External evaluation of the eBank project

Final report

Gráinne Conole
Email: g.c.conole@open.ac.uk
December 2006

CONTENTS

Exec summary

This document summarises the findings of the external evaluation of the JISC-funded eBank project. The key focus of the evaluation was on the pedagogic aspects of the eBank project as described in Work Package 3 of the overall project plan - ‘e-learning embedding and evaluation’; the evaluation also explored organisational and technical issues which arose from the project activities. Data was collected through document analysis, interviews with five members of the project, interviews with seven students involved in a trial using the eBank materials as part of one of their masters’ courses and observation of students undertaking their final assignment presentations as part of the same course.

The evaluation looked at: project aspirations and the origins of eBank, collaboration and inter-disciplinarity, links with related projects, key success factors and outcomes, dissemination mechanisms used by the project team, barriers and enablers to the uptake of the products produced, conceptual models underpinning the project, and the student experience of using eBank material. The project was closely aligned with a number of other initiatives in digital library work and computer science and in particular those associated with the e-Science initiative (such as CombeChem, R4L, e-Malaria, and other JISC digital repository projects). Cutting across these projects were a number of inter-linked issues about: benefits and issues associated with electronic research data, issues around the management of data (archiving, retrieval, dealing with data collections), mechanisms for improving laboratory working practice (reducing redundancy, increasing collaboration, improving information literacy skills), and the nature and format of published data and research papers.

The aims of eBank were threefold: i) to make data available through open access, ii) to link data to references and iii) to make research data available and applied in the learning context. The interviews fore grounded a series of key issues with current research practice processes and potential ways in which technologies might address these. These concerned the nature of electronic data and the way in which it is archived, managed and retrieved, as well as issues to do with the human and organisational aspects of the research process and how better use of technology might improve research processes.

The project achieved five inter-related achievements: i) a data repository of crystal structure data, ii) a metadata application profile, iii) an aggregator service, iv) integration within the RDN PSigate portal, and v) demonstration of the use of eBank-type material in a teaching context. In addition, two key success factors are evident: the productive nature of the interdisciplinary team involved and a comprehensive dissemination strategy with appropriate targeting of relevant stakeholders to ensure buy in and take up of the concepts underpinning the project. Identified barriers and enablers centred on nine key issues: ownerships, research practice, level of ICT skills, institutional infrastructures, publishers’ attitudes, technical issues, funding drivers, competing agendas and IPR issues.

External evaluation of the eBank project

Gráinne Conole
Email: g.c.conole@open.ac.uk
December 2006

Introduction

This report outlines the findings of the external evaluation of the JISC-funded eBank project (http://www.ukoln.ac.uk/projects/ebank-uk/), which was carried out between January – August 2006. eBank is a JISC-funded project, funded in three Phases, with funding for phase two running until September 2006. Funding for Phase 3 was confirmed post completion of the evaluation. The main partners were:

- UKOLN, University of Bath (Liz Lyon – Director, Rachel Heery /Sally Lewis – Project Manager, Monica Duke – Software Developer, and Traugott Koch – Researcher)
- University of Southampton (Jeremy Frey and Simon Cole – Chemistry; Les Carr – Electronics and Computer Science)
- University of Manchester, PSIgate - now part of INTUTE[1] (John Blunden-Ellis)

In addition the project worked with a number of other key stakeholders, including professional bodies and publishers. The focus of the external evaluation was on the pedagogic aspects of the eBank project as described in Work Package 3 of the overall project plan - 'e-learning embedding and evaluation'. The project documentation states that:

> This work package will develop the inclusion of e-research data in e-learning courses. The MChem course at Southampton has been identified as an appropriate masters course together with some courses for PhD students. Appropriate units which contribute to the Southampton chemistry courses for years 3 & 4 of the MChem course and for years 1 & 2 of the course have been identified (Chemical Crystallography, Chemical Informatics), where links to the primary research data are essential. This applies also to the project work undertaken by the chemists in years 3 & 4, chemists on the Chemical Informatics courses and those learning about x-ray crystallography in Se. Primary research data outputs will be embedded in learning materials in a number of ways e.g. through links in reading lists, through essay assignments, through analytical problems, through practical work, through RDN PSIgate links etc. The students who are enrolled on the course will use and test the materials. The response of students to the course materials using eBank primary data will be co-ordinated and evaluated by an external evaluator. Learning outcomes will be assessed against the learning objectives for these course modules and pedagogical benefits will be explored. An evaluation report will be produced.

Methodology

The evaluation methodology was designed to produce both a broad survey of the eBank project activities in relation to Work Package 3, coupled with a detailed picture of the development activities and the associated issues, barriers and enablers. Further

[1] http://www.intute.ac.uk/

details about the rationale and set up of the evaluation are outlined in the initial evaluation baseline document. The objectives were to

1. adopt a participatory evaluation approach that provides formative and summative feedback to the eBank project,
2. carry out an examination of the project activities and outputs in relation to Work Package 3,
3. evaluate the success of the project in terms of Work Package 3 against aims and objectives,
4. identify interventions: make suggestions about possible activities and approaches,
5. carry out a pedagogical analysis and evaluation,
6. explore strategies for sustainability and make recommendations for future actions in relation to the outputs of the project and the findings of the evaluation study,
7. produce a final report with recommendations (for best practice and possible further work).

The evaluation of the specific objectives was organised around three key aspects of the project; pedagogical, technical and organisational issues. The pedagogical evaluation looked into the pedagogical models and theory used in eBank in terms of course design, development and delivery of course material. Additionally the methods used to develop and deliver courses were compared with relevant research literature. The organisational evaluation addressed the overall organisational issues (both intra- and inter-) associated with the eBank project. In particular the focus was on the approach adopted in terms of collaboration, dissemination and working in an inter-disciplinary context. The evaluation of the technological aspects looked at the systems used to implement eBank and associated issues.

The main activities of the evaluation included the following:

- Documentary analysis of the project ie project plans and reports.
- Interviews with key members of the project team:
  - Liz Lyon (eBank Director, UKOLN) f-t-f, 1/08/06
  - Jeremy Frey (Chemistry, University of Southampton) f-t-f, 23/05/06
  - Simon Cole (Chemistry, University of Southampton) f-t-f, 23/05/06
  - Monica Duke (Software Developer, UKOLN) telephone, 1/08/06
  - Traugott Koch (Research, UKOLN) f-t-f, 1/08/06
- Observational analysis of students doing their final assignment presentations in the MChem Chemical Informatics module at Southampton University on the 23rd May 2006.
- Interviews with seven students taking part in the MChem Chemical Informatics module at Southampton University on the 23rd May 2006.

Document analysis provided contextual data about the project and its key activities and milestones. Interviews with project team members were held at the Universities of Bath and Southampton; all were face to face except for one which was done by phone. Interviews were semi-structured using a set of questions derived from the evaluation brief, document analysis and discussion with the eBank director. All interviews were audio recorded and transferred to a computer. Interviews were transcribed, focusing in particular on key segments which related to the principle evaluation questions. Field notes were also kept during the interviews and the student observations. Interviews were coded using an emergent theme approach which iteratively identified the key categories of interest. All quotes included in this report have been coded, for example [IntN] signifies data from Interviewee N.

Findings and discussion

The findings of the evaluation cover the following broad themes:

- Project aspirations
- Collaboration and inter-disciplinarity
- Links with related projects
- Key success factors and outcomes
- Dissemination mechanisms
- Barriers and enablers
- Conceptual models and pedagogical issues
- The student experience
- Future directions and recommendations

The following quote perhaps best encapsulates the core philosophy underpinning eBank:

> [The eBank philosophy is about] the separation of intellect and interpretations in a journal article from the underlying data, from the underpinning data, so that, which enables that data to be made openly available for reuse. [Int 2]

Project aspirations

The key aspirations underpinning eBank were:

- To make data available through open access, so that it could be disseminated more quickly
- To link data to derived references and enable demonstration of provenance
- To see research data available and applied in the learning context.

The origins of eBank can be traced back to 2002 (Lyon, 2002) although the work builds on conceptual ideas of the consortium members from before then. Lyon (2003) outlined a number of benefits to making data openly available: providing direct access to data, linking data to research publications, providing a mechanism for ensuring robustness, openness and provenance in the academic process, and enabling more rapid dissemination of scientific ideas. For teaching and learning she saw a benefit in terms of:

> …enriching the student experience, this approach [i.e. provide students with access to original data and linking data to research references] would help to develop their evaluation and critical skills, because they would be able to go back and look at the conclusions a researcher had derived from a set of data and they could analyse it themselves and think did they [the researcher] make the right decisions, how would I have done it, was the method correct, those sorts of things, so that was the pedagogical benefits I could see? [Int 10]

Within this broad framework of benefits, the eBank team members interviewed highlighted a range of interconnected aspirations for the project and reasons for its conception.

> EBank was one of the very early projects …. [We and others] wanted to solve all the problems related to storing research data and making it accessible. One key aim developed later …was

> creating links between the digital libraries and e-science areas and applying digital library results and technologies to e-Science outputs. [Int 11]

This quote demonstrates that the project was closely linked to both digital library and e-Science research and development work at the time. Within these broad initiatives a number of key 'problems' were becoming evident and researchers, such as those involved in eBank, were beginning to explore potential ways in which technology might offer solutions to these problems. The interviews with project team members fore grounded a number of these problems and potential solutions as described here and these formed the backbone of the focus for much of the eBank work:

- **Electronic data:** The increased prevalence of computers in the laboratory both to collect and store data in recent years has resulted in a fundamental shift in the research process. This gives rise to new possibilities about how data can be used, but also raises issues about how it could be found, manipulated, and archived.
- **Archiving and retrieval:** Previously research output, even when derived electronically, was primarily paper-based (an infra-red spectrum, a paper tape of diffraction data, a computer-generated diagram of a crystal structure). Researchers across different domains were beginning to consider the implications for research of harnessing technology more for research purposes. There was recognition that computers were becoming more powerful and hence storage was less of a problem, meaning it was possible to store experimental outputs electronically. But this data needed to be managed, raising questions about what were the best ways of archiving and retrieving this kind of data? How could it be protected against damage or loss? What would happen if the storage device became redundant or was overwritten or erased?
- **Collections:** At around this time there was a growing interest, particularly in the digital library community, in thinking about collections (and in particular collections of e-Science data) – how they could be described and made openly available through collection level descriptions. Inherent in this was the concept of thinking about data specifically as a collection (c.f. with other types of collections such as images or books) and how these collections could be described and made available.
- **Redundancy:** A recognised problem in laboratory practice was that there was an unnecessary level of inefficiency in the research process - scientists repeating experiments because they could no longer find the original data set, or a particular line of investigation being delayed when workflow and outputs were lost because an individual left the team. More systematically storing research outputs offered a means of reducing this inefficiency.
- **Collaboration:** Managing the data and research outputs from team laboratories was recognised as particularly problematic. How, for example, could one researcher find work carried out by another researcher who might have left the lab? Where would data from a particular experiment carried out a year ago be stored, would it still be possible to retrieve it? Practices for managing such collaboration varied, both in approach and degree of success, from laboratory to laboratory. However researchers began to see that storing data electronically offered new possibilities in terms of providing a more streamlined and coherent collaborative research management process.
- **Information literacy skills:** The increased prevalence of electronic research data meant that both researchers and trainee researchers/students needed to develop

new information literacy skills in terms of handing information (to deal with data in the laboratory as well as being able to find and use external data appropriately).

> The whole complexity of the Chemistry world now, of the difficulty of searching for information. We know students are not so well trained in keeping laboratory note books as they used to be and they are having to process a lot more, the speed with which things are moving. So anything that helps moving in a direction that makes it easier to find and recover… [Int 1]

- **Published data.** The internet offered a vehicle for making research outputs available beyond the laboratory, both informally through email to colleagues but also more formally through online electronic journals. Researchers were becoming increasingly aware of the limitations of paper-based data and the advantages of electronic data – not just in terms of the data generated by individual researchers but also that generated by others and made available – for example through a peer-reviewed journal. Access to research findings in the form of an electronic paper was one advantage but perhaps more powerful was the fact that a graph replicated in a journal paper would in fact be based on electronic data which meant that, in principle, it would be possible to provide wider access to that underlying data so that other researchers could manipulate and use it.

