# Jisc Project Plan

| Project Information | | | |
|---|---|---|---|
| **Project Title** | Jisc UK Research Data Discovery Service | | |
| **Impact area(s)** | | | |
| **Start Date** | November 2014 | **End Date** | 31 July 2016 |
| **Project Director** | Catherine Grout | | |
| **Project Manager** | Christopher Brown | | |
| **Partner Institutions** | Digital Curation Centre (University of Glasgow), UK Data Archive (University of Essex) | | |

| Document Information | | | |
|---|---|---|---|
| **Author(s)** | Christopher Brown | | |
| **Project Role(s)** | Project Manager | | |
| **Date** | 09-Feb-2015 | **Filename** | Jisc UK Research Data Discovery Service Project Plan v5 |

| Document History | | |
|---|---|---|
| **Version** | **Date** | **Comments** |
| 0.1 | 19-Sept-2014 | Initial draft |
| 1.0 | 10-Oct-2014 | Updated to include RB/CG feedback |
| 1.1 | 30-Oct-2014 | Include RB review |
| 2.0 | 08-Dec-2014 | Updated to include DCC and UKDA feedback and comments |
| 2.1 | 11-Dec-2014 | Minor corrections |
| 3.0 | 19-Feb-2015 | Updates from Jisc/DCC/UKDA meeting 04-Feb-2015 |
| 4.0 | 17-Mar-2015 | Work package updates and budget moved to separate document |
| 5.0 | 27-Mar-2015 | Additional work package for institutional implementation of Discovery Service |
| | | |

## Table of Contents

NB : This table of contents 'auto-populates' - to update the table of contents – place cursor in the table of contents, right-click your mouse, click 'update field', select appropriate option

# 1. Project Overview

## 1.1 Project Summary

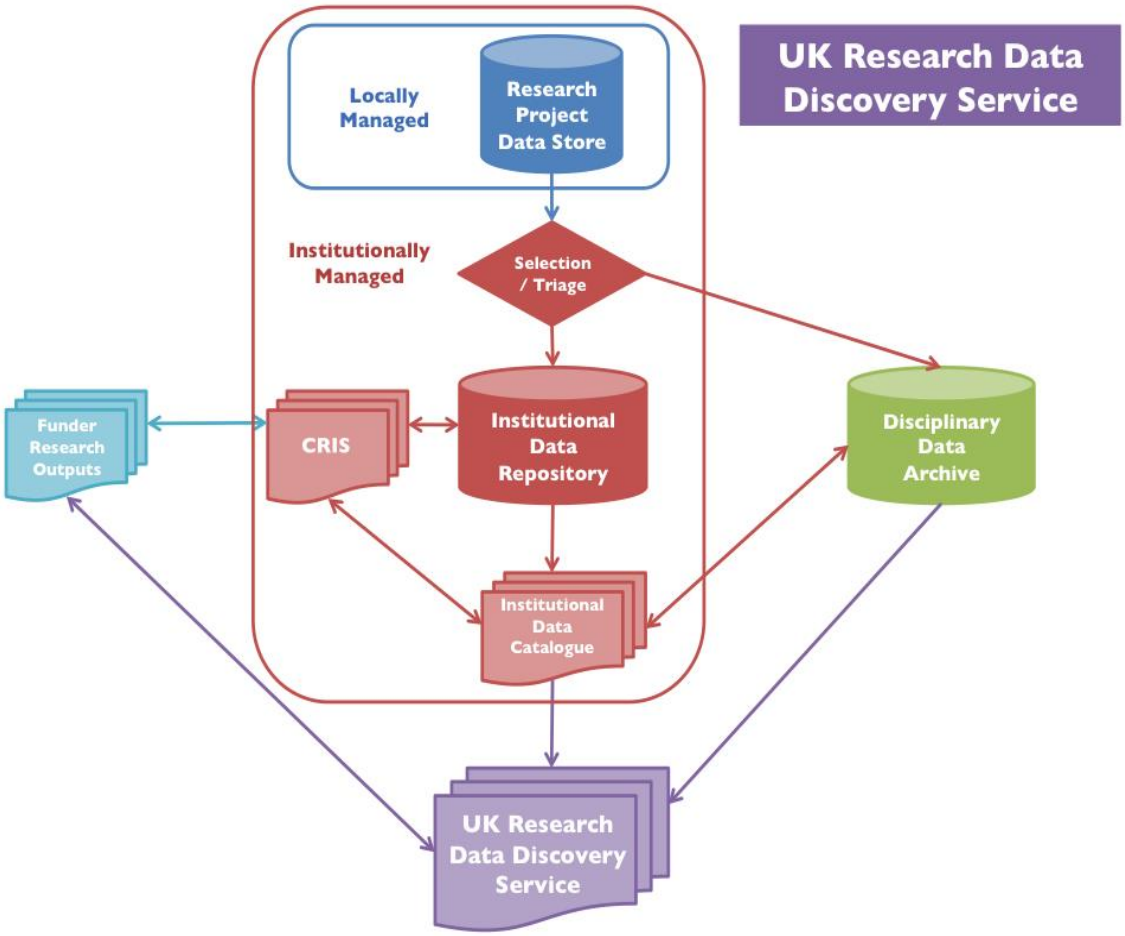
Following on from the initial pilot (phase 1) work, where the Digital Curation Centre (DCC) piloted an approach to a registry service aggregating metadata for research data held within UK universities and national, discipline specific data centres, this project will build on this pilot work with the aim of running a UK Research Data Discovery service. This project (phase 2) will lay firm foundations for the service, including a service operation plan and business case for its delivery into the future.

