

Cover Sheet for Proposals*(All sections must be completed)***Name of Capital Programme: Repositories and Preservation Programme****Bid for Calls :** (Please tick ONE BOX ONLY, as appropriate)**Discovery to Delivery and Interoperability Demonstrators (Strand C)****Call I – Interoperability Demonstrators** a) Interoperability demonstrators**Repository Start-Up and Enhancement (Strand D)****Call II – Repository Start-Up and Enhancement Projects** a) Repository start-up projects b) Repository enhancement projects c) Rapid innovation projects: enhancing repository content**Digital Preservation Across the Lifecycle (Strand H)****Call III – Digital Preservation Across the Lifecycle** a) Digital preservation across the lifecycle**Name of Lead Institution:** The University of Southampton**Name of Proposed Project:** eCrystals Federation**Name(s) of Project Partner(s):** UKOLN (University of Bath), DCC (Universities of Bath & Edinburgh) and Unilever Centre for Molecular Informatics (University of Cambridge),**Full Contact Details for Primary Contact:****Name:** Simon Coles**Position:** Manager, EPSRC National Crystallography Service**Email:** s.j.coles@soton.ac.uk**Address:** School of Chemistry,
University of Southampton,
Southampton, SO17 1BJ.**Tel No:** +44(0)2380596722**Fax No:** +44(0)2380596723**Length of Project:** 15 months**Project Start Date:** 1st Nov 2007 **Project End Date:** 31st Jan 2009**Total Funding Requested from JISC:** £199,999**Funding Broken Down over Financial Years (Mar – Apr):****Nov07 - Mar08**

Soton RA (5 months), UKOLN RA (5 months), Cambridge RA (3 months).

Apr08 – Jan09

Soton RA (10 months), UKOLN RA (10 months)

Total Institutional Contributions: £226,792

Outline Project Description

This project will begin to establish a global and sustainable federation of data repositories for crystallography: the **eCrystals Federation**, and builds on and is informed by three previous phases of work from the **eBank Project** (<http://www.ukoln.ac.uk/projects/ebank-uk/>). In particular eBank Phase 3 was a preliminary scoping study for the deployment, interoperation, harvesting and preservation of such a Federation. eCrystals addresses the stated need to “*create[ing] the technical conditions required for interoperability for resource discovery and access to resources held within digital repositories*”. The proposal is submitted to the Repositories and Preservation Programme strand D, section b) Repository enhancement projects. The 18-month project will be led by the University of Southampton, with the core partners at UKOLN (Bath), DCC (Bath/Edinburgh) and Cambridge.



The eBank team has had many significant achievements to date the most notable being:

- Proving the effectiveness of a data repository
- Demonstrating the importance of linking data in institutional repositories with scholarly articles
- Changing the way crystallographic data can be published
- Establishing a method for maintaining the provenance of a dataset in the publication process
- Producing outputs that are a product of both domain (Chemistry) and digital libraries (UKOLN) communities working together to produce a system that is resilient to the demands of the user
- Creating an interdisciplinary team that is aware of the wider issues, as well as the domain ones

The eCrystals Federation will make a major contribution to the development of a repository-based e-infrastructure for research data, and represents an instantiation of the “Federation Data Deposit Model” described in the *Dealing with Data Report* from a recent study completed by UKOLN. The project will establish a solid foundation of crystallography data repositories across an international group of partner sites, with metadata harvested by a number of aggregator services. Building on the eBank Phase 3 results, partners will work together to harmonise the metadata application profiles from repositories operating on different platforms (EPrints, DSpace, Fedora & ReciprocalNet), investigate aggregation issues arising from harvesting metadata from repositories within the networked information environments in other countries (EU, USA & Australia) and enable the Federation of institutional archives to interoperate with international subject archives (IUCr and CCDC) and other third party harvesters.

eCrystals will provide a testbed for the StORe middleware and the CLADDIER “Ping” mechanism to enable the linking of data to derived resources such as primary publications. The project will build on eBank-UK achievements already made in the areas of data curation, preservation and provenance by application to an operational Federation and through the provision of a testbed for the DRAMBORA Toolkit. Co-ordinated advocacy and training approaches, will be promoted through the Federation Framework, and these will directly inform the “data aspects” of the Repository Support Project, in which UKOLN and Southampton are partners.

The project team is highly interdisciplinary, with representatives from the digital library, computer science, chemistry and e-Research communities. This very successful collaboration, now formally includes participation from the wider chemistry community (Unilever Centre at Cambridge, the Science & Technology Facilities Council STFC, Glasgow University, Newcastle University, University of Sydney, ReciprocalNet at Indiana), digital libraries (UKOLN, StORe & ARCHER), digital preservation experts (Digital Curation Centre with STFC), publishers (International Union of Crystallography, IUCr, Royal Society of Chemistry RSC & Chemistry Central), the Chemical Database Service (CDS) and the key learned society & subject oriented bodies in the field (IUCr & the Cambridge Crystallographic Data Centre, CCDC).