These kinds of problems, and potential technological solutions, were part of a wider spectrum of issues being considered by the e-Science initiative.[2] Specifically in terms of application of e-Science philosophies to addressing research issues in Chemistry, the following scenario was suggested by one of the eBank:

> ... you looked up the literature and you found the reference and you go to that reference and you get the paper and from the paper you'd see a figure and from the figure you'd get the data and from the data you'd get the data it was derived from and so on .. and the most powerful incentive is that you'd do that for your own papers so that if somebody said to you 'did you think about analysing this this way?' you could actually do it because you could go back. [Int 1]

In terms of this vision, for the Chemists at Southampton University, the aim with eBank was to build on existing expertise in combinatorial Chemistry and x-ray crystallography developed as part of the Combechem project,[3] to produce a demonstrator project in x-ray crystallography which illustrated some of these ideas about maximising the use of electronic data in the research process. Crystallography was chosen because as a sub-discipline of Chemistry, it already had a long history of the use of computing as part of the research process and as such information associated with crystal data collection (raw data from the diffractometer, crystal co-ordinates and other structural parameters) was already available in digital format. The idea was:

> to build up some model [demonstrator] that would allow you to disseminate this in a way that people would know the data existed and would have a uniform (ish) path to get back to the original data. [Int 1]

However the development of such a demonstrator was far from trivial and was inter-dependent on a number of issues. How should such data be tagged and with what?

[2] http://www.rcuk.ac.uk/escience
[3] http://www.combechem.org/

How should the research data be stored? What were the most appropriate mechanisms for providing access to the data? What protocols for exchange between different systems would be needed?

> Therein lies one of its difficulties – you cant actually demonstrate its full power until you've got all the pieces of the cake. [Int 1]

Focusing on x-ray crystallography as a subject area helped, as the workflow processes were already partially in electronic format and there was a degree of common agreement amongst crystallographers about file formats and work processes which meant that it was possible to provide a reasonable demonstrator of the potential across the work flow process from crystal data collection to publication.

Collaboration and inter-disciplinarity

One of the early success factors of the project was the realisation that the Chemists would need to work with both Information specialists and Computer Scientists. A key benefit of the project was that they did adopt an inter-disciplinary approach from the start – drawing on the range of expertise from Chemists, Librarians, Information Scientists, and Computer Scientists. The involvement of the library community was crucial in terms of helping to address some of the complex information management issues which were arising out of this work - such as how could one document and follow the types of rich and complex digital workflow processes which occur in research laboratories? The significance of having an inter-disciplinary team was evident to the team:

> A meeting of cultures, which is quite an output in itself. Good to work at interface sometimes. [Int 2]

However development of a shared understanding across these different disciplines proved to be non-trivial. An early trigger appears to have occurred when the Director, Liz Lyon, and Jeremy Frey become aware of each others independent, but increasingly convergent, interests in this area:

> [Liz speaking to Jeremy sometime before the project] …he seemed to be coming from a similar, but slightly different track, which was about publication at source, which is how he described this, which is about making data opening available and it kinda clicked. [Int 10]

Networking at conferences and other joint events allowed the individuals from the different disciplines to begin to build a common interest and shared understanding – often triggered by ideas generated from presentations or chance conversations in between sessions. Although on the surface this may seem somewhat *ad. hoc.* in nature, this is not the case, because in fact it is evident that there are a range of complex and inter-woven interests amongst the consortium member which were applied to tackle essentially the same problem domain but from different perspectives. Without this underlying shared vision, no amount of networking or serendipitous opportunities would have provided enough of a trigger to spark inter-disciplinary collaboration. It was also about ensuring there was time to develop the shared understanding.

> I guess one output from the project has been learning a lot about how the other half live, if you like. The first year of it was really quite hard work because different communities use different jargon. It wasn't just a case of 'tomato' [American] 'tomato' [English], …

> completely picking up on the wrong thing. So quite a lot of misunderstanding happened initially, so has really promoted quite an involved understanding of how a different community works, it's a very good cross discipline project in that respect. [Int 2]

In addition to making the link between the Information Science and the Chemistry aspects, the project also built on existing interests in the development of digital repositories (which were happening in parallel) – a shared interest for both the Librarians and the Computer Science community. Conversations around this aspect made it clear that the problem being articulated, in what was to become eBank, would be a useful area to test out some of the ideas and developments occurring in digital repository work.

> And still I think there was an ambition to more systematically cooperate ... e-Science, digital library and IT people… quite a hard thing to do, the communities are not very eager to cooperate. .. it's a hard process, there are so many different interests and conceptions .. the terminology is so different. Hard to get three different communities to work together different – … on the other hand the crystallographers themselves are at least interested in doing the repository…. Which seems not to be the case in a lot of other disciplines. [Int 11]

Early conversations with the emergent e-Science community also proved important; in particular there seemed to be potential in applying expertise from the Information Science community, who had a lot of experience in developing digital library architectures and expertise in handling and managing information, to e-Science – ie the e-Science work was uncovering problems which the Information Science community could potentially help solve. Together these 'connections' illustrate the way in which the project made links across parallel, but related research activities. What is interesting is the way in which the project bought together a range of disparate parallel, technology-related activities which were occurring at the time and made connections between these and had a vision of seeing how they could be applied to a particular problem.

> We talked a bit more the whole idea of making data, putting data in some sort of a repository and making it available for learning through this scholarly knowledge cycle seemed to become a possibility. [Int 10]

> It was about making connections and seeing where activities that are happening in one area can be migrated, transformed and transferred into another area. [Int 10]

A key trigger appears to have been a workshop which was held at UKOLN,[4] which was the first workshop that bought together the e-science community with the digital library community.

Therefore, eBank developed out of a set of inter-connected interests in a shared problem space. It bought together subject experts from different disciplines with an underlying shared understanding (it is interesting to note for example that many of the consortium members had scientific backgrounds). This shared epistemological belief may have more significance than might at first appear. For example there appears to be a tacit, shared world view which aligns with a Scientific, positivist interpretation of

---

[4] Held at University of Bath on 5th August 2004, see the 'eBank UK Evaluation Report: Outcomes of eBank UK workshop' for more details available online at http://www.ukoln.ac.uk/projects/ebank-uk/workshop/evaluation-report.doc

the world (Oliver *et al.*, 2007). Trying to bring together disciplines with more disparate epistemological beliefs might be significantly more problematic.

> So all of this thinking has pre-dated eBank but kind of, I was interested in finding a way of bringing those communities together and demonstrating the convergence, the benefits, the added value of bringing together of escience people with the digital library people and eBank achieved that as well. So there are lots of treads that came together in eBank, lots of threads that had been brewing and developing for some time, that came together in eBank. [Int 10]

The benefit of this shared language across the different communities involved in eBank was recognised as important:

> EBank still has the advantage of talking to scientists in another discipline for a couple of years, this is an advantage other projects wont have. [Int 11]

However development of this shared language required effort; there were a number of implicit assumptions within each community which needed to be made visible and addressed:

> The digital library people especially made a lot of assumptions about that everybody would like to share their data and…. disciplines are basically similar in their traditional and in their working with data and the data would be … opening accessible and that different disciplines had a strong impetus to cooperate …and I think many of these assumptions are not quite true. [Int 11]

Transferring this approach to other disciplines would probably require additional effort, as the approach adopted in one discipline will not necessarily work in another.

Links with related projects

Projects which eBank either built on or linked closely with are listed here and visualised in figure one.

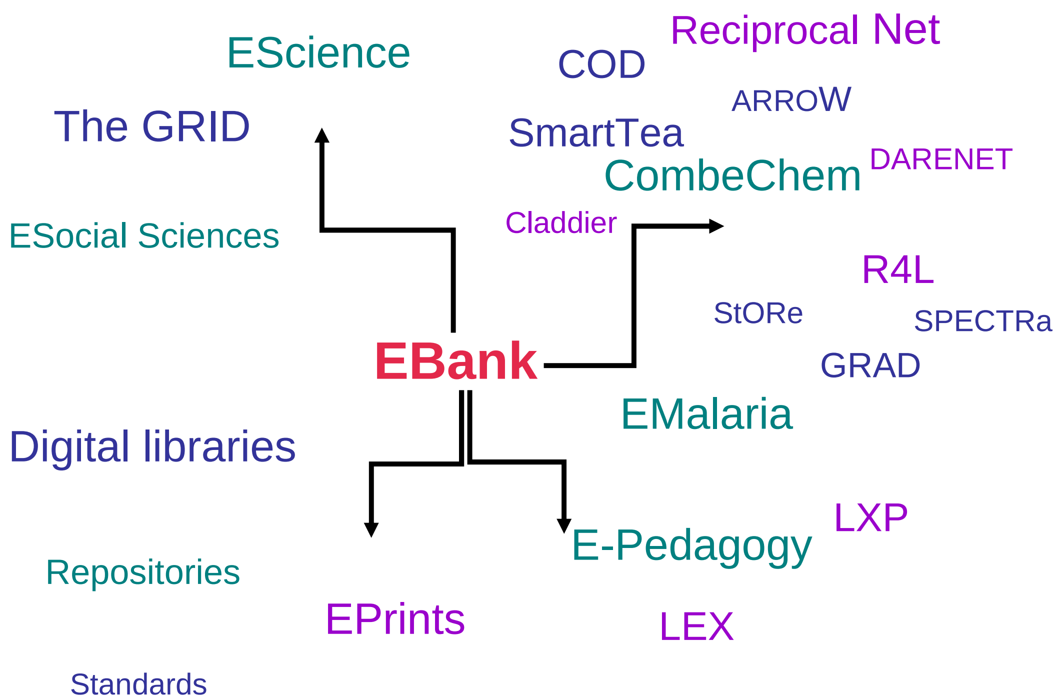

Figure one: visual representation of related projects

- **CombeChem** - provided a link between eBank and wider eScience initiatives. It 'developed an e-Science testbed that integrates existing structure and property data sources within a grid-based information-and knowledge-sharing environment'.[5]
- **R4L**[6] - focused on an earlier part of the research cycle, repositories in the laboratories. It was concerned with the management of data – such as linking repositories and dealing with multiple sources of information from a range of different Scientific instruments.
- **SmartTea and Mytea** – were concerned with electronic laboratory note books. Development activities included work on the electronic monitoring of laboratories (such as attempting to provide automatic information about the laboratory environment, people and data). This research links across to research on the use of mobile devices for learning (see Kukulska-Hulme and Traxler, 2005 for an overview). Smart Tea,[7] explored the use of mobile devices in the laboratory focusing on the use of hand held PDAs in Chemistry to support the research workflow process. The project, MyTea,[8] extended this work into the Bioinformatics area. These projects are important, and of relevance to eBank, because they provide insights into laboratory workflow process (of which eBank activities form one aspect), which needs to be better understood in order to developed technological systems for both research and learning.

[5] http://www.combechem.org/
[6] http://r4l.eprints.org/
[7] http://www.smarttea.org/
[8] http://mytea.org.uk/

- **Biosim** - the 'Biosim' grid project[9] was concerned with running very large simulations and then storing the trajectories of those simulations so that they could be manipulated. Key to both Biosim and eBank is the same idea of large repositories of data tagged with metadata to enable easy access, handling and extraction. The eBank team believed that the experience gained in these kinds of projects, about how to manage large amounts of research data and information, has potentially far reaching consequences and benefits for both research and teaching, particularly as electronic information becomes more prevalent.
- **EMalaria**[10] - was developed as a site for making electronic research outputs available for teaching purposes. Although primarily aimed at School children, the eMalaria site was used by undergraduates at Southampton as part of the eBank teaching pilot (see the section 'The student experience' for further details). Typical teaching scenarios for using eBank and eMalaria included getting students to retrieve small molecules from the eBank system and then putting them into the eMalaria site in order to undertake a range of manipulations and investigations. Such a linkage between eBank and eMalaria provides a demonstration of a means of explicitly integrating research activities with learning and teaching practice.
- **ePrints**[11] - is part of a wider movement concerned with making research publications and other research outputs electronically available, rather than locked into proprietary systems.
- **JISC Digital Repository projects** - There are also links with a number of other digital repository projects, such as the Claddier[12] project at Southampton and the GRADE,[13] and StORe[14] projects. Also related is the SPECTRa[15] work being carried out at Cambridge by Peter Murray-Rust and colleagues describing chemical data.
- **Electronic lab books** - Led by Jim Myers, the 'Electronic lab books'[16] project was part of a broader project, 'Collaboratory for Multi-scale Chemical Science'.[17] The project focused on labelling up experimental data. Other work includes developing mechanisms for storing reaction kinetics data. This work opens up the possibility of the development of repositories of research data across different areas of Chemistry, for example the potential to create an eKinetics equivalent of eCrystals.[18] What this emphases is that:

> Repositories of higher level data have now started to appear, driven from the US. [Int 1]

- **International projects** - the 'reciprocal net' initiative[19] from the States is more of an educational tool. It is a digital libraries project funded under the educational strand of NSDL. Other international projects of note include DARENET[20] from

[9] http://www.biosimgrid.org/
[10] http://emalaria.soton.ac.uk
[11] http://www.eprints.org/
[12] http://claddier.badc.ac.uk/trac
[13] http://edina.ac.uk/projects/grade/
[14] http://jiscstore.jot.com/WikiHome
[15] http://www.lib.cam.ac.uk/spectra/
[16] See for example http://collaboratory.emsl.pnl.gov/
[17] http://cmcs.ca.sandia.gov/
[18] http://ecrystals.chem.soton.ac.uk/
[19] http://www.reciprocalnet.org/
[20] http://www.darenet.nl/

the Netherlands and ARROW[21] in Australia. COD, [22] the Crystallography Open Database, enables members of the community to upload files to a central repository. It is primarily concerned with published files, no data chain is available and it doesn't give access to the raw data. Projects like COD do however raise issues about the relationship of solo repositories with more mainstream offerings and national data centres or repositories.

