1. NARRATIVE DESCRIPTION

1.1. EXPLANATION OF THE LIBRARY’S INNOVATION

- DataBank is the confluence of a series of developments that have taken place within Bodleian Digital Library Systems and Services department over a number of years that incorporate existing standards and practices within a technical framework that takes a significantly different approach to development in order to meet our goals of longevity and sustainability. In particular, we place an emphasis on simplicity which is manifest in different ways throughout the project:

  o Minimize Custom Code. Wherever possible, we re-use existing code libraries and avoid re-implementing functions that are already well developed elsewhere. Effort is put into the evaluation of code resources for stability and community support. For example, the POSIX file system provides a well tested platform and semantics for grouping, copying and linking files along with basic access control – this codebase is better tested and more scalable than any repository code – therefore it is used as the baseline storage model.

  o Modularity. Even with careful evaluation, software and hardware becomes obsolete. In order for the overall architecture to be robust and long-lived we have to be able to replace modules as this occurs. Effort is therefore put into defining interfaces and functionality so that components can be more easily replaced over the lifetime of the system. In some cases, it is wise to not use the full functionality of a component if this compromises its ultimate ‘replace-ability’ by creating too many dependencies.

  o Avoid Premature Decisions. The majority of the users for much of our digital content have not been born yet and we cannot realistically anticipate what their requirements may be. To this end, design decisions should always be evaluated in terms of the limitations or inflexibility that they introduce to the system. A key result of this approach is that we do not proscribe metadata standards or file formats more than is absolutely necessary for functional purposes.

  o Scalability. Growth is an inevitable characteristic of digital research outputs so scalability must be considered at the outset. Architecturally, this is most easily achieved through distributed systems - parallelizing storage and processing. However, this approach requires a refactoring of conventional workflows to shorten dependency chains, with an emphasis on structuring tasks that can be carried out asynchronously.

  o Genericness. The underlying data model for DataBank is based on a generalised version of the FEDORA-Commons model in order to accommodate varied content. Objects can comprise multiple files and these files can be in folders which are exposed via the REST API.
- Recoverability. DataBank aims to store complete objects with their metadata in a transparent, unpackaged and uncompressed format on disk. Complete objects can thus be located, extracted and understood simply reading the filesystem from storage nodes without additional software.

- The existence and strength of DataBank has positioned the University of Oxford well in preparation for the Research Data Management and Open Access initiatives and demands which are every increasing.

- The University of Oxford is committed to making the outputs of its research and scholarship available as widely as possible to as many people as possible, as quickly as possible. This aspiration is a key element of The University of Oxford Strategic Plan, which states that we shall ‘Ensure that the fruits of the University’s research activities are exploited and disseminated for the benefit of society and the economy.’ The University also states in its ‘Policy on the Management of research data and records’ the ‘The University of Oxford seeks to promote the highest standards in the management of research data and records as fundamental to both high quality research and academic integrity:’ and ‘The University recognises that accurate and retrievable research data are an essential component of any research project and necessary to verify and defend, when required, the process and outcomes of research. Research data are valuable to researchers for the duration of their research, and may well have long-term value for research, teaching and for wider exploitation for the public good, by individuals, government, business and other organisations, as a project develops and after research results have been published.’ These aspirations and intentions are in line with the current research data management environment and when compared with other higher education institutions, the University of Oxford is at the forefront of provision of an infrastructure to support research data management for its researchers and free access to as much research data as possible where appropriate. The recognition, by the University of Oxford, of the need for infrastructure such as DataBank to support research data management and open access has been instrumental in establishing this leading technology. The recent Research Councils UK and the Business Innovation and Skills funding granted to the University of Oxford, a portion of which has been allocated to the technical advancement of ORA and DataBank, is a further endorsement and investment in the sustainability of these invaluable tools for the University’s growth.

1.2. PUBLISHED MISSION STATEMENT OF THE NOMINATED INSTITUTION

- The mission of the University of Oxford is to achieve and sustain excellence in every area of its teaching and research, maintaining and developing its historical position as a worldclass university, and enriching the international, national, and regional communities through the fruits of its research, the skills of its alumni, and the publishing of academic and educational materials.