I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)	YES (See Appendix B) 	NO
I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)	YES 	NO

eCrystals Federation

1. Introduction

This project will begin to establish a global and sustainable federation of data repositories for crystallography: the **eCrystals Federation**, and builds on and is informed by three previous phases of work from the **eBank Project** (<http://www.ukoln.ac.uk/projects/ebank-uk/>). In particular eBank Phase 3 was a preliminary scoping study for the deployment, interoperation, harvesting and preservation of such a Federation. eCrystals addresses the stated need to “*create[ing] the technical conditions required for interoperability for resource discovery and access to resources held within digital repositories*”. The proposal is submitted to the Repositories and Preservation Programme strand D, section b) Repository enhancement projects. The 15-month project will be led by the University of Southampton, with the core partners at UKOLN (Bath), DCC (Edinburgh) and Cambridge. Funding is sought for **£199,999** from 1st November 2007.

The eCrystals Federation will make a major contribution to the development of a repository-based e-infrastructure for research data, and represents an instantiation of the “Federation Data Deposit Model” described in the *Dealing with Data Report* from a recent study completed by UKOLN. The project will establish a solid foundation of crystallography data repositories across an international group of partner sites, with metadata harvested by a number of aggregator services. Building on the eBank Phase 3 results, partners will work together to harmonise the metadata application profiles from repositories operating on different platforms (EPrints, DSpace, Fedora & ReciprocalNet), investigate aggregation issues arising from harvesting metadata from repositories within the networked information environments in other countries (EU, USA & Australia) and enable the Federation of institutional archives to interoperate with international subject archives (IUCr and CCDC) and other third party harvesters. eCrystals will provide a testbed for the StORe middleware and the CLADDIER “Ping” mechanism to enable the linking of data to derived resources such as primary publications. The project will build on eBank-UK achievements already made in the areas of data curation, preservation and provenance by application to an operational Federation and through the provision of a testbed for the DRAMBORA Toolkit. Co-ordinated advocacy and training approaches, will be promoted through the Federation Framework, and these will directly inform the “data aspects” of the Repository Support Project, in which UKOLN and Southampton are partners.

The project team is highly interdisciplinary, with representatives from the digital library, computer science, chemistry and e-Research communities. This very successful collaboration, now formally includes participation from the wider chemistry community (Unilever Centre at Cambridge, the Science & Technology Facilities Council STFC, Glasgow University, Newcastle University, University of Sydney, ReciprocalNet at Indiana), digital libraries (UKOLN, StORe & ARCHER), digital preservation experts (Digital Curation Centre with STFC), publishers (International Union of Crystallography, IUCr, Royal Society of Chemistry RSC & Chemistry Central), the Chemical Database Service (CDS) and the key learned society & subject oriented bodies in the field (IUCr & the Cambridge Crystallographic Data Centre, CCDC).

2. Project Description

2.1 Background and Objectives

Whilst institutional repository networks which provide managed storage and open access to the textual interpretations of research, are emerging e.g. SHERPA (<http://www.sherpa.ac.uk/>) and DRIVER (<http://www.driver-repository.eu/>), the data repository landscape within institutions is considerably less mature. Well-established community archives such as the UK Data Archive and the EBI sequence databanks for bio-informatics data, provide curated resources in certain disciplines. However, the technical infrastructure and associated support for research data remains fragmented and there are gaps in provision as exemplified by Open DOAR where out of the 76 recorded UK institutional repositories a mere 4 contain datasets. This is against a backdrop of an increasing “deluge” of data generated by both large-scale facilities and institution-based small-science. In addition, the highly social, participative, (and chaotic) constructs of the current Web environment are changing scholarly communications, and we are starting to see scientific data as well as textual information, being shared, discussed and evaluated in blogs and wikis e.g. within the associated R4L Project <http://www.jisc.ac.uk/conference2007/>. This is in contrast to the more formal standards-driven service-oriented architectural approach of the eFramework.

The pioneering JISC funded eBank-UK project (three phases since Sept 2003), has constructed an institutional repository that makes available the raw, derived and results data from a crystallographic experiment (<http://ecrystals.chem.soton.ac.uk>), developed the the eBank aggregator service for metadata harvesting by 3rd parties and promoted the linking from primary data to other research outputs within the

scholarly knowledge cycle (Lyon, Ariadne July 2003). Phase 3 also investigated preservation and curation aspects of the data repository and evaluated approaches to audit and certification. Phase 3 was positioned as a transitional scoping study for the proposed eCrystals Federation, and this bid describes the first stages of full implementation. The results from Phase 3 are currently being assimilated and collated into a series of reports, however there are a number of outcomes which are already evident:

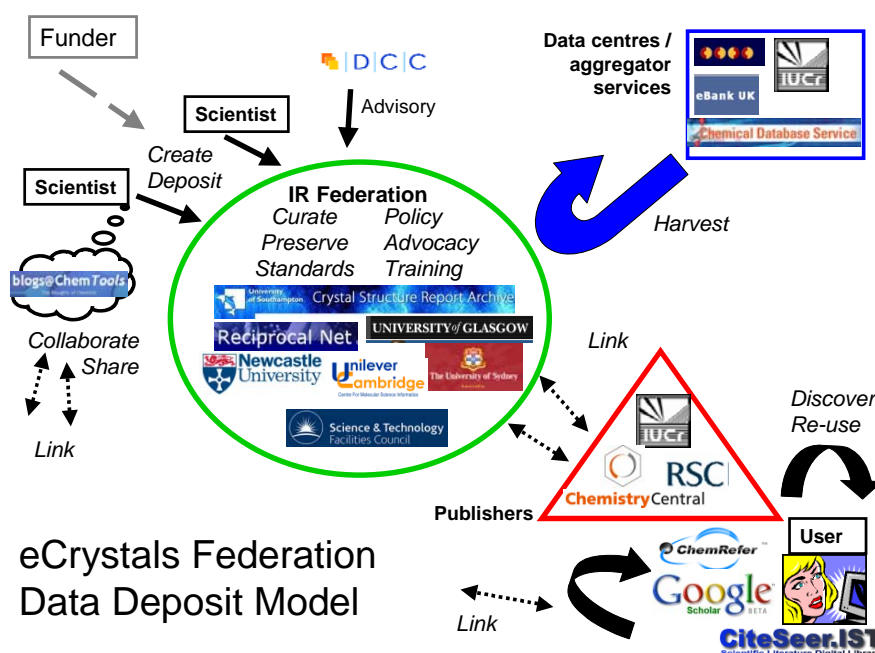
- Crystallographic laboratory practices are very varied, ranging from a more automated workflow with outputs handled and manipulated digitally, to a very “hands-on” process where an individual crystallographer oversees the process and maintains paper copies of results in a filing cabinet.
- This variation in laboratory practice has implications for the ease of adoption of a standard metadata schema such as the eBank Application Profile.
- Crystal structure data and associated information is complex, should be considered as compound objects and will require the use of a metadata packaging format such as METS or MPEG DIDL.
- There are likely to be a range of persistent identifiers in use within any discipline. The allocation of identifiers by the issuing agency must be efficient, reliable and scaleable.
- When considering preservation and curation, these aspects need to be addressed: audit and certification processes and procedures, representation information for crystallography data, preservation metadata for crystallography data, conformance to the OAIS Reference Model of repository software in use within the Federation.
- It is clear that preservation and curation issues will have to be addressed politically by both institutions and the community.
- Advocacy programmes will be essential to assist with populating the data repositories, since there is no established culture of sharing data within the chemistry domain.
- The implementation of a data embargo procedure/policy will be an important factor in encouraging searchers to deposit data destined for eventual open access.
- The pro-active support of professional societies, publishers, data centres and other key domain stakeholders is essential to achieve buy-in from the scholarly community.
- It is unclear as to the exact nature of the relationship between subject-based and institutional repositories and mechanisms for machine to machine interoperability will be necessary.

Building on these outcomes, the three Key Objectives of this project are:

1. To create an operational Federation of data repositories in the crystallography domain thereby testing the effectiveness of the Federation Data Deposit Model, both within the crystallography community, and as a potential framework for other disciplines.
2. To make recommendations on preservation good practice for institutional data repositories
3. To assess the socio-political and technical interactions between subject and institutional repositories and sustainable models for partnership.

The original (Phase 3) model for the eCrystals Federation has been revised (see Figure 1 below).

Figure 1: The eCrystals Federation Data Deposit Model.



In one sense, we can view the Federation as a form of Virtual Organisation (VO), which will develop its own set of governance agreements, policies, practice, behaviours, security and cultural values. It will be essential to develop these community aspects in parallel with the technical infrastructure.

Using the “roles” outlined in the *Dealing with Data Report* as a basis, the Federation partners can broadly be broken down into the following groups:

- 1) Institutions: Repository providers are the federation partners and have expressed their support (see Appendix). They comprise the Universities of Southampton, Cambridge, Glasgow, Newcastle, Indiana (USA, ReciprocalNet), Sydney and ARCHER (Australia) and STFC and represent institution-based repositories. The partners have been selected on the basis of their significance in crystallography, but also because they represent a truly global multi-platform data network.
- 2) Scientists: the individual crystallographers in the laboratory and practising chemists who create the crystal structures as part of their routine workflow.
- 3) Data centres: CCDC is a professional body with a subject repository for crystal data and CDS is a national service that provides federated searching across chemistry databases. They may be considered as the primary data harvesters of eCrystals.
- 4) Publishers: IUCr is the learned society representing crystallography, is a publisher of 8 journal titles and maintains standards for communicating and representing crystal structures. The RSC is a key publisher in the field and Chemistry Central is an emerging Open Access publisher who will operate a repository to store and link data relating to publications in their journals.
- 5) Users: scientists in related disciplines, students and other third parties who have a requirement to use crystallographic data as part of their research.