- **Digital library research -** eBank also had connections with the JISC's development of the 'Information Environment' and members of UKOLN were involved in this work from a digital library perspective, such as specification of an 'Information Environment Architecture' (Powell, 2005). Parallel work to specify a research environment was also being done at this time as part of the GRID developments (See for example, Berman, Fox and Hey (eds), 2003; and Atkinson *et al.*, 2005).

eBank was the first of the JISC data repository projects and was the first one that sought to link data to derived articles. Other digital repository projects (as part of the JISC digital repository programme, data cluster)[23] followed, such as Claddier, GRADE and StOoRe, looking at linking data to derived references and although these other projects were looking at holding data and making it available, none of them were looking at using the semantic labelling that was envisaged as part of the work of eBank and related projects. This extension of eBank in terms of semantic labelling and RDF work is an important area for future exploration.

Key success factors and outcomes

The culture of sharing is particularly strong amongst crystallographers so they were a good community with which to test out the vision inherent in eBank. Embedded in this, is the fact that part of the aspiration behind eBank was to provide a clear demonstrator which showed researchers the potential of this approach, by showing them how easy it was to find and use data, which could then be used for a multitude of both teaching and research purposes. The project achieved five inter-related achievements:

- The data repository of crystal structure data at Southampton as a proof of concept demonstrator, populated with about 100 data sets.
- The metadata application profile.
- The aggregator service at UKOLN which aggregates the metadata about crystal structures in the repositories and makes it available for third party access.
- The aggregator service has been embedded within the RDN PSigate portal, which provides an exemplar of how this work can be linked in the learning domain.
- Demonstration of the potential of use of eBank-type material in a teaching context.

One interviewee stated that one ingredient for success was to ensure that the focus of the project was on production of a good exemplar which could demonstrate the potential of the ideas inherent in the project.

---

[21] http://www.arrow.edu.au/
[22] http://www.crystallography.net/
[23] See http://www.jisc.ac.uk/whatwedo/programmes/programme_digital_repositories.aspx for details and links to projects

> Got a demonstrator there with some data in it… so … that can actually do dissemination … and people can see … and there is not much data, data set repositories so it was very innovative in that sense. [Int 12]

Another success ingredient noted (given the interdisciplinary focus of the work) was to ensure that all participants worked within an agreed existing framework:

> Yes and it is built with the JISC infrastructure… and the JISC information environment in mind. It uses the protocols that are approved by the JISC for the new search services. [Int 12]

Also highlighted as important was the need to articulate the benefits of this approach (ie open access to research data) both to individual researchers (so that they could disseminate their research findings more quickly) and to the wider research community. Traditionally research outputs have tended to be as ‘polished’, peer-reviewed, published data (often in the form of static, journal articles). eBank opened up the possibility of making ‘raw’ research data available – either for individual use, within a shared research work space or more widely across the research and teaching communities. One of the key points, stressed by those interviewed, was that this was about making data easy *to find* and easily *available*, enabling researchers to think of different ways in which they might then use the data. This however does potentially change the perceived ‘value’ or worth of such data – if it is ‘open’ and easily accessible. It also raises questions about an individual researcher’s ‘moral obligation’ in terms of making their research information available and shareable across the research community. Indeed the university sector fundamentally differs, it could be argued, from other Business sectors, in that its primary ‘product’ is the research and teaching outputs of individuals, their intellectual capital, rather than any tangible, physical product. This view of ‘information’ and intellectual outputs and its worth is an inherent characteristic of universities (both in terms of research and teaching) and is fundamentally different from the perspective of those in other sectors – such as commerce (Oliver *et al.*, 2007). It is unclear yet what the impact might be of making such intellectual capital more explicitly available in terms of its perceived value and worth.

Working with crystallographers was a key success factor because this was a community which already understood the value and importance of sharing information; however they were still in need of good technical solutions to achieve this. Another key success factor to emerge was the fact that the team was interdisciplinary in nature – bringing subject specialists and librarians together to tackle the same problems (see section ‘collaboration and interdisciplinarity). The importance of this type of collaborative approach is increasingly being recognized by funding councils as the number of types of multidisciplinary and interdisciplinary teams needed to address complex modern problems increases (see Taylor, 2004 for a recent example in e-learning specifically). Initially however there was a mismatch of expectations and perceptions, but gradually the different specialists began to understand and learn from each other:

> A lack of appreciation on our side as to why they were worrying about particular details which didn’t seem to make any sense… I think both sides realized that you couldn’t apply traditional library models to this data but we also realised why you needed to apply some rigour to this….[Int 1]

Underpinning this was the notion of these different communities working together on a shared vision – making data opening available:

> The whole idea … that you made your data available, you haven't given it to anyone, anybody, you've just made it available. Its still sitting on your site and you control it – but you've made it public. [Int 1]

Inherent in this is the issue of ownership of research output and the tension between academics having control instead of the journal publishers. eBank developed a close relationship with the ePrints initiative which is leading the movement to open access of research publications (See www.eprints.org and Hey, 1997). The concepts and ideas in eBank are now being integrated into ePrints developments. This provides a good example of the project feeding into a parallel initiative.

However it is important to note that the central vision of the project was not without its complexity; having potential huge impact on research practice and the research data cycle, leading to changes in the roles of the different stakeholders (students, tutors, researchers, publishers, professional bodies, etc) involved. The project therefore needed to maintain a careful balance of pushing the vision forward, of evangelising, whilst also being aware of and taking account of different stakeholder perspectives. It is evident that there are a range of political complexities and sensitivities associated with this type of development, not least because the essence of the project related to issues of ownership and control, and potential intervention/changing of standard establish practice and ways of doing things; the project needed to steer a careful path through this complex maelstrom:

> We have successfully steered our way through quite a complex area – which tends to get quite hot under the collar…open access publishing, open publishing of data and the way we have managed to separate out a conventional journal article from the underlying data and data separate, and publish the data via a separate open route so we've just about managed to steer our way through fairly complex territory … without treading on anyone's toes. [Int 2]

Despite these complexities the project did appear to achieve a remarkable degree of success. They were able to put in place mechanisms for getting the relevant stakeholders on board by understanding and working with their different agendas and finding a compromise that suited all. Working directly with the publishers and professional bodies was of particular note as this gave weight and credence (and sustainability) to the project outcomes. As the following quotes illustrate, financial commitment from the professional bodies to involvement in these activities is clear demonstration of the success and impact of the project.

> [we have] come to a very good agreement. Database and publishing people everything is reconciled.. are ok with this. We have crystal community body people on board, Cambridge and IUCr crystallographic society - [we] presented this to their executive committee who have accepted it and now its in their business plan that this can be done. [Int 2]

> It can actually be a little bit unusual for this informatics projects to come up with a working deliverable let alone one that looks like it will end up being embedded into community and into research practice so a bit unusual in that respect. A lot of these demonstrators are just sitting on selves gathering dust …this is been pushed forward both by the digital libraries people and by chemistry and crystallography… and in one form or other will become part of daily use. [Int 2]

Dissemination mechanisms

The purpose of dissemination was threefold: providing information about the project and its activities, evangelization of the broader underpinning aspirations and, related to this, a contribution to change management (in terms of how this might lead to changes in research and teaching practice).

The project team used a variety of standard routes to disseminate the findings of the project to the wider research community – both nationally and internationally (particularly in North America and mainland Europe). Conferences and word of mouth not surprisingly were deemed the most valuable communication channels, but also of note were journal articles and conference posters.

> Dissemination has been very thorough to various …communities because of the make up of the team. It has quite a good reputation with external people. We've had quite a good acceptance rate with publications and conferences. [Int 12]

Future ideas for dissemination include applying these ideas to other research communities and spreading the vision of electronically manipulating and sharing research data. This is a sensible strategy and follows Moore's adoption curve of moving from innovators to early adopters (Moore, 1991). A sensible strategy is to work on addressing specific research problems and community needs – for example researchers who need access to large amounts of data, or mechanisms of facilitating peer evaluation of data. The interdisciplinary nature of the team meant that it has been possible to target a wide range of different potential users and communities:

> Right from the start because we were a multi-disciplinary team, we had digital library people, we had chemists, also we had computer scientists, we have been able to address at least three different communities through the dissemination process, so we have spoken at, some of us have spoken at Chemistry conferences, some of us have spoken at digital library type conferences and some of us have spoken at more computer science type conferences…. So we have spread the word right across the canvas if you like.
> [Int 10]

The interviewees listed a range of communities who were interested in the findings for a variety of different reasons:

> We do get comments…. that this is of relevance, of interest to the scientific community because we need to hear more from the digital library community…[Int 12]

> [Interest to scientific or digital library community?] No I think its both because as I said in the digital library community there is not a lot of scientific data there, out there…. [Int 12]

In addition to the traditional academic conference and paper publications, the consortium were also able to tap into the popular media (including an article published about eBank in the Times Higher and reference to the eBank project by the Director of e-Science, Tony Hey, in an article in Science) which helped to broaden their coverage and in particular was valuable in terms of raising the awareness of the project with senior managers and policy makers. The project held a series of consultative workshops and meetings with some of the key professional bodies – notably the Royal Society of Chemistry (RSC), the Cambridge Crystallographic Data Centre (CCDC) and the International Union of Crystallography (IUCR). The involvement of these wider stakeholders was seen as key to ensuring wider adoption of the eBank ideas. This has been of significant benefit in terms of the project being taken seriously more widely and ultimately in terms of the vision being applied more generally. Engagement of these key stakeholders was seen as important, not only in terms of making them aware of the project, but also in terms of them endorsing it and potentially acting as champions:

> So the editor in chief of the International Union of Crystallography, not only the standards setters but they publish about 8 journal articles (sic), about 8 journal titles as well. And the editor in chief is now promoting this as the, as one of the ways forward to crystal structure publication. So if we've managed to get him that far I think it is going to be a reasonable success... And having publishers actually promote it for us that helps. [Int 2]

Getting beyond basic dissemination to actually changing practice is notoriously difficult, but a key aspiration of the eBank project was about 'changing practice through example', ie by showing the research community what was possible. This was a fundamental driver in terms of producing the eBank demonstrator, aggregator service and the link to PSIgate. As a portal, PSIgate is in essence a one stop shop which links through to many resources in the physical sciences so:

> they could be quite pivotal in helping us making links between eBank data and related data available on the internet. [Int 2]

In terms of core communities which need to be targeted in terms of taking up the findings of the project, five core communities were cited and discussed by the interviewees: students, tutors, researchers, librarians, and publishers. For all of these communities the concept of 'changing practice through example' is relevant, however additional factors relate to specific community needs. An illustrative example was provided by one interviewee with respect to librarians. The library community have been very successful at developing a niche speciality in terms of applying their skills in the e-learning domain. However they are currently less comfortable about how to collaborate in the e-research domain and indeed a lot of them are daunted by this prospect. However the need for a new hybrid profession in this area, equivalent to the 'learning technologist' role in e-learning (Beetham, 2002; Timmis, 2003, Gosling, 2001) is evident, as a recent NSF/NSB report on long-lived data collections acknowledges by identifying 'data scientists' as a new role/breed of professionals (NSF, 2005). Conole *et al.* (2007) as well as Beetham and others (*ibid*) have described the way in which technological change leads to organizational change and changes in practices, individual roles and identities. What is evident is that new roles continually emerge in terms of support, as we develop more sophisticated uses of technology.

