

The data imperative

Managing the UK's research data for future use



A summary of the report of the UK
Research Data Service feasibility study

The 2008 Research Assessment Exercise has confirmed that the UK's higher education research base remains world-class and in many respects world-leading. How are we to ensure that this national capacity for knowledge generation and innovation remains competitive, and that our researchers can continue to benefit from a high quality research infrastructure?

Contents

Foreword	1
Introduction	5
Key findings	7
Features and benefits of the cooperative service model	7
Cost-benefit assessment	9
Governance	11
Conclusions	11
Key recommendation	12



Foreword

The 2008 Research Assessment Exercise has confirmed that the UK's higher education research base remains world-class and in many respects world-leading. How are we to ensure that this national capacity for knowledge generation and innovation remains competitive, and that our researchers can continue to benefit from a high quality research infrastructure?

Part of the answer lies in the effective management of the rich haul of research data captured on a continuing basis by those researchers. Over recent years, the availability, development and application of advanced computing and information technologies have led to huge growth in the volumes of research data being generated. Research data have become a valuable resource that needs to be maintained for future access and re-use if we are to reap the full benefits of the UK's investment in research.

The rewards of managing this resource effectively include:

- The ability to share research data, minimising the need to repeat work in the laboratory, field or library, thus saving time and effort
- The ability to retrieve and compare data from multiple sources easily, leading to powerful new insights
- The identification of new research areas for collaborative study: research data will increasingly be the starting point for new research as well as a key output

The challenge, however, is not just about storage. Data must be structured to allow retrieval, preservation and re-use. This is the hardest part. It requires procedures for effective data management to be in place from the beginning of the research process, ensuring, for example, that appropriate metadata (data about the data) are generated and formatted.

Few research projects can escape this challenge on account of size or discipline. Although the data deluge has been most apparent in large-scale research projects that have exploited grid computing, some of them linked to the UK e-Science programme, small projects also generate important data that need effective management. And advances in technology are transforming research in the arts and humanities and the social sciences just as much as in the physical and biological sciences.

Research funders in the UK and elsewhere are increasingly recognising the value of making research data widely available. Many now require researchers to manage their data outputs proactively and deposit them for future access. The Organisation for Economic Co-operation and Development's (OECD) recent statement¹ on access to the outputs of publicly funded research resonates with this trend.

¹ *OECD principles and guidelines for access to research data from public funding*
Paris: OECD, 2007 <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

The UK government has also acknowledged the issue. The far-seeing report *Science and innovation investment framework 2004–14*² put the development of our national e-infrastructure centre stage, and the subsequent 2007 report from the then Office for Science and Innovation (OSI), *Developing the UK's e-infrastructure for science and innovation*³, scoped it in more detail. Crucially, the development of national capability and capacity for the management of digital research data emerged as a key component of the e-infrastructure. Since these reports were published, other advanced economies have started to make significant investments in this area.

In 2007, the Higher Education Funding Council for England (HEFCE) funded the UK Research Data Service feasibility study under its Shared Services programme, to show how the UK can turn the OSI's vision of a coherent data infrastructure into reality by building on existing investment and good practice. This document outlines the study's main findings and recommendations.

We believe the study has shown conclusively that investment in a national approach to the management and exploitation of our research data will add significantly to the value of our research effort and strengthen our research base in a cost-effective way.

- For the researcher, who may fear that the demands of data management will distract them from the core business of research, the national data service we propose will provide access to advice about data management and to data storage and curation facilities
- For UK higher education, the service will be cost-effective. It will leverage existing expertise and provision and avoid duplication of effort. Our existing data centres, such as the UK Data Archive and the network of Natural Environment Research Council (NERC) data centres, will be able to contribute their knowledge to the development of a truly national approach

We hope that this document will draw the attention of stakeholders and policy makers throughout UK higher education and government to the study, and encourage them to act on its recommendations.

The choices we make in the next few years about the management of our digital research data assets will help to determine how competitive the UK remains in the global knowledge economy.