A research data registry, such as Research Data Australia developed by ANDS, has the potential to encourage the discovery and reuse of those research data assets held by universities. It also helps individual researchers and universities comply with funder requirements to make research data assets – or their records – available, accessible and as discoverable as possible. In the UK context, it is essential that such a registry should also comprise data assets held in national disciplinary data centres, thereby further assisting discovery, establishing links between distributed data collections and mitigating the risk of hardening data silos. Equally importantly, by enhancing potential for discovery and providing an overview of research data produced in UK universities, such a service would be of significant use to research administrators, librarians, data managers, research funders, research policy makers and researchers themselves by providing an important component in Jisc's broader research information infrastructure.

Phase 1 of the pilot engaged the support of a number of Higher Education Institutes (HEIs) and Data Centres to test a Research Data Discovery Service. This six month pilot tested an existing data registry architecture based on the software and metadata requirements of Research Data Australia developed by the Australian National Data Service (ANDS). Its aims were to demonstrate the feasibility of a Research Data Discovery Service for the UK and to develop a better understanding of the optimal technical platform, metadata strategy and harvesting mechanism. An essential feature was to initiate the engagement with stakeholders from the HEIs and Data Centres to ensure the pilot was designed to meet stakeholder requirements.

As the main requirement of this phase is to develop a UK-wide discovery service, it is important to build on the stakeholder engagement, further evaluate the ANDS solution, evaluate an alternative such as CKAN, assess whether any other solutions are potential candidates, and continue the metadata standards work before moving the pilot to a suitable instantiation for a future service. This plan describes the work required that will ensure a service can be produced, that is fully informed by the assessment of these options, and is supporting the requirements of the stakeholders involved.

The diagram below, which was the initial diagram used for the prior pilot project, provides an illustration of the service and possible flows of information, but should not be seen as a precise architecture as implementation may vary in precise detail.