There are also two additional groups who are associated with the Federation:

- 6) Advisory services: the DCC will provide guidance on preservation and curation practice including the creation of preservation metadata and audit and certification tools. Institutional library and information services will play an important role in the sustainability and preservation of repositories and will be engaged in policy matters from the outset (Cambridge and Southampton University Libraries leading). In addition IUCr has an interest in the preservation of federation data and scoping the operation of a subject repository.
- 7) Third party services: it is expected that third party services will develop across repository federation infrastructure and the model is being developed with this in mind. Whilst such service development will be the subject of future proposals, integration with the StORe project middleware and the CLADDIER Ping mechanism, which provide services to link data and publications, will be implemented as part of this proposal and the eCrystals Federation will act as a testbed.

2.2 Fit to Programme Objectives and Outcomes

This proposal to create the eCrystals Federation of data repositories will contribute to the high-level Outcomes listed in the Circular (para 26), and most specifically address:

“Digital repositories ...to empower institutions to develop populated repositories”: a Report on Successful Advocacy for Open Data will facilitate awareness raising and cultural change amongst scientists.

“Digital preservation...defining services, roles and responsibilities”: The project will work with a range of domain stakeholders including publishers and learned societies to determine appropriate practice.

As an Enhancement project, eCrystals is addressing the statement (para F40) *“there may be specialist community requirements that need to further develop a supported distributed network of repositories for subject specific or media requirements”*, for data in the crystallography/chemistry domain.

2.3 Project Outcomes

The project expects to deliver against a number of broader outcomes:

- Raised awareness and understanding amongst the research community of the benefits of openly sharing data, and of data deposit within a managed IR: eCrystals will contribute through the Federation Advocacy Programme and Dissemination activities.
- More informed repository managers and developers who are/wish to implement data repositories within their institution: the eCrystals experience will be shared with the Repository Support Project and promoted more widely through this activity.
- Enhanced understanding of the critical relationships and interactions between IRs, subject repositories, learned societies and data centres: eCrystals will work with IUCr, CDS and CCDC to

provide mutually beneficial services and support, in addition to scoping the requirements for a subject based repository.

2.4 Value to the JISC Community

The eCrystals Federation is well placed to become the leading “best practice domain exemplar”, illustrating how institutional repositories can be used to effectively provide managed access and preservation capabilities for the stewardship of research data in an institutional context. The early pioneering work of the eBank Project has meant that eCrystals is able to provide a relatively mature data repository testbed, where service integration can be explored, scalability issues can be investigated and there is a discrete community of practising scientists. Because an open and collaborative approach has been adopted by eBank from the start, we are also positioned to build on excellent well-established working partnerships with key publishers, learned societies and domain data services. The Federation will provide JISC with valuable experience of the practical challenges and working solutions, to many of the issues facing institutions in a rapidly-changing scholarly landscape.

3. Work Plan

3.1 Work Package 1: Repository deployment and support

This workpackage will establish a Federation comprising a baseline of six partners with populated data repositories. Appropriate Federation agreements and policies will be established between partners and will be underpinned by a Federation Framework. A phased Roll-out Plan for practical deployment will be produced, with the Universities of Sydney, Glasgow and Newcastle in the first tranche, all using the eCrystals/ePrints software platform. The other partners (ReciprocalNet, Cambridge, STFC and ARCHER), operating repositories on different software platforms, will be introduced at a later stage.

The Southampton RA and project manager will perform site visits to install the e-Crystals software, provide systems administration training, demonstrate the deposit process and advise on methods and policies to ensure and encourage deposit. Uptake by depositors will rely on a number of crucial factors, particularly the establishment of an embargo process that will allow release of the data into the public domain at a time that may be controlled and timed to align with other (related) publishing processes. The publication policy of the National Crystallography Service (http://www.ncs.chem.soton.ac.uk/pub_pol.htm), developed in conjunction with the eBank project whilst using the eCrystals demonstrator repository, will demonstrate best practice and assist in advocacy matters. It is likely that further partners will be engaged during the course of the project, as a result of publicity and dissemination exercises and they will be accommodated as part of the exit strategy.

In order to enhance discovery of Federation repositories and data holdings, a description of the resources i.e the eCrystals data collections, and a list of providers, will be published in the Information Environment Service Registry (IESR). This will be implemented based on the IESR model, where the records are submitted by a registered contributor. For each dataset or collection, a record will be created to describe the data, plus a record for each method of accessing that data ('service' records). The 'agents' (agents own collections and/or administer services), will also be described. The role of record contributors (and editors), within the Federation will be examined, but it is anticipated that they will work closely with the IESR Content Manager (Leigh Morris). This will enable primary harvesters to interoperate with the Federation and also to identify when new members join / new collections are available.

WP1 Deliverables: Operational Federation of repositories. Federation Collections listed in IESR.

3.2 Work Package 2: Repository interoperability and linking services

One of the main aims of WP2 will be to establish a “core” Federation Application Profile, and to achieve full agreement and adherence to the Profile by all participating data repositories. The role of the Information Environment Metadata Schema Registry in this process will be investigated, and mappings between existing schema's will be explored.

