Access Protocol (LDAP [13]) have also become popular. Companies, such as Novell and Microsoft, are in the market with their enterprise directory products.

The problems faced in the deployment of organisation-wide identities are not only technical but organisational. Administrative processes in the organisation have to be justified to ensure smooth operation. For instance, if the user account is closed immediately after the working contract ends and the person is removed from the payroll system, the new contract has to be made in time if the employment still continues. Otherwise the user's account is closed and she is not able to do her work.

Co-operation between different organisational units inside the administration of a university and between the administration and the faculties is needed for example in the integration of student registry and user database. Sufficient level of trust between organisational units is necessary, which is challenging as shown by Allen [1].

2.2 Using Network Services Across Organisational Boundaries

So far the discussion has been limited to using services inside an organisation. However, co-operation between organisations is increasing, and as a result the user of a service does not necessarily belong to the same organisation as the service.
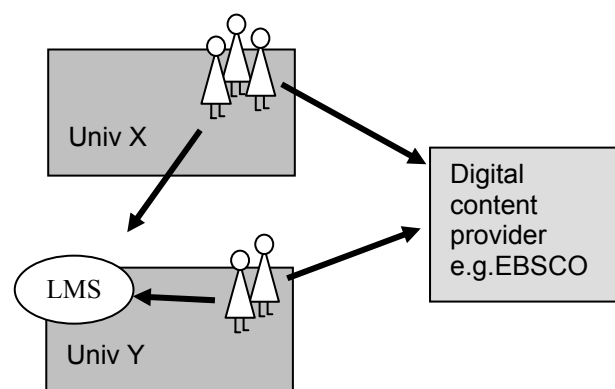
Figure 4. Cross-organisational use of network services.

For example, a student from university X ("origin site") may attend a course provided by university Y ("target site"), and the course may use some web based learning management system (LMS) such as WebCT (Figure 4). The target site somehow has to identify and authenticate the user, and obtain assurance of her authorisation for the service she is accessing. To avoid

assigning new identities and issuing new usernames for the user, her identity should be mediated from the origin site to the target site. This is called cross-organisational identity, or federated identity in short (top row in Figure 3).

However, it is not always necessary to uniquely identify the user. In some contexts, it is sufficient to make sure the user is authorised to access the service. For example, the university libraries may have subscribed certain digital content to all the researchers and students in the university. For the content provider (such as EBSCO, a provider of digital contents for university libraries, Figure 4), it is enough to know, that the user accessing the service is either a student or a researcher of the university. From the data protection point of view the content provider should not even get the identity of the user, only her role in the university.

2.3 Identification and Authentication of Users

The need for stronger authentication increases as the scope of user identity gets larger. As the user has the same identity in various network services in her organisation and even across organisational boundaries, the risk of an authentication failure gets bigger. There are more trusted components in which a security vulnerability can cause the security to fail, and once impersonation becomes possible for an attacker, there are more places in which the identity can be abused. Thus, deploying cross-organisational identities increases the demand for strong authentication (the arrow to the upper right corner in Figure 3).

There are different ways to implement strong authentication. Some European governments have plans on launching an identity card for the citizens. The identity cards contain a chip, which utilises PKI in authenticating the user for public and private network services. The chip can be inserted in a mobile phone as well, removing the requirement for an external smart card reader. Smart cards and PKI can be utilised for strong user authentication in the universities as well. Experiences on deployment of PKI based on smart cards are documented for example in [12].

Single sign-on is a commonly referenced concept related to user authentication. For the user, single sign-on means the ability to authenticate only once, and then have access to all the resources available without any further authentication. Single sign-on architectures have been studied for example by Clercq [3].

3 A Model for Cross-organisational Use of Personal Services

This chapter introduces a model for cross-organisational use of personal services. The model contains three entities (Figure 5): the origin site, the target site and the user in question.
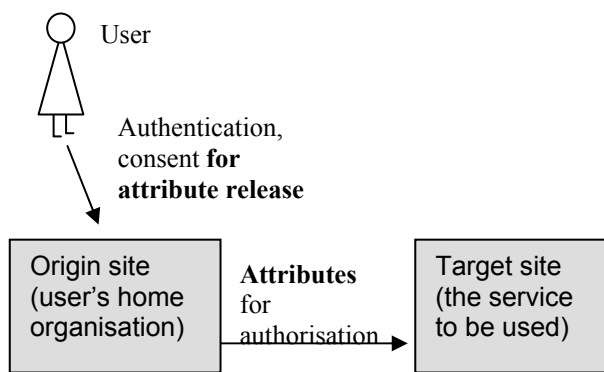
Figure 5. Entities in cross-organisational use of personal services.

3.1   Origin Site

The origin site is the entity that assigns an identity to a user. The identity is represented by an appropriate identifier, such as a user name, which is unique in one domain. Some architectures use identifiers that are globally unique (for example the phone numbers used for identifying a subscriber in the public switched telephone network), in other architectures federated local identities are enough (for example, the identity federation in the Liberty Alliance project to which we will return in Chapter 4). From the data protection point of view, federated local identifiers are preferred, because global identifiers make it easier to aggregate personal information from different sources, causing a violation of privacy.

If global user identifiers are used, a method for implementing global uniqueness is to introduce hierarchy to the namespace. Each institution administrates its local namespace, and some global unique identifiers such as domain names are used to distinguish between organisations. EduPerson [4] suggests that a new attribute eduPersonPrincipalName is introduced in higher education. Bob Smith, for instance, could be known as *bsmith@univ.edu*.

A drawback of hierarchy in the namespace is that once a person is for example a student in two universities, she has automatically two identities. In most cases, this is not a problem, as she probably is acting in some role in one of the two universities, such as as a distant course student in order to include the course in her studies in university X. However, in certain circumstances the two identities can be problematic or at least confusing for the user herself.

Revoking and reassigning unique identifiers in cross-organisational user administration is as problematic as in intra-organisational user administration. If Bob Smith leaves his university, can his unique identifier *bsmith@univ.edu* be later assigned to a Bill Smith starting his studies at the university? If yes, how can it be prevented that the "new" *bsmith@univ.edu* gets access to the information the previous one has left for example to a learning management system he has used?

The origin site is not only responsible for administrating the unique identifiers, but also other attributes belonging to the user. These include her name and other contact information, such as phone number and email address. If some of these change, the user has to remember to update her contacts only to the origin site, and the changes can be mediated automatically to the targets. The origin site also maintains attributes expressing the user's relationship to the home organisation, such as the information that she is a student, professor etc.

A special set of attributes are the credentials used for authentication, including for example passwords and certificates. Maintaining appropriate means for user authentication is the responsibility of the origin site. Some services may have higher requirements for the reliability of the authentication, making it necessary to maintain several credentials for one user; less sensitive services can be used anywhere, more sensitive only on a workstation with appropriate equipment such as a smart card reader.

3.2   Target Site

The target site is the organisation that controls the service the user wants to access. The target site can be, for example, another university whose learning management system is used in some distant course. In other words, a university can act both as a origin site and a target site. The target site can also be some national level organisation such as a national portal for researchers or students, or some commercial content provider, such as EBSCO.

It is expected that the target site wants to control who is able to access the service. The target site authorises the users based on the attributes provided by the origin site. For example the student portal lets only students access the service.

3.3   User

The user is a member (student, staff, faculty, etc in the context of universities) of an origin site. She uses the services provided by target sites and has access only to the services permitted for her.

The European Union Directive (95/46/EC) on data protection requires that in most cases a data subject has to give her unambiguous consent for dissemination of her personal data. User consent for transmission of personal information from the origin site to the target site is needed. If the user is not willing to release attributes that are necessary for the target site, access may be denied or granted only to some lower service level. If the target site gets only information about the role of the user (such as that she is a student at university X) but the identity of the user is not disclosed, the disseminated data is not considered as personal and problems related to data protection become easier.

4 Requirements for Cross-organisational Use of Services

This chapter introduces requirements for an architecture based on the model presented. The requirements incorporate an agreement about the protocol used in communications between the entities, trust between them, the schema used for attributes exchanged by the origin and target sites, and the security infrastructure used in securing the communications. The agreement is made between the entities involved in cross-organisational transactions, forming a community called a federation.

4.1 Protocols Used in the Communications

The traditional protocol for transferring personal information in the Internet is LDAP [13], that is based on X.500 directories. LDAP is commonly used in white page directories, which can be used like phone books to find contact information for people. LDAP is also widely deployed in user administration inside organisations, and products like Novell eDirectory and Microsoft Active Directory support it.

User

1. Authentication:
bsmith@univ.edu/D6hkRtqJ

Target site

4. Hi Bob…

2. LDAP query
bsmith/D6hkRtqJ

Univ X
Origin site   LDA

3. LDAP response
Authentication ok.
He is Bob Smith,
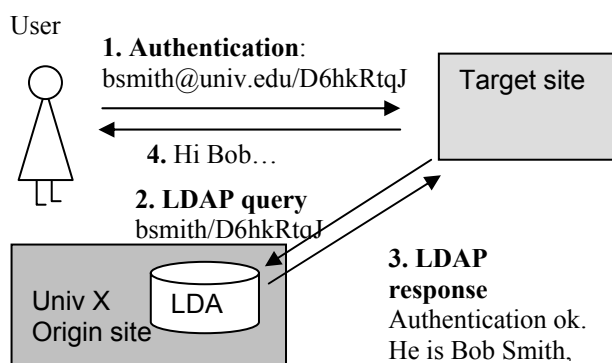
Figure 6. Use of LDAP in cross-organisational user administration.

LDAP can be used for cross-organisational use of services as depicted in Figure 6. The user (Bob Smith) gives his unique identifier (*bsmith@univ.edu*) and password (D6hkRtqJ) used in the origin site to the target site. The correctness of the password is checked against the LDAP directory of the origin site, and the directory provides attributes (e.g. Bob's name and role as a student) to the target site.

