e-Bank UK

A proposal submitted to the JISC by UKOLN, University of Bath with the University of Southampton, and PSIgate, University of Manchester, in response to Circular 8/02.

December 2002

Introduction

This proposal describes e-Bank UK, a 12-month project which will investigate the issues surrounding provenance and the use and re-use of original data for research and learning purposes, and will result in the development of an e-Bank UK pilot service for the benefit of the HE and FE communities. It builds on existing JISC-funded development of a repository for metadata relating to research articles (ePrints UK) and on outputs from the EPSRC-funded e-Science Combechem Project. The concepts underpinning the pilot bring together both research and learning & teaching activities in an innovative manner and they seek to demonstrate the direct relationship between the two areas, which results in the growth and development of rich knowledge structures. The project will also bring together members of the digital library and Grid/e-Science communities, both through technical development work and through wider dissemination activities. The project is to be led by **UKOLN** in partnership with the University of Southampton, and PSIgate, University of Manchester. Funding is sought for £129,497 based on a start date of 1st August 2003.

Project description

Conceptual framework

This project is designed to test the hypothesis proposed in earlier conference presentations[1] that research and learning processes are cyclical in nature, and that subsequent outputs which contribute to knowledge, are based on the continuous use and re-use of data and information. The creation of original data may be followed by one or more additional processes which in turn generate modified datasets e.g. aggregation of experimental data, selection of a particular subset, repetition of a laboratory experiment, statistical analysis, manipulation of a molecular structure, annotation of a diagram etc. This derived data is clearly related to the original data in some way and can be re-purposed through inclusion in a crystallography database, in a pre-print or in a peer-reviewed article. These secondary items can themselves be re-used through a citation in a related paper, by a reference in a reading list or in modular materials as part of an online course. This concept is illustrated using the chemistry domain and specifically relating the Combechem Project[2] which is an e-Science Grid-enabled initiative, to relevant online chemistry courses for undergraduate and postgraduate students.

Currently once an experiment is finished, the initial dissemination may be via a letter or communication, followed later by a more detailed explanation in a full paper describing a more in depth analysis and perhaps collecting up several related results. While reference data may be provided there is unlikely to be any link back to the raw (or even processed data). The existence of e-prints and ability to publish data at source (i.e. currently via a web site) suggests a different approach. To illustrate this consider an example from crystallography:

- A new compound or material is submitted for structure determination by x-ray diffraction by the NCS (National Crystallography Service).
- Once the experiment has been run (run time 10 min to several hours), the data is analysed and a structure determined and then validated (validation is a semi- automatic process, automation being developed further).

---

[1] Lyon, L. (2002) Developing Information Architectures… …to support Research, Learning and Teaching, UCISA Conference, March 2002. http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/presentations/dev-inf-arch/intro_files/frame.htm
Frey, J., De Roure D. and Carr L. (2002) Publishing at Source: Scientific Publication from a Web to a Data Grid, EuroWeb 2002 Conference, Oxford. Dec 2002. http://www.ecs.soton.ac.uk/~lac/publicationAtSource

[2] Combechem Project Web page http://www.combechem.org/

- At this stage it would be reasonable for an e-print system to carry the identification of the compound and the fact that a structure had been determined together with the unit cell dimensions (this is a basic characterisation of the material).
- If it is determined that the structure is routine and not worthy of special discussion (at this stage), the atomic coordinates (i.e. full structure) are made available via the e-print and submitted up to the crystallographic databases.
- The e-print system links to the raw data and other researchers could look at this to assess the validity of the work.
- If the material is worthy of discussion then the group who produced the material writes and submits the communication to a journal.
- The e-print access is altered at this stage to give access to the coordinates and the raw data to the referees who will validate the journal article. This leads to interesting issues of confidentiality of the referee.
- Once the journal article is published the e-print links are opened to the community.

