

A National Research Data Service for
the UK?
UK Research Data Service Feasibility
Study

John Milner
Project Manager



The potential

“Because digital data are so easily shared and replicated and so recombining, they present tremendous reuse opportunities, accelerating investigations already under way and taking advantage of past investments in science.”

Clifford Lynch, Executive Director of CNI, ‘Big data: How do your data grow?’ *Nature* 4 September 2008



The potential

- And it's not just science.....

The project

- HEFCE Shared Services programme
- JISC also contributed funding
- Joint initiative from RUGIT and RLUK (formerly CURL)
- SERCO appointed as consultants in February 2008
- Project Manager appointed in March 2008
- Governance provided by Steering Committee and Project Management Board (PRINCE2 principles applied)
- Over 40 stakeholder bodies identified
- Ongoing process of communication via mailing lists and personal briefings



What's the problem?

- Data deluge
 - Available/reliable storage
 - Distributed/variable management arrangements
- Research data: an untapped resource
 - Often unstructured and inaccessible
 - Lack of coherent policies and common standards
 - Many formats and software-dependencies
- Varying but increasing funder requirements for data management, preservation and sharing
- HE Library and IT services under pressure to provide solutions at institutional level

What's the problem?

- Whole data lifecycle, not just storage
 - Creation, selection, ingest, storage, metadata, retrieval, preservation
 - Enabling subsequent access, analysis, synthesis, reuse of data
- Link with the published output
- It is the *management* of the data that needs a UK-wide approach

What's the problem?

- Not all research data should be preserved and/or shared
- Examples of added value from preservation and sharing:
 - Clinical trials
 - Observation of a unique event e.g. census data
 - Cumulative value e.g. GenBank, BioBank
- Need for clear mechanisms/criteria for selection

Aims of the project

- Develop an understanding of the UK's current and future research data management needs
- Identify gaps in current services
- Test the feasibility of a UK-wide coordinated approach to the management of research data against fragmented approach
- Avoid reinventing the wheel in any proposed solution

Methodology – case studies

- Four case study universities: Bristol, Leeds, Leicester, Oxford (April to June 2008)
- Questionnaires and focus groups at first three
- Complementary internal project at Oxford dovetailing with UKRDS
- Total number of individuals consulted: 700+

What did we learn?

- Over 360% growth in data volume expected over the next 3 years
- c. 50% of data estimated to have a useful life of up to 10 years
- 26% seen as having indefinite retention value
- Most research data is currently held locally

What did we learn?

- 21% use a national or international facility
- Most share data within research teams
- 18% share data via a data centre
- 43% believe that their research could be improved by access to a wider range of data
- Those with no obvious access to a (inter)national facility are particularly keen on a UKRDS

Methodology – desk research

- Ongoing throughout the project
- Finding out what services already exist in UK
- Speaking to key service providers
- Following initiatives in other countries, notably USA, Canada, Australia, Europe
- Keeping track of rapidly developing area

What are we learning?

- UK well served with national data centres with considerable skills and resources which could be more broadly used
- Digital Curation Centre Life Cycle Model provides a useful approach to data management
- Data management plans (e.g. Wellcome and AHRC) necessary
- JISC's integrated environment and JANET essential infrastructure
- JISC and RIN studies (eg data handling skills, preservation costs) provide some insight and context