However, use of LDAP in cross-organisational user administration has a drawback. As long as authentication is based on a shared secret (such as a password) Bob has to reveal his password, a most sensitive piece of information, to the target site. This causes two risks:

1. If the security of the target site is not properly taken care of, Bob's password can be compromised. For example, if the target site uses basic authentication on top of plain HTTP (not HTTPS) in communications with Bob's web browser, his password is transmitted to the target site in cleartext and it can be sniffed from the network. The origin site has little chances to ensure that the security in each target site is up-to-date.

2. If the architecture above becomes a standard practice and dozens of services start to use it, Bob has no real chance to deduce which service is trustworthy and which is not. Entering the password to any untrusted service is risky. A fake service, whose only intention is to gather passwords from careless users, would probably be a success for a cracker.

Kerberos protocol has been a traditional solution to the problem. New protocols overcoming the problem have been introduced on the WWW. The authentication of the user is always done in the origin site, which provides assertions about the user's identity or other attributes to the target site. As Bob's password is never passed to the target site, a compromise in target's security does not reveal the password, and the damage is restricted to the particular target site. On the other hand, Bob is always authenticated by the familiar authentication server in his home organisation, and he can be told not to provide the password to any other web server.

The Shibboleth protocol specified and implemented by Internet2 is a notable example of such a protocol [9]. The protocol utilises SAML (Security Assertion Markup Language) and SOAP (Simple Object Access Protocol) in the communications between origin and target site. After piloting in some universities in the United States, the first versions of the open source implementation have been released. Attributes of the user are passed to the target site by the Shibboleth protocol, and the target makes the access control decision based on them. The implementation of Shibboleth also provides a mechanism for the user to give her consent for attribute release.

PAPI (Point of Access to Providers of Information) is another protocol used in cross-organisational use of resources [2]. The protocol implemented by RedIris is commonly used for accessing electronic resources in the Spanish higher education.

There are also activities outside the academic communities. The Liberty Alliance has defined a protocol for federating identities between organisations [11]. The focus of the Liberty Alliance project is to get rid of the several identities and related username/password pairs that a user has in electronic services in the Internet, without introduction of a centralised architecture and a globally unique identifier. In the Liberty architecture, the user is authenticated by an identity provider (origin site) and the identity is then federated to service providers (target site), providing a single sign-on experience to the user. In public, the decentralised Liberty protocol is considered to be a challenger for the Passport protocol, whose architecture is centralised around Microsoft.

4.2 Trust between the Entities

In the federation the target sites have to trust the origin sites, which maintain the identity and attributes for the users and the credentials necessary for authentication. The security of the user administration of the service relies on the assertions that the origin site has provided about the users. Thus, the user administration of an

organisation needs to be implemented properly before the organisation can enter the federation as an origin site. For instance, the origin site has to ensure that the user account of a certain person is closed when she leaves the organisation.

Also the origin sites have to trust the targets to properly handle the attributes released by the origin. A specially sensitive user attribute is the password used for authentication, if the origin or the user sends it to the target site as cleartext. The architectures presented above, where the user is always authenticated by the origin site, lower the required trust considerably because the passwords never reach the target site.

The user is concerned about her privacy, and she has to trust the origin site that it will not release attributes to the target site without her consent. Even if the user gives her consent for attribute release, only attributes necessary for the target site may be released. The user also has to trust the origin that the log files, which may contain sensitive information about services the user has accessed, are not used to violate her privacy. In most cases, the user privacy is protected by the data protection legislation.

Agreements between origin and target sites are expected to ensure a certain minimal level of security controls implemented by the sites. To avoid many-to-many relationships between the origin and target sites, the federation agrees on the minimal requirements for joining organisations. There may be several federations for services with varying sensitivity; for example, the accuracy of assertions on the user attributes for students is probably lower in library services than in health care services.

In Liberty Alliance, the federation for trust establishment is called a Circle of trust and in the Shibboleth project a Shibboleth club. The requirements for an organisation joining inCommon, the Shibboleth club formed in Internet2, are drafted in [10].

4.3   Schema for the User Attributes Exchanged

The schema describes the user attributes exchanged in the federation. It should cover the syntax and semantics of the attributes, including the vocabularies. The schema should provide attributes for identification, authentication, and authorisation of users.

A lot of schemas have been specified for the directories in the Internet. The basic set of attributes have been defined in Internet standards and are widely used in the phonebook-like white page directories. Attributes such as the given name, surname, postal address, email address, phone number and user password and certificate are specified in [14, 15].

However, in higher education there are requirements that the Internet standards do not fully cover. Most of them are related to attributes necessary for authorisation, because authorisation is usually based on the user's role in the organisation, and relevant roles vary from organisation to another. Usually organisations extend the schema with their own attributes, whose syntax and semantics are specific for the organisation. Enabling cross-organisational use of services, however, requires that the federation agrees on certain basic set of attributes required for authorisation in the target sites.

In the United States, Educause has defined a schema called eduPerson [4], which contains attributes specific for higher education. In eduPerson a new attribute for authorisation is eduPersonAffiliation, which expresses the person's relationship to the organisation. The controlled vocabulary contains values student, faculty, staff, employee, alumn, member, and affiliate. One person may have several roles, for example a post-graduate student in a laboratory has probably values student, faculty, employee, and member. One of them can be promoted to the primary one. European higher education has also had some interest for a similar schema [19].

The attribute eduPersonAffiliation provides a coarse basis for authorisation, as some services are provided only for students, some (such as the previous example of EBSCO) to all members, and so on. On the other hand, it opens up the problem of defining the semantics for each value. For instance, does 'student' cover only students aiming at a degree or should further education students, open university students, etc, be counted in as well?

The need for more fine-grained information about the role of a person in the organisation and the national differences in higher education have caused academic communities in many countries to specify attributes of their own, for example in Swizerland [16] and in Norway [5]. The national schemas present new attributes whose semantics utilise vocabularies maintained by national bodies, such as national statistical offices (for example in a vocabulary '311' means 'doctor of theology'). The higher education institutions already use the codes internally in the student registries.

4.4   Security Infrastructure

A security infrastructure, such as a public key infrastructure (PKI), is required to ensure the authenticity and integrity of the messages exchanged. The origin site and the target site require certificates for mutual authentication and the integrity check of the assertions exchanged. The certificates and SSL/TLS protocol are also used for authenticating the origin and target sites to the user.

For the time being, personal certificates are not widely used, and the authentication of the users cannot be based on PKI on a large scale. Passwords and other weaker means are used instead. However, as the authentication of a user is a local matter for each origin site, the use of strong authentication is not restricted by any design choice. Instead some services may require authentication that is stronger than passwords.

There are several commercial Certificate Authorities (CA) available, and in higher education some universities and research networks have also established a CA of their own. A small number of CAs trusted by the federation is expected to be used for server certificates in the origin and target sites.

5 Initiatives Going On in Higher Education in Europe

Intuitively, the problem of user identification and cross-organisational use of services appears to belong to the higher education institutions, because the users are usually students or employees in some institution. In Europe, the activities, however, are not driven by European University Information Systems association (EUNIS) but by the association of national research and education networks (TERENA). To ensure that the work done in national research networks fulfils the requirements for cross-organisational user administration, discussion and exchange of information between the two associations would be helpful.

In TERENA, related work is done in a specific Task Force TF-AACE (Authentication and Authorisation Coordination for Europe) [18]. TF-AACE is a gathered group of people from individual research networks, who have their own projects, e.g. in Spain (RedIRIS), Netherlands (Surfnet), Switzerland (Switch), Norway (Uninett) and Finland (Funet). TERENA has also trans-atlantic co-operation with Internet2, which develops the Shibboleth protocol and eduPerson schema and has also other related activities.

Switch has been the forerunner for Shibboleth in European research and education network. The AAI (Authentication and Authorisation Infrastructure) project [17] has also specified a schema for attributes used in Swiss higher education. The Norwegian FEIDE project has a schema for LDAP directories used in Mellon o Moria, the architecture designed for cross-organisational use of personal services in Norway [5].

In Finland, the HAKA project, a common project for Finnish higher education, has started pilots for cross-organisational user administration. In the pilots, Shibboleth protocol and funetEduPerson, the Finnish equivalent to eduPerson, is used for accessing services in the portal of the Finnish Virtual University, the Finnish Virtual Polytechnic and the Finnish Electronic Library. More information is available in [7].

6 Conclusions

Demand for middleware that mediates user identities and attributes between services inside an organisation and between organisations has increased. Driving forces are the growing number of personal services in the network, increasing co-operation between organisations, and requirements for flexible and easy use of services without compromises in the information security and privacy.

This paper outlined a model for cross-organisational use of personal services and the requirements implied by the model. In European universities, there are activities aiming at building an infrastructure for cross-organisational use of personal services. The challenges, however, are not only technical but also political and cultural, requiring a new kind of co-operation and trust between organisations and organisational units.

References

[1] D. Allen, "Information infrastructures, information behaviour and trust". Proceedings of EUNIS2002, the 8th International Confrence of European University Information systems. pp.167–178, (2002).

[2] R. Castro-Rojo, D. R. López. "The PAPI system: Point of Access to Providers of Information". Proceedings of TERENA Networking Conference, (2001).

[3] J. D. Clercq. "Single Sign-On Architectures". InfraSec 2002, LNCS 2437. pp. 40-58, (2002).