The full report will be available at www.ukrds.ac.uk



Professor John Wood CBE FREng

*Professor of Engineering, Imperial College
Chair, European Research Area Board
Chair, UKRDS feasibility study Steering Group*



Jean Sykes

*Librarian and Director of IT Services
London School of Economics
Project Director*

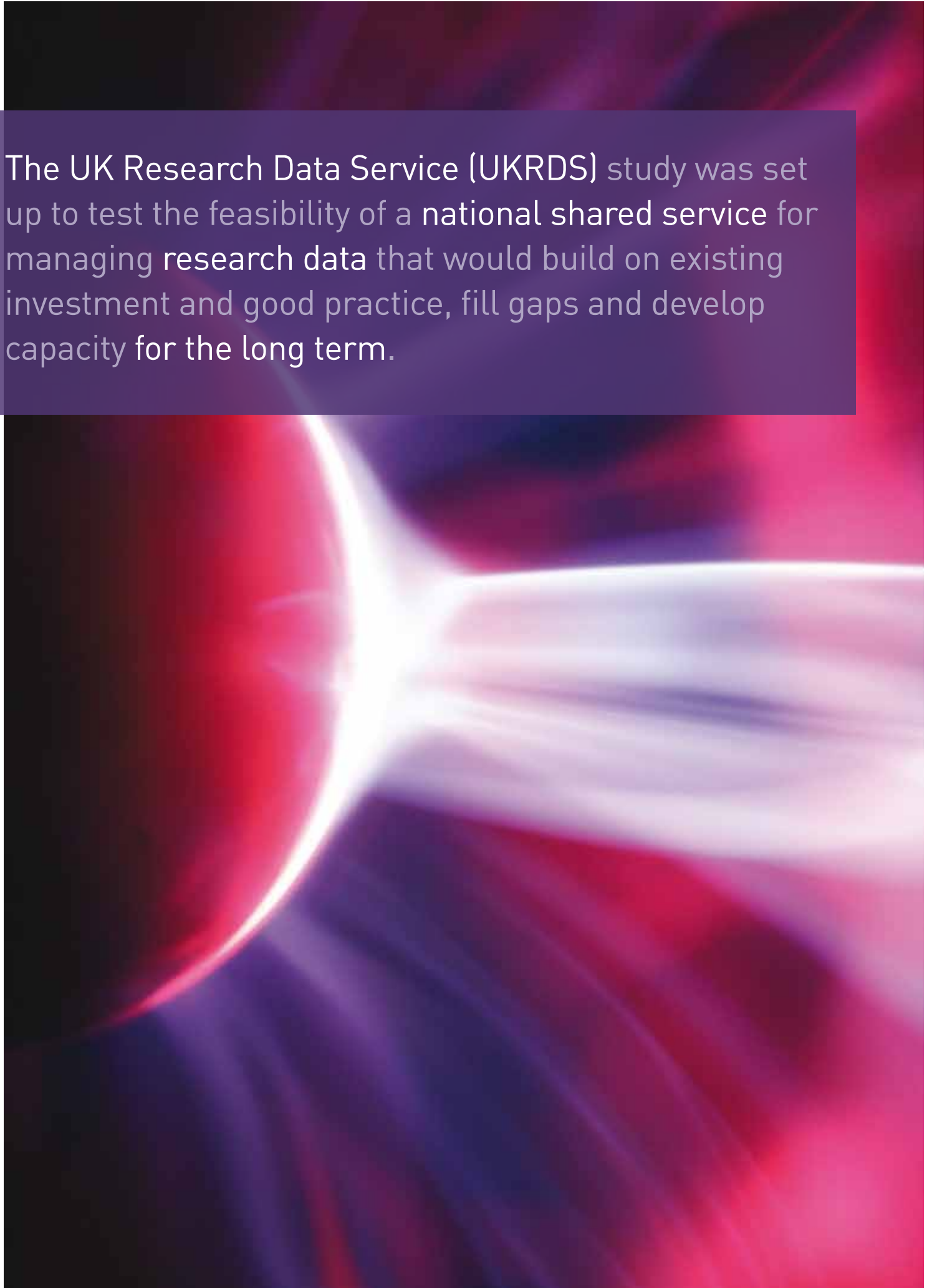
² *Science and innovation investment framework 2004-2014*, Her Majesty's Stationary Office, London, 2004

³ OSI e-infrastructure Working Group, 2007, *Developing the UK's e-infrastructure for Science and Innovation* <http://www.nesc.ac.uk/documents/OSI/report.pdf>

Data must be structured to allow retrieval, preservation and re-use. This is the hardest part. It requires procedures for effective data management to be in place from the beginning of the research process, ensuring, for example, that appropriate metadata (data about the data) are generated and formatted.



