Evolving a Digital Library Environment to the Changing Needs of its Users

Leslie Carr[1]
lac@ecs.soton.ac.uk

Timothy Miles-Board[1]
tmb@ecs.soton.ac.uk

Gary Wills[1]
gbw@ecs.soton.ac.uk

Guillermo Power[2]
gp2@ecs.soton.ac.uk

Christopher Bailey[2]
cpb@ecs.soton.ac.uk

Wendy Hall[1]
wh@ecs.soton.ac.uk

[1]Intelligence, Agents, Multimedia Group
[2]Learning Technologies Group
University of Southampton
Southampton, UK

Simon Grange
NHS Wessex Deanery
Orthopaedic Surgical Training,
UK
simon.grange@btopenworld.com

ABSTRACT

A digital archive, together with its users and its contents, does not exist in isolation - there is a cycle of activities which provides the context for the archive's existence. In arguing for the broadening of the traditional view of digital libraries as merely collections towards the processes of collecting and deploying, we have developed an "extended" digital library environment for orthopaedic surgeons which bridges the gap between the undertaking of experimental work and the dissemination of its results through electronic publication. However, in embracing such an approach, we must also consider that the archive should be able to evolve in accordance with the changing needs of its users — we cannot predict in advance the myriad different types of experiment that future users will want to carry out. This paper therefore discusses our recent efforts in addressing this issue through the implementation of a user-oriented Template Generation Toolkit.

Categories and Subject Descriptors

H.3.7 [**INFORMATION STORAGE AND RETRIEVAL**]: Digital Libraries; H.5.4 [**INFORMATION INTERFACES AND PRESENTATION**]: Hypertext/Hypermedia; I.7.4 [**DOCUMENT AND TEXT PROCESSING**]: Electronic Publishing; J.3 [**Computer Applications**]: LIFE AND MEDICAL SCIENCES—*Medical information systems*

General Terms

Human Factors

Keywords

EPrints, E-Science, Grid, Orthopaedics, Virtual Universities

1. INTRODUCTION

A digital library, together with its users and its contents, does not exist in isolated splendour. There is a cycle of activities, which provides the context for the library's existence, which the library supports through its various roles of information access, discovery, storage, dissemination and preservation. In the e-learning context the digital library has an important role in the undertaking of science, and with the recent developments of the Grid for computer-supported scientific collaboration [4] and Virtual Universities for computer-supported education [2], its role has increased.

Marchionini and Maurer assert that "digital libraries will allow learners of all types to share resources, time and energy and experience to their mutual benefit" [7]. In their proposed future of digital libraries, sharing resources becomes an important factor in supporting teaching; this includes the ability to share raw scientific data and other datasets. Many e-science projects have collected a vast amount of data: if the next generation of scientists are to go beyond the present position it is essential that they have access to the raw data in their research and training. Marchionini and Maurer also suggest that digital libraries should offer greater opportunities for users to deposit information. These early visions are slowly being realised, for example McGrath *et al.* have developed a system that will locate, browse and retrieve astronomy data across several databases [9], but there is still a need for those that have the technology skills, librarians, and users, to work together to provide appropriate tools for handling, manipulating and analysing these large datasets [10].

There are projects beginning to do this, for example the Digital Library for Earth Science Education (DLESE) project allows students to explore geospatial materials and Earth data sets; groups of students can then manually create reports using this data, and discuss them [8]. Weatherley et al. have proposed a model that will aid reviewers in reviewing complex material or a digital collection [11]. The peer review of collections and peer comment is a significant part of the dissemination process, which adds value to any collection. Lyon sees the digital library in the context of an information grid as consisting of a collection of resources for learning and teaching, data repositories for research purposes, or as archives of diverse cultural heritage materials [6]. While this is only a proposed scenario, Lyon recognises the need for researchers to undertake experiments, deposit raw data, and produce pre-prints using Web services.