## 1.1 Background and Context

The changing practice of research increasingly requires the data and other sources that constitute evidence underpinning findings to be made available for verification and reuse. While the facility to verify findings has always been a central principle of research integrity, the ability easily to communicate, reanalyse, combine and reuse digital data has led to calls for increasing openness in research practice, for example in the Royal Society's *Science as an Open Enterprise* report.[1] In step with this development, and keenness to see the greatest return on investment in research through the maximum use and reuse of data assets, research funders have issued policies requiring grant holders to make research data accessible to the greatest extent appropriate. Increasingly, too, scholarly journals have published policies recommending, and in many cases requiring as a condition of publication, that data substantiating published findings should be deposited in an appropriate database or archive, or otherwise made available.

Universities have an essential role in realizing these aspirations, and supporting researchers to adjust practice in response to such policies. There are many international, discipline-oriented databases for specific types of data that have been established in response to the needs of specific research communities. The UK has a particularly strong infrastructure of national, discipline-specific data archives (including the UKDA / UK Data Service and the various NERC data centres). However, significant gaps remain. The EPSRC Policy Framework on Research Data lays out

[1] Royal Society 2012, *Science as an Open Enterprise* http://royalsociety.org/policy/projects/science-public-enterprise/report/

expectations that place responsibility for ensuring the preservation and availability of research data with universities and other research organisations in receipt of EPSRC funding.[2] Even before this mandate, it was clear that universities face a significant challenge in supporting good practice management of digital research data through the lifecycle. There are concerns about the availability of robust storage and backup facilities and the capacity (skills, resources, infrastructure) to manage and preserve digital research assets.

Jisc has responded to these developments in a number of ways. At the core of the second Jisc Managing Research Data (JISCMRD) programme were seventeen large projects developing research data services (including institutional policies, support services, repository infrastructure and data catalogues).[3] In this way the programme developed practice and solutions that could be shared beyond these participating institutions. The Digital Curation Centre (DCC) provides guidance and support to UK universities, via a variety of mechanism including more tailored advice in a series of Institutional Engagements.[4] UMF investment allowed the development of tools and the offer of services in the cloud, brokered by Janet, and also the testing of technical solutions for a metadata registry for research data. In phase 1 of the project, the pilot discovery service undertook this work with added stakeholder engagement and support of capital from Jisc.

In order to be re-used, research data must be ***discoverable***. Universities are making research data assets available through repositories or other data portals. EPSRC requires research organisations to maintain a data catalogue. It is likely that some mechanism for aggregation will be necessary to increase visibility and to promote discovery and linking between datasets in related subject areas held in different institutions. Whereas document repositories can, in principle, make articles open to full-text searching by Google, this recourse is not available to data archives relying on metadata.

It is feasible that for some universities a discovery service will be able to fulfil their requirement for a local research data registry/catalogue. This issue has not yet been tested in the pilot, but will need to be assessed as part of this second phase.

A registry solution that aggregates simple, but textually rich, metadata records for research data assets held in Australian universities and data centres has been developed by the Australian National Data Service (ANDS).[5] Research Data Australia provides a discovery service for Australian research data collections.[6] Research Data Australia presents records as web pages and thus promotes the visibility of data resources to search engines. The information architecture establishes connections between data collections, thus promoting discovery. Two important related aims of Research Data Australia are:

1. to break down data silos, encouraging linking and reuse of related data collections, particularly in interdisciplinary research;
2. to facilitate linking data to other research outputs, making data citation and referencing easier, and thereby incorporating data in research achievements and impact.

The ANDS approach has demonstrated success and the software is relatively mature. As UK universities become more involved in the management of research data and capacity develops, the requirement for a UK Research Data Discovery Service has grown.