In many areas technical reports provide a method of disseminating information that is not thought worthy of publication in the primary literature. Currently within Chemistry such information is typically disseminated only via books, manuals and monographs, even though this information can be vital in actually repeating and using new techniques. The e-Bank system should enable the more rapid dissemination of such materials while still providing a distinctive institutional label to the information.
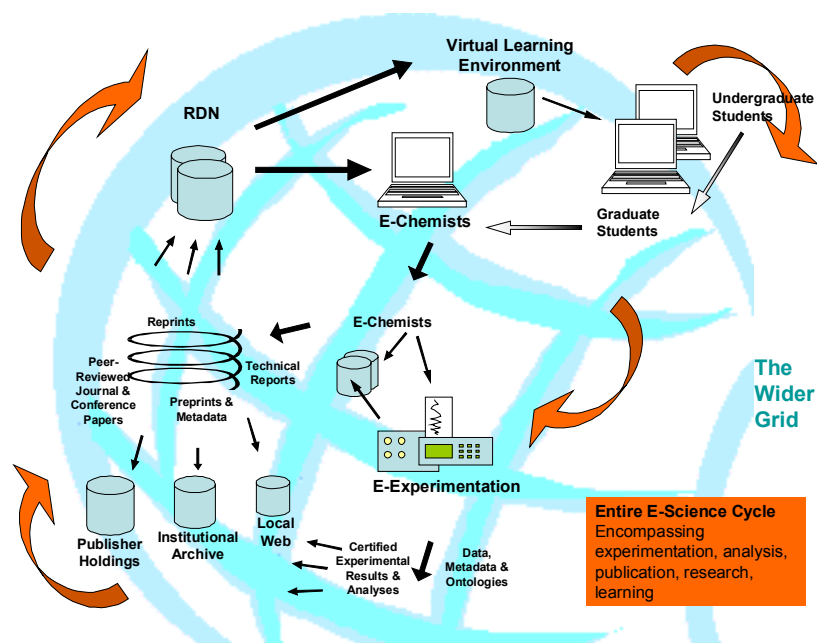
For a complete cycle to be implemented in either direction *i.e.* for discrete research data to be explicitly referenced in learning & teaching materials or for a student to "rollback" to the original research data from a secondary information resource such as a published article or from an online course delivered via a Learning Management System or Virtual Learning Environment (VLE), a number of assumptions must be made. These relate largely to the discovery process but are also closely linked to the requirement for essential data curation procedures. The assumptions are:

- The integrity of the original dataset is maintained.
- There is a shared understanding of the concept of provenance.
- The original dataset is adequately described using a metadata description framework based on agreed standards.
- A common ontology for the domain is understood.
- Each dataset and derived data and information is uniquely identified.
- Open linking technology is applied to the original dataset and the derived data and information.

We are proposing to test the hypothesis using the chemistry domain as a case study, with a view to assessing the feasibility of implementing the generic approach across other disciplines and other communities *e.g.* further education. The Combechem Project is an ideal research testbed since large quantities of varied data are generated including electronic lab books, crystallography data and physical chemistry data i.e. textual, numeric and 2/3D molecular structure datasets. Publication sources include fast track "Letters" formats referenced in postgraduate teaching modules, whilst peer-reviewed articles have been published which reference experimental data. These sources together with

metadata from ePrints UK, and the PSIgate Physical Sciences hub repositories, will form the basis of the pilot e-Bank UK service which is described in more detail in the Technical Architecture section below and in Work Packages 2-5. The pilot will also be informed by research into the requirements of stakeholders such as researchers, learners, tutors and content providers. These requirements will be defined based on a mix of desk-based research and interviews/focus groups with data/information users and creators and form the content of Work Package 1.

