JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

Virtual Research Environments: A Literature Review

Project

Project Acronym	CORE	**Project ID**	
Project Title	Collaborative Orthopaedic Research Environment		
Start Date	01 November 2004	**End Date**	31 October 2006
Lead Institution	University of Southampton		
Project Manager & contact details	Dr Gary Wills Intelligence, Agents, Multimedia Group School of Electronics and Computer Science University of Southampton Southampton, SO17 1BJ		
Project Web URL	www.core.ecs.soton.ac.uk		
Programme Name (and number)	*Virtual Research Environments (05/04) Strand III*		
Programme Manager	Dr Maia Dimitrova		

Document

Document Title	Virtual Research Environments: A Literature Review		
Reporting Period			
Author(s) & Project Role	Gary Wills (Project Manager), Lester Gilbert (Technical Manager), Yee Wai Sim (Co-Investigator), Chu Wang (Co-Investigator), Les Carr (Co-Investigator)		
Date	22/04/2005	**Filename**	ecstr_iam05_001.doc
ISBN	085432822X		
Technical Report Number	ECSTR-IAM05-001		
Access	☐ Project and JISC internal	■ General dissemination	

Document History

Version	Date	Comments
0a	8 April 2005	Initial draft of report
1a	22 April 2005	Final version

Virtual Research Environments: A Literature Review

G. B. Wills, L. Gilbert, Y. W. Sim, C. Wang, L. Carr
School of Electronics and Computer Science
University of Southampton
E-mail: {gbw, lg3, yws01, cw2}@ecs.soton.ac.uk
Technical report Number: ECSTR-IAM05-001
ISBN: 084532822X

1 Science, Publishing and Hypertext

The accepted role of scientific and scholarly publication is to record research activity in a timely fashion, keeping others in the research community up to date with the current developments. Until very recently, it has been the case that printed journals were the most efficient method for the dissemination and archival of research results. Technical advances in the past decade have allowed the process of scholarly communication to take other forms, particularly in the dissemination and storage of articles via the World Wide Web.

It is worthwhile pausing to consider that technological proposals for improving the dissemination of scientific knowledge have been suggested for some sixty years. Immediately prior to the Second World War, a scholar named H.G Wells proposed a microfilm-based index to all human thought and knowledge [33]. The experience of coordinating thousands of American scientists during that war led Vannevar Bush to propose a similar system complete with what we would now call "hypertext links" [3].

The technical advance which eventually made this possible in reality was, of course, the World Wide Web. Developed at CERN to facilitate "instantaneous information sharing between physicists working in different universities and institutes all over the world", it gave publishers a new medium for making their journal archives available [17]. On top of that, it also gave authors the means to break the so called "Faustian bargain" and directly distribute their articles in pre- or post-publication form from their own Web pages [15] or in organised "eprint archives" [10].

However, other factors beyond the development of the printing press in the late 15th century led to the production of the first Scientific Journal in 1665 [30]. The emergence of a reliable postal system and the development of experimental methods in the 16th century also had a significant role.

Similarly, it may not be simply the technical support for reproducing and distributing articles electronically (electronic publishing, eprint archiving and digital libraries), but also the emergence of technical support for improving human communication in the form of highly collaborative, large scale activities and analyses (i.e. the Grid and Virtual Universities). Such support is likely to precipitate significant change in the field of scientific communication and significant changes in the way its communications are produced, curated and disseminated [22]. For example, the old medium allowed a paper to be published as the summarisation of a scientific activity. The discarded raw observations that led to the article's conclusions are replaced by a description of the method for recreating the experiment. However, researchers are becoming more interested in the potential of the new medium for preserving experimental data as well as experimental conclusions: the ability to provide hypertext links between the article and the data to create an audit trail for reviewers and thus facilitate further analyses and meta analyses.

1.1 The Grid

The Grid, as depicted in Figure 1, is a computing infrastructure for undertaking "big science" [9]. Beginning as a mechanism for applying scientific computation to large scale experimental procedures, it has developed to encompass large scale human collaboration [10], the kind of distributed collaboration which now characterises many areas of scientific endeavour (from particle accelerators to genomic experiments). This kind of support might be in the form of videoconferencing, meeting facilitation or even decision rationale and group memory capture rather than simply large scale computation and peta-scale data access.