[Source: University of Oxford Strategic Plan 2008-13]
1.3. HISTORY OF DEVELOPMENT AND IMPLEMENTATION OF THE PROGRAM IN BRIEF

- DataBank originated in 2008 as an adjunct to Oxford University Research Archive (ORA), our institutional repository, as we had the need to host and disseminate a small number of data addenda to research publications. ORA was FEDORA-based, so it was felt that a similar REST API would be useful to enable easy integration with the ORA user interface – which would be the only route to access DataBank content. Data did not strictly fit the remit of ORA (as it was then constituted) so could not be accommodated within the existing repository. However, for a relatively small volume of content, it was felt that a complete FEDORA implementation was not warranted. At that time California Digital Library was developing its Microservices approach to repositories and the Bodleian extended their approach to produce a more FEDORA-like result. Subsequently, DataBank was used as an underlying component in the 2009 JISC-funded EIDCSR project that looked at curation requirements for research data, the 2010 ADMIRAL project that developed a prototype lightweight data management tool (DataStage) which archived material in DataBank and the 2011 HEFCE/UMF-funded DataFlow initiative to productionise DataBank and DataStage. Most recently, DataBank provides the underpinning of the DataFinder catalogue that is a key component of the DAMARO (DAta MAnagement Rollout at Oxford).

1.4. INTENDED CLIENTELE

- DataBank is a mainstream tool for managing valuable University of Oxford research data assets for a wide variety of users. There are three main categories of user: i) data producers who use DataBank for storage, curation, visibility, citation and delivery of their data; ii) data consumers whose main uses are discovery of and access to Oxford datasets; and iii) administrators and funders who need DataBank for data management, compliance and reporting purposes. Data producers mainly comprise researchers of the University, including postgraduate research students. Data consumers covers a broad group including internal and external researchers, but also any external users and potential collaborators with an interest in Oxford data.

- DataBank is a key services for a) enabling the University to know and understand its research outputs and b) retain a local copy that is not at the risk of reliance on external services for access and preservation and for business intelligence, and c) retains a copy of the University’s IP within the institution.

- Work is underway to identify and assess the benefits and impact of DataBank, one of the main vehicles being the recently run Oxford RDM Survey which will be repeated at intervals. Potential benefits have been recorded together with the evidence required to assess whether or not they have been realized. (see http://blogs.oucs.ox.ac.uk/damaro/2012/12/05/jisc-mrd-benefits-workshop/)

- DataBank will help researchers comply with funders’ policies on storage and access to data. Based on the Oxford RDM Survey of researchers, we anticipate quantities of data deposited in DataBank in the first year of operation to be in region of 1 Petabyte or more. Figures calculated from the responses indicate that around 2400 Oxford researchers are likely to require DataBank services.

- DataBank enables the Bodleian Libraries to deliver a key service to its readers, both Oxford researchers and the wider global research community. Decisions taken by the service designers
about core metadata and item description are key to the future success of the service. Pragmatic solutions have been implemented for mandatory metadata fields to enable discovery and compliance with research funder policies on dataset description.

- Bodleian staff are at the forefront of research data developments, a rapidly evolving area, and as such, are learning and contributing to new practices concerning research data management. This not only involves technical staff, but others such as those involved with provision of a helpdesk service and subject librarians.

1.5. PRINCIPLE PLAYERS

The following members of Bodleian Libraries staff have been instrumental in the development and implementation of DataBank.

- Neil Jefferies Neil Jefferies (MA, MBA) is Research and Development Project Manager for the Bodleian Libraries, responsible for the development and delivery of new projects and services. He was involved with the initial setup of the Eprints and Fedora Repositories at Oxford and is now working on the implementation of a long-term digital archive platform. Neil has been involved in a number of other successful Bodleian Digital Library projects such as the Mellon-funded Cultures of Knowledge Project. Neil is Technical Director of the IMPaCT project, was Co-PI on the DataFlow project and he is an invited contributor to the SharedCanvas DMSTech and Fedora Futures projects. He also represents the Bodleian on the JISC PALS (Publisher and Library Solutions) panel. Previously, he has worked in a broad range of computer-related fields ranging from chip design and parallel algorithm development for Nortel, writing anti-virus software for Dr Solomon’s, and developing corporate IT solutions and systems for several major blue-chips.

- Anusha Ranganathan Anusha Ranganathan has been with the Bodleian Digital Library for 4 years. She is the lead developer of ORA (Oxford University Research Archive) and develops and maintains the core technical components of the digital library infrastructure. Anusha has worked on the design, development and support of Databank, an RDF-enhanced open source data repository which is aimed at the long-term preservation of research data. Having a background in Digital Signal Processing and system testing, she brings in a wealth of experience in developing systems with the aim of archiving and publishing digital assets including various special collections held at the University of Oxford, with adherence to semantic web principles. Previously Anusha worked as a software test engineer while at Oxford Instruments and Box Telematics. Anusha has a B.E. in Electronics Engineering from Bangalore University, India and an M.S. in the same field from Syracuse University, USA.