Two Supporting Studies exploring 1) an overview of current research in the area of provenance including considering definitions of provenance, and 2) the feasibility of a common framework for the description of datasets, are described in Work Package 6. Each study will produce a substantive report for dissemination. The project outputs will be evaluated against the Stakeholder Requirements specification in a consultative process centred around a workshop involving key players (Work Package 7). Results and feedback will be summarised in an Evaluation Report and in recommendations for future work, as this project is viewed as the first phase of a longer-term strategy to enhance access to a more comprehensive digital library of research data and related information. A Final Report will be produced as part of Work Package 8, to summarise the project achievements and ongoing challenges.

Technical Architecture

This project will build on the technical architecture currently being deployed within the context of the ePrints UK project[3]. The ePrints UK project is developing an architecture that supports the harvesting of metadata from eprint archives in UK academic institutions and elsewhere using the OAI Protocol for Metadata Harvesting (OAI-PMH)[4]. The harvested metadata records will be enhanced using three Web services offered by OCLC and the University of Southampton (automatically assigned subject classification terms to each metadata record, applying name authority control to any names in each record and undertaking some citation analysis within the full-text of each eprint). Enhanced metadata records will be stored in a central ePrints UK database upon which a number of subject-focused Web interfaces will be developed. Finally, each interface will be embedded into the gateway/portal services currently being offered by each of the RDN hubs.

The e-Bank UK project will augment this work by also harvesting metadata about research data from institutional 'e-data repositories' (initially those made available by Combechem and University of Bath during the project lifetime but others in the longer term). The eprints.org software will be adapted to provide storage for and metadata descriptions of the research data output from the project (so-called 'e-data' archives) as well as the research publications. The research data will be described using the schema developed during Work Package 2. Metadata records harvested from e-data repositories will be stored in the central database alongside the enhanced eprint metadata records gathered as part of the ePrints UK project. The e-Bank UK project will make recommendations for how eprints should cite the research data on which they are based, probably by using a URI based on the unique identifier that is assigned to the research data when it is deposited in the e-data archive.

An enhanced end-user interface for ePrints UK, targeted for delivery through the RDN PSIgate Hub, will be developed. It will offer navigation from eprint metadata records to research data metadata records and vice versa. The project will also investigate the technical possibilities for inferring which subject classification terms may be associated with research data, based on knowledge of the terms that have been automatically assigned to the eprints which cite those data resources.

Work Package 1 Stakeholder requirements

This work package will identify the wider issues and requirements of the various stakeholders involved in the research and learning processes that relate to the discovery, access, use and re-use of research data, Grid datasets and derived information contained in digital libraries and virtual learning environments. It will draw on the views and experience of individuals within the chemistry domain and the Combechem project, and will include views from researchers, students, lecturers, tutors, librarians/information scientists, instructional designers/learning technologists and content providers. Technical, pedagogical, informational and practical management aspects will be considered. The information will be gathered through a number of focus groups, interviews and desk research. The findings will be co-ordinated and used to draft a Requirements Specification which will inform the development of the Pilot service.

---

[3] ePrints UK Project Web site http://www.rdn.ac.uk/projects/eprints-uk/

[4] Open Archives Initiative http://www.openarchives.org/
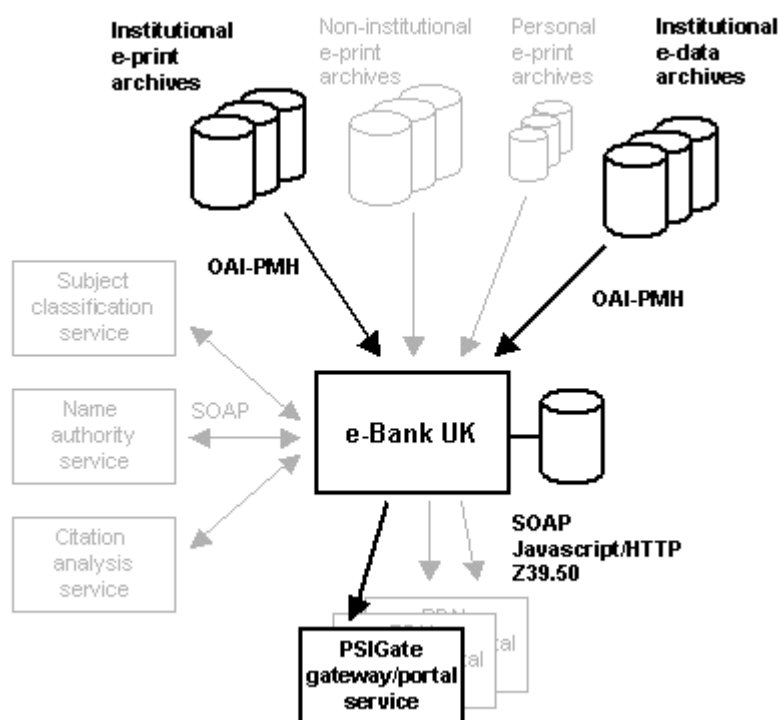
Work Package 2 Pilot development