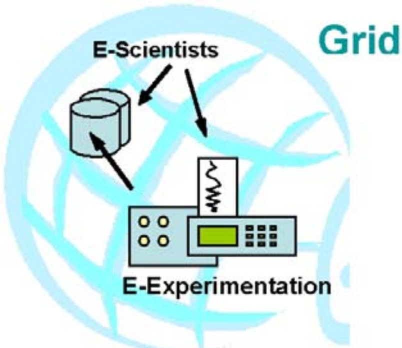

Figure 1: Insert description here

In the UK, the accepted term for scientific activities augmented by the Grid is "e-Science". This paper will adopt that term when we wish to discuss the scientific activity without unduly focusing on the enabling computational technology.

1.2 Virtual Universities

A Virtual University is a distance education programme that is delivered across the Internet: education in which students and teacher are not in the same place or the same time [5]. Virtual universities are also about large scale short-to-medium term collaborations, comprising groups of people with shared educational objectives. The particular role of education in the scientific process is discussed in the next section.

1.3 Open Archiving and Sky Writing

A third string to the e-scientists' slowly evolving bow is becoming increasingly familiar within the digital libraries community. Open Archiving [19] started with the aim of increasing the sharing and reuse of scientific information by promoting the development of interoperable archives of scientific literature. The most prominent example is the High Energy Physics (HEP) archive which currently has over 220,000 articles and 12,000 users a day.

However, increased access for e-scientists to an article is not the final goal of an Open Archive. It also seeks to improve the impact of that article (the take-up of its ideas and the subsequent use, refinement or generalisation of its results). Running in parallel with the (sometimes lengthy) publication process and avoiding the toll-based access of journal subscription, the HEP archive allows physicists to increase the tempo of their literature by reducing the delay between the appearance of an article and the appearance of a citation to it to less than a month [1]. The majority of these articles are subsequently refereed and appear in print journals, but the use of this ‘direct dissemination’ mechanism increases the speed of access to the latest scientific results and decreases the “impact delay” between projects. Even the unstructured dissemination of researchers’ home pages have promoted their research by increasing access to their writing and hence, increasing their impact [20].

This phenomenon described above is fancifully described as *sky writing* [16]: the potential of returning the speed of scientific communication from the year-on-year turnover of journal articles to a tempo more closely related to human conversation.

Although some have seen open archives as fulfilling a kind of vigilante role distributing a “grey literature”, they have been a niche form of digital library, usually run by a research community, for the members of that community. Now that institutional archives are becoming more established, they are often being managed by professional librarians on behalf of the various local communities that their institutions represent.

2 Diachronic Science

The technological aids reported earlier in this paper, wonderful as they are, suffer from being disjoint. The current vision for the Grid focuses only on the immediate aspects of E-Science, i.e. the experiments, analyses and meetings which occur over the duration of a project. As well as these synchronic aspects, any scientific effort (and e-Scientific efforts in particular) will have diachronic features, those collaborative activities which extend through time, enabling the influence of the project to carry on beyond its funded timescale and disseminating its knowledge beyond the boundaries of the original collaboration.

These activities are a well known part of the scientist’s profession (publishing papers, publishing data, giving seminars, re-running experiments and checking others’ results, comparing approaches from different projects, generalising or specialising the work of others, and, of course, teaching).

Publishing activities (the writing of reports, workshop papers, and refereed conference and journal articles) have a significant effect through time. The individual publications may be the result of the immediate collaboration, with working drafts exchanged between each of the project partners; the collected (linked) publications of a project show signs of the extended collaboration, as more results, analyses and conclusions emerge building successively on each previous publication. The details (ordering and preferences) of how these processes result in a combination of reports, presentations, conference papers, journal articles and a publicly accessible record of pre-prints and re-prints, change with the discipline in which the project is being carried out, but the general pattern remains the same, as show in Figure 2.