The second aim of WP2 will be the formalisation of the relationship between data in federation and subject repositories and scholarly articles. At the June 2007 Digital Deluge conference in Manchester, the JISC confirmed its intention to support a joint development by the StORe and CLADDIER projects and it is proposed that the Federation will directly support this initiative by acting as a testbed for the StORe middleware and the CLADDIER Ping mechanism. For StORe this will provide a second domain in which to trial pilot middleware previously demonstrated only within Social Sciences; it will also afford an opportunity to investigate the functionality of bi-directional links with a 'publisher's repository', which the StORe survey confirmed was a high priority deliverable amongst scientific researchers. For CLADDIER, the Federation will enable a structured environment in which to develop a robust schematic for supporting links between datasets and journal articles. In partnership with domain publishers such as IUCr, RSC and Chemistry

Central, the project will investigate embedding these links into the publishing process, including the potential for reviewers to assess data sets under controlled conditions. Further, the project will explore the use of the notification mechanism from Cladder to determine its potential to propagate cross-reference information between repositories and publishers.

We will also seek to work with the ORE/Pathways Project to scope their machine to machine interoperability protocols as a possible mechanism of interaction between repositories, in particular that involving federation members and a subject repository (Liz Lyon and Les Carr are on the Advisory Committee and the Southampton team are in close collaboration, along with Microsoft, to accelerate the ORE agenda).

WP2 Deliverables: Publication of the Federation Application Profile. A report with an evaluation of the effectiveness of the StORe middleware in the chemistry domain, an analysis of the issues and impact of bi-directional links with a publisher repository, and future options for integrating the StORe middleware in the publishing and research data processes.

3.3 Work Package 3: Repository interactions with data centres, publishers and learned societies

Appropriate harvesting and aggregator services will be tested in partnership with publishers and data centres. eCrystals will work with IUCr, RSC, CCDC and CDS to establish scaleable harvesting mechanisms for data repository records.

To embed the eCrystals methodology into established publishing procedures it will be necessary to work with scholarly literature publishers in the chemistry domain, -predominantly the RSC, but also Chemistry Central and IUCr. This work will establish protocols for associating datasets with publications, timing the publication of these related outputs and mechanisms for validation and/or review. Additionally, the deployment of a data repository, operating on the DSpace platform, by Chemistry Central will be investigated. This will probe the possibility of adding value to the Open Access publishing model by enabling authors to deposit data related to submitted articles.

The project will collaborate with CCDC, who operate as a data centre and form of subject repository, to ensure that the application profile provides sufficient information for accurate, thorough and automated harvesting and the service registry enables their harvesters to identify when a new and appropriate member joins the Federation by introducing an appropriately populated repository. The Chemical Database Service currently provides a federated searching mechanism for all existing small molecule crystallography databases and will harvest data from the Federation. This will provide invaluable feedback into the application profile and harvesting process and provide an exemplar of the possibilities for third party services to be built on the basis of federated data repositories.

Recommendations for good practice will be identified, focussing on their applicability to other domains. The findings will be published as a Disciplinary Case Study. We will also seek to collaborate with the Research Information Network in running a workshop for learned society and publisher representatives to raise awareness of data repositories and data publication beyond crystallography.

WP3 Deliverables: Repository Partnerships: Case Study Report for Crystallography. Publishers workshop.

3.4 Work Package 4: Repositories, Preservation and Sustainability

WP4 will build on the Data Curation and Preservation Work Package Deliverable from eBank Phase 3, (*A scoping study on the curation and preservation issues in the eCrystals Repository and Federation*), which identified several areas relevant to preservation and curation which need to be addressed. These include: audit and certification for long-term repositories; identification of critical Representation Information (which encompasses anything that may be required to convert data into information, including standards, data dictionaries, semantics and software); the identification of critical preservation metadata; and conformance of repository software to the OAIS Reference Model. One of the Report's main recommendations, is that a self-audit be undertaken in collaboration with the DCC using the latest audit instrument to be published, the DRAMBORA Toolkit. Using the eCrystals Federation repositories as a testbed, an assessment of the applicability of the Toolkit will be published as a Report. Appropriate preservation metadata will also be created for use in the Federation testbed based on PREMIS, and other relevant metadata schemas. This metadata will be used to enhance the eCrystals Application Profile. In addition, work will continue to identify and document critical Representation Information for the crystallography domain, as well as to undertake an analysis of repository software platforms in use in the Federation, to understand the extent of their conformance to the OAIS Reference Model (ePrints, DSpace, Fedora, ReciprocalNet platforms).

The University of Southampton has been progressing a co-ordinated approach to data curation policy formulation within the institution through a Working Group, which includes representation of all key institutional stakeholders. We will use the outcomes as a Case Study for the development and promotion of

institutional Data Preservation and Management Policies across the Federation partners. In a complementary study, an investigation of the positioning and sustainable business models for a subject repository will be performed in conjunction with Southampton and Cambridge University libraries. This will develop a role and a roadmap for a self-sustaining subject repository in the preservation process, in collaboration with IUCr. Additionally, building on the technical outcomes of WP2, the replication of repository content with other Federation members for preservation purposes, following the LOCKSS model, will be investigated

WP4 Deliverables: DRAMBORA Assessment Report. Report on preservation metadata within the Federation. Report on representation information for crystallography repositories. Scoping report on the sustainability and delivery of a subject repository for crystallography data.