The UK Research Data Service (UKRDS) study was set up to test the feasibility of a national shared service for managing research data that would build on existing investment and good practice, fill gaps and develop capacity for the long term.





INTRODUCTION

The UK Research Data Service (UKRDS) study was set up to test the feasibility of a national shared service for managing research data that would build on existing investment and good practice, fill gaps and develop capacity for the long term.

The project team identified three principal options for the future management of the UK's research data outputs:

- 1. No change** The current situation would remain in place with some disciplines well provided for and others not. A UKRDS would not be established, and any additional capacity for managing research data would be provided by individual universities
- 2. Highly centralised** A new, centralised agency would be created, with responsibilities in every area of data management, and with direct responsibility for the provision and management of all new capacity
- 3. Cooperative service** On this model, a UKRDS would be an enabling framework, working with the many UK stakeholders. It would act as a catalyst for new services and partnerships, as a centre of

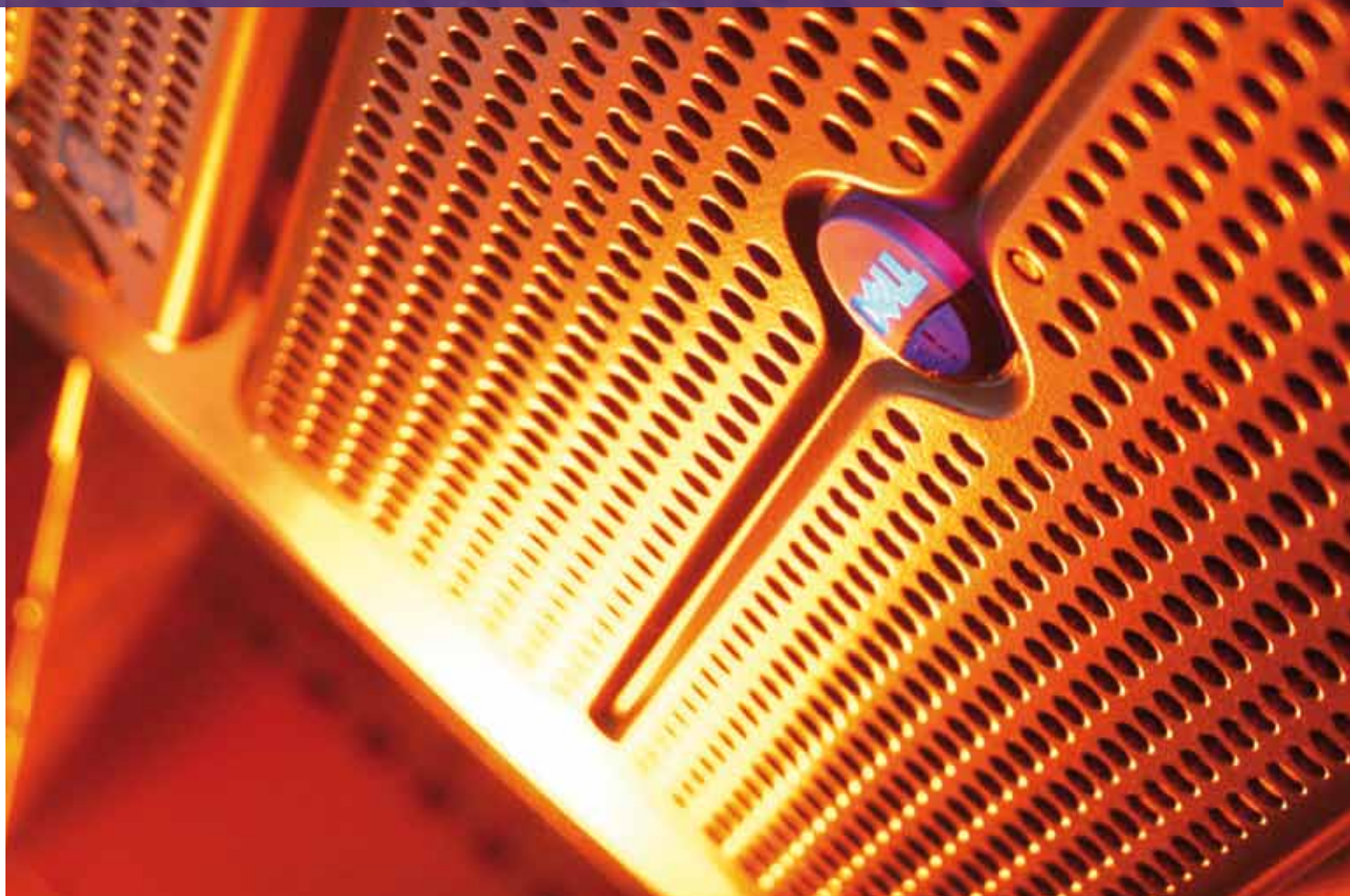
excellence, as a standards-guiding body and as a source of expertise and information about data management and repositories, commissioning additional capacity and building on current best practice and facilities