We have developed a Web-based environment to support such activities in the context of the Virtual Orthopaedic European University digital library archive: the Dynamic Review Journal (DRJ) [1, 3, 5]. Within the DRJ environment, orthopaedic trainees and surgeons not only collaboratively develop and disseminate the documents which are subsequently managed by the archive, but are also supported in the cycle of activities leading up to the production of these documents - the management of surgical trials, collation and analysis of experimental results, and organisation of internal project discussions.

However, in order for this approach to succeed, the digital library environment needs to be able to evolve to meet the changing needs of its user community — it is unreasonable to expect the developers of such a system to predict and encode every possible experimental procedure, every type of data that can be collated and analysed, and every possible dissemination route that users will follow to publish their results to the wider community. A fundamental aspect of the design of the DRJ is therefore the *Schema Space* — the mechanism by which the DRJ can be configured to the requirements of a community (in the case of the VOEU, the community of orthopaedic surgeons and trainees within the EU), a particular group of users, or even an individual. Until recently, this mechanism was only available to expert users with the appropriate technical nous — our most recent work has therefore focused on building a toolkit to allow individual non-expert users to configure the DRJ to the particular requirements of their experimental activities. After examining the DRJ and its Schema Space mechanism in more detail, this paper introduces the *Template Generation Toolkit*, shares our experiences in developing and deploying this new functionality, and discusses the new issues arising from this approach.

2. THE DYNAMIC REVIEW JOURNAL

The Dynamic Review Journal was originally conceived as part of the Virtual Orthopaedic European University (VOEU) project, which aimed to build a Virtual University dedicated to the training of Higher Surgical trainees and the ongoing professional education of orthopaedic surgeons across the European Community. The VOEU-managed learning environment for training surgeons consists of hypermedia educational material (including problem cases and assessment), interactive simulators, and communication tools (moderated and asynchronous message boards) together with the Dynamic Review Journal. The objective of the DRJ was to provide integrated computer support across the research and educational cycles because these activities are intrinsically coupled as a part of the requirements of a surgeon's Continuing Professional Development — research must be undertaken and papers published to achieve goals under the learning contracts with their Professional Colleges.

Within this context, the Dynamic Review Journal has two main functions: to aid surgeons in preparing findings for publication and to support the educational process. Orthopaedic surgical trials typically run for extended periods (up to 2 years), with postoperative assessment results being collected regularly. The collated results are then analysed and discussed by a team of surgeons before being disseminated to the wider orthopaedic community. The DRJ also allows surgical trainees to gain experience in the research process by immersing themselves directly in this research environment — trainees can, under supervision from a tutor or team leader, analyse data from existing trials, investigate hypotheses, discuss archived articles, and even prepare and submit their own reports for assessment. In addition, tutors will be able capitalise on these reviews to include the corresponding declarative and procedural knowledge in the educational modules.

Figure 1 illustrates the major activity spaces in the DRJ and the work-flow supported within these spaces.

2.1 Schema Space

In more general terms, beyond the specific application of the VOEU, the Schema Space is the mechanism by which the Dynamic Review Journal can be configured to the evolving needs of a particular scientific/learning community, through the formal specification of experimentation procedures relevant to that community. This configuration is currently achieved using three different types of schema:

- Data schemas describe the exact nature of the experimental data (for example, specification of variable names, types, and possible values). In the VOEU context, there are a number of data schemas for collecting orthopaedic clinical trial data, including shoulder and hip operation data and post-operative mobility test scores.

- Experiment schemas describe experimental procedures or protocols. For example, in VOEU a protocol could specify that a surgeon conducting a trial of type X needs to record an experiment description, statement of purpose and an outcome hypothesis. Human-readable guidelines are also included, to help users meet the requirements of the protocol and to help reviewers to ensure that the requirements have been met.

- Publication schemas describe the required format for submitting experimental results to relevant journals/ conferences (for example, Abstract, Introduction, Background, Experimental Methods, Results, and Conclusions). As with experiment schemas, human-readable guidelines are also included in publication schemas.