This work package will develop a version of the eprints.org software that is capable of providing effective access to research data as well as research publications. It will also develop a metadata schema for describing research data (data sets) based on existing work in the Comb*e*chem project. The resulting schema will be described in a human-readable document and as machine-readable XML and RDF schemas. The XML schema is required in order that metadata records conforming to the schema can be exchanged using the OAI-PMH. The RDF schema will support use of the schema in the context of the Semantic Web/Grid. The work package will also develop a detailed technical architecture for e-Bank UK. It is likely that the central database of metadata records and associated Web interfaces will be based on the Cheshire XML information retrieval system[5] in line with current plans for the ePrints UK project.

Work Package 3 Pilot service

This work package will develop a Web interface to the central database of eprint and research data metadata records. The Web interface will be hosted on the RDN Web site[6] and will form the basis of the interface that will be embedded into the PSIgate Web site. Embedding will be achieved using the CGI-based mechanism that was developed for RDN-Include.[7]

Work Package 4 Pilot testing & embedding

This work package will use the interface developed during Work Package 3 and embed it into the PSIgate Web site. This will require the installation and testing of a CGI script at the PSIgate Web server, configuration to conform with the look-and-feel of the PSIgate Web site and the addition of appropriate links from other parts of the PSIgate service.

Work Package 5 Repository deployment & compliance

Having developed the software, the key issue is to fill the repository with metadata relating to publications and data. The objective of this work package is to acquire chemical preprints, post-prints and (more importantly) experimental data identified by the partners. To provide information from a variety of Chemical sources to enable abstraction of the important general features of raw and analysed data, together with the corresponding metadata, a range of experimental situations within Chemistry will be utilised. These range from the archetypal synthetic chemistry laboratory (white

---

[5] Cheshire http://cheshire.berkeley.edu/

[6] RDN http://www.rdn.ac.uk/

[7] RDN-Include http://www.rdn.ac.uk/rdn-i/

coats in front of fume cupboards), to the equipment & computer dominated environment of physical chemistry (closer perhaps to physics and "big science"). Similarly information generation within virtual experiments based on computer simulation will be compared and contrasted with the laboratory-based work. In all cases methods of integration of new results with existing literature and other information sources (databases) will be considered.

Work Package 6 Supporting studies

Study 1 - Provenance: review of current research, issues and challenges.

Provenance is a well-established concept in the art world where the lineage, pedigree or origins of a painting are critical to determining its authenticity and value. [8] [9] It is of equal importance in science where the provenance or origin of a particular set of data is essential to determining the likely accuracy, currency and validity of derived information and any assumptions, hypotheses or further work based on that information.[10] Significant research has been carried out on describing the provenance of scientific data in molecular genetics databases SWISSPROT and OMIM [11] and in collaborative multi-scale chemistry initiatives.[12] The topic has recently been explored in a workshop at the latest Global Grid Forum (GGF6) in relation to Grid data[13] and the relationship of provenance to the Semantic Web has been noted.[14] The Open Archives Initiative has also carried out some work to describe the provenance of harvested metadata records[15] and the concept is included as an element in the administrative metadata which is part of the METS metadata standard[16].