> And I think there are some interesting possibilities there looking at the Librarian training and curriculum, library schools. To look at potential of equipping them with those sort of skills to enable them to work closer with the scientists and researchers to help them manage and add value to their data. [Int 10]

Some similar process may also be needed for teachers, who could be targeted through professional development programmes. An added complication here is the distinction between solely teaching focused staff and those with a teaching/research remit, where the later will be equipped with the necessary research data handling skills as part of their research training, whereas the former won't. This may be a particular issue in the FE sector:

…worried about the learning tutors, the lecturers because once again they may not have those sort of research skills, the data handling skills because those would be held by the research staff and whilst some do both, not all do. And … there may be differences between HE and FE. [Int 10]

Young researchers and scientists were also highlighted by one of the interviewees as an important group to target. In fact there are issues in terms of how far back in the educational cycle one needs to consider, which raises issues about skills integration and training across the whole educational spectrum:

We need to reach the young scientists, the people doing PhDs, or undergraduates actually .. its much earlier than that, we probably need to get to them at school. If we think about the vertical learning environment, at some point we certainly need to think about how we equip them with data handling skills more effectively and with data analysis skills so that when they come into college and university and this stuff is going to be out there, they know what to do with it. Now I suspect that will happen quite seamlessly actually… [Int 10]

Barriers and enablers

What might be the longer term impact of projects like eBank on institutions? In terms of the division between teaching and research? In terms of the balance between individual/central ownership and control of research data? What does it mean to be a librarian, a researcher, in this new context? The barriers cited can be grouped as essentially either cultural/social or technical.

[The] Main barrier is a socio-political one - its a matter of a change of cultures, some people are embracing that as I have said, others are abit stand off-ish, you know they like the mystique of publishing and that sort of thing. [Int 2]

In terms of cultural/social barriers there were a set of issues raised by the interviewees about ownership of data and/or community sharing. One interviewee posited that whether a researcher is happy to share their data depended on the perceived value of the raw data, the length of time for which it might have value and how reasonable it was to make that data available. They went on to suggest that some researchers may not actually want to release their data to the wider community because of commercial or competitive advantage. There may be concerns because they don't feel the data is 'robust' or 'comprehensive' or they may simply want to release some, not all of the data.

The approach adopted by the eBank project was about making the research process more explicit, hence exposing previously hidden processes and practices. But what does this exposure actually tell us about the research process? Is it an accurate reflection of actual practice? Who is this information for and how might it be used? There are parallels here with concerns about the use of technology for teaching (Nobles, 2002) and the move from the essentially private, hidden teaching practice of the classroom to rarefied, exposed practice through lectures on the web, podcasts, or through teacher contributions in discussion forums.

And we are opening up rather than just presenting a result which is what's done currently in journals we are opening ourselves up 100% we are showing this is how we have arrived at this result so any of the magic-ory under the bonnet is exposed which I can foresee there been a few people ,,, hiding how they have done things. [Int 2]

The ideas embedded in eBank have the potential to fundamentally change research practice. However, research communities vary in their approaches and research working practices and these will impact therefore on how successfully ‘eBank-type’ ideas are taken up. The appropriateness of making research data opening available is also dependent on the type of data. For example making laboratory books available is more contentious than other forms of laboratory data, because they often contain ‘rough’ information and thoughts, generated in situ in the laboratory; ‘incorrect’ information that researchers, not surprisingly, might be reluctant to share. In cases such as this it may be more appropriate to provide some kind of filtering mechanism to the data. Having data available in a repository and/or making it available publicly are all different degrees of access which need to be considered by each research community in terms of their core values and in particular what they feel comfortable with.

The technical barriers cited by the interviewees were those associated with: getting different types of data into a schema and/or a repository, decisions about the type of repository, as well as technical issues around developing, running and maintaining repositories. For researchers not knowing whether or not a repository is going to continue to exist over time was cited as a potential major barrier. One interviewee suggested that host institutions needed to provide some form of formal commitment to repositories, offering a notional ‘commitment’ figure of five to ten years. There are also associated issues about the attitudes of journals and publishers to this new form of making data available, for example what is their policy on agreeing to provide electronic links from published articles back to the original data stored in an institutional repository? In general there are real issues with initiatives of this kind in terms of the level of awareness amongst the different stakeholders – or rather lack of it. Interviewees felt that until researchers were aware of the possibilities and convinced of the benefits, they will be unable to take full advantage of initiatives like eBank and unwilling to change their practice. A specific technical issue cited concerned tools development.

> …we have got the architecture to get, store and retrieve, but it’s the educational tools which could double up as research tools as well [which are needed]. [Int 2]

> So it’s the tools to link to related articles or objects and half the problem there is that there aren’t that many areas that are openly publishing, certainly not by this mechanism so we have to use some trickier to try and make those links. [Int 2]

This demonstrates that eBank is only part of a larger picture in terms of how the whole research cycle might be transformed.

> Well first of all we have to have some critical mass….. So we definitely need more critical mass of data …because at the moment its not clear to me at least what the plan is for making it incremental … not clear where the funding would come from to keep populating [And] until there is a number of different repositories we cannot really, can’t evaluate what we have done and whether its applicable to others until that happens and you cant build cross-search services across them. [Int 12]

Nonetheless the project has provided a very valuable starting point; a means of showing what was possible and a means of the different communities thinking about how to take this forward.

> So we’ve got a little something, so that people can actually look at and start to make suggestions… until they really start to implement more of these things .. people wont use in their day to day work. [Int 12]

Other barriers, cited by the project team, included: barriers to making data open and accessible, issues about data storage and maintenance, institutions lacking the necessary infrastructure or support facilities to develop and maintain data sets. Funding barriers (lack of finances to support the development or maintenance of a data set; competing institutional demands on resources) were also cited. Local agendas and politics are also likely to influence success, as is evident in other examples of large-scale technological intervention (Conole *et al.*, 2007). Finally, as with other examples of open access (Hylen, ND) or reuse initiatives (Littlejohn, 2004), there may be a host of legal and Intellectual Property Right (IPR) barriers.

> For an individual researcher it's the worry of having to get their hands dirty with the software I would think and for the libraries it's the fact that they are now forced to come up with a data preservation policy. .. And even here where our library is kinda forward thinking they have had to sit back and think right ok if we do take this on board we are setting a precedent here where we are storing research data essentially which is not something they have been used to doing and its unsure territory. [Int 2]

This raises issues about the impact of projects like eBank on changing roles and organisational structures. It points to the blurring of the boundaries of ownership and control and of what constitutes 'research' and 'teaching'.

Other barriers cited were those around the development of a shared language and identifiable benefits. There is still a lot of work needed in terms of developing a shared understanding and practice:

> I think that people need to agree on things like the terms used, author, how to format author names, author…. Vocabularies of this kind of data. [Int 12]

And of getting the benefit across:

> … convincing people of the value of doing it to start with … [Int 12]

This suggests that there needs to be a significant cultural change in order to both convince people of the benefits of these kinds of innovations and more significantly getting them engaged enough to want to change practice. This further evidences the role and importance of tangible demonstrators in such projects to show the potential. However there is a delicate balance needed in terms of iteration between demonstrator and practice – individuals will not necessarily know what they want to do until they see what's possible and that might not be possible until all the other pieces of the 'future technical' vision are in place.

> And we've also talked in the project about whether there needs to be …to local practice and workflow, so then before the training we need to do some finding out exercises. [Int 12]

Another feature to emerge was that a lot more is needed in terms of understanding local practice; research practice is complex, messy and difficult to articulate. Furthermore there are also huge local variations in practice and different sub-cultures within disciplines, institutions and in individual research groups – a very similar procedure may be carried out entirely differently in two different research groups.

> I think the investment in time to do the deposition which seems to be an issue ….with publications in many fields and Chemistry does not actually have a good history of giving

> open access (if you compare them say to the physicists who were the pioneers) so… what [would be] recommended for the publications is a mandate which and that would be something that would apply to the data as well. [Int 12]

Many of the research barriers outlined by the interviewees were also thought to be equally applicable from a teaching perspective. One specific barrier for learning and teaching cited was the need for tutors and students to develop the appropriate range of 'e-skills' to design and deliver (from the tutor perspective) and to use (from the student perspective) resources like eBank. There are a number of other issues suggested with respect to the student perspective. We are dealing with a new generation of students, who are comfortable with technology and don't see it as an innovation:

> For students its new technology but they embrace new technologies faster than the growns ups these days so .. once the tools are there I don't thing there will be many barriers at all, in fact they will probably take to it easier than the current one, which is currently very disparate tools, its not very joined up … how the research cycle works. [int 2]

This view echoes the findings emerging from the JISC LXP study which examined students' experiences of the use of technology (Conole *et al.*, 2006). Nonetheless there are barriers for students in terms of the current status of the tools, which are not quite integrated or seamless. The tools are still at the level of there being a need to understand the underlying 'mechanics' and almost a conceptual model in order to be able to apply them. In a sense this is like the early days of computer use when users needed to have an understanding of the hardware and code; in contrast nowadays users have intuitive interfaces designed to support specific user tasks, which shows that the conceptual models have been fully developed and integration achieved.

> The lifecycle of data and how it fits into the bigger picture is not clear so we have to teach them lots of different bits and pieces and showing them how everything links together is almost conceptual they find that pretty difficult to grasp… You know using software and data mining tools, general searching, you know that kinda stuff they are kinda ok with it but conceptually how it all plugs together is abit difficult. They are quite happy to get their hand dirty at the key board. [Int 2]

Many of the above can actually be turned into enablers. For example, if an institution does have an appropriate technical infrastructure, support models and structures in place, this is likely to facilitate uptake. Warburton (2006) describes a similar set of barriers and enablers associated with institutional uptake of e-assessment. Professional bodies can also provide support in terms of agreeing guidelines and standardized mechanisms for publication.

Many academics are worried about potentially litigious outcomes of making data publicly available (OECD, 2003) and a number of the interviewees stated that clarity over legal issues and IPR would make a significant difference in terms of the broader uptake of projects like eBank. There is evidence that publishers are beginning to see the importance of this and are working with others to develop appropriate guidelines (see for example SHERPA,[24] and the associated RoMEO[25] and JULIET[26] sites). The Research Assessment Exercise (RAE) as a driver is also important, in particular in

[24] http://www.sherpadp.org.uk/index.html
[25] http://www.sherpa.ac.uk/romeo.php
[26] http://www.sherpa.ac.uk/juliet/

terms of what might constitute as research worthy in the future. A significant enabler would be if future RAEs looked beyond the traditional paper, peer-reviewed articles and accepted open access materials and data sets as valid research output. This raises fundamental issues about what constitutes and counts as a research output, as well as wider issues about the role and worth of the peer review process and the overall quality assurance of research. JISC is acting as an enabler in terms of awareness raising – both in terms of the possibilities and the associated issues with this kind of research and development work. Of particular relevance here for eBank is the JISC's digital repositories programme and its capital funding programme.

Getting institutional and/or departmental buy-in was also considered crucial by those interviewed. The role of champions was also seen as critical – especially champions who are well regarded by their peers. Another enabler cited was appropriate awareness raising, promotion of best practice and associated support materials. These barriers and enablers are similar to those which have been identified to support successful implementation of other research and educational innovations (Conole *et al.*, 2007; Salmon, 2006).

A further concern raised was about ensuring the embedding and sustainability of such innovations beyond initial uptake. One interviewee suggested that there was a need to demonstrate not only the technical sustainability but also the economic sustainability stating that researchers and teachers are unlikely to invest in depositing materials in a digital repository or spending time thinking about how the materials can be used for either teaching or research purposes if they lack confidence in the longer term viability of the data set (will it still be there in five years time? Will it still be possible to access the data?).