[4] Educause. "eduPerson Object Class". http://www.educause.edu/eduperson/

[5] FEIDE project. http://www.feide.no/index.en.html

[6] Greater Nordic Middleware Symposium. "GNOMIS". http://www.uninett.no/arrangement/gnomis/

[7] HAKA project. http://www.csc.fi/suomi/funet/middleware/english/index.phtml

[8] Internet2 Middleware Initiative. "Identifiers, Authentication, and Directories: Best Practices for Higher Education", (2000). http://middleware.internet2.edu/internet2-mi-best-practices-00.html

[9] Internet2/MACE. "Shibboleth Project". http://shibboleth.internet2.edu/

[10] N. Klingenstein. "Draft Club Shib outlines", (2002) http://shibboleth.internet2.edu/docs/draft-internet2-mace-shibboleth-club-shib-guidelines-01.txt

[11] Liberty Alliance Project. http://www.projectliberty.org/

[12] M. Linden, P. Linna, M. Kivilompolo, J. Kanner. "Lessons Learned in PKI Implementation in Higher Education". Proceedings of EUNIS2002, the 8th International Confrence of European University Information systems. pp.246-251, (2002).

[13] RFC 2251. "Lightweight Directory Access Protocol (v3)". Internet Engineering Task Force, (1997).

[14] RFC 2256. "A Summary of the X.500(96) User Schema for use with LDAPv3". Internet Engineering Task Force, (1997).

[15] RFC 2798. "Definition of the inetOrgPerson LDAP Object Class", Internet Engineering Task Force, (2000).

[16] The Swiss Education and Research Network. "Authorization Attribute Specification", (2002). http://www.switch.ch/aai/docs/AAI_Attr_Specs.pdf

[17] The Swiss Education and Research Network. "Authentication and Authorization Infrastructure". http://www.switch.ch/aai/

[18] Trans-European-Research and Education Networking Association. "TF-AACE: Authentication, authorisation coordination for Europe". http://www.terena.nl/tech/task-forces/tf-aace/

[19] Trans-European-Research and Education Networking Association. "TF-LSD: LDAP Services Deployment". http://www.terena.nl/tech/task-forces/tf-lsd/

Providing Quality of Service in Wide Area Networks

Ursula Hilgers and Peter Holleczek
Regionales Rechenzentrum der FAU Erlangen-Nürnberg, Martensstraße 1, 91058 Erlangen
{hilgers,holleczek}@rrze.uni-erlangen.de

Richard Hofmann
Universität Erlangen-Nürnberg, Informatik 7, Martensstraße 3, 91058 Erlangen
rhofmann@informatik.uni-erlangen.de

IP networks are more and more used for transporting distributed applications. Examples are teleconferences or high quality video and audio transmission. Adding such services to IP networks continuously raises the quality of service requirements. As the "best effort" service of the IP protocol is not able to fulfill these demands, new mechanisms must be provided in network components. In particular, they must allow data transport with different transmission characteristics. We have investigated these mechanisms by measurements on current network devices like routers and switches in order to find out, what quality of service can be expected in existing networks. Based on the results of theses measurements, we developed a network architecture that is able to fulfil the quality of service demands of all except the most demanding applications on current IP network technology.

1 Introduction

Service integration into current IP networks requires transporting data of distributed applications with different service requirements over currently available IP infrastructure. In addition to that, service requirements of applications grow continuously. E.g. multimedia applications require end-to-end performance guarantees, sometimes even with hard realtime behaviour. On the other hand, planning of Wide Area Networks is often done without considering these Quality of Service (QoS) requirements of distributed applications. Because of this, a network architecture must be developed which allows to transmit distributed applications with different requirements without any performance degradation.

This paper is organized as follows: In section one, we develop a service class concept that defines the QoS characteristics of different distributed applications. In section two, we introduce those mechanisms implemented in network components that are destined to provide different service behaviour in networks. Although our measurements show that these mechanisms suffer from performance lacks, they can be used for providing QoS services in current IP networks. In the last section we present a QoS architecture based on these mechanisms that takes care of their restrictions and hence allows to provide the specified services in WANs.

2 QoS requirements of applications

Quality of service (QoS) is defined by the ITU–T as the collective effort of service performance and as such determines the overall degree of satisfaction of a user with a service [31].

QoS parameters are needed to specify the requirements of applications concerning the behaviour of the commmunication infrastructure between two end systems. There are many parameters that are mentioned in literature [28, 30]. In this work, we will focus on the following objective parameters:

– The *transfer rate* describes the bandwidth requirements of an application. In contrast to applications that send a constant data stream, there are others with bursty sending behaviour. The latter ones produce a traffic stream characterized by a mean rate, a peak rate and a burst size.

– The *delay* defines the latency of the transmission between a sender and a receiver [41]. It is calculated as the sum of the latency on the transmission links, in the network components and in the end systems. In this work, the latter shall not be considered here but is a focus in [24].

– The *jitter (delay variation)* describes the variance of the delay in a network [13], e.g. the quality of the synchronisation of video frames in a data stream: If the jitter is two big, the voice seams to be chopped and there are artefacts in the video.

– *Loss Rates* should be low in networks. Loss rates include the loss of complete data units as well as single bit errors.

– High quality multimedia applications require hard realtime behaviour. This means that the value of an action decreases against $-\infty$ if a given deadline is exceeded. With soft realtime behaviour the value of an action decreases continuously without any strong impact [36].

– The *availability* is the probability that a system acts without an error at a certain point of time. In this paper, we will regard on the one hand applications which require an availability near one (although this is hard to realise in computer networks). One example is the transmission of applications in telemedicine. On the other hand there are applications which do not suffer too hard if there are little failures in the network.

In addition to these parameters, the *subjective transmission quality* is an important parameter to describe the QoS requirements of an application. There a human spectator evaluates the perceived quality of a considered application. The rating can be done according to the Mean Opinion Score (MOS) which is presented by the ITU–T [32]. The MOS has the categories "excellent", "good", "fair", "poor" and "bad".

In the following section, applications with different characteristics are of interest e.g. text based data transmission, Voice over IP applications, teleconferencing applications or high quality audio and video applications. In a lab environment we determined the QoS requirements for the following applications by measurements.

2.1 Text-based applications

Most of the applications transported over the internet are text-based applications. They use the Internet Protocol (IP) for data transmission [11]. IP is a network-layer protocol that contains addressing information and some control information for routing packets in networks. On top of IP, transport protocols like TCP and UDP are implemented.

2.1.1 TCP

The Transmission Control Protocol (TCP) is responsible for verifying the correct delivery of data from a sender to a receiver [12]. TCP detects errors or lost data and triggers retransmission via an acknowledgement algorithm until the data is known to be received correctly and completely. Furtheron, mechanisms to react on congestion in networks are added in the protocol [34]: the slow start and congestion avoidance algorithm.

These algorithms assume that packet loss caused by real damage is very small (much less than 1%). Instead, the loss of a packet signals congestion somewhere in the network between the source and destination [46]. If a packet gets lost, the slow start mechanism reduces the transmission rate

of the sender, decreasing congestion in the network. After such a reduction, the congestion avoidance algorithm again increases the transmission rate of the sending application step by step. If a packet gets lost on all TCP connections in the network, all TCP senders begin a slow start and the network throughput is extremly reduced. This effect is called global synchronisation.

To test the quality of service requirements of TCP applications in a lab evironment we used the test setup in figure 1 [26]. Two Ultra 60 SUN workstations exchange TCP traffic. They are both linked to a Cisco router. The connection between the Cisco routers is an ATM link. Into this link, an impairment generator is inserted. This tool is able to simulate "realistic" network behaviour as it can generate bit errors and it can delay packets of a traffic stream.
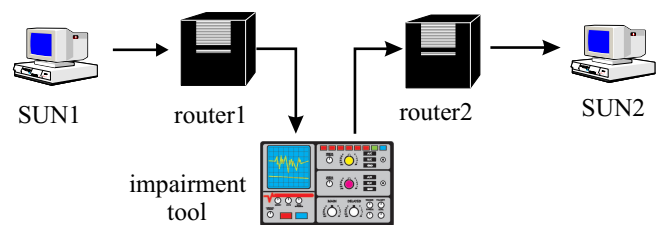

Figure 1: Test setup to test the QoS requirements of text–based applications.

As an example for an application with only small bandwidth requirements the sender sends TCP traffic with the tool `ttcp` [7]. The maximum throughput between the two workstations in this test setup is 65 MBit/s with TCP packets of 429 bytes. (This is the average packet size in the german scientific network [15]).

First, a constant delay is inserted in the TCP stream. It can be recognized, that with a delay of 10 ms the throughput between the two SUNs is reduced to 10 Mbit/s. The throughput decreases because for each TCP segment the sender waits for the acknowlegement of the receiver. New packets are not sent before the old ones are acknowledged. With a delay greater than 300 ms the bandwidth even reduces to a value smaller than 1,5 Mbit/s and the subjective quality is poor.

In the next step, bit errors were produced by the impairment tool. With a bit error rate more than 10^{-5}, the interaction of the user with the `ttcp` tool is no longer satisfactory because the response time of the tool `ttcp` is too big [26]. The reasons for that behaviour are the retransmissions of the TCP segments. With a high error rate, the TCP protocol has to retransmit packets several times until the sender receives the acknowledgements of the receiver.

The same tests were redone for the applications `ftp` which uses a transfer rate about 20 MBit/s. The tests show, that the delay should be less or equal to 200 ms to obtain a satisfactory transmission quality. The bit error rate should be less than 10^{-6} [26].

2.1.2   UDP

The User Datagram Protocol (UDP) defines a mechanism to transport data over networks with a minimum of protocol mechanism [43]. The protocol is transaction oriented, and delivery and duplicate protection are not guaranteed. One example for an application which transports traffic over UDP is the network management protocol SNMP [9].

To check the QoS requirements of UDP traffic, a UDP traffic source sends data from SUN1 to SUN2 (see figure 1). As UDP is not a protocol with an acknowledgement mechanism, an increasing delay has no influence on the transmission quality. But if an error rate is generated by the impairment tool to degrade the quality of the UDP stream, it can be seen that the bit error rate should be less or equal to 10^{-6} to get a satisfactory transmission quality [26].