This study will bring together in a Report the body of work on provenance, review the observed trends and directions, identify gaps in work in this area and present some conclusions and recommendations for future activities for the JISC. The study will particularly focus on the cyclical nature of the research and learning process described earlier, and the relationship between the creation, curation and management of research data and its integration into published information resources which are contained in digital libraries.

Study 2 – Feasibility Report on Dataset Description and Schema

This study will consider the description of (multimedia) datasets within the chemistry domain particularly with regard to the perceived hierarchy of data and metadata from raw data up to "published results". To ensure the linking of this hierarchy a variety of issues will be addresses, including identifying common attributes of a dataset and relating these to domain-specific characteristics, managing legacy data, metadata created at source by laboratory equipment and the relationship to data curation activities. The Comb*e*chem project will be used as a case study and metadata from three sources (*e*-Lab book, crystallography data and physical chemistry data) will be investigated. The findings will be presented in a Report. One further outcome of the study will be a draft schema for describing chemistry datasets. The work of the CMCS will also be relevant to this Study.

Work Package 7 Evaluation & Recommendations

WP7 will seek feedback from the community through a consultative workshop which will inform the evaluation of the the pilot service against the Requirements Specification (WP1). The Workshop will be organised by UKOLN through the Events Manager, and will be open to members of the digital library and Grid communities. It will be particularly important to consult with a wide representation from different groups including researchers, students, lecturers, tutors, librarians/information scientists, instructional designers/learning technologists and content providers and not to restrict to individuals from the chemistry domain. The results of this workshop will be distilled and an Evaluation Report produced which will form a part of the project Final Report. In addition, a number of recommendations for future work will be presented which will include prospective technical developments to the pilot and also comment on the generic applicability of the Project findings to other disciplines and subject areas.

---

[8] Getty Provenance Index http://piedi.getty.edu/

[9] Museum of Modern Art, New York. Provenance Project http://www.moma.org/provenance/

[10] Peter Buneman et al http://db.cis.upenn.edu/DL/whywhere.pdf

[11] Peter Buneman et al http://www.cis.upenn.edu/~wctan/papers/02/sigmod02.pdf

[12] Collaboratory for the Multi-scale Chemical Sciences CMCS http://cmcs.ca.sandia.gov/index.php

[13] Data derivation and provenance workshop GGF6 http://www-fp.mcs.anl.gov/~foster/provenance/

[14] Carole Goble Position Statement at GGF6 Workshop.

[15] Open Archives Initiative guidelines http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm

[16] METS http://www.loc.gov/standards/mets/

Work Package 8 Project Management

Project management and partner co-ordination will be provided by UKOLN and will be achieved by an initial project start-up face-to-face meeting with all partners, a similar mid-term meeting and a project closure meeting. Communication between partners will be supported by a dedicated project discussion list and informal methods. Project staff at UKOLN will be members of the Distributed Systems & Services team led by Andy Powell with additional strategic direction provided by the Director, Liz Lyon. UKOLN research effort will be provided by the Research & Development Team. Financial reports will be supplied by the UKOLN Resources Co-ordinator and a Summary Final Report will be produced at the end of the Project.

Impact

The outcomes of the project will provide immediate benefits to the e-Science community and to the wider research, and learning & teaching communities in higher and further education, but will also have the potential for very significant long-term impact. The availability of original data, (whether created in a laboratory, in a recording studio or in a field survey), together with the ability to track its use in subsequent research work, publications or learning materials, will have major impacts in a number of areas:

- Tracking the protocols, mechanisms and workflows integral to the research process
- Improving access to research outputs
- Validating the accuracy and authenticity of derived work
- Supporting the explicit referencing and acknowledgment of original contributors
- Raising the accepted standards associated with publication of research outputs
- Increasing the speed of dissemination of research activity
- Expanding the ways in which data and information is used and re-used for a range of purposes
- Enhancing the quality and richness of materials created for learning and teaching.
- Enabling learners to examine the original data underpinning published work.