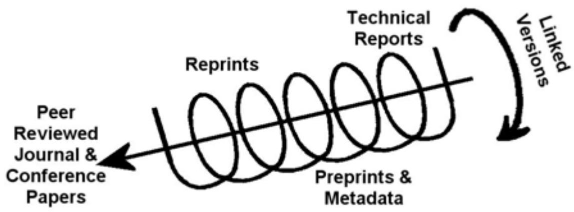

Figure 2: Part of the publishing cycle

This immediate collaboration generates "the literature" which is subsequently read by other scientists in other projects, and whose work it informs and inspires. These scientists may be minded to re-run the original experiments, perhaps adapting the process in some way, to check results, or to specialise or generalise a reported principle. They, in turn, write articles which link to the original work, thus demonstrating its impact, a quality prized by tenure committees and funding organisations alike. This loose-coupled, diachronic collaboration afforded by publishing is a principal foundation of the scientific process, whereby we "stand on the shoulders of giants".

This is where digital libraries stand; mediating hypertext access both to publishers' post-print certified collections and the communities' preprint archives. However, focused as they are on acquiring, maintaining and preserving collections (within a strict budget) and then providing information discovery services to their clients, the researchers and academics, they are blind to (or neutral about) the processes by which this information is created, crafted, evaluated, gathered, distilled, expressed, reviewed and then exploited.

Publication provides a mechanism to extend research collaboration beyond the confines of the original activities, incrementally advancing the scope of scientific knowledge. The current deployment of web technologies increases the effectiveness of this loose-coupled collaboration. The role of the digital library is to focus the various channels (archives, publishers' websites and aggregation agents) into a single portal which mediates these strands of diachronic collaboration.

There is also a looser coupling that exists between researchers as educators and their students, particularly in the context of higher education. Here the output of research activities is used to form and inform the next generation of scientists and e-Scientists. The deployment and adaptation of scientific materials into study packs, modules or learning objects is a crucial part of the educational process, and a *raison d' être* of university education. It may be that the "distance" between cutting edge research and the classroom may be different for postdoctoral education (as demonstrated in a later section) and undergraduate education, but the goal of passing new information on is a kind of collaboration (in the extreme, an extended diachronic collaboration with a subsequent generation).

It seems reasonable that the accepted picture of e-science (see Figure 1) could be enlarged from its current focus on experimentation and analysis to feature these processes of wider significance; since without these aspects of diachronic collaboration, there would be no ongoing science and, indeed, no scientists.

3 Related Work

Dalgaard expands the notion of scholarly hypertext away from hypertext being merely intertextual relationships between articles to the relationship between text and archive [8]. He points out that from its very inception, hypertext was thought of both at the level of the text and at the level of the network, arguing that in the context of the Web, hypertext has become the paradigmatic rhetorical structure of a global and distributed archive. Accordingly, a scholarly archive is the collection of scholarly texts, and the catalogues and reference works giving access to them. Dalgaard observes that most navigational options are presented as texts (lists of works, authors' name, references, etc). This is the hypertextualization of the scholarly archive.

The historical image that many people have of the scientific process is that of a lone scientist or small team working in a basement laboratory. A similar picture appears for the use of libraries, where researchers ferret away in dusty books for vital missing bits of information. Levy and Marshal have examined the early underlying assumptions and how they affected digital library development [21]. In their article they challenge these images, especially the assumption that digital libraries are used by individuals working alone. They point out that the work carried out by both research staff (in conducting research) and library staff (in providing the service), is one of collaboration, and that digital libraries should support formal and informal collaboration and communications.

Similarly Marchionini and Maurer point out that "digital libraries will allow learners of all types to share resources, time and energy and experience to their mutual benefit" [24]. 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 e-scientists are to go beyond the present position it is essential that they have access to the raw data in their research and training. 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 [26], 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 [29].

Marchionini and Maurer also suggested that digital libraries should offer greater opportunities for users to deposit information. 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, provides services to create and use materials, and facilitates peer reviewed teaching and learning resources [25]. The peer review of collections and peer comment is a significant part of the disseminating process, which adds value to any collection. Weatherley *et al.* have proposed a model that will aid reviewers in reviewing complex material or a digital collection [32]. 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 [23]. In her proposed scenario, researchers would undertake experiments, deposit raw data, and produce pre-prints using web services.