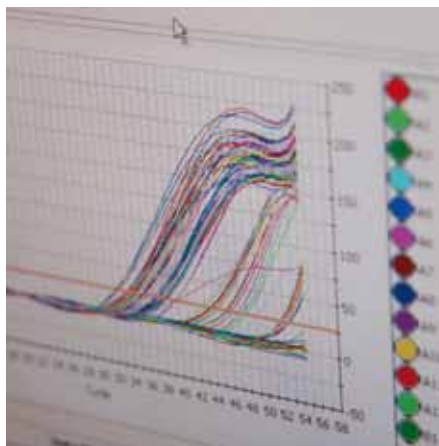
The study embarked on three key areas of work to inform its recommendations:

- 1.** It identified four case study universities (Bristol, Leeds, Leicester and Oxford, which were considered to be representative of research-intensive UK universities) and assessed researchers' requirements through the use of questionnaires, interviews and workshops
- 2.** It engaged with a wide range of stakeholder groups including: major funding bodies, archives and libraries, existing facility and service providers, and others including those involved in this work internationally
- 3.** Wide-ranging desk research was undertaken, aimed particularly at assessing the UK provision in an international context

What researchers said

- An increasing number of disciplines are producing electronic research data
- Research data is difficult to retain and manage once the funding associated with the project ceases
- Most researchers share data – only about 12% do not make their data available in any way. Informal peer exchange networks within research teams and with collaborators are predominant
- 21% of researchers use a national or international facility
- Most research data is stored at faculty or department level
- Only about 18% deposit data with a data centre. However, a much higher percentage (about 43%) expressed the need to access other researchers' data
- Those who did not have access to an established facility were particularly keen on a UKRDS





KEY FINDINGS

A number of issues emerged from consultations with groups representing approximately 700 researchers at the four case study universities (see findings from researchers opposite).

The engagement with stakeholders and desk research indicated that although there is substantial infrastructure and expertise in the UK, it exists in 'islands' with limited coherence and cross-communication because each island was established independently to address a particular problem. These include: significant investments in JISC activities, for example, JISC Services and Collections, the JANET network, federated access management, the Digital Curation Centre and institutional repositories; facilities such as the data centres already operated by some of the Research Councils; and the Research Information Network, which is providing an increasingly important evidence base of researchers' needs.

It is clear that the UK has a sound basis on which to build, but it needs to act soon if it is to maintain its competitive position. Work is already under way in Europe, the USA, Canada and Australia. The study found that the cooperative service model (the third option identified) offers the optimum solution for effective management of the UK's research data outputs.