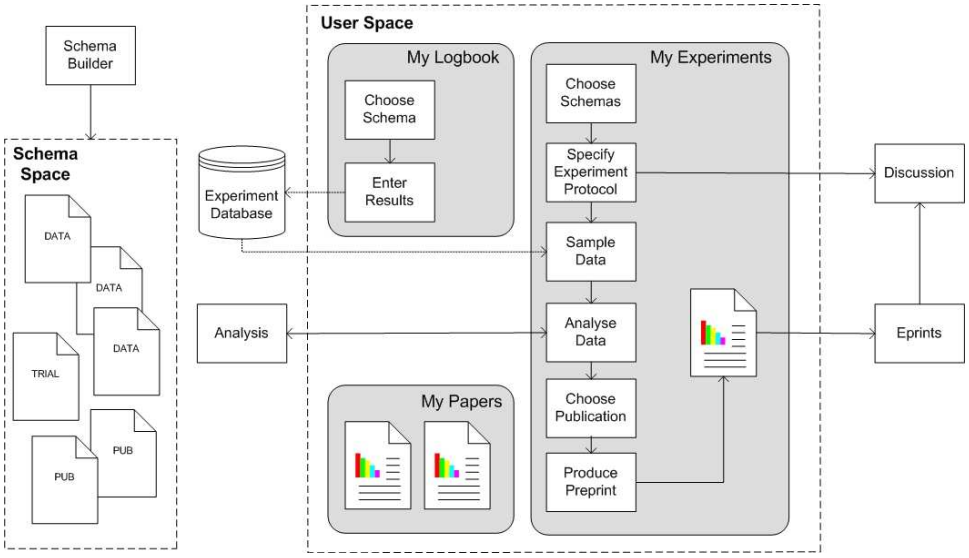

Figure 1: Activity spaces and work-flow.

In the VOEU context, there are currently two publication schemas describing the submission formats for the Journal of Bone & Joint Surgery (JBJS) and the British Medical Journal (BMJ).

All three types of schema will eventually be created using the *Template Generation Toolkit*, introduced in Section 3.

2.2 User Space

The *User Space* is where users use the schema space to orchestrate practical data entry and collation, experimentation, and dissemination. The User Space is further subdivided into three personalised areas: My Logbook, My Experiments, and My Papers (Figure 2).

My Logbook is an experiment logbook, in which experimental results can be entered (in accordance with a selected data schema). Logbook entries are subsequently added to the community database, making data available[1] to other community members.

My Experiments is a workspace for the experiments which the user works on. A user may be involved an experiment in the capacity of lead investigator (initiates experiment and acts as coordinator and contact for duration of experiment), associate investigator (assistant), or reviewer (monitors the progress of the experiment and reviews its outcomes according to agreed guidelines). When a new experiment is initiated, a discussion facility is automatically set up to facilitate and record communication between the users involved (this is also the means by which reviewers can give feed back the practitioners). Figure 1 outlines the work process facilitated by My Experiments:

1. **Define/Choose experiment and data schemas** — The lead investigator chooses from the Schema Space the experiment and data schemas which best describe the procedure to be carried out and the data to be collated. If no suitable schemas are available, the lead investigator can create a new schema specifically tailored to the requirements of the experiment. In the case of trainees, a tutor or group leader takes the role of lead investigator.
2. **Specify experiment protocol** — The lead investigator enters the specifics of the experiment protocol (in accordance with the chosen schema), including assigning associate investigators and reviewers to the study. In the case of trainees, the tutor assigns the roles of associate investigators and reviewers to students.
3. **Sample data** — Investigators create a dataset for the experiment, either by importing their own records from the My Logbook area, or by searching the community database.
4. **Analyse data** — Investigators perform a series of analyses on the dataset, using a distributed analysis service, to test the experiment hypothesis.
5. **Define/Choose publication schema** — To initiate the publication cycle, an investigator first chooses the publication schema corresponding to the target conference/journal. Again, if no existing schema meets the requirements of the desired dissemination route, a new schema can be created. In both cases, this produces a 'skeleton' paper. In the case of trainees, the chosen publication schema will usually correspond to a 'assignment report' format.
6. **Produce outline paper** — The investigators proceed to 'flesh out' the paper to produce a basic pre-print. Results from the dataset analyses can be selectively included in the paper.
7. **Submit to Eprints** — The completed pre-print can be previewed before being automatically submitted to the the Eprints digital library component. Subsequent