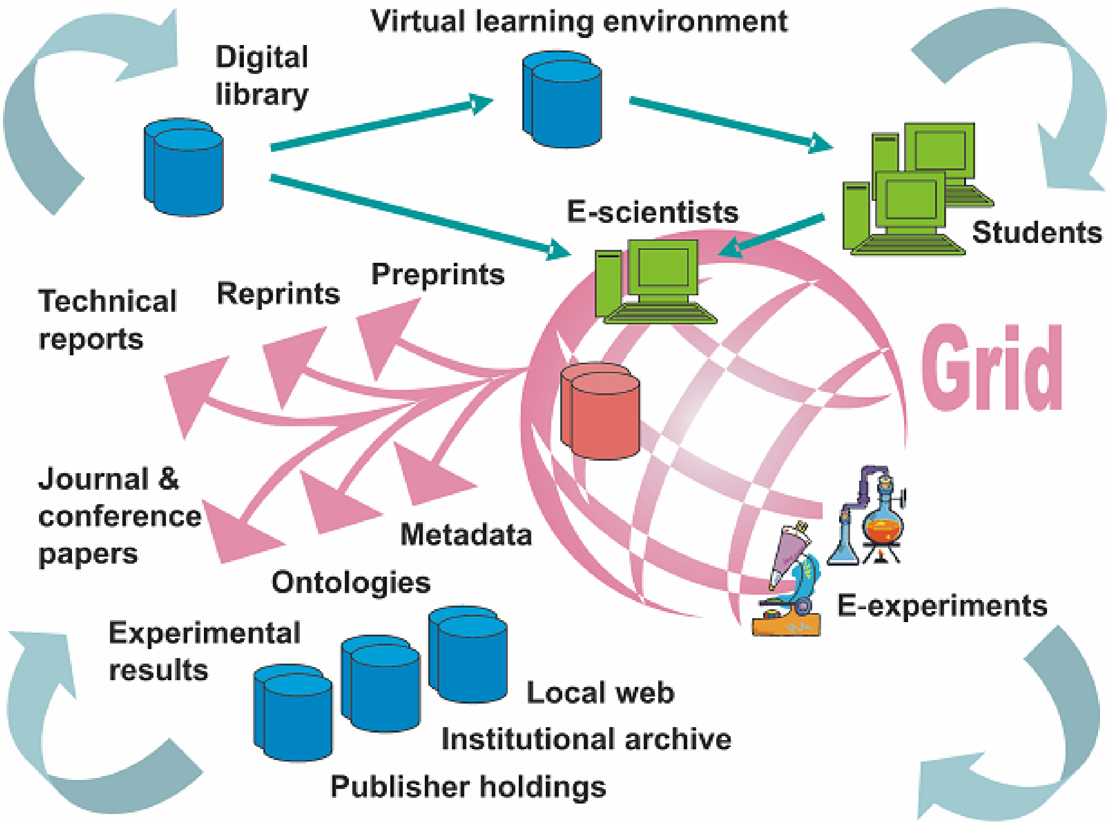

Figure 3: Complete cycle of e-Science

Recent advances in Web technologies allow such concepts to be realised. The Dynamic Review Journal (DRJ) has been implemented as a Web-based environment for supporting a critical subset of the e-science cycle (Figure 3): the collation and analysis of experimental results, the organization of internal project discussions, and the production of appropriate outline documents depending on the requirements of conferences and journals selected for dissemination [4, 6, 13]. Within the tightly coupled 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.

3.1 Other Medical VREs

The large e-Science initiative here in the UK [28] has seen a number of projects focused on supporting the collaboration of science using grid technologies. This section will concentrate on project that uses the Grid to aid scientific enhancement through Grid technology.

The Integrative Biology Project [18] aims to provide an infrastructure to support the investigation of heart disease and cancer. They are developing a customized Grid framework to run large-scale organ simulations, to manage a database of simulation results and to support collaborative analysis and visualization [11]. The project will build a number of demonstrators to aid the understanding and validation of such as Grid system.