One key driver identified by the interviews was the importance of working with the publishers, and in particular of understanding what motivates and drives individual researchers:

> Or the other aspect that we've been moving forwards is agreement with the publishers that rather than submit the data to them directly, from the publication you can link to data in a repository, if that took off that could be quite a major driver for people to have their data locally and linked from the official publications and especially if that made the publications turn over faster … for example. [Int 12]

This quote illustrates that there are a range of issues in terms of ownership and control, but points to the benefit of moving towards a distributed and flexible system. What is evident from the interviews is that the barriers and enablers are multiple, complex and inter-connected:

> I think the barriers need to be addressed from different angles really all at the same time. [Int 10]

It was suggested that cultural differences also have an impact on uptake – the degree to which different individual researchers and communities buy into and take up the concepts depends very much on local practice and agendas. Geographical variations were evident in terms of the interest in eBank and its underpinning philosophy. The Australians it appears were very enthusiastic about the concepts, as were most of those the project team encountered in the UK. However the Americans appeared to be a little more guarded and it has taken more time to get the concepts and benefits of eBank over to them. These Geographical differences are in part due to the different role and status of crystallography in the different countries and the historical context within which developments like this have occurred. It may also perhaps be in part to do with more ingrained cultural issues and differences (see Hutton, 2002 for a comparison of UK and US perspectives on economy and Conole (forthcoming, a) on an international comparison of policy drivers and their impact on practice).

> There is a more, a bigger open access community in Europe and Australia I think as opposed to the States. The States are only just getting on board with the programme if you like Their cyberinfrastructure programme is only just taking off and that is all about Grid computing, its not really about open access. Whereas places like the UK and Holland are huge. And even though that's not you know .. digital libraries … and that sort of attitude does seem to filter down into other departments if you like… much more accepting of this way of doing things. Maybe its because we are seeing more repositories in these Geographical regions or for whatever reason. [Int 2]

This highlights the importance of cultural issues and the importance of awareness raising and providing test examples to push up against existing systems, which can demonstrate potential to move ideas beyond the earlier adopters. It also illustrates the importance of getting endorsement from what are seen as the key authority figures in the area – in this case the publishers and the professional bodies. These aspects of the project also highlight the importance and influence of wider macro-directions (influenced in turn by socio-cultural and historical local perspectives – see for example Conole and Dyke, forthcoming and Conole, 2006) and their impact on practice and how this can directly hinder or help particular small (in the scale of things) interventions.

One interviewee felt that there was also an issue in terms of who the end users might actually be, who would be interested in the data made available through these kinds of services? They also speculated to what extent this concept might be applicable to other areas of science.

> I think its still quite a question which communities would use the crystallography … because we've really talking to the people who produce the crystallography not the people who actually use the results an applied area. So far we haven't had much contact with those people … there's not been an in-depth study of their requirements. [Int 12]

Conceptual models and pedagogical issues

A feature to emerge from the evaluation was the importance of the conceptual models underpinning eBank and how these models help conceptualise and shape the project vision. What is also evident is that development and in particular clear articulation of such models requires time – to develop and refine the language used, establish the models and to think about how they can be applied. A number of models were evident (figure two).

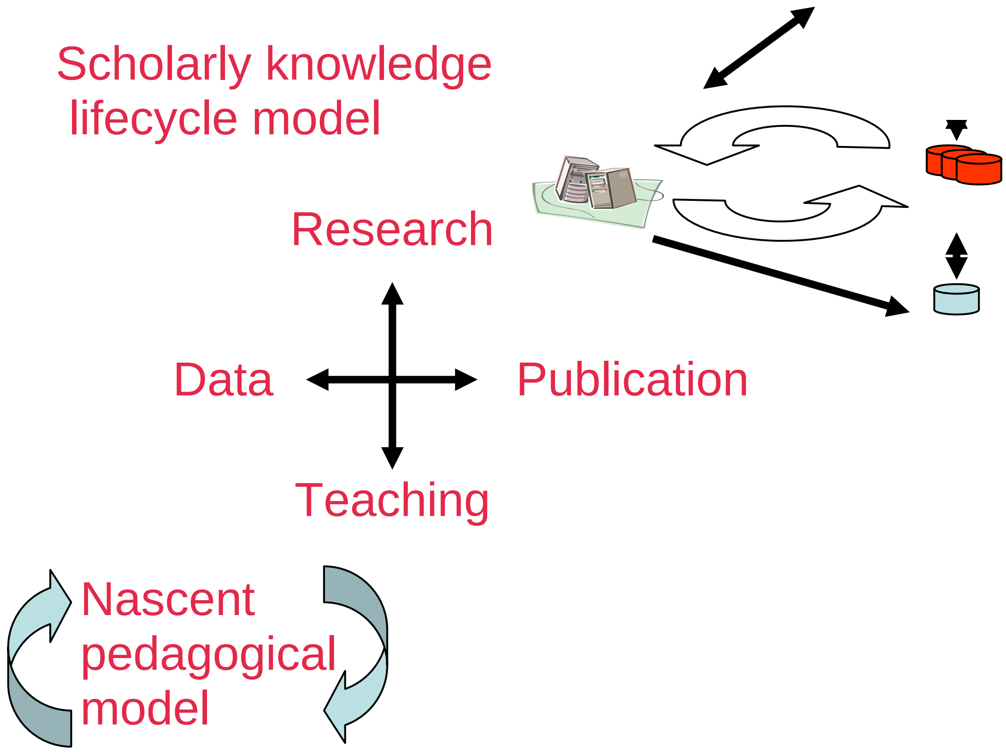

Figure two: inherent models underpinning eBank.

Firstly there is the scholarly knowledge lifecycle model (Lyon, 2003). Secondly there is a model around the notion of providing the link from the data through to publication and vice versa and thirdly there is a nascent model about articulating a set of pedagogical approaches which might be applied to capitalise on the potential of this approach. More difficult to articulate but also evidently important is the issue about shared language and the evolution of definitions as the consortium worked towards developing some shared understanding. Part of the issue was finding 'labels' and ways of describing some of the abstract processes involved:

> There was the concept of thinking about data as a collection – collections of images, collections of books, but thinking of data as collections that you could describe and make available [Int 10]

These concepts - 'scholarly knowledge cycle' (as a means of describing the process) and 'data as collections' (as a means of describing a particular aspect of or property of data) – can be viewed as examples of how the researchers were struggling with, attempting to articulate and make sense of the problem space they were working in. Holyfield outlines a similar problematisation of conceptual models to describe technological developments in relation to the use of the concept of 'Managed Learning Environments' (Holyfield, 2002).

In terms of the application of the ideas in eBank for teaching and learning purposes there did not appear to be an explicit pedagogical model, although those interviewed did have some specific ideas of how the outputs from eBank might be used for teaching. The development of new pedagogical models which harness the potential of technologies is notoriously difficult and work in this area is still in its infancy (see Conole *et al.*, 2005 for a critique of existing pedagogical models and their impact), however without a clear explicit pedagogical model it is difficult to guide teachers in designing new activities which make use of technical innovations. A book does not encourage problem-based or reflective learning; it's about how this is used within a structured learning activity that enables these approaches to be adopted. Similarly the outputs from eBank in isolation are of little use for teaching purposes unless they are applied within a clear pedagogical model to achieve specific learning outcomes. There is a nascent pedagogical model implicit in the scholarly cycle, what's needed is to turn this into an explicit model which can be used in teaching. The development of specific learning scenarios or patterns (Goodyear *et al.*, 2006; McAndrew, 2006) built around the use of eBank data might be the basis for developing a pedagogical model.

Furthermore what would 'count' as an innovative pedagogical model in this respect? For the crystallography students, access to material electronically is no longer innovative (they are used to and constantly exposed to a myriad of electronic data to support their studies). And students involved in crystallography are particularly comfortable as research data has been available in electronic format for a long time in crystallography. To innovate the pedagogical model would need to draw out the new learning opportunities access to data in this format might provide for the students; ways in which it might enable them to do things such as interrogate and manipulate the data in ways that they haven't been able to before.

However, it is worth noting that other areas of Chemistry and other Science disciplines are generally not so technically literate as the Crystallographers and hence providing access to data electronically in a teaching context would be valuable. There is therefore a balance between about the degree of readiness of a discipline to take ideas forward and apply them versus the level of technical expertise – if they are already using technology extensively, it will take more to demonstrate innovation.

The student experience

Embedding eBank material in teaching

One of the three aspirations underpinning eBank was to explore how making research data electronically accessible and linked to published references could be utilised in a teaching context. To explore this, the consortium trialled the use of eBank material in the MChem course at the University of Southampton, within the final year module on Chemical Informatics (Chem 6016). The module accounted for about 12.5% of the overall course and consisted of 24 lectures plus workshops and exercises. The co-ordinator was Jeremy Frey; the other tutor was Jonathan Essex. In addition there were invited guest speakers and speakers from industry.

Typical activities used in the course included searching the eBank database for molecules and then getting the students to make links with the Cambridge Crystallographic Data Centre. A key pedagogical aspiration of the module was to get students to understand and manipulate chemical research data. The module was designed so that students had ample opportunities to work collaboratively in workshops, but they also had the opportunity to work through problems individually. It was an optional module, running in the second semester until the last week of May. The course was accompanied by a Blackboard site which contained course material and was used for course administration and dissemination. All the students had previously completed a six-month work placement in industry. The face-to-face components of the course ran on Tuesdays and Wednesdays. During 2005-2006 the course had 15 students registered. In the previous year it was also combined as an option for third-year students, which increased the numbers taking the course. Some postgraduates also did the course (all postgraduates need to get 120 credits during the first year of their PhD, which is the equivalent of 4 units).

Chemistry students at Southampton are exposed to technology at other times of their programme. Of particular relevance is the use of technology in the 'IT for Chemists' module which is a first-year course (Chem 1004, Simon Cole – co-ordinator). This includes some use of databases as well as the eBank data. In general there are small elements of IT use throughout the undergraduate Chemistry course, including use in lab classes – particularly in the Physical Chemistry laboratories. The team have discussed the potential of extending the use of eBank to other courses, for example the Crystallography postgraduate course.

The students were provided with direct access to the eBank material via the Blackboard VLE site associated with the course and they then used this as a basis for some of their project work. This use of eBank type material in a learning context proved a useful starting point and seems to have been successful (students were able to access the data, use it and found it useful), but it is also evident that this is very much an initial exploration of its potential for learning. What is unclear is exactly what benefit students will derive from having exposure to such data, how they might use it and from a teaching perspective what type of structured learning activities need to be developed which can bring out the kinds of aspirations (such as developing critical and evaluative skills, experiential learning, etc) envisaged as potential possible applications of this to learning. Also providing students with a link back to data underpinning publications offers potential, but as yet, unclear, unexplored potential.

> I think we are less clear about the value of the linking from the data to the derived publications and what we can do with that information I think there is a huge amount of work to do there. [Int 10]

There are evident opportunities though which need exploring (see future research) as the following quote illustrates:

> And one can think about how you might further develop your courses around the ability to do that and how you might exploit that possibility. And in the future you can envisage that new parts to courses may develop, courses may develop in different ways and the learning process may change, as a result of being able to access the crystal data in eBank case but other data in other disciplines and there may be differences in disciplines in how that works and that's another interesting area. [Int 10]

Rationale for taking the course

Students appeared to have a good understanding of the focus of the course:

> An intro to what Chemical Informatics is, what it does and the basics of like doing some sort of computer based analysis of chemical structures and … in this case the binding sort of enzyme bindings but its looking to try and see how you can use the data to improve a system without having to go and do the wet lab chemistry. [Int 8]

> So if have a library of molecules already synthesised then you can test that against a new receptor or something that you have found so that makes like a lot easier for synthetic chemists. [Int 9]

The unit tended to attract those students with a particular interest in the technological aspects of Chemistry. Student descriptions of the course illustrated some of the reasons why they were interested in taking the course and highlighted a number of key pedagogical benefits, such as: providing students with access to real authentic data, the ability to build skills and understanding progressively through interrogating and interaction with the data and the ability to learn through doing by manipulating data:

> First looking at simple models, modelling the data and then the more complicated ones that they use in Industry [Int 4]

> Use statistic methods to run experiments in the workshops… use excel to perform multilinear regression .. for making models for the presentation we've just done…its all about data gathering and plotting, plotting what can be left out, what can be left in…[Int 6]

The course was designed so that there was a progressive building up of complexity and use of real and authentic data, clearly linked and of relevance to work-based learning, which was timely as these students had just completed their six-month work placement.