THE FEATURES AND BENEFITS OF THE COOPERATIVE SERVICE MODEL

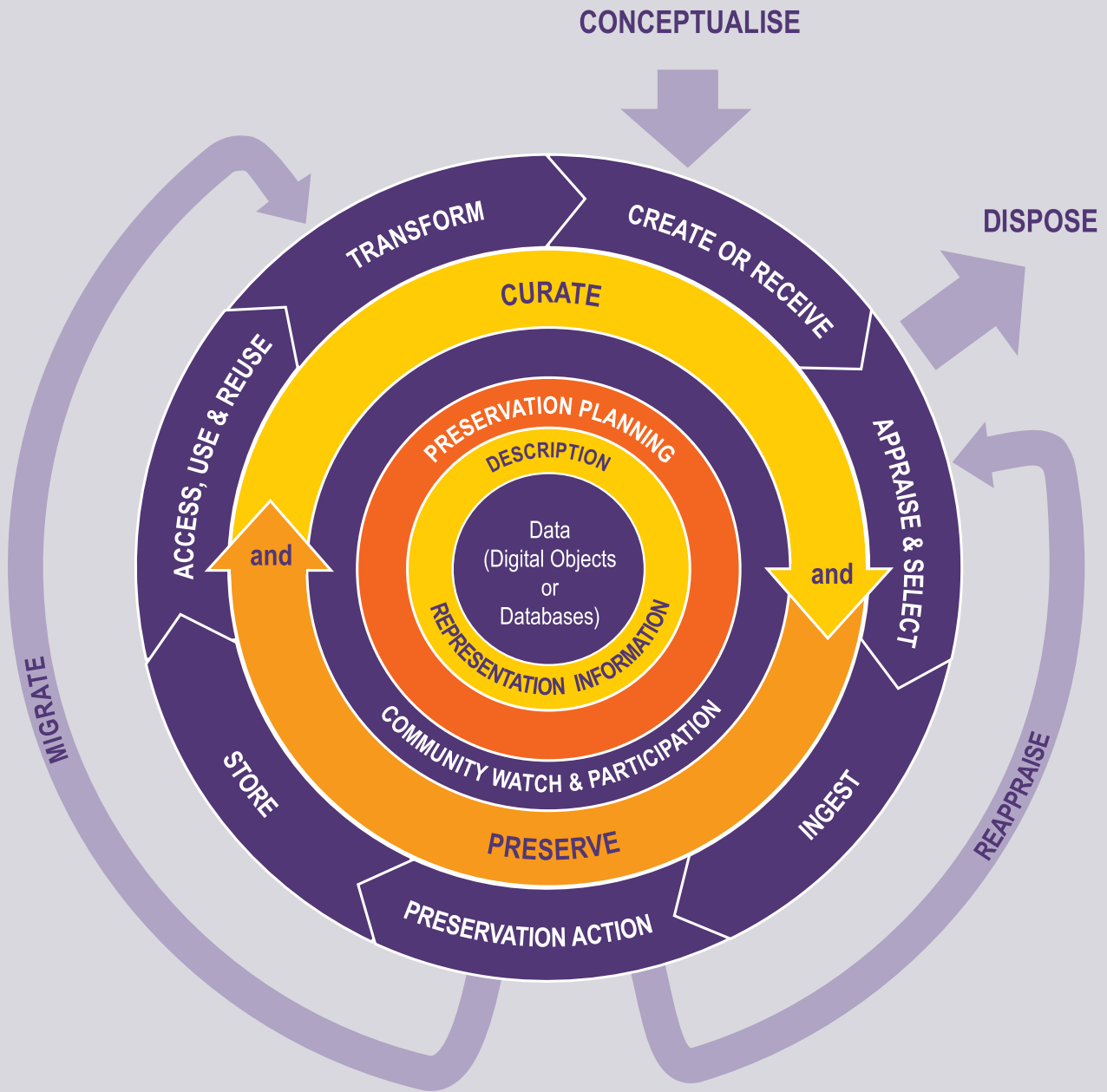
Research costs are growing, and the management of research data is a significant cost. A shared service approach to data management holds promise of minimising the long-term impact and adding value through better exploitation of research data.

Central to the cooperative service model is the development of data management plans by researchers, based on the data life cycle as described by the Digital Curation Centre (DCC) (see over), and the development of a central registry of such plans. This approach would allow a UKRDS to maximise exploitation of existing facilities within the UK, and to identify and fill gaps in current provision.