3.5 Work Package 5: Advocacy and Dissemination

Advocacy services and other support, will be developed by the Federation, informed by an analysis of researcher requirements. This will build on the results from the SPECTRA and R4L Survey, and will complement Repository Support Project advocacy work. Each partner will set realistic targets for populating their repository with data sets, and practical experience in successful advocacy methods and approaches will be shared, and will contribute to wider policy development. A particular focus will be an assessment of the applicability of the chosen approaches for use in other disciplines and communities of practice. The application profile, software and public project documentation and reports will be made openly available via the JISC repository and project website.

An "Open Data Roadshow" will include advocacy workshops for partners focussing on data issues at each site to engage local research teams, which will include members of chemistry departments outside the crystallography unit. A workshop for partners and interested parties will be held in order to evaluate the Federation and discuss approaches for modification and improvement of protocols. An "Open Data Workshop" for the wider community, in particular the learned societies whose membership is the practising scientist and who are also involved in the publishing process, will be held in association with the Research Information Network (RIN) as a component of their on going series of workshops in this area. The outcomes of the advocacy work will be synthesised in a formal Report.

The eBank Project has an outstanding dissemination record and eCrystals will continue this pattern of high profile, high impact presentations at conferences, journal articles and articles in community/society publications. We are specifically targeting the following conferences: ECDL, CNI Taskforce meeting, OAI6, eScience AHM, DCC, BCA, ACA and IUCr congress and an article is planned for *Nature*. The Federation will also have a Web presence, and appropriate social networking software will be utilised for communications and dissemination activities.

WP5 Deliverables: Open Data Roadshow and Workshops. Report on Successful Advocacy for Open Data

3.6 Work Package 6: Project Management

Project administration and management will be carried out by the Project Manager at the University of Southampton. A Project start-up meeting and quarterly face-to-face team meetings will be held with partners, in addition to fortnightly teleconferences/Access Grid sessions to monitor progress. A project plan will be developed and progress reports and updates will be produced for JISC as required.

Strategic direction will primarily be supplied by Simon Coles and Liz Lyon, working with other key domain stakeholders. Crystallographic issues will be primarily overseen by Simon Coles and Mike Hursthouse, e-Science issues will be advised on by Jeremy Frey, technical repository matters will be dealt with by Les Carr and digital libraries concerns will be managed by Liz Lyon.

WP6 Deliverables: Project Plan. Progress Reports. Final Report.

3.7 IPR

The project will comply with the terms of the JISC Funding Agreement. It is expected that whilst most (if not all), outputs will be openly available (with Creative Commons licenses where appropriate), on the project Web site, any intellectual property from resulting research outputs, will be subject to the Copyright, Designs and Patents Act 1988. IPR issues for the Federation will be considered as part of WP1, and will inform the subsequent development of any Federation Framework.

3.8 Sustainability Issues

Sustainability is the main focus of WP4. The Federation is seeking to build a self-sustaining repository Federation model where community participation and engagement provide the primary drivers for data sharing and deposit in an IR. To this end, eCrystals will be proactively supporting the development of both

institutional data preservation policies and domain/community-based approaches, in close collaboration with publishers and learned societies.

3.9 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

4. Project Deliverables and Timetable

WP	Description	Deliverables	Month	Lead Effort + partners
1	Repository deployment	Operational Federation of repositories. Collections listed in IESR.	1-15	Soton, + Cambridge, USyd, RNet, ARCHER, Glas, Newc, STFC
2	Repository interoperability and linking services	Publication of the Federation Application Profile. Publications linking Demonstrator with integrated StORe middleware and CLADDIER Ping.	3-15	Soton, + UKOLN, Cambridge, StORe, CLADDIER
3	Repository interactions with data centres, publishers and learned societies.	Repository Partnerships: Case Study Report for Crystallography. Publisher Workshop	3-15	Soton + IUCr, CCDC, CDS, RSC, UKOLN
4	Preservation, curation & sustainability	DRAMBORA Assessment Report. Report on preservation metadata. Report on representation information. Scoping report on sustainable subject repository.	1-15	DCC/UKOLN + Soton, Cambridge
5	Advocacy and Dissemination	Open Data Roadshow & Workshops. Report on Advocacy for Open Data	1-15	UKOLN + Soton, RIN
6	Project management	Project Plan. Progress Reports. Final Report.	1-15	Soton

5. Risk Assessment

Risk	Level	Likelihood	Contingency
Recruitment difficulties	Medium	Low	Existing staff will be re-assigned where possible.
Loss of a team member	High	Medium	Multiple staff at each site have expertise and skills required.
Lack of engagement of partners	High	Low	Involve partner representatives in project meetings and workshops. Advocacy programme at all sites.
Socio-legal issues	Medium	High	Seek expert legal advice
Problems with eprints.org / e-Crystals and other software	Low	Medium	Re-allocate resource to software development
Aggregation delayed due to insufficient data in repositories.	Medium	High	Identify additional Federation partners.