> Its quite an interesting course, its quite different to a lot of courses I've done… basically its comprised of …you had set lectures, then you had workshops and you also had this kind of an assignment which was very much kind of do it yourself. Its like what was a project … at the time when I first got given it I didn't think it would take up as much time as it did. I mean really it did take up a lot of time. [Int 5]

Some of the quotes from students are very encouraging, with the students echoing some of the key underpinning aspirations for the project (outlined earlier in the section 'project aspirations'), articulated by the project team as illustrated earlier:

> There were several parts to the course – We started off with how to get 2D and 3D representations of molecules onto a computer using a one-dimensional format, a SMILE string …so just ways of like getting data into a format so that it can be easily shared between different computers or different people without having to change lots of things [Int 7]

> Another quite nice section of the course, involved databases … searching databases and getting more use out of databases and how the best way to go about this and also how to put information into a database so if you come up with say a crystal structure ..how to get that into a format that the database will accept so that its easily accessible by lots of other people. [Int 7]

Students' ICT skills

It was clear from the interviews that all the students were comfortable with using technology and that many already have fairly good IT skills. In a sense this is not surprisingly, Chemical Informatics is optional and is one of the more computer-focused units on the course and hence is likely to attract students who have an interest in and initial skill in computing.

> I think I probably did more work on building the website than I probably needed. I really enjoyed the course from the point of view of doing it online… I like web-based work I've done. I spend a lot of time on the Internet.. this was what attracted me to the course. [Int 4]

> I taught myself [html] over the years … started building websites in school, stopped for a couple of years and just started up again. [Int 4]

From the interviews it was evident that the students use technology a lot to support their learning, with numerous references to the ways in which they were using email, the internet and other technologies. Most were self taught in terms of their IT skills, which were variable - some are evidently more experienced and interested in the technology, others adopted a more utilitarian attitude. In general the students' attitude was pragmatic – some have taught themselves html, whilst others created web pages using facilities available in Word.

> Apparently you can pop it in Word, prepare the document in a Word page and then save it as HTML … it's a cheater's way… [Int 6]

Students are exposed to a range of technologies in their course – from use of basic Office tools such as Word and Excel, manipulation of research data in databases, through to use of specialised chemistry software such as ChemDraw. In addition many use technology routinely for other aspects of their lives:

> I use it [technology] for typing up reports. Most of the time they expect us to type, sometimes we are given the option but I always type. [Int 4]

> Yes … I play an awful lot of video games, far too much. [Int 6]

> I use the computer, I write as a hobby so I use it to wordprocess, I don't like writing because my handwriting is awful and I type so much faster. I use it for playing games, talking, communication… got friends in various places like Singapore and Australia so keep in contact via the Internet. [Int 4]

The Chemical Informatics module had a number of computer-based workshops where the students were exposed to facilities like the Cambridge Crystallography Data and were able to manipulate and experiment with the data. They were also given general support on experimental design and the underpinning mathematical techniques which they would need to understand and use in order to undertake the modelling as part of the course.

> We used computers a lot …. for workshops and things like that. We have two or three computer-based workshops, one on … hmm crystals … CCD .. and looked at molval. The 3D… what's it called … connection tables and bond distances. [Int 3]

> We also had a workshop on designing experiments, using regression and mathematical principles. How it can help you to … more effectively optimising experiments and things like that… [Int 3]

> There were several workshops, where we searched several databases, one of which was the Cambridge Crystallographic database. [Int 7]

The prior use of technology earlier in the Chemistry course, such as students' exposure to Excel in the first and second year physical chemistry labs, provided a useful grounding for using technology in the Chemical Informatics course.

> I suppose undergraduate physical labs taught use to use Excel… and linear regression and things like that. [Int 3]

> We used to have, certainly in the Physical labs …used to do regression in Excel, we used to have .. but very much now I'm in the fourth year if you don't know it now you ask people.. [Int 5]

> Within Physical labs themselves as you sort of did the labs there're sort of step through guides .. how to do regressional analysis, how to do various techniques associated with the experiments you were doing. [Int 8]

Nonetheless, some did find the technology daunting at first and would have liked more training and support – in particular on using the internet and finding relevant information and using online journals.

> If it's [Information from the Internet] relevant, yes, it depends because Chemistry is such a specialised subject you cant find much online. But certainly with this course I have found a lot more data online than I realised was there prior to the course. [Int 4]

> It was quite daunting to start off with … looking for journals .. the problem with the Internet .you find something and then there is so much [information] and then you need to try and filter it out to get the good stuff. [Int 5]

Benefits

The importance of developing understanding through experience and by doing was evident:

> I think basically for me its being clarifying and really actually now understanding things I've been using for a while. Something like linear regression .. I used without really understanding, whereas now … I understand now its not complete magic … yeh hands on experience … now understand a bit about it, wouldn't say I understand it completely, but given me a better understanding. [Int 3]

> Model building … different ways, different ways you can test your data. And also very good at case studies, all the theory you learn. [Int 6]

Access to data and 'real' results, enabled students to interact and hence develop their own understanding, following the principles of constructivist learning:

> Before the course I hadn't really considered how the computer actually does it but … interesting to see how that works and then there is a part of the course where .. they taught you… how to interpret data and build models .. and that's probably quite a useful part of that project use the model building … its all very well people telling you this is a …peptide but until you actually use it, you cant really visualise it. [Int 5]

They valued the real-life, authentic nature of the tasks included in the course and found this motivating. A number of students commented on the value of being able to access real and authentic data sets online and also online research journals – a number also referred to the benefit of using of their Athens accounts to access material online from home.

Students valued the external input in the form of guest speakers from Industry and found this motivating – stating that it helped to contextualise the course and relate it to real work situations, as well as providing them with up to date input from industry experts.

> They actually had people coming in from Ingenta, giving us talks about how they use model building in their library building. [Int 6]

> Somebody came in from IBM I think and talked about like using websites using XML and stuff like that coz … I taught myself how to do a website, an HTML one, so like looking at someone

who knew what they were talking about, talking about .. ID things like I didn't understand I was doing them and showing like better ways of doing things .. put chemical structures up on the web .. ways that other people can view it and no matter what country or what language .. however they write then they can see the chemical structure the way that its supposed to be seen. [Int 9]

They also cited the value of being able to continue to communicate with tutors, fellow students and friends via email and chat whilst on industrial placement.

The course appears to be unique and quite different from the others students have taken. They commented on the fact that they liked the ability to be able to download and manipulate datasets over the internet and also the ability to see the translation of information from one format to another.

Well basically I've done nothing like it before, so it's the first time I've sort of delved into computing or computational chemistry … quite nice, quite enjoyed starting off with just like a string of data and pop it into say a database, just a flat string of numbers basically and then come out with a crystal structure, which is exactly what it should represent which is quite cool. [Int 7]

Students commented that this was a completely different module to the others on their course which they found interesting and useful as it taught them how to interpret and manipulate data.

Its [the e-Malaria site] very good at changing 2-D structures into the 3-D, … where you can use the graphics to rotate it, spin it, its quite fun. [Int 6]

Use of technology

Technology is clearly used extensively by the students to support their studies and appears to be explicitly required in some of the courses:

[Use technology in course] Yes we use it a lot – we use it, every report we have to do we use the computer to process, whether its organic processing, NMR, physical, processing graphs… and also drawing molecules [using chemdraw?] yes [Int 6]

They therefore use computing to support their study activities extensively, one student commented on the fact that his use of technology had increased since starting the course:

Certainly I think since I've come to Uni I have certainly used my computer a lot more… for report writing, for research.. for everything… [Int 5]

The students use word processing to support all aspects of their study, indeed a number commented that they now find writing by hand difficult and also that the word processor is essential in terms of spell checking. The use of technology has increased as the course has progressed and many students are now constantly online and connected at home, often via broadband.

I'm connected to broadband, constantly online, [that's at home?] yeh. I mean it helps using the Internet, researching, finding stuff out … and I think you can, there was a stage when you could write up your lab reports by hand but now its basically presumed that people do it on the computer, so all assignments, all lab reports, everything is written up [on the computer] … which is good for me because I am pretty bad at spelling. [Int 5]

Seems to get more and more, even through first year …Physical lab work all your analysis … was done using the computer .. so all the way through its being part of … and everything you hand in

> as coursework is typed, its not hand written at all [better isn't it?] particular when you have handwriting like mine!!! [Int 8]

Many of the students said that that they don't particularly use MSN chat for work purposes but do use it a lot with friends. Most of the students stated that they used computers a lot for other aspects of their life – to communicate with friends, to buy things or to play games. Interestingly some of these students expressed a preference to using chat with friends rather than their mobile phone.

> [Use of mobile phone]Not really I usually use the computer messenger. I do have a mobile phone which I do use but not much … turned off at the bottom of my bag! [Int 4]

They commented that they found the Blackboard lecture notes useful as they could constantly refer back to them.

> I did yes [use Blackboard], that had backups of all the documents we received and very helpful.. they put up all the data for our projects…we'd look up this spreadsheet and …which data set was for us. [Int 6]

> Yes you basically had the list of all the lecture notes so you could go back and look at them, maybe if you missed a lecture or if you don't really understand something and we had the list of peptides as well and information we needed for the mini-project. [Int 9]

Some students also appeared to use these as active learning tools, annotating and progressively adding to them to aid understanding. Students commented that the lecturers used Powerpoint to support their lectures extensively, which they print out and annotate during and after the lectures.

> Used the blackboard site more in this course than others – given copies of all Powerpoints for the course and handouts and can make annotations on these to aid understanding. [Int 3]

> They're [Powerpoint handouts} just given to us… in all courses, given to us, a copy…well all the ones that use Powerpoint, yeh. .. all course give out handouts, its either the Powerpoint slides the lecturer is using…It might be to go with say Organic Chemistry they use the board and they also use handouts too… but you have to make annotations. He [one of the lecturers] doesn't put everything on there so you need to annotate. [Int 3]

There were examples of very individual, personal ways in which the students were using the technologies. For example, one student uses Powerpoint as a way of revising – he created 'self test' slides, which he then completes!

> But also personally I like to make a good presentation where I've got formulas and have a test…its quite sad.. as a test.. so I have one slide saying Hess' law and the next slide .. and I have to write it down [you use it to do your own tests?] yes exactly it's the only way I learn. [Int 6]

Others used a range of technologies for specific personal reasons:

> I use the mobile phone as a calendar … knowing kinda deadlines.. and that's quite useful. [Int 5]

> I use it [email] to send my reports home, to look at later, yes I use email in that sense. [Int 5]

Students valued being able to communicate with the tutors via email and appreciated quick responses as well as one-to-one support and input:

[Use of email, for] Contacting lecturers with deadlines, etc That was actually a big thing with my supervisor on the project… he was sending me an email every day .. ooh got to reply .. make sure I am doing this, this this.. its good welcome to check up on you. [Int 6]

They use email to contact tutors, generally with queries about work deadlines or assignments; and it was evident that they appreciated this personal direct contact.

[Email]Being able to communicate with the lecturer, … I emailed [the lecturer] on Saturday night at eight o'clock and he replied at midnight…. I don't know what that says about him or me, but… certainly email.. easy to get in touch with the rest of the class… and him. [Int 3]

[Use email] Sending any problems to tutors, receiving answers, sending ideas to friends .. see what they think of it … obviously messenger-type things for friends. [Int 8]

Interestingly there was evidence of peer pressure in terms of students changing the way in which they used technology:

Just started using that actually [MSN chat] [with friends?] yes yes, I got told to use it … 'I've had enough of phoning you up its costing me a bomb get MSN its free'. [Int 6]

The students appear to use the internet extensively and as a first 'port of call' to find information and get access to online research papers. They also use it as a way of checking understanding by looking up alternative descriptions of terminology on the web.

I use the Internet a lot … to do … research, something I don't understand I might have a look on the Internet and see the different explanations to help …. [Int 3]

And any information you want about the university you just get from the website. [Int 4]

Interesting there is evidence, probably as a result of this course, that the students are now beginning to access and use other research data sets over the internet:

I use the internet for research, erhm, also I've used it to obtain molecules and calculate data but mainly, mainly research. [Int 6]

Obviously there's the e-malaria system should have been used to get practical data we can use ourselves and there are various other sort of web-based programmes that we've used to try and build these models … molinspication I think which… is a property calculator which allows you to… draw in molecules or input SMILE strings .. and then produces the basic properties of the molecules…molecular weights, volumes, and so on and so forth. [Int 8]

Not only do the students see the internet as their primary source of information, but they are using a range of searching and evaluation skills to find appropriate resources for their particular needs, beyond those that they are directed do formally as part of the course:

Some other university lecturers' sites … there was one at .. cant remember, Strathclyde university I think, one of their guys had done a basic summary of like regression and … help you understand the basic principles behind … not just for this course but for the course in general. [Found the course] ….In Google, typed in… linear regression.
[Int 3]

Blackboard seems to be used fairly extensively across the course in terms of general administration and depositing of course materials and lecture notes – however it is used more in this course because it is also used to host the datasets. Students said that they found it useful to have materials available from the Blackboard site and in particular to be able to download the data sets.