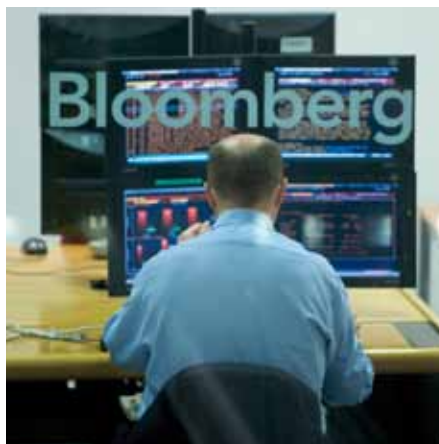
A cooperative service model would have a number of benefits. It would help to:

- Protect and extract greater value from research investment
- Preserve opportunities for future research
- Promote the work of the institution and researcher
- Inform the strategic development of the research infrastructure
- Reduce research data duplication, re-creation and errors, and unplanned data loss

The DCC Curation Lifecycle Model



Digital Curation Centre (DCC) curation lifecycle model This model has already been influential in shaping thinking about the actions required to achieve effective management of digital research data. It has informed the study's work on the capabilities required of a national research data service. www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf



- Plan volume growth/capacity more effectively
- Provide more opportunity for re-use, cross-reference and dataset integration
- Target retention and disposal more appropriately
- Share skills, giving better coverage and productivity in both service providers and researchers
- Provide an effective focus for best practice in data curation

Additional direct benefits to the institution, to the researcher and to funders could include:

- Guidance on which repository to get research data from and a gateway to approved service providers
- Help with the use of data management plans to facilitate life cycle management of datasets
- The opportunity to inform strategic development of the research infrastructure at local and national levels, and work with stakeholders to inform policy and resourcing of post-project long-term data management
- The opportunity to commission new services to fill gaps in data management provision