Following successful proof-of-concept, the approach advocated in this study would complement parallel work on promoting the creation of e-print repositories and subsequent harvesting of associated metadata into aggregator services. It is anticipated that there will be similar issues of sustainability, buy-in from academics, security, privacy and other cultural barriers, which will need to be addressed at the next stage.

Accessibility issues

Full account will be taken of issues relating to accessibility of Web-based systems and software and the outputs of this project will conform to published standards and guidelines. UKOLN hosts the UK Web Focus which is pro-active in promoting these principles.

Project Deliverables and Timetable

Work Package	Description	Deliverables	Months	Lead Effort + partners
1	Stakeholder Requirements	Requirements specification	1-9	UKOLN + Southampton
2a	Pilot development	Technical requirements Technical specification & schema	1-2	UKOLN + Southampton
2b	Pilot development	e-Prints software technical requirements & development	1-4	Southampton
3	Pilot service	Demonstration service	3-8	UKOLN
4	Pilot testing & embedding	Interoperability with PSIgate	9-10	PSIgate
5	Repository deployment & compliance	Depositing eprints and data and ensuring data providers OAI-compliance	3-9	Southampton + Bath
6	Supporting studies	Report on provenance Feasibility Report on dataset description and Schema	1-4	Southampton + UKOLN

7	Evaluation & Recommendations	Consultation workshop Evaluation report Recommendations for future work	11-12	UKOLN + Southampton
8	Project management	Summary Final Report	1-12	UKOLN

Dissemination of Project Outputs

Dissemination of information outcomes from project activities will be achieved in a number of ways. A Project Web site will be created at the start of the project and will contain current information on activities, reports from the Supporting Studies and the Evaluation Report. Links to relevant articles and projects relating to the e-Bank UK project will be added. Presentations and publications derived from project work will also be available on the site.

Project progress will be presented at relevant conferences e.g. GGF, and workshops, including any specific JISC Development Programme events. Particular attention will be paid to disseminating the work across both digital library and Grid/e-Science communities and to engage both groups in recognising the importance of provenance and its value in validating the processes of research and learning. The Consultation Workshop will also serve as a channel for disseminating the results of the project and it is hoped to reach a broad audience through this event.

Partnerships and Supporting Organisations

This bid is submitted by UKOLN, University of Bath, in partnership with:

- Intelligence, Agents & Multimedia Group, Department of Electronics & Computer Science, and the e-Science Comb*e*chem project, Department of Chemistry, University of Southampton.

- PSIgate, University of Manchester

Letters of support from the project partners (Southampton) and hub (Manchester) are attached (see Appendix). Note that UKOLN was unable to obtain a letter of support from the University of Bath, as negotiations are still taking place between JISC and the University, concerning future UKOLN funding arrangements and JISC is aware of this omission.

The project team will continue to pro-actively liaise with other groups who may be working in this area including David Boyd and the e-Science team at CLRC, who are developing the CLRC Integrated e-Science Environment and DataPortal.

Project staff will work in partnership with members of relevant JISC teams, provide progress reports as required and participate in Programme evaluation activities. It is envisaged that the work will inform the development of the JISC Information Environment, and the project team will collaborate closely with the IE Support work based at UKOLN. The project will also contribute to any relevant JISC Development Programme events and future JISC strategic development activities. The potential for long-term impact on the activities and outputs of the JISC and its communities has been noted earlier.