2.2   Multimedia applications

As the number of applications increase which transports video and audio data over IP networks, the QoS requirements of these applications are of interest. Voice over IP, adaptive applications and applications to transmit video and audio are regarded.

2.2.1   Voice over IP

Voice over IP applications transport the voice packets packet switched over point to point network connections. First, a signalling protocol, e.g. *H.323* [33], establishes a connection. The IP packets are transported with the Real Time Transport Protocol (*RTP*) [44] on top of UDP. RTP allows to transport data with realtime characteristics and supports unicast and multicast data streams. Each IP packet is preceeded by an RTP header that contains timestamps. This feature allows the receiver to compensate jitter produced in the network.

To test the QoS requirements of Voice over IP applications, two telephones are connected over a network, in which the impairment tool is integrated. Also in this test, the impairment tool generates bit errors and a delay to reduce the transmission quality of the network connection.

The test demonstrates, that the bit error rate should be lower than 10^{-6} to get a good transmission quality of the voice traffic [18, 26].

If the voice packets are delayed by the impairment tool, the latency should be below 200 ms so that the speakers do not keep interrupting each other. The jitter should be lower than 120 ms to get a good quality [47].

2.2.2   Adaptive applications

Adaptive applications continuously control the bandwidth at their disposal. Data is transported over RTP. Using the information in the RTP header the application can see the available bandwidth in the network. Corresponding to the available bandwidth, the sending bitrate of the application is adjusted. Examples for adaptive applications are the teleconferencing tools `vat` and `vic`.

Tests in our lab environment showed that adaptive applications must have a bandwidth not lower than 0.2 MB/s to obtain a good transmission quality [16, 26]. The bit error rate should not be higher than 10^{-6}. The delay should be lower than 200 ms and the jitter lower than 120 ms [47].

2.2.3   High quality audio and video

In [16, 38] tests are presented that define the QoS requirements of hardware codecs which are able to transport high quality video. Hardware codecs compress the video frames with different compression algorithms before they are transmitted over networks. The video frames are transmitted between Coder and Decoder in an AAL5 ATM stream.

Our measurements showed that, as soon as the bandwidth for the video stream is reduced, the objective quality of the stream is not longer acceptable. These objective test results correspond to the results of subjective tests. As soon as the numver of error-free video frames decreases, the transmission quality reduces.

MJPEG codecs can be used e.g. for teleteaching applications. They require a transmission rate of about 11 MBit/s [27]. The tests to define the QoS requirements in networks showed, that the error rate should not be greater than 10^{-6} and the delay lower than 200 ms. In addition to that, the jitter should not be greater than 120 ms [47].

To get a better picture quality, MPEG codecs with MPEG4:2:2P@ML are used [37]. If e.g. applications in telemedicine are considered, the loss rates should be smaller or equal to 10^{-11}. Delay should be lower than 150 ms and the jitter lower than 80 ms.

2.3   Class concept

With these results of the previous paragraph, a service class concept can be introduced:

- The first class $\mathcal{C}_1$ allows the transmission of applications with very high quality and high timing and availability requirements. Soft realtime behaviour is sufficient. The delay should be less than 150 ms and the jitter lower than 80 ms. The bit error rate should be lower that 10^{-11}.

- Class $\mathcal{C}_2$ transports multimedia data with good tranmisssion quality up to a bandwidth of 15 MBit/s. The bit error rate should be lower than 10^{-8} and the jitter lower than 120 ms. E.g. streaming applications belong to this class. Because these applications transport data only in one direction, no requirements concerning the delay are defined. Hard realtime behaviour is required and the availability can be lower than 1.

- Class $\mathcal{C}_3$ and class $\mathcal{C}_4$ transmit voice data with bandwidths lower than 0,064 Mbit/s. The availability

should be near one. The delay should be lower than 200 ms, the jitter below 120 ms and the bit error rate lower than 10^{-6}. Soft realtime behaviour is sufficient. The difference between C_3 and C_4 is that C_4 includes applications with bursty sending characteristics whereas class C_3 considers applications with constant transfer rates.

– Class C_5 transports network management traffic with only small transfer rates. No realtime and latency requirements are necessary but a high availability. The bit error rate should be lower than 10^{-6}.

– Class C_6 and C_7 contain text based applications. They require no hard realtime behaviour and the availability can be lower than 1. C_6 contains applications with high bandwidth requirements like `ftp`. The bit error rate should be lower than 10^{-6} and the delay lower than 200 ms. Text based applications with interactive character send data with small transfer rates. For those applications, the bit error rate should be lower than 10^{-5} and the delay lower than 300 ms.

– Class C_8 contains all applications with no QoS requirements, the "best effort" traffic.

With this class concept any application can be assigned to a class that fits the best to its requirements. Hence, these classes form a class concept, which encloses all possible applications. In addition, it helps users to describe the QoS characteristics of their applications. They are able to specify the required transmission service and this can help to negotiate service level agreements (SLA) with providers.

3   Providing QoS in networks

As the "best effort" service of the IP protocol is not able to provide different service classes, more sophisticated mechanisms must be activated in network components. In this section, we first investigate implementations of such mechanisms by measurements on currently available routers. With the results of these experiments, the implementation is compared to the required functionality and performance of the mechanisms.

3.1   Classification

In order to provide different service classes in IP packets, the basic issue is to classify the data stream. Then packets of different service classes can be distinguished in the network components and treated in a different way.

The classification of IP packets is based on flows. A flow is a set of all packets from a single application session. They have the same sender or receiver IP address and port number [48]. All packets of a flow request the same sevice in the network.

To classify IP packets, the ToS byte in the IP headers is used (figure 2). Cisco routers use the first three bits of the field, the precedence bits, to classify packets [10].
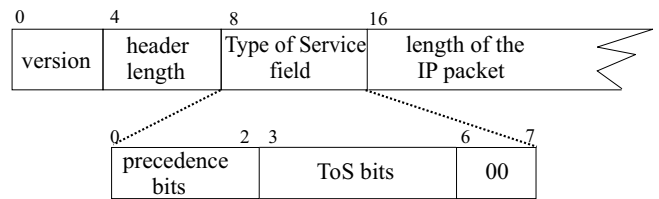

Figure 2: The first fields of the IP header with the Type of service field (ToS) [11].

Since applications are not able to mark the ToS byte, network components must change the ToS byte before a packet enters the network.

In [23] it has been shown, how the CPU load of an interface of a 7507 Cisco router is influenced by the classification task which is applied to an increasing number of flows. If there are 30 flows entering the system, relabelling the packets generates such a large CPU load that packet losses appear. This is a very cirtical issue, as usually there are thousands of flows on an entrance at the boundary of a network.

This effect is caused, because the classification task on the considered interfaces is performed in software. To activate this mechanism on interfaces with high bandwidths, this task should be executed on router interfaces by high–performance ASICs.

3.2   Congestion Management

To increase the availability and functionality of networks, they must react as soon as the load in the network grows. As a sign of congestion buffers and queues fills up and the delay in the network increases. In the end, packets are dropped. Because of that, mechanisms to control the network load and to avoid traffic congestion must be implemented in network components.

3.2.1   Traffic control

To reduce congestion mechanism are needed to control and reduce the load of traffic streams. When the traffic is controlled at the entrance of a network, packets can be dropped, e.g. if agreed service rates, which are specified in service contracts, are exceeded. By this, resources in the network are protected.

There are two mechanisms to control and regulate the transmission rate of traffic streams: policing and traffic shaping. The goal of both mechanisms is to reduce a traffic profile to given bounds. As policing, traffic shaping reduces the transfer rate of a traffic stream, but the packets are not dropped at once. Instead they are buffered to smoothe bursty traffic.

Both mechanisms are implemented by token buckets. A token bucket specification consists of two parameters: a token rate R and a bucket size B. The token rate R specifies the continually sustainable data rate. The bucket size B

specifies the amount by which the data rate can exceed R for a short time.

Tests in our lab environment showed, that policing and traffic shaping are available on network components nowadays [26], but the control mechanism must be implemented in special ASICS. Otherwise we can notice the same performance impact which we have seen in paragraph 3.1.

A possibility to influence delay, jitter and loss rates by this mechanism is to vary the bucket size. If it is increased, the loss rate of bursty traffic streams can be reduced because they are buffered instead of dropping them. But delay and jitter can increase because the queue length grows. Considering our class concept, traffic control should not be activated for traffic of applications with hard realtime requirements (e.g. class C_1) because then the loss rate increases.

On Cisco routers, there exists an implementation with one seperate token bucket definition for each traffic class. In congestion situation, this behaviour allows to drop low priority packets while high priority packets are saved.

3.2.2 Congestion avoidance

Congestion avoidance algorithms like Random Early Detect (RED) remove congestion as soon as an overload situation appears [20, 45]. To achieve this goal these algorithms continuously watch the average queue length avg on the outgoing interface of a network component. It is compared to two thresholds. If avg is less than a lower threshold, congestion is assumed to be minimal and the packet is queued. If avg is greater than an upper threshold, congestion is assumed to be serious and the packet is dropped. If the average queue length is between the two thresholds, then the packet is dropped with a calculated probability, which increases the closer avg gets to the upper threshold.

Protocols like TCP/IP react to packet loss by reducing the sending rate and entering a slow start phase. Because only packets of some TCP connections are dropped, global synchronisation does not appear.

By this mechanism, congestion in networks is avoided. Packets are dropped if congestion is assumed to be minimal. Because of this only some connections are notified to back off in contrast to standard tail drop behaviour where in the worst case packets are dropped from all connections. In addition to that, the queue size is controlled, so that jitter and delay are bounded, i.e. they are lower than with tail drop behaviour.

All these characteristics of the mechanisms could be observed when the mechanism was tested with TCP sources in our lab environment [25]: The throughput especially for bursty traffic sources increased and the delay decreased. But when activating RED on Cisco routers, it can be observed that the CPU load increases about 40 %. This must be considered.