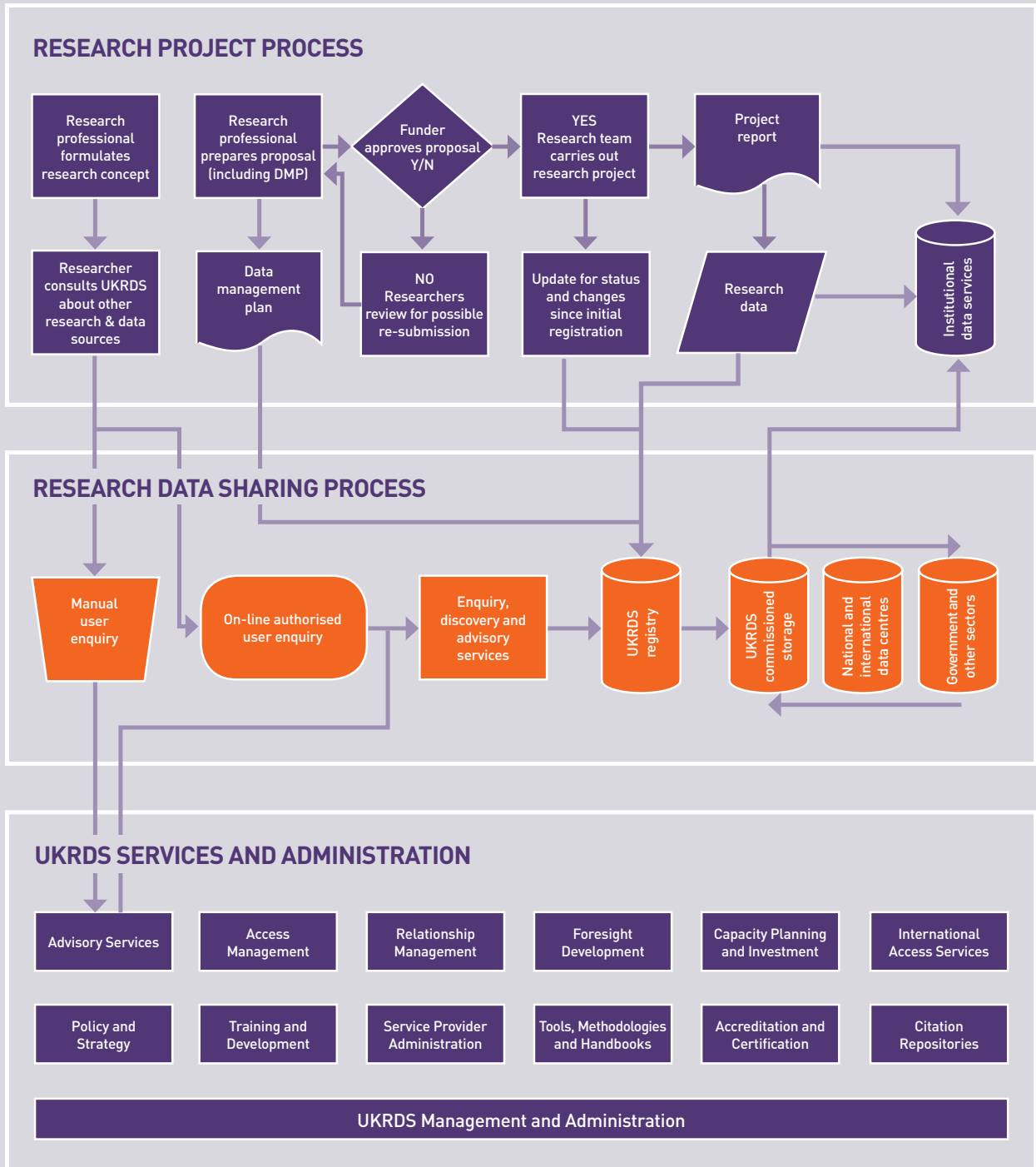
COST-BENEFIT ASSESSMENT

The study developed a model for a phased implementation that would enhance the knowledge base required to scale to a national service.

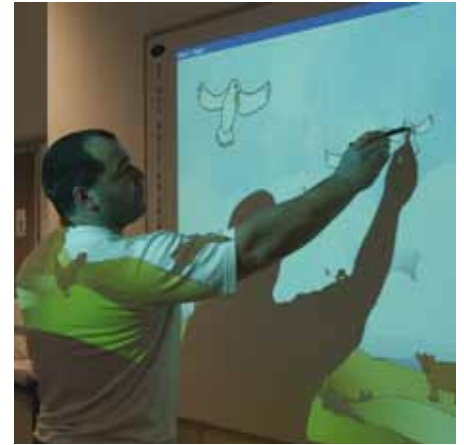
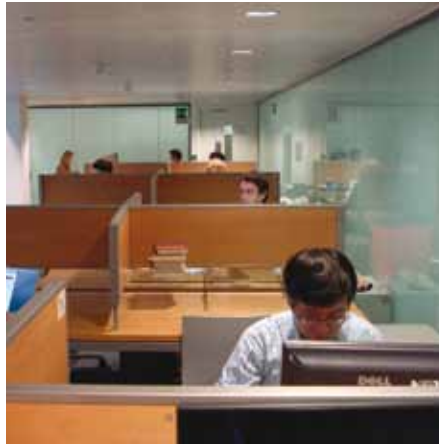
This is based on an initial Pathfinder phase, lasting two years, and involving a relatively small group of key stakeholders, including universities, one or more research funders, and at least one existing national data service. The Pathfinder concept represents an integrated approach rather than a pilot, and implies that a complete service can be implemented with a limited set of stakeholders to test the practical implications of the service delivery infrastructure before scaling up to the whole sector. Scale-up will be driven by demand and constrained by available resources.

The costs and benefits analysis in the full report shows how cost-benefit will build up from the foundations laid in the Pathfinder stage. This analysis shows that, even on the basis of conservative assumptions, a scaled-up UKRDS based on the cooperative service model would generate significant savings to the sector compared with the costs of developing the necessary capacity at each individual institution.

UKRDS Basic Processes



UKRDS process diagram Developed by the UKRDS consultants, this diagram places UKRDS capabilities into a fuller context, including the role played by data management plans (DMPs)



GOVERNANCE

The study recommends that the most effective long-term structure is likely to be a Company Limited by Guarantee with a Board made up of stakeholder representatives. Development of the most appropriate governance model will be progressed further during the Pathfinder phase in consultation with funders and other stakeholders.

CONCLUSIONS

The UKRDS feasibility study has clearly established:

1. The need for a UK-wide approach to research data management
2. The existence of significant gaps in provision that require filling
3. That there are significant building blocks in place with which to develop a UK-wide shared service for maximum cost-effectiveness and efficiency
4. That other countries are making considerable efforts to surface and exploit research data on a national or regional scale
5. That a UKRDS based on the cooperative service model (option 3 above) is feasible and would offer the following strategic benefits:
 - Capitalising on existing investments by providing coherence to extract maximum value from infrastructures and services already in place
 - Providing a shared service across the dual support system and thereby supporting Department for Innovation, Universities and Skills (DIUS) strategies
 - Achieving leverage of research value and output, and increasing the UK's global competitiveness in research
 - A cost-effective approach to the delivery of additional capacity to fill gaps in existing provision
 - Avoiding the opportunity costs of leaving individual higher education institutions to manage rapidly increasing volumes of research data