[1] anonymously and with the patients consent

versions of the paper leading to submission, peer review and reprint are managed by the Eprints server. Investigators continue to discuss the development of the paper in the discussion forum.

It should be noted that the work-flow is not enforced as a linear progression from experiment protocol to pre-print; investigators can make changes to the experiment protocol as the experiment progresses (for example, bringing a new associate investigator on board, inviting a new student to join the exercise), return to the dataset at any point to add/remove experimental results or perform more analyses, and produce many different pre-prints describing different aspects of the experiment.

Finally, *My Papers* provides a simple shortcut allowing the user to quickly access all the papers/reports produced by the various experiments/exercises worked on.

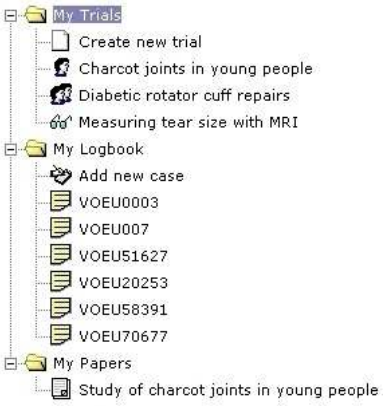

Figure 2: DRJ user space.

3. DEFINING DATA SCHEMAS

The Dynamic Review Journal stores patient case data from hundreds of thousands of operations, each case being described by one of several different schemas; it therefore made sense for us to focus our initial efforts on facilitating user generation of *data* schemas, before considering the experiment and publication schemas. The schemas in the Schema Space provide an interface between the data stored by the system and the users' view of that data - in the case of the data schema, the type and format of sets of data variables are specified and subsequently used to dynamically generate interfaces for entering, viewing, and performing statistical analyses of the stored data (since the data definition is abstracted in the schema, new views of existing data can also be added with little effort). Therefore, in order to introduce different datasets to the DRJ environment, the user must first formally describe the data structure and add it to the Schema Space.

Until recently, this mechanism was only available to expert users with the appropriate technical knowledge to formalise their data requirements using the XML Schema grammar. Our aim was therefore to develop a schema-building tool which would enable the wider (non-expert) DRJ user community to configure their working environment to their changing needs. The first barrier to achieving this goal was the

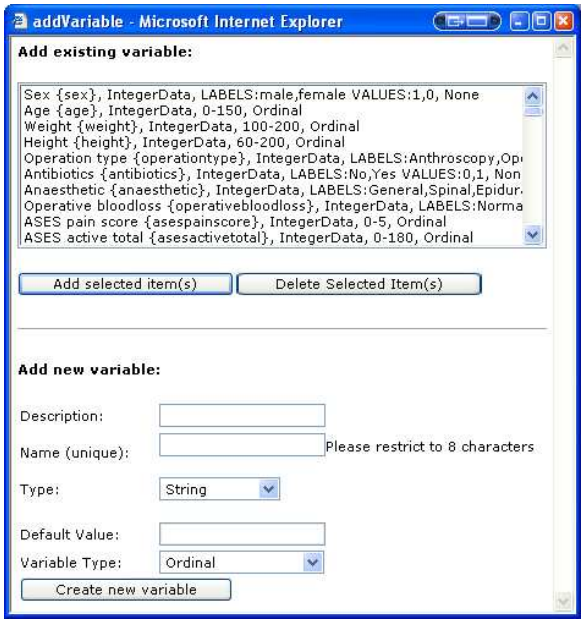

Figure 3: Adding a new variable to a schema in the Template Generation Toolkit.

use of the term 'schema' - for surgeons, this is an alien term within their day-to-day work experience, so the more familiar term 'template' was chosen in its place. This change of name from schema to template also emphasised the evolutionary nature of most schemas. Rather than start with a blank schema and add new items, the majority of new schemas would be modifications of existing ones. Therefore an existing schema would form a template around which new ones could be created. The DRJ's schema-building component thus became known as the *Template Generation Toolkit*.