However, with the activation of RED on an egress router interface there are no throughput guarantees in overload situations for high quality streams. If e.g. there are many UDP traffic sources, they do not reduce their load on the network when exposed to traffic drops. This mechanism, too, should not be activated for high quality traffic with hard realtime requirements.

On Cisco routers, there exists an implementation of RED that allows to configure different discard thresholds for different service classes [4]. This mechanism is called Weighted RED. Packets with lower QoS requirements are discarded at a very low average queue length (e.g C_8) whereas the discarding of packets which belong e.g. to class C_6 does not start until the queue is nearly filled.

3.3 Call admission control and resource management

Algorithms to manage network resources must exist to provide guaranteed QoS in networks. Call setups are accepted by resource management functions, if the call admission control (CAC) decides, that there are enough free resources in the network to fulfil the QoS requirements of the application.

The Resource Reservation Setup protocol RSVP is a unicast and multicast signalling protocol designed to install and maintain reservation state information at each router along the path for a traffic flow [6]. Because of this overhead [35] states that RSVP is no mechanism which can be implemented on interfaces with high bandwidths. So there is no possibility to do per flow reservation in IP networks nowadays.

On the other hand, applications with high QoS requirements e.g. hard realtime behaviour should not bests of two parameters: a token rate R and a bucket size B. The token rate R specifies transported over networks, where there is no resource reservation.

3.4 Scheduling

Scheduling algorithms manage the buffer size at the outgoing interfaces of network components. They are able to support several outgoing queues with different behaviour. Because of this, they can provide traffic classes with different service characteristics. They define the outgoing bandwidth and the amount of buffer space for each queue, that means service class.

The default scheduling mechanism at outgoing interfaces of routers is FiFo queueing (First in First out). Packets are transmitted in the sequence in which they arrive. If queues are congested packets are tail dropped.

To provide different service classes other scheduling mechanisms are needed. One example for such a scheduling algorithm is Generalized Processor Sharing which is known as Weighted Fair Queueing (WFQ). This mechanism is able to provide different amounts of capacity to different queues. With WFQ, each queue is assigned a weight that determines how many packets are transmitted from that queue [42].

GPS provides a way of guaranteeing that delays for token bucket regulated flows do not exceed some bounds [40]. In addition to that, flows in different queues are isolated from each other [14]. In addition to that the service rate for each class can be calculated: If there are N queues with weights ω_α, $1 \leq \alpha \leq N$. R is the physical outgoing bandwidth on the considered interface. Then the bound for the service rate T_α of queue α with weight ω_α is:

$$T_\alpha \geq R * \frac{\omega_\alpha}{\sum\limits_{i=1}^{N} \omega_i}, \qquad \sum_i \omega_i \leq 100. \tag{1}$$

Thus, the sum of all weights should be less or equal to 100.

The throughput of a service class depends on the number of classes which are configured on an interface and their weights. If there is one class, which send not so much traffic as it is allowed to, the remaining bandwidth is divided into equal parts on the other queues.
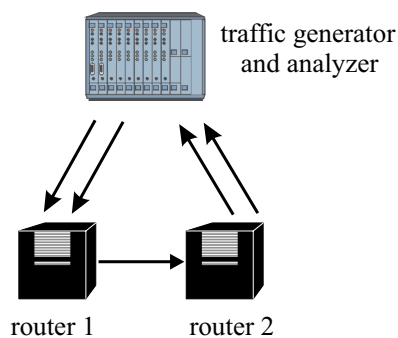
Figure 3: Test setup to check the throughput of the WFQ mechanism.

The following tests show the functionality of the WFQ implementation on Cisco routers. Two Cisco 7507 routers with operating system 12.1(3a)T1 are linked with a PoS connection (155 MBit/s). A traffic generator is connected to the routers. It sends two flows which belong to two different service classes. The sending bandwidth of each flow is 97 Mbit/s. By this, an overload situation is created at the outgoing interface of `router1`. Each flow is assigned to a separate queue and every queue receives a different weight.

With these values, first the theoretical throughput according to equation 1 (called nominal value) is calculated. This value is compared with the one, obtained by the test. By this, the functionality and the accuracy of the implementation of WFQ on the considered Cisco routers can be checked.

In the first test, the weight for flow 1 is 20 and the one for flow 2 is 40. The nominal service rate and the measured values are as follows: 50 MBit/s (nominal) and 49.6 MBit/s (measured) for flow 1 and 99.8 Mbit/s and 99.2 Mbit/s for flow 2. There is only a small difference between the nominal value and the measured one. The error is lower than 0.8 %. The test is redone with other weights and the error between the two values is about the same.

In the next test the traffic generator sent five flows which belong to the same traffic class and are stored in the same outgoing queue. We want to see how the bandwidth is divided between the five competeing flows. It can be seen that each flow of this class obtains the same part of the outgoing bandwidth.

Unfortunately, as we have seen with the other mechanism in the preceeding paragraphs, the CPU load increases about 50 % when scheduling is activated.

3.5 Routing protocols

Routing protocols forward IP packets along a route that is preferred under a performance criterion. This is a scalar metric assigned to each router-to-router hop in the network. If there are redundant paths in the network, the availability in cases of failures can be increased by routing algorithms because they are able to find alternate paths.

Constraint Based Routing is a routing mechanism, where in addition to the standard metric other parameters are considered to calculate the best path. This algorithm can be used to transport services with high quality requirements, when e.g. parameters like bandwidth or delay should be minimized on the way through the network.

4 Evaluation of a QoS architecture

As shown in the preceding chapter, there exist mechanisms to provide differentiated services in networks. The measurements show that the currently available network devices cannot provide QoS by theirselves, but they have sufficient capabilities to build a QoS architecture. In the remainder of this paper we propose a QoS architecture that allows to provide the specified services in WANs on the basis of current network infrastructure despite the performance lacks detected in our measurements.

4.1 Existing approaches

Before our proposal for an architecture is presented, existing and standardized QoS modells shall be introduced.

4.1.1 Differentiated Services Architecture

The Differentiated services architecture (DiffServ) is an architecture for implementing scalable service differentiation in the Internet [3]. IP packets are classified and aggregated to a few service classes with different QoS characteristics. The classification of the IP packets is done according to the first six bits in the Type of Service field. Classification, marking, policing, and shaping operations are only implemented at network boundaries or hosts. The marked packets receive a particular per-hop forwarding behavior on nodes along their path. This includes scheduling and congestion avoidance algorithms.

For the classification, marking, policing, and shaping operations all flows of a service class are aggregated and the

mechanisms are applied on them. This is done to avoid scalability problems in the network components. Because no resource reservation protocol with connection admission control is included in this approach, there are no quality of service guarantees for data streams with high quality of service requirements like applications which belong to classes $\mathcal{C}_1$ and $\mathcal{C}_2$.

4.1.2 Integrated Services Architecture

The Integrated Services Architecture (ISA) is a model to transport realtime services over the internet [5]. The ISA includes three service classes: a "controlled-load", a "guaranteed" and a best effort service. The controlled-load service provides the data flow with a quality of service closely approximating the QoS that same flow would receive from an unloaded network element, but uses a call admission control to assure that this service is received even when the network elements are overloaded [49]. Guaranteed service provides bounds on queueing delays between end systems. In addition it guarantees the bandwidth [50].

As the DiffServ approach the ISA uses classification, traffic control and congestion avoidance. As scheduling mechanisms WFQ is proposed. A resource reservation protocol is a key building block to manage the resources in the network. However, as all these mechanisms should be applied on single flows, this model can not be realised nowadays for performance reasons, because activation of the mechanisms that operate on a per flow basis leads to performance problems.

4.1.3 Asynchronous Transfer Mode

The Asynchronous Transfer Mode (ATM) is a link layer protocol based on connection oriented communications [21]. Data is transmitted in cells with a size of 53 Byte. To achieve connections with different QoS characteristics, the ATM Forum has defined serveral service categories [1]. Four of the most common categories are presented here. Constant Bit Rate (CBR) traffic requires a fixed data rate and a predictable response time. It can be used for video conferencing and interactive audio (e.g. telephone). Real-time Variable Bit Rate (rt-VBR) is intended for traffic streams with timing constraints and a variable bit rate. Compressed video streams with image frames of varying size can be transmitted by rt-VBR connections. CBR and rt-VBR can be used to transport applications with high priority hard realtime requirements.

The non-realtime Variable Bit Rate (nrt-VBR) category guarantees an average transmission rate with bursty character. There is no time synchronisation between traffic source and destination. Rt-VBR is used for multimedia applications which tolerate a small amount of losses like adaptive applications. The Unspecified Bit Rate (UBR) category can be used by applications which can tolerate variable delay and cell losses, e.g. file transfer and email.

The service category must be specified by the user before a connection is established. This is done by a service contract containing the service requirements of the categorie. During the connection setup, the network checks with the Call Admission Control (CAC) function if there are enough resources available on the way the connection will take through the network. The CAC function will then accept or reject the call. If the call is accepted, the required performance and QoS is guaranteed during the lifetime of the connection.

Apart from the CAC function, there are traffic management functions which control the network resources and avoid congestion in network components [1]. The Usage Parameter Control (UPC) monitors and controls the traffic and the validity of a connection. The traffic profile is compared with the parameters of the traffic contract. If the contract is violated, cells may be passed, discarded or tagged. In this way, the UPC function protects network resources by active data stream regulation. The Traffic Shaping mechanism can be used to modify the traffic profile according to the traffic contract by delaying the cell transmission or by discarding cells.

4.2 Architecture

As the specified class concept cannot be provided by the presented QoS modells for IP networks, a new concept shall be developed.

4.2.1 Scalability

To build the architecture, mechanisms with scalable behaviour should be used. This means, that the overhead in network components does not depend on the number of flows in the network or connected end systems but the number of network components or lines. Performance lacks because of such a behaviour were demonstrated in paragraph 3.1. There the CPU load grew critically with an increasing number of flows. Additionally, resource reservation algorithms like RSVP can not do reservation per flow on links with high bandwidths (see paragraphe 3.3).