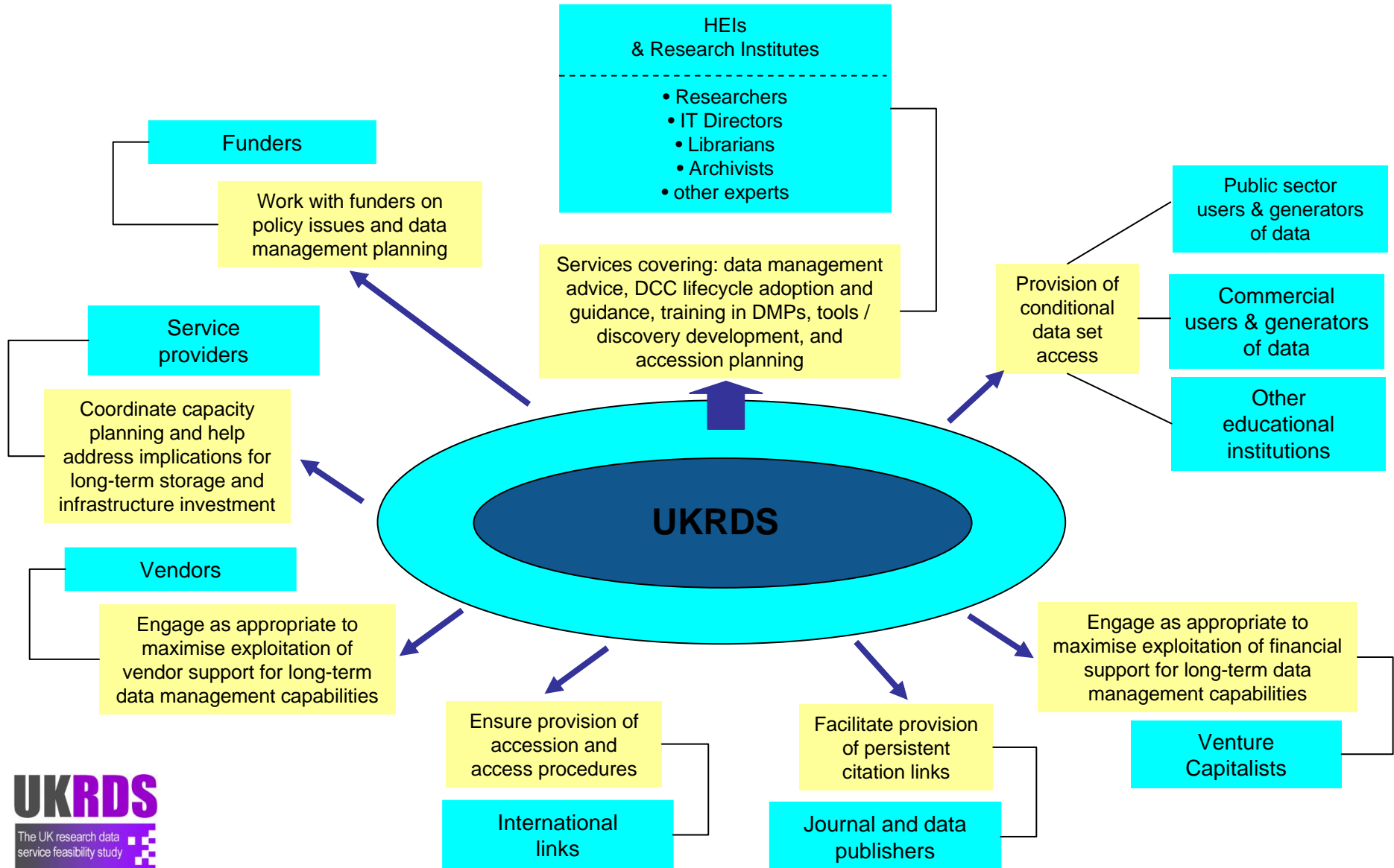
What are we learning?

- The thesis posited in the bid to HEFCE is borne out by the case studies and desk research
- Research data needs managing
- There are major gaps to be filled
- A coherent UK-wide approach involving existing services is feasible
- UKRDS has the potential to deliver greater value to the research community than a fragmented approach

Initial Report Recommendations

- UKRDS could add coherence to what is already there and:
 - Be clear about responsibilities
 - Be governed by a set of policies
 - Deliver a set of processes
 - Support a set of standards
 - Be clear about IPR and reuse permissions
 - Be measured by an agreed set of KPIs
 - Be funded to ensure it can be relied on long term
- Steering Committee approved a “co-ordination” model as opposed to highly devolved or highly centralised

Communities and headline processes



Communities

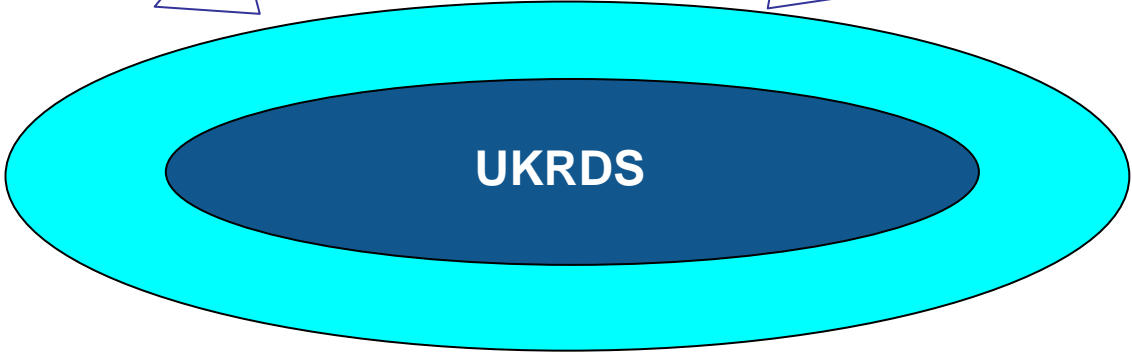
Early Adopters

HEIs & Research Institutes (academic status)	Funders
<ul style="list-style-type: none">• Researchers• IT Directors• Librarians• Archivists• other experts	Service providers
	Vendors
	International links

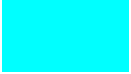
Later Adopters

Commercial users & generators of data	Public sector users & generators of data
Other educational institutions	Venture Capitalists
Journal and data publishers	Others