6. Engagement with the Community, Dissemination and Evaluation

eBank phases 1, 2 & 3 had an outstanding track record for dissemination, with references in articles in the *Times Higher* and in *Science* magazine, papers or contributions at most leading international conferences in the field (AHM, JCDL, CNI, ACS, ECM, BCA, IUCr Congress & Bielefeld) and an article in a special issue of the *Journal of Chemical Information and Modelling*. The outcomes of Phase 3 have informed the *Dealing with Data Report*, JISC Digital Repository Infrastructure development and the Repositories Support Project (where both UKOLN and Southampton are partners). Full details of all journal articles, posters, conference papers, presentations, technical specifications and reports are available on the eBank Project Web site <http://www.ukoln.ac.uk/projects/ebank-uk/>.

WP5 will focus on dissemination and advocacy. Engagement will include contributions to JISC Development events and participation in JISC Programme activities, in addition to promoting eCrystals at international conferences, via the Web and in a range of publications. Full engagement with the crystallographic community is vital to the success of the e-Crystals Federation and social networking software will be used to help to build and develop community participation in addition to publicity and papers at critical conferences; BCA Spring Meeting (Apr 2008; SJC on program committee) and the IUCr International Congress (Osaka, Aug 2008). A self-evaluation of the Federation will be a significant element of the Final Report, focussing particularly on the degree to which advocacy programmes have been successful in creating populated data repositories, the level of interoperability and integration achieved and the collaborative partnerships with other domain stakeholders.

7. Impact

eBank has had a significant impact on promoting open access approaches for research data outputs. The innovative project spawned four JISC-funded projects within the Digital Repository Programme working with data repositories and implementing scholarly linking. The scholarly knowledge cycle diagram (Lyon, Ariadne 2003) was re-used by Lorcan Dempsey at OCLC in the *Pattern Recognition Annual Review* in 2003 and has subsequently been quoted in many keynote presentations. The eBank project was described in an article on e-Science by Professor Tony Hey in *Science* magazine in July 2005 and has featured in the *Times Higher e-Supplement* in May 2006, and in the *Dealing with Data Report*. The eCrystals Federation will build on this track record by:

- Producing a domain exemplar of a federated structure for data repositories – an area of increasing attention and importance.
- These new mechanisms for publishing crystallographic data have been discussed with the publishing community and are likely to be considered by numerous journals as a requirement for publication of related scholarly articles.
- Support for the eBank approach has been forthcoming from the IUCr executive and eCrystals will build on this to make a global impact by developing a sanctioned example of best practice in data publication.
- The Federation is a primary contributor to a ‘white paper’ on preservation of crystallographic raw data through its CrystalGrid participation and will, again, be a demonstration of best practice.
- eCrystals will provide an exemplary approach to the publication of data for other analytical chemistry fields and domains outside the discipline.

8. Budget

8.1 Standard budget

Details relating to wages removed.

JISC contributions			Institutional contributions	
Total Costs:		£199,999	Total Costs:	£226,592

The budget is presented with details of fEC contributions from the core partner institutions and the matched funding provided by both direct staffing allocation (NCS, DCC & StORe/CLADDIER) and work in kind from supporting partners.

8.2 Value for Money and Matched Funding

The eBank Project has delivered excellent value-for-money to date. On three relatively small project grants, eBank has achieved significant impact and been highly influential within the JISC community. Resources have been used efficiently and effectively, leveraging other related activities where appropriate (e.g. R4L, CrystalGrid, DCC), and this pattern will continue in the development of the e-Crystals Federation.

The e-Crystals Federation will comprise four directly-funded core partners and 14 supporting partners. We have obtained quantitative statements of financial support from many of these, whilst others have agreed unspecified staff effort, time and other contributions, which taken together will provide the matched funding required by the Call. By dedicating staff effort to interacting with our project team and working with us on the design, deployment, technical support, management, promotion, population and preservation of their repositories, all of these partners will be contributing a significant amount of resources and development work that in total, equate to a sum that matches the funding from JISC.

Details relating to wages removed.

In addition follow on proposals are currently in preparation and will be submitted in the early phase of the eCrystals project. Discussions are currently being held with Microsoft and NSF and approaches will be made to learned societies such as IUCr, RSC, ACS and IoP. A proposal to form an international committee concerned with the standards required for the communication of chemistry data and built on eBank / eCrystals experience is in preparation for submission to IUPAC (£25K).