[2] EPSRC Policy Framework on Research Data http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx
[3] JISCMRD Programme, Research Data Management Infrastructure Projects
http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/infrastructure.aspx
[4] DCC http://www.dcc.ac.uk/
[5] ANDS http://www.ands.org.au/
[6] Research Data Australia http://researchdata.ands.org.au/

Additionally, a data registry has the potential of forming a key part in the infrastructure for measuring and understanding data reuse. Thomson Reuters have launched a Data Citation Index. They are keen to be able to harvest quality metadata from a single source and are already collaborating with ANDS to use Research Data Australia. Thomson Reuters may be keen to collaborate with the Discovery Service pilot in the future. Through the Jisc Research at Risk initiative the possibility of shared research data services has arisen and the development work in this area will need to be considered in this project. Similarly there are other parts of the infrastructure that can facilitate research data re-use, for example research equipment or software. The UK Research Data Discovery Service pilot will need to fit within this context and consider connections between these related pieces of work. However, we must be careful to avoid scope creep and to make links and connections in a planned way.

### 1.2 Aim and Objectives

The main aim of this project (phase 2) is to develop a UK Research Data Discovery Service that enables the discovery of UK research data and meets Jisc's customer requirements. At the end of the pilot, it should be in a position to be taken forward into a service run by Jisc. Phase 2 work will need to ensure that any future service is of high quality and based on a robust architecture and service proposition.

To achieve this aim the project will undertake the following:

- Develop and agree sector requirements for a UK Research Data Discovery Service.
- Develop key use cases, both for human and system interfaces, for the use of the discovery service. These will be developed through the lifetime of the project to test and guide its scope, architecture and functional needs.
- Ensure development is steered by the UK research data community through the project's governance structure.
- Evaluate the ANDS and CKAN software as a potential solution to developing a research data discovery service. This will include the investigation of other potential alternative solutions.
- Collaborate closely with the HEIs and Data Centres[7] from phase 1, where they are willing to participate, to ensure their metadata is harvested and successfully imported into the registry through an easy-to-use interface.
- Identify and finalise the agreement on the metadata standard/profile that is appropriate for a successful cross disciplinary service.
- Identify the architecture that a UK service could operate on and develop a functioning service instance, including the opportunity for institutions to have a localised view into the registry.
- Ingest metadata into a functioning service instance for all participating Data Centres and HEIs.
- Develop the business case for a UK Research Data Discovery Service including evidence based market research and cost-benefit analysis.
- Establish and run stakeholder groups to engage with the community to understand their needs and to help to build an effective solution.
- Evaluate the role of this service as providing institutional infrastructure for data discovery and how it works with universities.
- Ensure the service and user interface has undergone comprehensive usability tests.

[7] Appendix C – HEI and Data Centres Selected

- Produce toolkits, guidance and advice, as appropriate, on implementation.
- Clear articulation of where the UK Research Data Discovery Service sits within other elements of research data infrastructure.

### 1.3 Anticipated Outputs and Outcomes

It's the intention for the project to be run in an iterative and agile way, where work and outputs are reviewed and changed to meet needs. This includes the potential for agile project sprints in the management and method of development.

| **Output / Outcome Type** (e.g. report, publication, software, knowledge built) | **Brief Description** |
|---|---|
| Sector Requirements Report | A report on wider sector requirements for a research data discovery service. |
| Use Cases | Use cases demonstrating the use of the Research Data Discovery Service, reviewed and updated during the project. |
| Workshop reports | Interim reports from workshops held during the project for HEIs and Data Centres to report on issues, progress, etc. |
| Software Evaluation report | Report describing the investigation of alternative solutions and the evaluation of ANDS and CKAN software, detailing any adaptations to UK circumstances and appraising how the software could underpin a future UK research data discovery service |
| Software implementation | Working implementation of chosen software |
| Expert engagement: metadata and technical processes | Expert engagement through the Technical Advisory Group and the Metadata Advisory Group to establish working agreement for the pilot, around relevant software development and implementation, metadata, interchange format and crosswalks. |
| Metadata Development reports | Report summarising agreement around metadata issues (agreed metadata standard/profile, exchange format, crosswalks etc.). |
| Engagement: key UK data centres | Engagement with key UK data centres to harvest metadata records to the pilot Research Data Discovery Service. |
| Data Centre Pilots Implementation report | Report on Data Centre Pilots Implementation. This includes agreed metadata crosswalks, metadata harvest from existing Data Centre catalogue, agreed harvesting workflows and business processes. |
| Engagement: selection of UK university research data repositories | Engagement, through a series of funded phase 2 pilots with a selection of UK universities to harvest metadata records from data collections to the Research Data Discovery Service. |
| HEI Pilots Implementation report | Report on HEI Pilots Implementation. This includes agreed metadata crosswalks, metadata harvest from existing research data repositories, agreed harvesting workflows and business processes. |