Processes defined by needs of the Communities

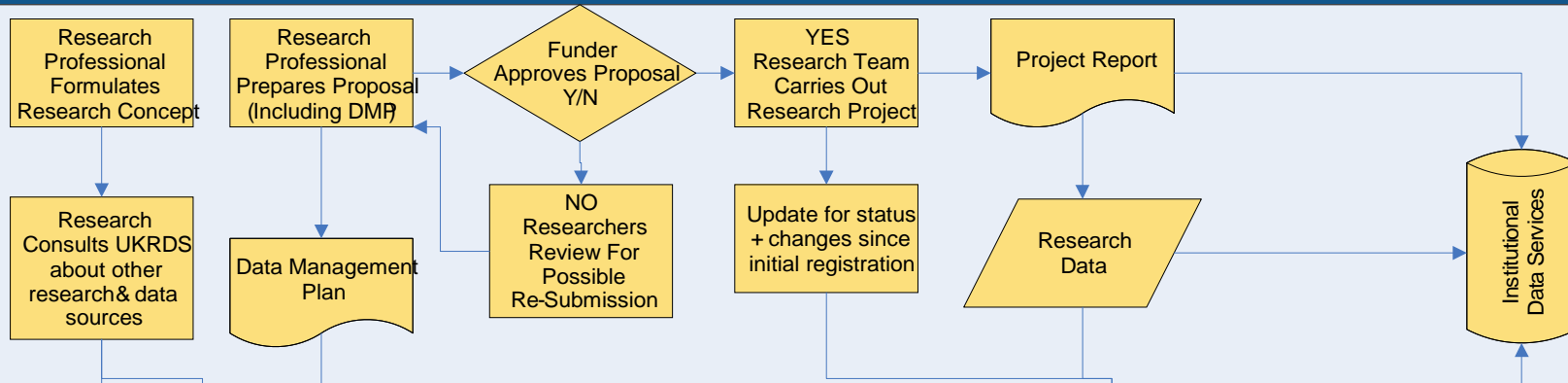


Communities of interest

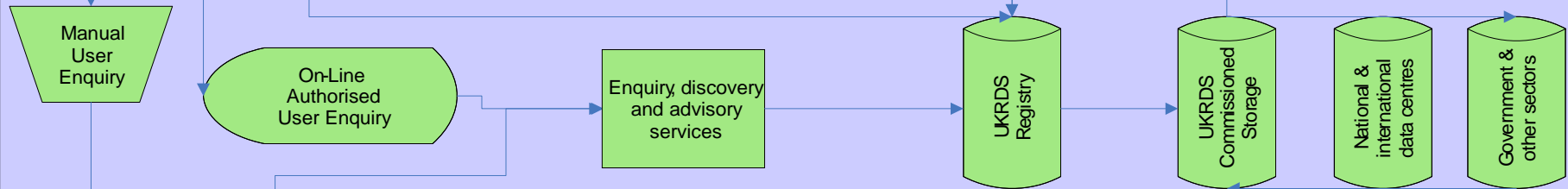


UKRDS Basic Processes

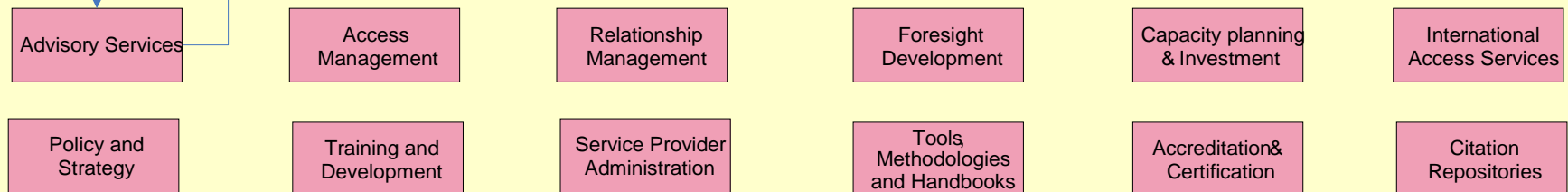
Research Project Process



Research Data Sharing Process



UKRDS Services and Administration



UKRDS Management and administration

Are we in step?

- US model, distributed and NSF-funded: 5 large 'Datanets' (consortia of universities) to build data stewardship capabilities : \$100m over 5 years
- Australian model, centralised and top-down from government: ANDS (Australian National Data Service) with AUS\$ 24m over 3 years
- Similar national initiatives in Canada (Research Data Canada) and Germany

The next steps

- Seek interim HEFCE/JISC funding in 2009 for a 'Pathfinder' service in co-operation with case study institutions and some existing providers (probably DCC RIN and UKDA)
- Hold an international conference 26 February 2009 to promote the proposed service and highlight developments in the UK and abroad

What's the Pathfinder?

- A UKRDS complete but in miniature
- Core is a searchable data management plan registry
- Registry provides focus for long term data set management and basic discovery service
- Avoids risk of trying to “boil the ocean” while testing the concept reasonably realistically

Key messages

- The study has put forward a business case for what researchers are expected to need
- It is not really about storage there's lots of that and it's relatively cheap
- It is about the management of the whole data lifecycle
- It is about selection for preservation and sharing

Key messages

- Many excellent facilities are already in place
- A UKRDS must embrace rather than replace existing facilities adding coherence to the landscape
- There are significant gaps to be filled in support of researchers who don't have a dedicated subject based facility

Conclusion

- A UKRDS is likely to be feasible both practically and financially
- Need support of funders, HEIs and other key stakeholders to get it established and stable for the long term
- The Pathfinder approach is the opportunity for HEI's and established services such as UKDA to make it happen