9. Partnerships and Supporting Organisations

This bid is led by the University of Southampton, with core partners at UKOLN (University of Bath), the Digital Curation Centre and the Unilever Centre (University of Cambridge). **Supporting partners are:**

The International Union of Crystallography (<http://www.iucr.org/>), a body serving the world community of crystallographers, promoting international standardisation and cooperation in crystallography and enabling international publication of crystallographic research; **The Cambridge Crystallographic Data Centre** (<http://www.ccdc.cam.ac.uk/>), operator of the world repository for all small-molecule organic and metal-organic crystallographic data published in journal articles; **The Royal Society of Chemistry** (<http://www.rsc.org/>), the largest organisation in Europe for advancing the chemical sciences supported by a worldwide network of members and an international publishing business; **The Chemical Database Service** (<http://www.cds.dl.ac.uk/>), provider of a single interface to search all available crystallographic databases (CSD, ICSD, CrystMet); **Chemistry Central** (<http://www.chemistrycentral.com>), an emergent Open Access series of journals, covering all aspects of the chemistry domain; **StORe** (<http://jiscstore.jot.com/WikiHome>), a JISC Digital Repositories Programme project, involved in linking 'source' and 'output' repositories; **CLADDIER** (<http://claddier.badc.ac.uk/trac>), a JISC Digital Repositories Programme project, concerned with linking data to publications; **The Research Information Network** (<http://www.rin.ac.uk/>), an organisation funded to assess and coordinate the provision of research information in the UK; **ReciprocalNet** (<http://www.reciprocalnet.org/>), a well established consortium of partners (US based but also including USyd and NCS) sharing and publishing crystallographic data; **ARCHER** (<http://archer.edu.au/>), a DEST project based on data repositories supporting the research process; **The University of Sydney** (<http://mmsn.net.au/>), a lead partner in the Molecular and Materials Structure Network, (an Australian Research Council funded project); **The Science and Technology Facilities Council** (<http://www.scitech.ac.uk/>), a multidisciplinary research organisation supporting scientists and engineers world-wide through operation of large-scale research facilities and provision of strategic advice to the government on their development; **University of Glasgow** (<http://www.chem.gla.ac.uk/xtal/>), a UK Chemistry department with a strong crystallographic section that supports a wide range of research; **Newcastle University** (<http://www.ncl.ac.uk/xraycry/>), a well established crystallography laboratory with world leading experience in operating a synchrotron based national service and a strong background in data-based publishing.

10. Previous Experience, Capabilities, Key Personnel

The core members of the e-Crystals Federation Project Team represent a solid proven partnership that has been successfully working together on the eBank-UK project since 2003 and has a demonstrable track record for delivering project outputs.

School of Chemistry, University of Southampton

Simon Coles is the Manager of the internationally renowned UK National Crystallography Service (EPSRC) and has led the NCS involvement in the UK e-Science program from the CombeChem (EPSRC) project. SJC is a founder member of the CrystalGrid Collaboratory (e-Science Sisters project, EPSRC) and a co-investigator on the digital repository based projects eBank-UK and Repository for the Laboratory (JISC). SJC is crystallographic editor for three chemistry journals and coauthor of over 300 peer reviewed (primarily data-based) publications. SJC serves on the advisory board of the Research Information Network (Engineering and Physical Sciences) and the council of the British Crystallographic Association and is a member of the Southampton University working group to establish an institutional data preservation policy. **Jeremy Frey** is a Professor in the School of Chemistry and will provide expertise and direction in Smart Laboratories and Notebooks, e-Science and Informatics. **Mike Hursthouse** is a Professor of Chemistry and world leader in crystallography research and service provision and will provide consultation and expertise in crystallographic matters. **Les Carr** is a Reader in the School of Electronics and Computer Science and will provide technical advice on ePrints repository software and innovations.

UKOLN, University of Bath

Liz Lyon is Director of UKOLN, where she supports the development and implementation of the Information Environment, promoting synergies between digital libraries and e-research. She has led the eBank UK project, and is Associate Director (Community Development) of the UK Digital Curation Centre, in which UKOLN is a partner. Her academic background was in Biological Sciences and she has a doctorate in cellular biochemistry.

The Digital Curation Centre

Manjula Patel holds a PhD in computer graphics and animation and has worked on projects related to: digital libraries; metadata; interoperability; digital curation and preservation; the Semantic Web; agent technologies; metadata for cultural heritage and e-Learning; virtual reality and multimedia technologies.

Unilever Centre, University of Cambridge

Peter Murray-Rust is Reader in Molecular Informatics in the Unilever Centre at the Department of Chemistry, University of Cambridge. PMR and his group have worked for several years with leading publishers and related organisations to develop new ways of using primary chemistry manuscripts and are responsible for developing Chemical Markup Language. PMR is Co-investigator on a 4-year EPSRC/eScience project on "Extracting the Science from Scientific Publications" and the JISC SPECTRa-T project.

Key Contact

Dr Simon Coles
Manager
EPSRC National Crystallography Service
University of Southampton
Southampton, SO17 1BJ

Tel: 02380 596722
Fax: 02380 596723
Email: s.j.coles@soton.ac.uk
<http://www.ncs.chem.soton.ac.uk/>

Appendix A: Letters of Support

Letters of support from all partners. **Chemistry Central and University of Sydney to follow.**

Appendix B: FOI form

Freedom of Information checklist.