Capabilities and Experience

UKOLN

UKOLN is a national centre of expertise in digital information management. It provides services to the library, information and cultural heritage communities. Its goals are to:

- Influence policy and inform practice
- Promote community building and consensus-making by actively raising awareness
- Advance knowledge through research and development
- Build innovative systems and services based on Web technologies

UKOLN has an established track record for successfully delivering quality project outputs and disseminating this information to the wider community through a mix of high profile events, electronic and hard copy publications and the Web site. UKOLN has worked on the development of the technical architecture that underpins the JISC Information Environment, is a partner in building the IE Service Registry and continues to provide a technical advisory service to JISC for this infrastructure development. UKOLN is involved with a number of JISC and EC-funded projects, including the 5/99 "Subject Portals Project" and the Open Archives Forum. UKOLN has initiated joint activities with

the Grid/e-Science community through workshops with the National e-Science Centre, Edinburgh and is actively engaging with other key players through collaborative activities.

UKOLN is responsible for the hosting of the RDN Web site and for overall technical coordination and interoperability across the RDN. The RDN is a free Internet service dedicated to providing effective access to high quality Internet resources for the learning, teaching and research community (further and higher education). The RDN provides access to a series of Internet resource catalogues containing descriptions of high quality Internet sites, selected and described by specialists from within UK academia and affiliated organisations. The RDN is a co-operative network consisting of a central organisation, the RDNC and a number of independent service providers called hubs (see below). The RDNC co-ordinates RDN activity and creates the environment in which all the partners involved in RDN work. It is the central body, responsible for the overall development of the service. RDNC activities include: setting service standards, creating the collection development policy for the network, promoting the RDN, developing strategic partnerships and supporting hub work.

Intelligence, Agents, Multimedia Research Group, University of Southampton

The Intelligence, Agents, Multimedia Research Group (formerly the Multimedia Research Group) in the Department of Electronics and Computer Science is a major research group of 100 people with an international reputation in the area of open hypermedia and its application to distributed multimedia systems and digital libraries. Previous research has resulted commercial hypermedia systems (Microcosm and the Distributed Link Service). Recent digital library projects relating directly to this proposal include the Open Journals Framework (OJF) JISC ELib project that finished in 1998, and the recently completed Open Citation (OpCit) JISC/NSF Digital Libraries project. The EPrints archival software on which this bid is predicated was developed within the group, benefiting from JISC funding from the OpCit and OPSiS projects. The group is active in the digital libraries, open archiving, hypermedia, multimedia and Web research communities, is a member of W3C and in 1997 hosted the ACM Hypertext conference.

Department of Chemistry, University of Southampton

The Chemistry department is a grade 5 research department covering the whole range of chemistry research from the interface with Physics to that with the biological and life sciences. The Department has a high profile in Combinatorial Chemistry being the site of the Combinatorial Centre of Excellence (£6M, JIF funding), and hosts the EPSRC National Crystallography Centre (http://www.soton.ac.uk/~xservice/).   The department has extensive contacts with the UK chemical and pharmaceutical industries. Its Comb*e*chem project (conducted in collaboration with the IAM group, above) is one of the UK's national e-science projects. The project is working on Grid-enabled combinatorial chemistry, concentrating on crystallography, laser and surface chemistry and development of *e*-Lab, an environment using pervasive computing technology to record detailed information on all aspects of laboratory work. The goal of the project is to develop an e-Science test-bed that integrates existing structure and property data sources, and augments them within a grid-based information and knowledge environment. The synthesis of new chemical compounds by combinatorial methods provides major opportunities for the generation of large volumes of new chemical knowledge and is the principal drive behind the project. An extensive range of primary data needs to be accumulated, integrated and relationships modelled, so that maximum knowledge can be derived. The service-based grid-computing infrastructure extends to devices in the laboratory and involves enriched systems, (including multimedia and live metadata), full support for provenance and innovative techniques for automation throughout the environment.