3.1 The Template Generation Toolkit

The Template Generation Toolkit (TGT) enables non-expert users to quickly and easily add new schema specifications, or 'templates' to the DRJ Schema Space, in accordance with their specific requirements. The TGT has been developed in the form of a *pedagogical toolkit* — such toolkits guide users through the process of articulating their information needs to produce an information plan for a particular task. Pedagogical toolkits lie between restrictive wizards which force a user through a particular route (for example, a printer installation wizard) and more open frameworks which provide unrestricted access to a different parts of a creation process. The TGT therefore, is designed to provide a structured way of creating a new data schema and specifying the variables associated with it.

To aid this task, variables can be grouped together into sets based on the type of data they record. For example, a patient data set could include age, weight, height, and ethnic origin, while a data set for recording the range of movement in the elbow could include range and degree of motion in various directions, pain experienced, and muscle response. New schemas can then be built using sets defined by existing schemas, and 'fine-tuned' by adding, deleting, or modifying

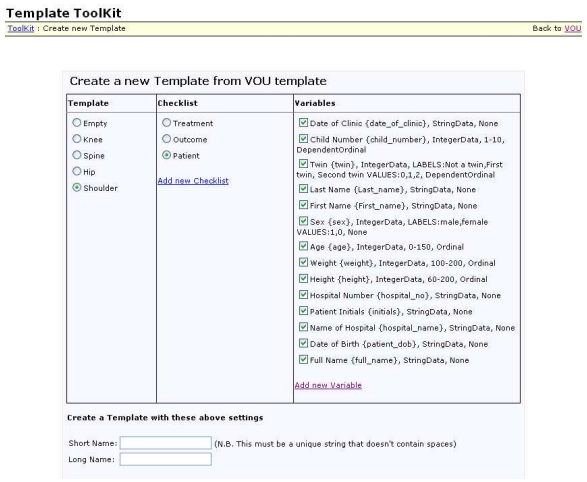

Figure 4: Building a new schema from existing variable sets in the Template Generation Toolkit.

individual variables (Figure 4). To add a variable to a new schema, the user can either select an existing variable from a list of previously defined variables, or create a new variable. When creating a new variable, details such as description, data type, default value and variable type must be specified. The variable type defines the *statistical affordance* of the variable; the user selects from a list of variable types such as dependent interval, independent ordinal, ratio etc (Figure 3).

3.2 Using Data Schemas

Once a data schema has been created using the Template Generation Toolkit, it is immediately available to members of the VOEU community to use to input new case data via the entry forms generated from the new schema (Figure 5). When a suitable number of cases based on the new schema have subsequently been entered into the VOEU database, a group of surgeons may choose to carry out a trial to identify trends in the data. The new schema is selected as the data source in the trial protocol, allowing the entered cases to be added to the trial dataset. Relevant cases can be extracted from the database by using a search form generated from the schema (Figure 6).

The surgeons can then choose between a number of statistical analysis methods and invoke them on the dataset. Since the statistical affordance of each variable was specified at the time of schema creation, the DRJ can generate an interface which hides the variables which are inappropriate to the selected statistical analysis.

3.3 Issues Raised

The development work on the Template Generation Toolkit has raised some issues within the context of the VOEU community, which have wider implications that must be borne in mind as we move towards our goal of more widespread adoption of a Dynamic Review Journal-based infrastructure supporting the experimental work of other on-line scientific communities.