One solution is the concept of MPLS trunks. MPLS is a technology for using label switching and for the implementation of label-switched paths over various link-level technologies. This includes procedures and protocols for the distribution of labels between routers [17]. MPLS trunks aggregate all flows which belong to a traffic class and the mechainsms are applied to the flow aggregate [2]. They are configured according to the network topologie between the endpoints of the network. Then all communication peers can exchange packets of a special service class over these trunks.

With this solution, no guarantees for individual flows can be given. Because of that we propose to use the following idea, which can be realised with the network components available nowadays: At the network boundary, classification, traffic control, policing and shaping is done per flow to check, that the specified service level agreements are not violated. This should be possible because near the

sources, the number of flows is not so large as in the core of a network. Within the network, flows are not treated as individuals but they are aggregated to service classes (MPLS trunks) on which the network mechanisms are applied. This aggregation reduces the complexity in the core network components and because of this a scalable architecture can be proposed.

4.3 Realization of the class concept

One requirement of the QoS architecture is that all service classes of the developed concept should be provided bidirectionally between end systems. But we have seen above that the IP protocol is not able to fulfil hard realtime behaviour [22]. As a result, our proposal is to transport such applications with QoS requirements over an ATM infrastructure. ATM is able to fulfill hard timing requirements [38, 23]. For all other classes C_2 until C_8 an IP network is managed.

The management of two concurrent networks causes more planning and administration work and on the same time higher costs, too. But customers who need very good transmission quality pay for such a service [15].

In the following, the building blocks of an IP architecture are presented that realises the class concept. The proposal is based on the network mechanisms presented.

Classification

To mark the 7 classes of the class concept (the eigth'th class is realised by an ATM network) we take the first three bits of the Type of Service field, the precedence bits (see figure 2). The value zero mark packets of class C_0. According to this, the values one to six are used for the other classes C_2 up to C_7.

As the classification must be done for each flow, it should be activated in network components as near at the source as possible, as near the source, the number of flows is substantially lower than in the network. Hence, a scalable architecture can be implemented and the performance problems can be avoided.

To provide these classes in the whole internet, the interpretation of the precedence bits for a special class must be the same in every administrative domain. Provider of different networks need not to remark packets if they rely on the marks in the IP packets which they receive from their neighbour. This can be agreed upon in service level agreements. However, this solution is only possible if accouting mechanism are able to count the traffic of each service class so that the charging for each individual service class is fair.

Resource management, traffic control and regulation

Unfortunately, no resource management algorithm for individual flows is implemented on routers nowadays. So, other mechanisms must be used to control the traffic which is sent to a network. Because of this, traffic received by a boundary router of a domain must be controlled, so that only such an amount of traffic enters the network that has been considered when planning the capacities of the links.

As in the case of the classification, traffic control should be done per flow and because of scalability concerns as close to the source as possible. The bounds for the transfer rates are configured as static values at network boundaries. They must be changed if service level agreements change. If packets violate the traffic contract they should be dropped. This mechanism should substitute the missing resourcen reservation algorithm.

At the boundary of the network traffic shaping per flow can be activated, too. By this, the burstiness of the traffic is reduced so that queues do not overflow so often. When configuring traffic shaping and policing on the network boundary, the value with which the conformity of the traffic profile is compared should correspond to the maximal amount of traffic which can be send by the customer in this service class. Also the burst size of the buckets should not be selected not too low for classes with bursty sending characteristics (class C_2, C_4 to C_8).

Congestion avoidance

According to the results in section 3.2.2, RED should be activated in the whole network for the traffic classes C_6, C_7 and C_8. As has been shown the throughput of the network can be increased by RED. The threshholds of the class C_8 should be configured such that packets are discarded at a very low average queue length, whereas the discarding of the packets of class C_6 and C_7 does not start until the queue is nearly filled.

Scheduling

As scheduling algorithm WFQ is proposed. This mechanism is able to assign each service class with its individual requirements to a seperate queue. By this, each class can send different transfer rates and the bandwidth of one service class is divided equally between the different flows of one class. In addition to that, the scheduling algorithm guarantees an upper bound on the delay. WFQ is implemented on routers of different vendors nowadays.

The weigths $\omega_{\mathcal{K}_i}$ for each service class to define the outgoing bandwidth can be configured with the transfer rates T_i of each class according to equation 4.3 (see paragraph 3.4):

$$\omega_{\mathcal{C}_\alpha} = \frac{1}{\sum\limits_{i=2}^{8} T_{\mathcal{C}_i}} * T_{\mathcal{C}_\alpha}, \quad \sum_{i=2}^{8} \omega_{\mathcal{C}_i} = 1.$$

The outgoing capacity R of each interface should not be smaller than the following value:

$$R > \sum_{i=2}^{8} T_{\mathcal{C}_i}.$$

Routing

To provide classes which require a high service quality in the network, the links in the network should be as high that even in situations of a network failure traffic can be routed.

4.3.1 Management

To increase the functionality of the network, configuration and management tasks should be done automatically. For this, policy systems can be used which are even able to support dynamic service level agreements.

In addition to that traffic flows of each class should be monitored to get fair charges for every customer. Finally, measurements must be performed in the network so that customers can get statistics about the quality of the different service classes for which they pay [41, 24, 8].

4.4 Capacity of links

The capacity of the links in the network should correspond to the bandwidth requirements of the service classes even in situations of a network failure. Then the traffic of classes which require low loss rates can be rerouted in case of failures.

To get information for the planning process of the capacities it is useful to collect accounting information. The bandwidth requirements of the classes are known because of the specified SLAs. But the capapcity of the links in the core of the networks can be determined only with complex algorithms [19].

4.5 Service level agreements at network boundaries

It must be ensured that the specified services are guaranteed in the whole network from end to end, particularly at network boundaries. Service providers must agree concerning the interpretation of the precedence bits in the ToS byte. In addition to that mechanisms which are implemented on network components of different providers must be interoperable.
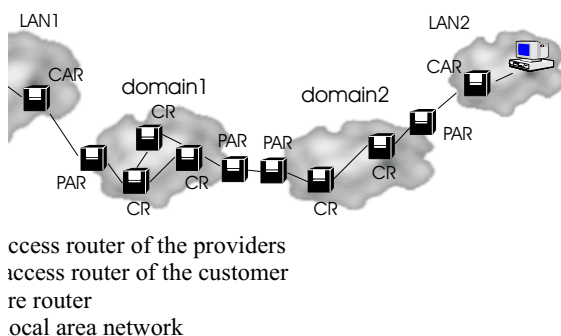
Figure 4: Internet router architecture.

In figure 4 an abstract internet architecture to realise a QoS architecture is shown [51]. The architecture consists of several components: At the border there are customer networks, `LAN1` und `LAN2`, and the core consists of two different administrative domains of different providers, `domain1` und `domain2`. Customer access routers (`CAR`) and provider access routers (`PAR`) connect the customer and provider networks. In the core packets are forwarded by core routers (`CR`) according to the required service. At the interface which are connected to the network of the provider `PAR` and `CAR` must provide the same functionality as the core routers.

On the `access routers` at boundaries between a customer and a provider network all mechanisms which can be seen in figure 5 should be activated. All incoming packets should be classified, if it is not done in the neighbour domain. For every flow the conformance must be checked according to the specified contracts. With this, resources in the network are protected. Non conformant packets are dropped. On every outgoing interface traffic shaping, policing, WFQ and RED should be activated as described above. In the `Core Routers` of the network,
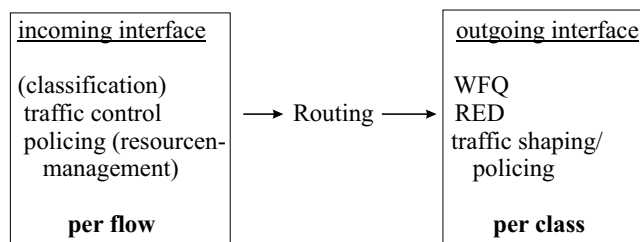

Figure 5: Mechanisms which should be activated in network components at network boundaries.

an aggregate of flows is subject to classification, traffic control, and traffic shaping. On the outgoing interfaces, WFQ and RED should be activated.

5 Summary

Users must be able to define the QoS requirements of the applications they use. In order to provide this capability, we introduced a class concept that allows to assign every application to this service class which fits the best to the requirements of that regarded application. As networks must be able to provide these required services existing mechanisms to provide different service classes in networks were presented in the second section.

Although the presented mechanisms suffer from performance lacks, a network architecture based on these mechanisms was developed that fulfills all requirements. This could be achieved by thorough planning of network traffic and rigourous traffic shaping on the borders of networks. On the one hand, the architecture takes into consideration the results which are obtained by the investigation of the

network components. On the other hand, it regards the end-to-end QoS requirements of different applications.

References

[1]   ATM Forum, "Traffic Management Specification 4.0", The ATM Forum, Technical Committee, (1996).

[2]   D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, J. McManus, "Requirements for Traffic Engineering Over MPLS". *Request for Comments 2702*, (1999).

[3]   S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", *Request for Comments 2475*, (1998).

[4]   U. Bodin, O. Schelén, S. Pink, "Load–tolerant Differentiation with Active Queue Management", *Proceedings ACM SIGCOMM*, (2000).

[5]   B. Braden, D. Clark, S. Shenker, "Integrated Services in the Internet Architecture: an Overview", *Request for Comments 1633*, (1994).

[6]   B. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, "Resource ReSerVation Protocol (RSVP) — Version 1 Functional Specification", *Request for Comments 2205*, (1997).

[7]   "Homepage of Caida", *http://www.caida.com*, (2002).