PSIgate - RDN hub for the Physical Sciences, University of Manchester

Both PSIgate and GEsource are led by the University of Manchester, and are being taken forward by a joint team. During the creation of both hubs, the team has developed considerable expertise in a number of areas, including: Z39.50, XML, Perl, HTML, Web site infrastructure design, and database editing. Additionally, PSIgate has been heavily involved in importing data from external sources (British Geological Survey; Institute of Physics Publishing) and incorporating these into a PSIgate look and feel database. Considerable development work is being undertaken using Java, Tomcat, and other portal related software. Recently PSIgate was selected by ISI to be included in ISI Current Web Contents.

Key Personnel

UKOLN

Liz Lyon has been Director of UKOLN since October 2000. She is involved with the development and implementation of the integrated JISC Information Environment, which includes building architectural models for distributed digital libraries. She has led collaborative activities with UK e-Science and Grid initiatives, has promoted the potential of

knowledge transfer between IE and Grid initiatives, and is a member of JCSR. Originally trained as a scientist, her Ph.D. is in cellular biochemistry.

Andy Powell, is Assistant Director and Distributed Systems & Services Team Leader. Andy received a first class honours degree in Software Engineering from the University of Birmingham in 1984. Between 1984 and 1996 Andy was employed as a Computer Officer in Bath University Computing Services working on a variety of development and support activities and moved to UKOLN in 1996. Andy's main areas of work are to provide technical consultancy to the JISC Information Environment Development Team and to coordinate technical support and interoperability across the Resource Discovery Network. Andy has previously worked on several Internet resource discovery related research projects. These have included ROADS, DESIRE, PRIDE, BIBLINK, TF-CHIC, NewsAgent, Gate-Z and (in an advisory capacity) the Subject Portals Project. Andy has been active in the development of the Dublin Core, attending a number of international workshops and meetings, he co-chairs the DCMI Collection Description Working Group, and is a member of the Dublin Core Advisory Board and the DC Usage Board. He developed the metadata generator DC-dot, which has been widely used internationally, the metadata help utility DC-assist and is co-editor of the DCMI Namespace Policy.

University of Southampton

David De Roure is a Professor of Computer Science in the 'Intelligence Agents Multimedia' Group in the Department of Electronics and Computer Science. Within the e-Science programme he is investigating the use of knowledge technologies to enhance collaboration between e-Scientists, and his application of Semantic Web technologies to grid computing applications has led to the successful 'Semantic Grid' initiative. David has worked for many years with distributed information systems and is a member of the Advisory Committee of the World Wide Web Consortium; he was Programme Co-Chair of the WWW2002 Conference and is involved in current Web standards activities including the Web Ontology Language.

Jeremy Frey is a Reader in the Department of Chemistry. His research in physical chemistry is in the application of laser techniques to the investigation of molecular properties (PI of Laser Driven Scanning tunnelling Microscopy, EPSRC £460k and nanoscale single molecule x-ray scattering, Basic Technology £2.4M). He has been involved in a number of inter and multi-disciplinary projects and this prompted his interests in data management and distribution. He is the PI of the CombeChem EPSRC e-Science project on the application of e-Science and Grid technologies to chemical structure and property prediction.

Mike Hursthouse is a Professor of Chemistry and a crystallographer of long standing who regularly is in the top 10 cited physical scientists (ISI). He is the director of the National Crystallography Service (NCS) and involved in a major part of the CombeChem project as well as the Centre of Excellence for Combinatorial Chemistry. He has a strong interest in the automation of single crystal x-ray structure determination and the corresponding automation of the structure refinement and validation process and is looking a new models for the dissemination of this information.

Leslie Carr is a Senior Lecturer in the 'Intelligence Agents Multimedia' Group in the Department of Electronics and Computer Science where he co-ordinates the various EPrints research and development activities. His research interests include the application of hypermedia technologies to digital libraries and the deployment of engineering methodologies to produce effective and reliable information systems. He is a Programme CoChair for the 2003 ACM Hypertext conference.

Key Contact

Dr Liz Lyon   Tel: 01225 386580
Director
UKOLN   Fax: 01225 386838
University of Bath
Bath, BA2 7AY   Email: e.lyon@ukoln.ac.uk