The MIAKT (Medical Imaging and Advanced Knowledge Technologies) project [27] provides support for medical decision making during a multi-disciplinary meeting for breast-cancer screening, diagnosis and treatment. The MIAKT project also has an implementation of a Task Invocation Framework which uses web-based services to provide discrete and disparate functionality to a generic client application [31].

GEMSS: Grid-Infrastructure for Medical Service Provision [14] is a project to develop grid middleware specifically to ensure the privacy of patient data and used to run a number of different types of medical simulation [1].

4 JISC's VRE Programme

JISC[1] have a Virtual Research Environments Programme [34] aimed to engage the research community in:

- Using available tools and frameworks to build and deploy Virtual Research Environments (VRE) and where necessary developing or integrating new tools and frameworks;
- Evaluating the effectiveness of VREs in supporting the research process and how they might fulfil the future requirements of researchers in the UK.

The programme recognises the complex range of tasks undertaken by researchers and the need to support research activities within and across disciplinary boundaries. A VRE will not only support the underlying process of research in both the small and large scales but will also not restrict the access to resources required by the research team.

JISC do point out that there is no precise definition of a VRE and the programme encourages the project to develop frameworks, influence standards and to populate the VREs with services appropriate to the domain.

5 Conclusions

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. A digital library lies at the heart of every VRE. In the e-research 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 [10] and Virtual Universities for computer-supported education [5], its role has increased. 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

[1] Joint Information Systems Committee: http://www.jisc.ac.uk/

in carrying out, and therefore a VRE must be able to evolve in accordance with the changing needs of its users.

Reference

1. Benkner, S., Berti, G., Engelbrecht, G., Fingberg, J., Kohring, G., Middleton, S. E. and Schmidt, R. (2004) GEMSS: Grid-Infrastructure for medical service provision. In *Proceedings of HealthGRID 2004*, Clermont-Ferrand, France.
2. Brody, T., Carr, L. and Harnad, S. (2002) Evidence of hypertext in the scholarly archive. In *Proceedings of the Thirteenth ACM Conference on Hypertext*, University of Maryland, pp. 54-59.
3. Bush, V. (1945) As we may think. Atlantic Monthly, 176(1), pp. 101-108.
4. Carr, L., Miles-Board, T., Wills, G., Power, G., Bailey, C., Hall, W. and Grange, S. (2004) Extending the role of digital library: computer support for creating articles. In *Proceedings of the Fifteenth ACM Conference on Hypertext and Hypermedia*, University of California, Santa Cruz, USA, pp. 12-21.
5. Carswell, L. (1998) The “virtual university”: toward an internet paradigm? In *Proceedings of 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*, University of Ireland, Dublin, pp. 46–50.
6. Conole, G., Wills, G., Carr, L., Vadcard, L., Hall, W. and Grange, S. (2003) Building a virtual university for orthopaedics. In *Proceedings of the EDMEDIA World Conference on Educational Multimedia, Hypermedia & Telecommunications,* Hawaii, pp. 22-27.
7. Crowder, R., Wills, G. and Hall, W. (2003) Evaluation of a hypermedia maintenance support application. *International Journal Computers in Industry*, 51(3), pp. 327–344.
8. Dalgaard, R. (2001) Hypertext and the scholarly archive: Intertext, paratexts and metatext at work. In *Proceedings of the Twelfth ACM Conference on Hypertext and Hypermedia*, Arhus, Denmark, pp. 175–184.
9. Foster, I. and Kesselman, C. (eds.) (1999) *The Grid: Blueprint for a new computing infrastructure*. Morgan Kaufmann.
10. Foster, I., Kesselman, C. and Tuecke, S. (2001) The anatomy of the Grid: Enabling scalable virtual organisations. *International Journal of Superconductor Applications and High Performance Computing*, 15(3), PP.200-222.
11. Gavaghan, D. J., Lloyd, S., Boyd, D. R. S., Jeffreys, P. W., Simpson, A. C., Mac Randal, D. F., Sastry, L. and Kleese van Dam, K. (2004) Integrative biology - exploiting e-Science to combat fatal diseases. In *Proceedings of the UK e-Science Programme All Hands Meeting 2004*, Nottingham, UK.
12. Ginsparg, P. (1996) Winners and Losers in the Global Research Village. In *Proceedings of the ’96 UNESCO Conference,* Paris. Available from: http://xxx.lanl.gov/blurb/pg96unesco.html.
13. Grange, S., Wills, G., Power, G., MilesBoard, T., Carr, L. and Hall, W. (2003) 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*, Marbella, Spain, pp. 122–123.