[8]   G. Carle, S. Zander, T. Zseby, "Policy–basiertes Metering f"ur IP–Netze", *Kommunikation in Verteilten Systemen (KIVS) — 12. GI/ITG–Fachtagung*, U. Killat, W. Lamersdorf (Hrsg.), (2001).

[9]   J. Case, M. Fedor, M. Schoffstall, J. Davin, "A Simple Network Management Protocol (SNMP)", *Request for Comments 1157*, (1990).

[10] "Homepage von Cisco Systems", *http://www.cisco.com*, (2002).

[11] Defense Advanced Research Projects Agency, "Internet Protocol", *Request for Comments 791*, (1981).

[12] Defense Advanced Research Projects Agency, "Transmission Control Protocol, DARPA Internet Program Protocol Specification", *Request for Comments 793*, (1981).

[13] C. Demichelis, P. Chimento, "IP Packet Delay Variation Metric for IPPM", *Request for Comments 3393*, (2002).

[14] A. Demers, S. Keshav, S. Shenker, "Analysis and Simulation of a Fair Queueing Algorithm", *Internetworking, Research and Experience*, (1990).

[15] "Homepage of DFN", *http://www.dfn.de*, (2003).

[16] F. Dressler, U. Hilgers, S. Naegele–Jackson, K. Liebl, "Untersuchung von Dienstqualitäten bei echtzeitorientierten multimedialen Datenübertragungen", *Pearl 99*, Peter Holleczek (Hrsg.), Springer, (1999).

[17] B. Davie, Y. Rekhter, "MPLS", Academic Press, (2000).

[18] Falko Dressler, Ursula Hilgers, Peter Holleczek, "Voice over IP in Weitverkehrsnetzen?", *Anwendungs– und System–Management im Zeichen von Multimedia und E–Business*, Fachtagung der GI-Fachgruppe 3.4 Betrieb von Informations- und Kommunikationssystemen (BIK 2001), Tübingen, Germany, (2001). *Request for Comments 3270*, (2002).

[19] V. G. Fischer, "Evolutionary Design of Corporate Net works under Uncertainty", Dissertation, Fakultät für Informatik, Technische Universität München, (2000).

[20] S. Floyd, V. Jacobson, "Random Early Detection Gateways for Congestion Avoidance", *IEEE/ACM Transactions on Networking*, (1993).

[21] W. J. Goralski, "Introduction to ATM Networking", McGraw–Hill, (1995).

[22] U. Hilgers, F. Dressler, "Echtzeit–Datenverkehr "uber IP–basierte Datennetze", *Pearl 01*, Peter Holleczek (Hrsg.), Springer, (2001).

[23] U. Hilgers, R. Hofmann, "QoS — ATM versus Differentiated Services", *Proceedings EUNIS 2001*, J. Knop, P. Schirmbacher (Hrsg.), (2001).

[24] U. Hilgers, "QoS von IP–Verbindungen unter Realzeitbedingungen", *Pearl 98*, Peter Holleczek (Hrsg.), Springer, (1998).

[25] U. Hilgers, R. Hofmann, P. Holleczek, "Differentiated Services — Konzepte und erste Erfahrungen", *Praxis der Informationsverarbeitung und Kommunikation*, (**2**), (2000).

[26] Ursula Hilgers, "Dienstgüteunterstuetzung in Weitverkehrsnetzen", *Arbeitsberichte des Institus für Informatik*, Friedrich–Alexander–Universität Erlangen-Nürnberg, (**35**)6, (2002).

[27] International Organization for Standardization, "Digital Compression and Coding of Continuoustone Still Images, Part 1, Requirements and Guidelines", *ISO/IEC JTC1 Draft International Standard 10918–1*, (1991).

[28] ITU–T Recommendation E.430, "Quality of Service Framework", (1992).

[29] ITU–T Recommendation G.711, "Pulse Code Modulation (PCM) of Voice Frequencies", (1993).

[30] ITU–T Recommendation I.350, "General Aspects of QoS and Network Performance in Digital Networks, including ISDNs", (1993).

[31] ITU–T Recommendation E.800, "Terms and Definitions related to QoS and Network Performance including Dependability", (1994).

[32] ITU–T Recommendation P.800, "Methods for Subjective Determination of Transmission Quality", (1996).

[33] ITU–T Recommendation H.323, "Visual Telephone Systems and Equipment for Local Area Networks which provide a non–Guaranteed Quality of Service", (1998).

[34] V. Jacobson, "Congestion Avoidance and Control", *Computer Communication Review*, (1988).

[35] A. Mankin, F. Baker, B. Braden, S. Bradner, M. O'Dell, A. Romanow, A. Weinrib, L. Zhang, " Resource ReSerVation Protocol (RSVP) — Version 1 Applicability Statement — Some Guidelines on Deployment", *Request for Comments 2208*, (1997).

[36] C. W. Mercer, "An Introduction to Real–Time Operating Systems: Scheduling Theory", *http://www.cs.cmu.edu/afs/cs.cmu.edu/user/cwm/www/publications.html*, (1992).

[37] J. L. Mitchell, "MPEG video compression standard", Kluwer, Boston, (2001).

[38] S. Naegele-Jackson, U. Hilgers, P. Holleczek, "Evaluation of Codec Behavior in IP and ATM Networks", *Special Issue of Informatica*, (**25**)2, (2001).

[39] K. Nichols, V. Jacobson, L. Zhang, "A Two–bit Differentiated Services Architecture for the Internet", *Request for Comments 2638*, (1999).

[40] A. K. J. Parekh, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks", *LIDS–TH–2089*, MIT Laboratory for Information and Decision Systems, Camebridge, Mass., (1992).

[41] V. Paxson, G. Almes, J. Mahdavi, M. Mathis, "Framework for IP Performance Metrics", *Request for Comments 2330*, (1998).

[42] A. K. Parekh, R. G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single–Node Case", *IEEE/ACM Transactions on Networking*, **(1)**3, (1993).

[43] J. Postel, "User Datagram Protocol", *Request for Comments 768*, (1980).

[44] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, "RTP: A Transport Protocol for Real–Time Applications", *Request for Comments 1889*, (1999).

[45] W. Stallings, "High–Speed Networks". Prentice Hall, (1998).

[46] W. Stevens, "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", *Request for Comments 2001*, (1997).

[47] R. Steinmetz, "Multimedia–Technologie", *Springer–Verlag Berlin Heidelberg*, (2000).

[48] S. Shenker, J. Wroclawski, "Network Element Service Specification Template", *Request for Comments 2216*, (1997).

[49] J. Wroclawski, "Specification of the Controlled/Load Network Element Service", *Request for Comments 2211*, (1997).

[50] J. Wroclawski, "Specification of Guaranteed Quality of Service", *Request for Comments 2212*, (1997).

[51] X. Xiao, A. Hannan, B. Bailey, "Traffic Engineering with MPLS in the Internet", *IEEE Network*, (2000).

JOŽEF STEFAN INSTITUTE

Jožef Stefan (1835-1893) was one of the most prominent physicists of the 19th century. Born to Slovene parents, he obtained his Ph.D. at Vienna University, where he was later Director of the Physics Institute, Vice-President of the Vienna Academy of Sciences and a member of several scientific institutions in Europe. Stefan explored many areas in hydrodynamics, optics, acoustics, electricity, magnetism and the kinetic theory of gases. Among other things, he originated the law that the total radiation from a black body is proportional to the 4th power of its absolute temperature, known as the Stefan–Boltzmann law.

The Jožef Stefan Institute (JSI) is the leading independent scientific research institution in Slovenia, covering a broad spectrum of fundamental and applied research in the fields of physics, chemistry and biochemistry, electronics and information science, nuclear science technology, energy research and environmental science.

The Jožef Stefan Institute (JSI) is a research organisation for pure and applied research in the natural sciences and technology. Both are closely interconnected in research departments composed of different task teams. Emphasis in basic research is given to the development and education of young scientists, while applied research and development serve for the transfer of advanced knowledge, contributing to the development of the national economy and society in general.

At present the Institute, with a total of about 700 staff, has 500 researchers, about 250 of whom are postgraduates, over 200 of whom have doctorates (Ph.D.), and around 150 of whom have permanent professorships or temporary teaching assignments at the Universities.

In view of its activities and status, the JSI plays the role of a national institute, complementing the role of the universities and bridging the gap between basic science and applications.

Research at the JSI includes the following major fields: physics; chemistry; electronics, informatics and computer sciences; biochemistry; ecology; reactor technology; applied mathematics. Most of the activities are more or less closely connected to information sciences, in particular computer sciences, artificial intelligence, language and speech technologies, computer-aided design, computer architectures, biocybernetics and robotics, computer automation and control, professional electronics, digital communications and networks, and applied mathematics.

The Institute is located in Ljubljana, the capital of the independent state of **Slove**nia (or S♡nia). The capital today is considered a crossroad between East, West and Mediter-

ranean Europe, offering excellent productive capabilities and solid business opportunities, with strong international connections. Ljubljana is connected to important centers such as Prague, Budapest, Vienna, Zagreb, Milan, Rome, Monaco, Nice, Bern and Munich, all within a radius of 600 km.

In the last year on the site of the Jožef Stefan Institute, the Technology park "Ljubljana" has been proposed as part of the national strategy for technological development to foster synergies between research and industry, to promote joint ventures between university bodies, research institutes and innovative industry, to act as an incubator for high-tech initiatives and to accelerate the development cycle of innovative products.

At the present time, part of the Institute is being reorganized into several high-tech units supported by and connected within the Technology park at the Jožef Stefan Institute, established as the beginning of a regional Technology park "Ljubljana". The project is being developed at a particularly historical moment, characterized by the process of state reorganisation, privatisation and private initiative. The national Technology Park will take the form of a shareholding company and will host an independent venture-capital institution.