> Use Blackboard, not so much for discussion or emailing but for the lecturers but up their lecture notes most of the time and tutorials, workshops, questions and answers go up there. And then most contact is done through undergraduate email. [Int 4]

> [Use of Blackboard] Yeh I did, all the ehm, handouts and lectures and coursework involved with the course was all placed on that and the … dipeptides … easy access, good access. .. I used it more in this course than any of my other courses, that might be the nature of you know… others [courses] haven't got the same information management they don't have to … details on the internet so much, they wouldn't have lots of molecules information… [Int 3]

> [Use of Blackboard] Initially for getting the actually projects from, but also a lot of the data … molecules .. on Blackboard.. instructions for using e-Malaria were also on Blackboard. [Int 7]

> The Blackboard contained all the information we needed to do the project at any point for doing the project so a list of tasks and a list of websites which might be useful, so the NCI database web link was on there, molinsipiration, although I'd already had that from my previous project. [Int 8]

Students used the site in different ways and complemented it with other means of finding out relevant information:

> [Use of Blackboard] A reasonable amount, went on to download hardcopies of everything that was on there, so not sort of regularly but did use everything that was on there at some point and any other problems just sort of emailed Jeremy [the tutor]. [Int 8]

Students appeared to be aware of the distinction in terms of how this course was using technology to a greater extent that some of their other courses, but were also aware of the importance and function of each of the different parts of the course – for example the fact that lecture notes were available on Blackboard was understood as being supplementary resources; not a substitute to actually going to the lectures.

> There was a Blackboard website which had kinda had the assignments and lecture notes, but I went to lectures so it wasn't a problem, but I think they …more. had this e-Malaria site which is where basically you can input your molecules and was a key website and it kinda helped me with some of the background for my project… its aimed at well younger age groups but its still good to get some basic background. [Int5]

They also developed their own learning strategies for using the material, a number for example said that they annotated lecture notes and materials as a way of aiding learning.

> [Blackboard] yeh that's used for all the lectures, its just whether…. Usually the lecturers put up their lecture notes and sometimes they may put up model answers… And so that's if you miss a lecture you can still get hold of the lecture notes…. Well I mean they put them up after the lectures, buts its, I mean, the notes they give you and the notes they put up it isn't really enough to fully understand them, you've got to actually go to lectures, so that you can then annotate the notes … its not like some courses where they give you a whole book at the beginning. [Int 5]

The value of access to online journals was evident, although it was unclear to what extent students had received formal training on library skills and information handling, one student for example had only recently discovered that he could access the online journals off site using his Athens account:

> I mean the library is quite good …got online journals .. I didn't realise …it was only in the last six months that I realised you can get access to the journals off campus….yeh so … I gotta go to the library … its been quite useful really because …all the journals are now online .. and you can go to them, search, … yeh opens that up abit. [Int 5]

> The library website which is pretty good for getting past papers, past exam papers sorry and also journal articles as well. Its certainly much easier than going into the library and photocopying.. [Int 7]

> [Use of technology] Literature search, mainly, using ChemDraw and data analysis …very you can access university databases of journals on the internet rather than going into library each time you need a journal article … you can access it at home [Int 9]

The interviews suggested that the purpose and function of the 'physical' library for many students is changing; many find the virtual library or resources generally available over the internet more useful. The physical effort of actually having to go to a library to get information is problematic, students are increasingly used to having access to whatever information they need from their computer:

> Use the internet lots and lots …find it easier coz then don't have to go into the library and things all the time… sit at home .. when you are looking for information just pull the journals off the internet … and saves a lot of time walking around … normally would be dead time otherwise. [Int 8]

The interviews revealed a rich variety of uses of technology – some fairly general trends such as the use of MSN chat with friends and email with peers and tutor; as well as some specific individual uses (for example the use of powerpoint for self assessment, or the mobile phone calendar function to remind student of deadlines).

> [Use of chat] I do with my friends,… nothing to do with academic stuff. [Int 3]

> Use it [chat] simultaneous [for work and with friends] … can also send file through … essentially instantaneous email [Int 8]

What is evident is that using technology is an integral and invaluable part of their study – students find the internet easy to use, are comfortable with the technology and find it easy to find relevant information.

> Internet is invaluable, …for finding out stuff, you know at short notice. Its so easy now, convenient, … yeh so I do use the Internet a lot. [Int 3]

> I seem to be constantly at my PC, either writing up reports or surfing the web finding various papers. [Int 7]

They mentioned using a range of technologies to support both the learning/work and leisure activities; games, the internet, chat, mobiles, mp3 players, etc. Use is clearly highly personalised and individual, tailored to specific needs and interests.

> [I]use the internet a little bit but not as much as some people. [Int 3]

What is clear however is that use of technology is integral, widespread and essential. Students appeared almost bemused to be asked about what they used and why, precisely because it was such a fundamental part of their practice. There are also suggestions however that their use of technology is changing the way in which they are learning; capitalising on the communicative and collaborative affordances (Conole and Dyke, 2004) of technologies, so that the students are part of a wider, distributed community of learners:

> Sort of messenger based things are quite useful if you need abit of help with some problem and you can just ask some one what they think about …have a discussion about the best ways of going about solving a problem. [Int 7]

However, students were also aware of the limitations of particular tools, for example one student commented that MSN Chat was limited in terms of its use for chemistry purposes as you can't draw chemical structures on it and indeed having to draw chemical structure generally on a computer was more cumbersome than quickly sketching it by hand.

> Messenger is sort of limited though … if you've got a problem with an organic problem you cant really draw out a load of curly arrows … and [using a computer] for drawing structures and things is very slow … and at the end of the day you can do it by hand. [Int 8]

As is inevitable with a course of this kind, there were some technical problems, which led to frustration on the part of the students, but overall these did not seem to deter from the students overall positive evaluation of the course.

> [E-Malaria site] I did use it but it was quite frustrating… I think its licence ran out… but problem was I didn't realise anything was wrong with it, it hadn't been mentioned… I might have been the first person to try and do it… loaded all my molecules in… done half of it that was the thing … I didn't realise anything was wrong.. and then it wouldn't work out …. Which was the important bit… and so that was frustrating. [Int 3]

Pedagogical benefits

A core aspiration of the project was that by providing a link between published references and research data, it would then be possible to make an explicit link between the data and the final published material in a learning context. A student reading references supplied on a course reading list would then actually be able to go and retrieve the associated data, to manipulate and interrogate it according to some defined learning activity. More broadly, Lyon articulates a number of potential benefits for learning (Lyon, 2003), such as providing access to authentic and up-to-date real research data and helping students to develop their evaluative and critical skills, Therefore students could:

> go back and look at the conclusions a researcher derived from a set of data and they could analyse it themselves and think did they make the right decisions, how would I have done it, was the method correct, those sorts of things, so that was the pedagogical benefits I could see. [Int 10]

This quote shows that there were some valuable pedagogical ideas being developed about how the concepts of eBank could be applied for learning, however these are nascent and have not yet been translated because there isn't a pedagogical model to show how this can be achieved. The difficulty of capture such tacit practice is well known and a number of approaches are being considered to try and address this (see McAndrew *et al.*, 2006 - on a comparison between patterns and learning design, Goodyear *et al.*, 2006 - on pedagogical patterns, Conole, forthcoming, b - on learning activities and Jones and Conole, 2006 - on capturing practice). Translation of these pedagogical ideas into learning scenarios or patterns[27] would be a valuable and fruitful area for further research.

In terms of work balance the students commented that the module had less time allocated to lectures in comparison with their other modules but was more intensive than other modules in terms of the workshops and the self-directed activities associated with the module – which is typical of a course which adopts a more constructivist pedagogical approach.

> I think to be honest its been less work than other courses, lecture wise but actually, there's been more workshop-based things involving searching databases and making models. [Int 7]

Active engagement with datasets enabled students to understand difficult concepts for the first time – by being able to experiment and use – for example linear regression – use in context enabled them to be able to see the relevance and value of such techniques.

Future directions and recommendations

The following have been identified as areas which are either planned to be carried out or are potential areas of work which could be explored.

1. Federation model related to the technical architecture and publishing landscape in the research domain.
    a. Getting more content into the data repository and building in more links
    b. Look at distributed federated models and explore the issues which arise with multiple institutional data repositories on different platforms.
    c. Look at how the data in an institutional data repository relates to the data curated, professional data centres like CCDC
    d. Working with publishers and professional bodies IUCR, RSC and open publishers such as the newly proposed Chemistry Central
    e. Exploring differences in workflows in different labs and different research practices in different labs, different sites; what are the technical implications are for metadata profiles, for normalisation of the metadata, for harvesting?
2. Further exploration of the potential of open research data in the learning dimension: In particular to explore the potential of open research data in the learning area and pedagogical benefits in far more depth.
    a. PSIgate is planning to undertake an evaluation of eBank embedded into INTUTE in the Autumn, working with students at Manchester University. They plan to focus on how students access and use the eBank crystal data and make use of the associated linkages.
    b. The initial use of eBank in the Chemical informatics module could be developed.
3. Also text mining and data mining – knowledge extraction, value added part of making the data opening available.

As the vision scenario quoted earlier depicts, eBank is part of a much broader and imaginative vision of the future potential of providing open access to research data

[27] See for example http://www.pedagogicalpatterns.org/

and the way in which technology can unlock this potential in novel and exciting ways, resulting in a potential shift in both the way in which we view and value information and the way in which we communicate and share information.

> A repository somewhere where you just put things and store them and get them back so its all the software that's built around … I am hoping that it will help crystal structure searching and change how that links with publications. The idea is to get the repository and the tools that are built around it to support the whole sort of crystal structural chemistry research cycle…so be involved in all sorts of aspects of teaching and learning when it comes to that cycle. [Int 1]

So there is a vision here of a complete cycle and a mechanism of linking the different components of that cycle together. This provides a model for the technical developments which is likely to be an important contributor to the success of the project, as discussed in the section in 'Conceptual models'.

The initial ideas and concepts embedded in eBank are now starting to propagate and extend. This is most clearly evident in the projects most closely aligned to eBank such as the application of the ideas from eBank to Kinetics:

> These kinetics people, who can see, having seen what can be done with the crystal structures can see how it can be used elsewhere. ref

In addition some of the ideas (of how Chemistry and Libraries can interact) which have arisen as a result of the project, have now been taken up elsewhere.

One of the key central aspects is the notion of providing enough structure to the data and making it available to others via a repository so that they can use it for their own research purposes. The demonstration of making particular data searchable via something like Google is now giving researchers ideas for what else could be made available in this way.

> If you can do it for this data we should at least do something like it [Int 1]

> So that even if we are not formally using a repository or running it through an ePrints systems and so on I think more and more people are beginning to think about lets at least make this data set available and structure it [Int 1]

There is also a knock on effect of this for journals in terms of how they make data associated with articles available and how they structure it and the fact that the data 'is not just a set of data linked off a paper because that doesn't allow you to search for it: you already have to know if exists'.

The notions inherent in eBank and projects like it, offer the potential for a radical future in which both research and teaching practice are fundamentally changed because of the way in which technology is used.

> It's making some of the paradigms of the new e-research and e-science principles possible, I think. So the idea being that in the future the way that scientific research is carried out will change. And rather than doing some wet lab experimental data generation… and creation …a scientist will have an idea and will go to a dataset that already exists base and run various programs, algorithms across it to test their hypothesis and based on the results of that may repeat that process, .. do something different. They may mine distributed data sets using grid processing, grid technology. They may run text mining programmes to look at connections between data and the derived textual corpora data. [Int 10]

This quote suggestions a vision of the future which involves new ways of viewing the connections between different types of research output and new ways of generating and viewing knowledge creation, which might lead to a paradigm shift (Kuhn, 1996) in terms of the way we view, value and undertake research.

The vision for learning is not as well understood at the moment, but the potential is clearly there, in terms of access to hitherto unavailable information, the ability to visualize and manipulate data in different ways and new ways of interacting with the data (for example annotation of data sets, crystal structure, research papers, etc). However, this raises questions about the development of appropriate skills for students (and their tutors) in order to be able to fulfil this potential:

> Could assume … that a student would need to develop more numerate skills if they are looking at data sets more regularly as a matter of course. It we assume that a learner in the future will be able to look at data, research data sets as a matter of course in say their final year and MScs, they will need to develop some new skills as will the researchers to be able to do that. And the classic combination is domain skills, IT skills and statistical skills, that combination of three things to develop the new e-researcher of the future or the new, maybe the new, e-science learner or whatever of the future. [Int 10]

Until we understand better what is possible, as well as what the students actually do, it is difficult to build these aspirations into curriculum development. In a sense there is an interactive co-development needed between evaluation and needs analysis alongside curriculum change.