Conclusion

A UKRDS based on the cooperative service model (option 3) is feasible and would offer the following strategic benefits:

- Capitalising on existing investments by providing coherence to extract maximum value from infrastructures and services already in place
- Providing a shared service across the dual support system and thereby supporting Department for Innovation, Universities and Skills (DIUS) strategies
- Achieving leverage of research value and output, and increasing the UK's global competitiveness in research
- A cost-effective approach to the delivery of additional capacity to fill gaps in existing provision
- Avoiding the opportunity costs of leaving individual higher education institutions to manage rapidly increasing volumes of research data

Key Recommendation

A UKRDS based on the cooperative service model should be considered for funding over a period of at least five years. In the first instance a 2-year Pathfinder phase should be funded.

UKRDS Steering Committee

Richard Boulderstone

Director of e-Strategy and Information Systems, The British Library

Linda Bradley

Head of Research Policy, Department of Employment and Learning (DEL NI)

Mario Campolargo (Corresponding member)

Head of Unit for Research Infrastructure, European Commission

John Coggins

Vice-Principal of Life Sciences, Clinical Medicine and Veterinary Medicine, University of Glasgow, and member of Research Information Network (RIN) Advisory Board

Ian Diamond

Chief Executive of Economic and Social Research Council (ESRC), and Chair of Research Councils UK (RCUK)

David Gani

Director of Research Policy and Strategy, Scottish Funding Council (SFC)

Neil Geddes

Director of e-Science, Science and Technology Facilities Council (STFC)

David Grant

Vice-Chancellor, Cardiff University

Paul Hubbard

Head of Research Policy, Higher Education Funding Council for England (HEFCE)

David Ingram

Director of the Centre for Health Informatics and Multiprofessional Education (CHIME), University College London, and member of Research Information Network (RIN) Advisory Board

Michael Jubb

Director, Research Information Network (RIN)

Simon Kerridge

Head of Graduate Research Support and Assistant Director (Research), Academic Services, University of Sunderland, and Secretary to The Professional Association of Research Managers and Administrators (UK) (ARMA) Board of Directors

David Lynn

Head of Strategic Planning and Policy, Wellcome Trust

Liz Lyon

Director, UK Office for Library Networking (UKOLN)

Tim Marshall

Chief Executive Officer, JANET (UK)

John Milner (Secretary)

UKRDS Project Manager

Wendy Piatt

Director-General, The Russell Group of Universities

Chris Rawlings

Head of Biomathematics and Bioinformatics, Rothamsted Research

Malcolm Read

Executive Secretary, Joint Information Systems Committee (JISC)

Linda Tiller

Senior Research Manager, Research, Business and Communities, Higher Education Funding Council for Wales (HEFCW)

Anne Trefethen

Director of the Oxford e-Research Centre, University of Oxford

Astrid Wissenburg

Director of Communications and Information, Economic and Social Research Council (ESRC)

John Wood (Chair)

Professor of Engineering, Imperial College London

UKRDS Project Management Board

Matthew Dovey

Programme Director, e-Research, Joint Information Systems Committee (JISC)

Martin Lewis

Director of Library Services and University Librarian, University of Sheffield

John Milner

UKRDS Project Manager

Tim Phillips

Director of Information Systems and Computing, University of Bristol

Stephen Pinfield

Chief Information Officer, University of Nottingham

Jean Sykes (Chair)

Librarian and Director of IT Services, London School of Economics and Political Science

Lynne Tucker

Chief Technology Officer, King's College London

Gina Woodward

Associate Director Strategy, Information Services, Cardiff University

We'd like thank all those who have contributed to this work

- Our principal funder, the Higher Education Funding Council for England (HEFCE)
- Our supporting funders, the Joint Information Systems Committee (JISC), Research Libraries UK and the Russell Group IT Directors' Group (RUGIT)
- Serco plc, who undertook much of the detailed research
- Our four case study institutions, the universities of Bristol, Leeds, Leicester and Oxford
- The Project Steering Group and Project Management Board listed above
- Many other researchers, data managers and policy makers who gave their time generously