| | |
|---|---|
| Business Case for implementing a Discovery Service | Development of a business model presenting options and costs for running a sustainable UK Research Data Discovery Service. With service operational plan. |
| Dissemination outputs | Disseminated outputs to the community via blog posts, Jisc/partner websites, community engagement, presentations, events, etc. |
| Final report | Project's final report including evaluation of the pilot and outputs. |

Note: In addition to the wider engagement with partners resulting from phase 1 liaison activity, there may also be engagement with a wider range of HEIs via Jisc's Research at Risk co-design theme. This will be developed further as the project progresses.

| **Outcome Type** | |
|---|---|
| Improved knowledge: Jisc and stakeholder community | Improved knowledge, in Jisc, its partners and the stakeholder community based on evidence, of requirements, capacity and feasibility of UK research data discovery service. |
| Stakeholder engagement and agreement around metadata schema | Stakeholder engagement and agreement around the development of a national research data discovery service and the metadata schema, other technical standards, and agreements required to underpin it. |
| Increased capacity for Research Data Management: UK universities | Increased technical and organisational capacity in HEIs, for implementation of services that are prerequisites to participation in a UK research data discovery service. |
| Improved knowledge for evidence based decision making | Evidence of likely benefits of research data discovery service, e.g. improved discoverability of research data held in institutional repositories, on which to base business case for service development. |
| Development of a business case for a sustainable production service | Proven customer demand and meeting a market need. Will lead to world class research through the discovery of HEIs and Data Centres' research data. |

### 1.4 Overall Approach

As described under Aims and Objectives above, the main aim of this project (phase 2) is to develop a UK Research Data Discovery Service that enables the discovery of UK research data and meets Jisc's customer requirements. This project will lay firm foundations for the service, including a service operation plan and business case for its delivery into the future. It should be run in an open, agile and iterative way and be user focused.

It is a planned that an agile approach will be used for the overall management of the project and potential service development. Characteristics of this approach will be:

- The use of users stories to define requirements
- An iterative approach to the development of the service
- The production of an indicative release and testing schedule
- Management by sprints - these may be variable in length

Overall governance will use more formal methods to manage issues such as risk and financial control.

In summary, the key objectives of the project are to further evaluate the ANDS software, evaluate the CKAN software (working closely with both development teams), investigate possible alternatives, engage HEIs and Data Centres to harvest their metadata and implement a working discovery service.

Jisc will undertake this project in partnership with the Digital Curation Centre (DCC) and the UK Data Archive. Any service that results from this project will be served as part of the Jisc research data infrastructure.

The initial six month pilot (phase 1), which ran from October 2013 to March 2014, established the initial partnerships and technical requirements for such a service. In the first phase HEIs and Data Centres volunteered to work with the pilot to allow their data to be harvested and populated in the registry. Funds are available, within the second phase of the project, to ensure these groups have a more formal relationship with the project, rather than relying on the goodwill of these organisations. It is important to include those institutions from the first pilot (phase 1) in this phase, where they are willing to participate, but wider engagement is important to ensure adoption of a future service and to ensure the project develops a viable product.

The following organisations were involved in the first phase:

HEIs
University of Edinburgh
University of Glasgow
University of Hull
University of Leeds
University of Lincoln
University of Oxford
Oxford Brookes University
University of Southampton
University of St Andrews

Data Centres

UK Data Archive (one of the partners in the pilot project)
Archaeology Data Centre
The active participation of the seven NERC data centres was achieved via the NERC Data Catalogue Service.