Furthermore there are evidently technical issues, in terms of linking up different systems developed at different times and hence the difficultly of achieving integration, which are bound to impact on how it is utilised in teaching:

> In the Informatics course we are trying to embed it in but essentially I am using current tools which is pulling us here.. there… bits and pieces from here there and everywhere and plugging them together, the idea that you could. [Int 2]

There is potential in terms of the ideas of eBank being developed across other aspects of the Scientific community. Importantly the team feel that the potential of this transferability is linked with the fact that they see the underlying philosophical approach they have adopted as appropriate:

> [The Philosophy is important] It [it the ideas in eBank] could be used in other domains, so I think the whole concept is absolutely correct – how its transferable depends on community and social issues and people being prepared to build software tools based on our philosophies.. this way is applicable in many more fields …. [could] adopt our kind of approach. [Int 2]

Across Chemistry it is evident that the principles of eBank are transferable to a wide range of areas – in particular the architecture developed is applicable for reasonably small data sets. It could be applied to a range of analytical areas which produce digital data and have reasonably identifiable work flows associated with them. So it would then be easy for anyone to follow the process and see easily how a result was derived and hence assess its validity. The related R4L project is taking aspects of this forward by looking at twelve different chemistry techniques on small instruments and looking at applying the eBank ideas across these.

> So rather than accumulating data here there and everywhere and then uploading to a database the experimenter is continually adds to database in real time and dynamically. – the database is building as the experiment is being done. .Yes so then that really will be a change in practice. [Int 2]

Recommendations

A number of key recommendations are evident and of relevance to a number of different stakeholders – in particular the JISC, the Research Councils, the DfES, and publishers.

1. The JISC and the Research Councils need to think about how they can mandate open access to the results of research funded by public money and in particular to the data in an appropriate way and provide infrastructure to enable that to happen [JISC and Research Councils]
2. Some research is needed to investigate what the implications and potential is of allowing learners at whatever level to engage with this data and to explore how it can enrich the learning experience [DfES].
3. Some research is needed into thinking about the skills the research and learning community need to be equipped with or need to gain to enable them to do this and how those skills will be provided – funding investing in each of these areas [DfES]
4. The metadata application profile is one of the primary outputs of project because it is the way in which repositories and services built on them will talk to each other. It is important because it facilitates the publication process and growth of databases. Firstly eBank demonstrates a way of storing and archiving data, secondly it demonstrates a way of publishing and maintaining ownership and thirdly (potentially) a mechanism for helping data repositories to grow. The recommendation is that institutions should install such repositories but make sure they adhere to these standards. [JISC, Institutions]
5. To pursue a conversation with the publishers to see if the ideas embodied in eBank are feasible, in particular in terms of linking to data at source from published papers. [Publishers]
6. Sustainability issues of initiatives of this kind are a concern which need to be further investigated; it is not just about the setting up of these types of resources, but issues about keeping the service running and adding metadata. [JISC]
7. It is still early days in terms of the development of standards and interoperability in this area, work on this needs to occur in parallel with the developments of related repositories.[JISC]

Conclusion

> But the new generation of researchers, when your Net Gen (Lippincott, 2006) become researchers in their own right or tutors, lecturers in their own right or whatever the word will be then, I don't think it should be lecturer, then the whole thing may take a step change. Web 2.0 I very much see the open access movement and the Web 2.0 social software movement as complementary to each other and if you have open access to the content .. and you've got the open software tools, the web services, the two come together beautifully. [Int 10]

Research-led teaching is a familiar part of the rhetoric of modern education in the UK (Brew, 2006) – particularly for the research-led universities, but providing evidence of examples of specific instantiations of ways in which research actually does impact on teaching is more difficult. The eBank project does offer a real and tangible example:

> eBank is a great way of demonstrating research-led teaching and learning. It's a way of very visibly demonstrating, a way of showing how research in a department is underpinning the teaching [how it] is embedded clearly into curriculum. [Research-led teaching is] a mantra for them [research intensive institutions] but you have to demonstrate that in real practice. I think it is important because the learning and teaching courses can't stand still, they need to be based on the latest research outputs and knowledge and hence need to be constantly involving in terms of content and this is one way of achieving that. [Int 10]

The experience of the eBank project highlights the potential innovative applications of technologies for both teaching and research. It raises a raft of critical questions about not only how we develop, manage and share information, but about the very nature of core concepts associated with education – what constitutes research, how is data valued and used, and what is the relationship between research and practice?

Acknowledgements

Thanks go to members of the eBank project for their cooperation throughout the evaluation, for the staff and students who agreed to be interviewed and to Liz Lyon, Jeremy Frey and Simon Coles for comments on the draft evaluation report.

References

Atkinson, M., Karasavvas, K.A., Antonioletti, M., Baxter, R., Borley, A., Chue, N.P., Hume, A., Jackson, M., Krause, A., Laws, S., Paton, N.W., Schopf, J.M., Sudgen. T/, Tourlas, K., and Watson, P. (2005), 'A new Architecture for OGSA-DAI', Proceedings of the UK e-Science All Hands Conference 2005, Engineering and Physical Sciences Research Council, available online at http://www.allhands.org.uk/2005/proceedings/papers/432.pdf [accessed 20/11/06].

Beetham (2002a) 'The Learning Technology Career Development Scoping Study', available online at http://www.jisc.ac.uk/index.cfm?name=project_career, last accessed 29th July 2003.

Beetham, H. (2002b) 'Developing learning technology networks through shared representations of practice', *Proceedings of the 9th International Improving Student Learning Symposium*, 421-434, Oxford: OCSLD.

Berman, F., Fox, G. and Hey, T. (Eds) (2003), *Grid Computing: Making the Global Infrastructure a Reality*, London: Wiley, outline available at http://www.grid2002.org/ [accessed 20/11/06].

Brew, A. (2006), *Research and Teaching: Beyond the Divide (Universities Into the 21st Century*, New York: Palgrave MacMillan.

Conole (2006), 'What impact are technologies having and how are they changing practice?', in McNay (ed), *Beyond Mass Higher Education: Building on Experience, The Society for Research into Higher Education*, Open University Press/ McGraw-Hill Education, 81-95.

Conole, G. (forthcoming, a), ‘An international comparison of the relationship between policy and practice in e-learning’ in Andrews and Haythornthwaite (Eds), *Handbook of e-learning research*, Sage: London.

Conole, G. (forthcoming, b), ‘The role of technologies in learning activity design’, in Learning Design, H. Beetham and R. Sharpe (eds), RoutledgeFalmer.

Conole, G. and M. Dyke (2004). "What are the affordances of information and communication technologies?" Alt J. **12**(2): 113-124.

Conole, G. and Oliver, M. (2007), *Contemporary perspectives in E-Learning Research: Themes, tensions and impact on practice,* RoutledgeFalmer.

Conole, G., de Laat, M., Dillon, T. and Darby, J. (2006), An in-depth case study of students’ experiences of e-Learning – how is learning changing?, paper submitted to the ASCILITE conference, Sydney, December, 2006.

Conole, G., Littlejohn, A. Falconer, I., and Jeffery, A. (2005) ‘Pedagogical review of learning activities and use cases’, *LADIE project report*, JISC funded project under the e-Frameworks programme.

Conole, G., Oliver, M. and White, S. (2007), ‘The impact of e-learning on organisational roles and structures’ in G. Conole and M. Oliver (ed), *Contemporary perspectives in E-Learning Research: Themes, tensions and impact on practice,* RoutledgeFalmer.

Conole, G., Smith, J. and White, S. (2007), ‘A critique of the impact of policy and funding’, in G. Conole and M. Oliver (ed), *Contemporary perspectives in E-Learning Research: Themes, tensions and impact on practice* , RoutledgeFalmer.

Conole. G. (2006), ‘Complexity and interconnection: Current directions and future challenges for e-learning – a macro-perspective’, Positional paper for the OU 2005/2007 Working Group, 12th May 2006.

Goodyear, P., de Laat, M. and Laly, V. (2006), ‘Using pattern languages to mediate theory-praxis conversations in design for networked learning’, ALT-J 14.3.

Gosling, D. (2001) ‘Educational development unit in the UK – what are they doing five years on?’, *The International Journal for Academic Development*, Vol. 6, No. 1, 74-90.

Hey, J. (1997), ‘E-journals for Research: the User Perspective’, *Serials* 10(1):65-68.

Holyfield, S. (2002) ‘Developing a Shared Understanding of the Managed Learning Environment – the role of diagramming and requirements gathering’ Commissioned JISC report, http://www.jisc.ac.uk/index.cfm?name=mle_diagram_report

Hutton, W. (2002). *The world we’re in*, Abacus

Hylen, J. (ND), 'Open Educational Resources: Opportunities and Challenges', OECD's Centre for Educational Research and Innovation, report as part of the CERI OER project, available online at http://www.oecd.org/dataoecd/5/47/37351085.pdf [16/12/06]

Jones, C. and Conole, G. (2006), 'Sharing practice in networked tertiary education: problems and alternative solutions', proceedings of the ASCILITE conference, Sydney, December, 2006, available online at http://www.ascilite.org.au/conferences/sydney06/proceeding/pdf_papers/p95.pdf [accessed 16/12/06].

Kuhn, T. S. (1996). The structure of scientific revolutions. Chicago ; London, University of Chicago Press.

Kukulska-Hulme, A., Traxler, J. (2005) Mobile Learning: A Handbook for Educators and Trainers, Routledge, London

Lippincott, J.K. (2006), 'Net Gen students: Learning, technology and libraries', *JISC/CNI conference,* 7th July 2006.

Littlejohn , A. (Ed) (2003), *Reusing Online Resources: A Sustainable Approach to eLearning*, London: Kogan Page

Lyon, L. (2002), 'Developing information architectures: to support research, learning and teaching', *UCISA conference,* March 2002.

Lyon, L. (2003), 'eBank UK: Building the links between research data, scholarly communication and learning', Aridane, Issue 36, available online at http://www.ariadne.ac.uk/issue36/lyon/ [Accessed 20/11/06]

McAndrew, P., Goodyear, P. & Dalziel, J. (in press) Patterns, designs and activities: unifying descriptions of learning structures, *International Journal of Learning Technology*. http://kn.open.ac.uk/document.cfm?documentid=5295

Moore, G.A. (1991) *Crossing the chasm*, New York: HarperBusiness.

Nobles, D. (2001), 'Digital diploma mills – the automation of higher education', FirstMonday, Issue 3, No. 1, available online at http://www.firstmonday.org/issues/issue3_1/noble/ [accessed 20/11/06].

NSF (2005), 'Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century', National Science Foundation report NSB-05-40, available online at http://www.nsf.gov/pubs/2005/nsb0540/ [accessed 20/11/06].

OECD (2003), 'Promoting access to pubic research data for Scientific, Economic and Social Development', OECD follow up group on issues of access to publicly funded Research data, March, 2003, available online at http://dataaccess.ucsd.edu/Final_Report_2003.pdf [accessed 16/12/06]

Oliver, M., Roberts, G, Beetham, H., Ingraham, B., Dyke, M. and Levy, P., 'Knowledge, society and perspectives on learning technology', in G. Conole and M. Oliver (eds), *Contemporary perspectives on e-learning research, London: RoutledgeFalmer*

Powell, A. (2005), 'A 'service-orientated' view of the JISC Information Environment', JISC commissioned report, UKOLN: University of Bath, available online at http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/soa/jisc-ie-soa.pdf [accessed 20/11/06].

Salmon, G. (2005) 'Flying not flapping: A strategic framework for e-learning and pedagogical innovation in higher education institutions', *ALT-J*, 13(3)

Taylor, J, (2004) *E-learning research consultation*, available online at http://kn.open.ac.uk/public/index.cfm?wpid=2843

Timmis, S. (2003) 'ELTI workshop: Institutional audits to inform the embedding of learning technologies and promote cross boundary working in our institutions', *ALT-C 2003*, workshop, Sheffield, available online at http://www.shef.ac.uk/alt/abstracts/90work/1/ws3-211.htm, last accessed 29 March 2006.

Warburton, B. (2006), 'Barriers and enablers to institutional uptake of CAA', PhD thesis, Southampton: University of Southampton.