14. Grid-enabled Medical Simulation Services. Information available at http://www.ccrl-nece.de/gemss/
15. Harnad, S. (1995) Electronic Scholarly Publication: Quo Vadis? *Serials Review*, 21(1), pp. 70–72.
16. Harnad, S. (2001) The future of scholarly skywriting. In: Scammell, A. (ed.) *i in the Sky: Visions of the information future*, pp. 216–218. Aslib.
17. Hitchcock, S., Carr, L. and Hall, W. (1996) A Survey of STM Online Journals 199095: the Calm before the Storm. In Mogge, D. (ed.) *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, 6th edition, Association of Research Libraries, Washington , Washington, D. C., pp. 7–32.
18. Integrative Biology Project. Information available from: http://www.integrativebiology.ac.uk/
19. Lagoze, C. and de Somple, H. V. (2001) The open archives initiative: building a low barrier interoperability framework. In *Proceedings of the First ACM/IEEECS joint conference on Digital libraries*, Cornell University, Ithaca, NY, pp. 54–62.
20. Lawrence, S. (2001) As we may think. *Nature*, 411(6837), pp. 521.
21. Levy, D. M. and Marshall C. C. (1995) Going digital: a look at assumptions underlying digital libraries. *Communications of the ACM*, 38(4), pp. 77–84.
22. Lucier, R. (1990) Knowledge management: Refining roles in scientific communication. *EDUCOM Review*, 25(3), pp. 21-27.
23. Lyon, L. (2002) Emerging information architectures for distributed digital libraries. In *Proceedings of the International Conference of Digital Library–IT Opportunities and Challenges in the New Millennium*, Beijing, China, July 2002.
24. Marchionini, G. and Maurer, H. (1995) The roles of digital libraries in teaching and learning. *Communications of the ACM*, 38(4), pp. 67-75.
25. Marlino, M., Sumner, T., Fulker, D., Manduca, C. and Mogk, D. (2001) The digital library for earth system education: building community, building the library. *Communications of the ACM*, 44(5), pp.80–81.
26. McGrath, R. E., Futrelle, J., Plante, R. and Guillaume. D. (1999) Digital library technology for locating and accessing scientific data. In *Proceedings of the fourth ACM conference on Digital libraries*, pp 188–194.
27. MIAKT project. Information available from: http://www.aktors.org/miakt/
28. National e-Science Centre. Information available from: http://www.nesc.ac.uk/
29. NevillManning, C. (2001) The biological digital library. *Communications of the ACM*, 44(5), pp. 41–42.
30. Schaffner, A. (1994) The future of scientific journals: Lessons from the past. *Information Technology and Libraries*, 13(4), pp. 239–247.

31. Shadbolt, N., Lewis, P., Dasmahapatra, S., Dupplaw, D., Hu, B. and Lewis, H. (2004) MIAKT: Combining Grid and Web Services for Collaborative Medical Decision Making. *e-Science All Hands Meeting 2004*, Nottingham, UK
32. Weatherley, J., Sumner, T., Khoo, M., Wright, M. and Hoffmann, M. (2002) Partnership reviewing: a cooperative approach for peer review of complex educational resources. In *Proceeding of the second ACM/IEEECS joint conference on Digital libraries*, Portland, Oregon, USA, pp. 106–114.
33. Wells, H. (1937) World Brain: The Idea of a Permanent World Encyclopaedia. In: *Encyclopédie Française*, Aise.
34. Virtual Research Environments Programme. Information available from: http://www.jisc.ac.uk/index.cfm?name=programme_vre