One issue that was faced concerned the lack of presentation

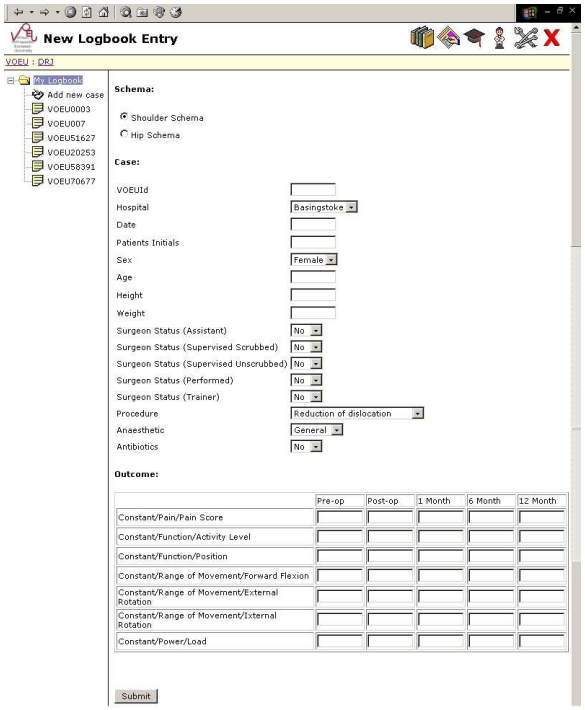

Figure 5: Entering new case data using input form generated from data schema.

information defined within the data schema. Without this information, the interfaces generated from the data schema (including data entry forms, data viewing tables, search forms, and data analysis forms) are simplistic and simply list each variable in the order that they are defined in the schema. While this was sufficient in most cases, in specific cases we were asked by surgeons to adjust the appearance of the forms so that they mirrored the paper-based counterparts which the surgeons were used to. This issue potentially hints at a much larger one — that in large-scale trials involving participants from multiple institutes throughout the EU there is a much greater need for customised presentation of information then would be required in the case of a smaller

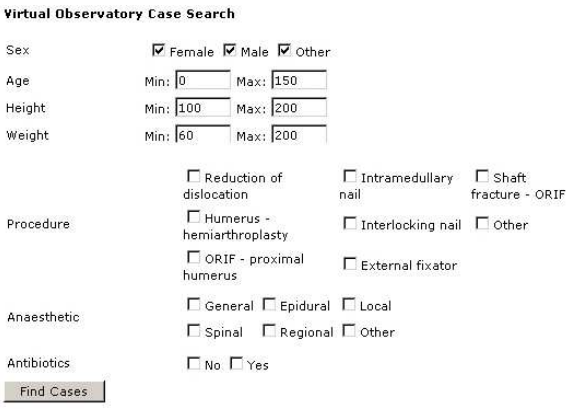

Figure 6: Searching for cases using search form generated from data schema.

trial whose participants are all part of the same institute. Issues such as cultural, linguistic and political differences all need to be properly addressed. This requires a larger development effort; in the current DRJ implementation few of these issues have been addressed. However, VOEU does currently hold a 'user profile' information for each registered user of the system, which provides a useful starting point for future development in providing customised interfaces in different languages or presentation styles. For example, user-specified style sheets could be used to render schema-based views in such a way as to mimic each institution's paper-based forms — this may include structural definitions such as defining the order in which a list of variables should appear, or purely presentational devices such as logos and fonts in order that the user's interface adheres to their institute's corporate identity.

Another issue facing multi-national trials is the potential difference in local policies between participating institutions; a VOEU-specific example is that of different policies used in different regions for assessing a patient's pre- and post-operative motor facilities — surgeons in the US use the AIPES system, whereas in the EU the CONS system is used. A mechanism to enable the DRJ to map between the different types of assessment could be defined in this instance, allowing scores from both types of system to be compared and analysed side by side; however, there may be some cases in which a measurement in one system has no equivalent in the system to which it is being compared.