- Bhavana Ananda Bhavana joined the University of Oxford as part of the ADMIRAL and DataFlow projects in 2010 and moved onto the Bodleian Digital Library team in May 2012 to work on DataFinder as well as continuing her work on DataBank and DataStage as part of the DaMaRO Project. She is an experienced software developer with wide range of technical abilities in web technologies, systems tools development for mainframes, applications development for financial enterprise solutions and software solutions for embedded systems. She has worked in most of the aspects of the software development life cycle from gathering requirements and development to maintenance. She has previously worked at Oracle-India and MicroFocus-UK as a software/applications developer. Bhavana has a Bachelors Degree in Computer Science Engineering
from Visvesvaraya Technological University, India and an MSc in Web Technologies from Oxford Brookes University, UK.

### 1.6. FUNCTIONAL SPECIFICATIONS AND REQUIREMENTS

- The full functional specification for DataBank can be found at ['Fundamental functionality of Databank'](http://www.dataflow.ox.ac.uk/index.php/databank).
- Databank is open source and the software is available for installation by other institutions for evaluation. At present, DataBank is not in use in production at any other institutions but the following organizations are evaluating the software for possible use in the future – Leeds University, Monash University, Queen Mary University of London and Glasgow University. DataBank also featured prominently in the Future Fedora discussions in late 2012.

### 1.7. URLS, PHOTOS, VIDEOS, OTHER MEDIA TO UNDERSTAND THE INNOVATION

- DataBank
- DataFlow
- DataStage
- DaMaRO
- [Open Access Oxford](http://www.dataflow.ox.ac.uk/index.php/databank) and more information to be published [here](http://www.dataflow.ox.ac.uk/index.php/databank) in the near future
- **Databank Poster** - presented at Open Repository 2012 (OR12) conference, Edinburgh July 2012
- **eScience All Hands Presentation** by David Shotton, Dept of Zoology, University of Oxford Sep 2011

### 1.8. PRESS COVERAGE

- N/A

### 1.9. USER DOCUMENTATION


### 2. NOMINATOR’S STATEMENT

**Nominating statement from Dr. Glenn Swafford, Director of Research Services, University of Oxford**

As Director of Research Services, I chair the University of Oxford’s Research Information Management Sub-Committee and also the steering group for two major projects that are developing services to build the research data management (RDM) infrastructure for the University.
As such, I am fully aware of the requirements and challenges surrounding research data management within a large and complex higher education institution.

The Bodleian Libraries have been instrumental in building services and systems to support Oxford’s researchers, particularly in the development of RDM systems. DataBank will be a critical central service that enables this world-class research-intensive university to store, preserve and manage its considerable data outputs. I feel confident having discussed RDM with colleagues at other institutions that Oxford is amongst the vanguard concerning development of systems for managing research data outputs. The innovative nature of this work has been recognised by several competitive funding awards from JISC, a UK national body charged with ‘advancing UK research, education and the student experience through the use of modern digital technology.’

One of the main benefits of DataBank is that it enables Oxford to store a myriad of different types of data outputs which is vital for Oxford researchers. It is not limited to certain data types or to methods of describing data. This is vital for the wide variety of disciplines and outputs that we deal with here at Oxford. In this way DataBank does not act as a barrier to innovation, allowing researchers easily to deposit datasets of any type and constructed as the research requires (such as single datasets or datasets comprising many different types of data as a single package). DataBank allows for any subject descriptions, something which is becoming increasingly important as our researchers employ a wide range of emerging subject specific terminology and descriptions. This facility allows for expansion across new subjects including cross-disciplinary areas.

DataBank provides the means for direct access to Oxford digital research data outputs: the University is committed to the widest possible dissemination of its research and therefore DataBank supports this strategic aim.

Development of DataBank has enabled the university to react promptly to changes in funder requirements concerning research data outputs, knowing that the storage and management requirements will be met. For me as a senior university manager, knowing that DataBank has the potential to interoperate easily with other services and systems as part of a robust infrastructure for research data management is crucial.

Provision of DataBank extends the Bodleian Libraries international reputation for collecting and curating research resources by adding digital research data to its existing collections.

I fully support the Bodleian Libraries’ nomination for this prize and strongly commend DataBank as a world-leading innovative service worthy of this prestigious award.

3. LISTING OF PUBLICATIONS OR REFERENCES
   ❖ N/A

4. LETTERS OF SUPPORT AND/OR TESTIMONIALS MAY BE SUBMITTED BY READERS/USERS, OTHER RESEARCH LIBRARIES, AND OTHERS.
   ❖ N/A