The promoters and operational entities of the project are the Republic of Slovenia, Ministry of Science and Technology and the Jožef Stefan Institute. The framework of the operation also includes the University of Ljubljana, the National Institute of Chemistry, the Institute for Electronics and Vacuum Technology and the Institute for Materials and Construction Research among others. In addition, the project is supported by the Ministry of Economic Relations and Development, the National Chamber of Economy and the City of Ljubljana.

Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel.:+386 1 4773 900, Fax.:+386 1 219 385
Tlx.:31 296 JOSTIN SI
WWW: http://www.ijs.si
E-mail: matjaz.gams@ijs.si
Contact person for the Park: Iztok Lesjak, M.Sc.
Public relations: Natalija Polenec

INFORMATICA

AN INTERNATIONAL JOURNAL OF COMPUTING AND INFORMATICS

INVITATION, COOPERATION

Submissions and Refereeing

Please submit three copies of the manuscript with good copies of the figures and photographs to one of the editors from the Editorial Board or to the Contact Person. At least two referees outside the author's country will examine it, and they are invited to make as many remarks as possible directly on the manuscript, from typing errors to global philosophical disagreements. The chosen editor will send the author copies with remarks. If the paper is accepted, the editor will also send copies to the Contact Person. The Executive Board will inform the author that the paper has been accepted, in which case it will be published within one year of receipt of e-mails with the text in Informatica LaTeX format and figures in `.eps` format. The original figures can also be sent on separate sheets. Style and examples of papers can be obtained by e-mail from the Contact Person or from FTP or WWW (see the last page of Informatica).

Opinions, news, calls for conferences, calls for papers, etc. should be sent directly to the Contact Person.

QUESTIONNAIRE

☐ Send Informatica free of charge

☐ Yes, we subscribe

Please, complete the order form and send it to Dr. Drago Torkar, Informatica, Institut Jožef Stefan, Jamova 39, 1111 Ljubljana, Slovenia.

Since 1977, Informatica has been a major Slovenian scientific journal of computing and informatics, including telecommunications, automation and other related areas. In its 16th year (more than ten years ago) it became truly international, although it still remains connected to Central Europe. The basic aim of Informatica is to impose intellectual values (science, engineering) in a distributed organisation.

Informatica is a journal primarily covering the European computer science and informatics community - scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the Refereeing Board.

Informatica is free of charge for major scientific, educational and governmental institutions. Others should subscribe (see the last page of Informatica).

ORDER FORM – INFORMATICA

Name: ....

Title and Profession (optional): ....

....

Home Address and Telephone (optional): ....

....

Office Address and Telephone (optional): ....

....

E-mail Address (optional): ....

Signature and Date: ....

Informatica WWW:

http://ai.ijs.si/informatica/
http://orca.st.usm.edu/informatica/

Referees:

Witold Abramowicz, David Abramson, Adel Adi, Kenneth Aizawa, Suad Alagić, Mohamad Alam, Dia Ali, Alan Aliu, Richard Amoroso, John Anderson, Hans-Jurgen Appelrath, Iván Araujo, Vladimir Bajič, Michel Barbeau, Grzegorz Bartoszewicz, Catriel Beeri, Daniel Beech, Fevzi Belli, Simon Beloglavec, Sondes Bennasri, Francesco Bergadano, Istvan Berkeley, Azer Bestavros, Andraž Bežek, Balaji Bharadwaj, Ralph Bisland, Jacek Blazewicz, Laszlo Boeszoermenyi, Damjan Bojadžijev, Jeff Bone, Ivan Bratko, Pavel Brazdil, Bostjan Brumen, Jerzy Brzezinski, Marian Bubak, Davide Bugali, Troy Bull, Leslie Burkholder, Frada Burstein, Wojciech Buszkowski, Rajkumar Bvyya, Netiva Caftori, Particia Carando, Robert Cattral, Jason Ceddia, Ryszard Choras, Wojciech Cellary, Wojciech Chybowski, Andrzej Ciepielewski, Vic Ciesielski, Mel Ó Cinnéide, David Cliff, Maria Cobb, Jean-Pierre Corriveau, Travis Craig, Noel Craske, Matthew Crocker, Tadeusz Czachorski, Milan Češka, Honghua Dai, Bart de Decker, Deborah Dent, Andrej Dobnikar, Sait Dogru, Peter Dolog, Georg Dorfner, Ludoslaw Drelichowski, Matija Drobnič, Maciej Drozdowski, Marek Druzdzel, Marjan Družovec, Jozo Dujmović, Pavol Ďuriš, Amnon Eden, Johann Eder, Hesham El-Rewini, Darrell Ferguson, Warren Fergusson, David Flater, Pierre Flener, Wojciech Fliegner, Vladimir A. Fomichov, Terrence Forgarty, Hans Fraaije, Hugo de Garis, Eugeniusz Gatnar, Grant Gayed, James Geller, Michael Georgiopolus, Michael Gertz, Jan Goliński, Janusz Gorski, Georg Gottlob, David Green, Herbert Groiss, Jozsef Gyorkos, Marten Haglind, Abdelwahab Hamou-Lhadj, Inman Harvey, Jaak Henno, Marjan Hericko, Elke Hochmueller, Jack Hodges, Doug Howe, Rod Howell, Tomáš Hruška, Don Huch, Simone Fischer-Huebner, Alexey Ippa, Hannu Jaakkola, Sushil Jajodia, Ryszard Jakubowski, Piotr Jedrzejowicz, A. Milton Jenkins, Eric Johnson, Polina Jordanova, Djani Juričič, Marko Juvancic, Sabhash Kak, Li-Shan Kang, Ivan Kapustøk, Orlando Karam, Roland Kaschek, Jacek Kierzenka, Jan Kniat, Stavros Kokkotos, Fabio Kon, Kevin Korb, Gilad Koren, Andrej Krajnc, Henryk Krawczyk, Ben Kroese, Zbyszko Krolikowski, Benjamin Kuipers, Matjaž Kukar, Aarre Laakso, Les Labuschagne, Ivan Lah, Phil Laplante, Bud Lawson, Herbert Leitold, Ulrike Leopold-Wildburger, Timothy C. Lethbridge, Joseph Y-T. Leung, Barry Levine, Xuefeng Li, Alexander Linkevich, Raymond Lister, Doug Locke, Peter Lockeman, Matija Lokar, Jason Lowder, Kim Teng Lua, Ann Macintosh, Bernardo Magnini, Andrzej Małachowski, Peter Marcer, Andrzej Marciniak, Witold Marciszewski, Vladimir Marik, Jacek Martinek, Tomasz Maruszewski, Florian Matthes, Daniel Memmi, Timothy Menzies, Dieter Merkl, Zbigniew Michalewicz, Gautam Mitra, Roland Mittermeir, Madhav Moganti, Reinhard Moller, Tadeusz Morzy, Daniel Mossé, John Mueller, Jari Multisilta, Hari Narayanan, Jerzy Nawrocki, Rance Necaise, Elzbieta Niedzielska, Marian Niedq'zwiedziński, Jaroslav Nieplocha, Oscar Nierstrasz, Roumen Nikolov, Mark Nissen, Jerzy Nogieć, Stefano Nolfi, Franc Novak, Antoni Nowakowski, Adam Nowicki, Tadeusz Nowicki, Daniel Olejar, Hubert Österle, Wojciech Olejniczak, Jerzy Olszewski, Cherry Owen, Mieczyslaw Owoc, Tadeusz Pankowski, Jens Penberg, William C. Perkins, Warren Persons, Mitja Peruš, Stephen Pike, Niki Pissinou, Aleksander Pivk, Ullin Place, Gabika Polčicová, Gustav Pomberger, James Pomykalski, Dimithu Prasanna, Gary Preckshot, Dejan Rakovič, Cveta Razdevšek Pučko, Ke Qiu, Michael Quinn, Gerald Quirchmayer, Vojislav D. Radonjic, Luc de Raedt, Ewaryst Rafajlowicz, Sita Ramakrishnan, Kai Rannenberg, Wolf Rauch, Peter Rechenberg, Felix Redmill, James Edward Ries, David Robertson, Marko Robnik, Colette Rolland, Wilhelm Rossak, Ingrid Russel, A.S.M. Sajeev, Kimmo Salmenjoki, Pierangela Samarati, Bo Sanden, P. G. Sarang, Vivek Sarin, Iztok Savnik, Ichiro Satoh, Walter Schempp, Wolfgang Schreiner, Guenter Schmidt, Heinz Schmidt, Dennis Sewer, Zhongzhi Shi, Mária Smolárová, Carine Souveyet, William Spears, Hartmut Stadtler, Olivero Stock, Janusz Stokłosa, Przemysław Stpiczyński, Andrej Stritar, Maciej Stroinski, Leon Strous, Tomasz Szmuc, Zdzislaw Szyjewski, Jure Šilc, Metod Škarja, Jiřı Šlechta, Chew Lim Tan, Zahir Tari, Jurij Tasič, Gheorge Tecuci, Piotr Teczynski, Stephanie Teufel, Ken Tindell, A Min Tjoa, Vladimir Tosic, Wieslaw Traczyk, Roman Trobec, Marek Tudruj, Andrej Ule, Amjad Umar, Andrzej Urbanski, Marko Uršič, Tadeusz Usowicz, Romana Vajde Horvat, Elisabeth Valentine, Kanonkluk Vanapipat, Alexander P. Vazhenin, Jan Verschuren, Zygmunt Vetulani, Olivier de Vel, Valentino Vranić, Jozef Vyskoc, Eugene Wallingford, Matthew Warren, John Weckert, Michael Weiss, Tatjana Welzer, Lee White, Gerhard Widmer, Stefan Wrobel, Stanislaw Wrycza, Janusz Zalewski, Damir Zazula, Yanchun Zhang, Ales Zivkovic, Zonling Zhou, Robert Zorc, Anton P. Železnikar

Informatica

An International Journal of Computing and Informatics

Informatica

An International Journal of Computing and Informatics