4. CONCLUSIONS

In answer to the question *What is a digital library?*, the University of California Digital Library FAQ states "systems that support the collections of the University" . Our continuing work on the Virtual Orthopaedic European University project argues for a broadening of this view, out from the collections themselves, to the process of collecting and deploying. We have presented our contribution to this ongoing effort, the Dynamic Review Journal (DRJ), and described its integration and use within a Virtual University learning environment as an example of a system which deliberately crosses the barriers between these areas (experimentation, analysis, publishing, dissemination, discussion, and education).

However, in embracing such an approach, we realise that it is impossible to predict in advance the myriad different types of scientific activity that future users of such a system will want to be supported in carrying out and therefore the DRJ must be able to evolve in accordance with the changing needs of its users. This paper has discussed the DRJ's Schema Space mechanism which allows new experimentation protocols, data formats, and dissemination results to be dynamically introduced to the system as needs arise. Our most recent work has focused on the user-oriented Template Generation Toolkit which empowers non-expert users to tailor the DRJ environment to their specific experimental activities. Lessons learnt from the development of this toolkit include the importance of *personalised* user views which could be achieved through the use of user-selected style-sheets, and the problems of managing large-scale scientific collaborations when the local policies of the participating scientists differ.

5. REFERENCES

[1] Leslie Carr, Timothy Miles-Board, Gary Wills, Guillermo Power, Christopher Bailey, Wendy Hall, and Simon Grange. Extending the role of the digital library: Computer support for creating articles. In *Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia, Santa Cruz, USA*, pages 12–21, August 2004.

[2] Linda Carswell. The Virtual University: toward an Internet paradigm. In *The 6th Annual Conference on the Teaching of Computing and the 3rd Annual Conference on Integrating Technology into Computer Science Education: Changing the Delivery of Computer Science Education*, pages 46–50. ACM, 1998.

[3] Grainne Conole, Gary Wills, Leslie Carr, Lucile Vadcard, Wendy Hall, and Simon Grange. Building a Virtual University for orthopaedics. In *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications (ED-MEDIA), Hawaii*, pages 22–27, 2003.

[4] Ian Foster, Carl Kesselmann, and Sam Tuecke. The anatomy of the Grid: Enabling scalable virtual organisations. *International Journal of Superconductor Applications and High Performance Computing*, 2001.

[5] Simon Grange, Gary Wills, Guillermo Power, Timothy Miles-Board, Leslie Carr, and Wendy Hall. Building a Dynamic Review Journal: Extending the role of the Virtual Orthopaedic University. In *Proceedings of the 3rd Annual Meeting of the International Society for Computer Assisted Surgery (COAS International), Marbella, Spain*, pages 122–123, 2003.

[6] Liz Lyon. Emerging information architectures for distributed digital libraries. In *Proceedings of Digital Library: IT Opportunities and Challenges in the New Millennium, Beijing, China*, July 2002.

[7] Gary Marchionini and Hermann Maurer. The roles of digital libraries in teaching and learning. *Communications of the ACM*, 38(4):67–75, April 1995.

[8] Mary Marlino, Tamara Sumner, David Fulker, Cathryn Manduca, and David Mogk. The digital library for earth system education: building community, building the library. *Communications of the ACM*, 44(5):80–81, May 2001.

[9] Robert E. McGrath, Joe Futrelle, Ray Plante, and Damien Guillaume. Digital library technology for locating and accessing scientific data. In *Proceedings of the fourth ACM conference on digital libraries*, pages 188–194, 1999.

[10] Craig Nevill-Manning. The biological digital library. *Communications of the ACM*, 44(5):41–42, May 2001.

[11] John Weatherley, Tamara Sumner, Michael Khoo, Michael Wright, and Marcel Hoffmann. Partnership reviewing: a cooperative approach for peer review of complex educational resources. In *Proceedings of the second ACM/IEEE-CS joint conference on digital libraries, 2002*, pages 106–114, 2002.