Although these organisations contributed voluntarily in the first phase, a certain amount of funding is available to support their engagement in the next phase and to gather clear requirements from them for the project.

For further information on participants in the Research Data Discovery Service development please refer to Appendix C – HEI and Data Centres Selected.

The governance structure of the project is described in *Section o*

The project team will have access to relevant sources of information and expertise across other Jisc services for appropriate deployment in the development of project outputs.

The project will identify and consider the options for an appropriate destination for the continued delivery of any service developed, which may include Jisc services such as Mimas. Service transfer issues, resilience and infrastructure and guides (code, etc.) may require clarification and agreement to ensure a successful handover of the service for ongoing delivery. Resource should be allocated specifically to these activities in the project
Governance Structure.

The work packages that comprise the project can be found in

Appendix B – Work packages.

### 1.5 Anticipated Impact

The impact from this phase can be divided into immediate and longer term benefits.

Immediate benefits for participants:

- increased knowledge and capacity for implementation to support the registry;
- agreement around a metadata profile;
- supporting development into service;
- any identified cost savings and efficiencies in adapting an existing software solution as opposed to developing the service from scratch.

Longer term benefits for a wider set of stakeholders:

- greater discoverability of research data held in UK universities;
- better compliance with funder policy to make research data easily available and discoverable.

| Impact | Anticipated Impact Description |
|---|---|
| Researchers | A UK-wide service which meets researcher needs may allow increased individual profile within discipline area and may also provide increased research opportunities and / or higher quality research as a result of improved ability to locate and access relevant research datasets. |
| Jisc | Delivering a UK-wide service will ensure Jisc is meeting a key customer requirement for promoting and enabling the discovery of UK research data. Knowledge gained through development of the pilot to a service. Contributing to UK and international metadata standards development. |
| HE institutions | Harvested metadata to increase visibility and discovery of their research data. Increased research profile through engagement as early adopters. Potential benefit to the whole sector from the service, not just the participants. |
| Data Centres | Harvested metadata to increase visibility and discovery of their research data. Increased research profile through engagement as early adopters. |
| Partners | Enhanced reputation, knowledge and skills through developing the Discovery Service. Contributing to UK and international metadata standards development. |

### 1.6 Stakeholder Analysis

| Stakeholder | Interest / stake | Importance (H/M/L) |
|---|---|---|
| Jisc | A research data registry and discovery service would be an essential component in a national research data infrastructure | H |

| | | |
|---|---|---|
| | and one that is currently missing by comparison to ANDS. As operator of the final service ensures sustainability of the discovery service. | |
| Digital Curation Centre (DCC) | Enhance their knowledge and expertise in this critical area of research data accessibility and re-use and be able to deliver related advice. | H |
| UK Data Service (UKDS) | Enhance their knowledge and expertise in this critical area of research data accessibility, discovery and re-use, increased cross-disciplinary data expertise and be able to deliver related advice. | H |
| HEIs holding or developing research data collections | Greater visibility and discoverability of data held in UK universities. | H |
| Discipline-specific Data Centres | Potential greater visibility and discoverability of data held by data centres and promotion of cross-disciplinary reuse. A Research Data discovery service would direct users to data centres rather than take them away. | H |
| Institutional research data managers and research support officers | Improved national coordination of metadata approach helps these stakeholders; supporting universities in better meeting funder requirements and promoting data assets. | H |
| Researchers and their representative bodies | Greater visibility for research data; promotes easier association of traditional publications and research data as called for by the Royal Society and research funders. | H |
| Research Funders | Better overview of data outputs from funded research, focus on detailed information about research data is complimentary to existing research outputs/outcomes systems (Gateway to Research, RCUK Research Outcomes System, etc); potential for increased impact through data reuse. | M |
| Non-HE consumers of research data in for-profit and non-profit enterprises | Greater discoverability of research data held in UK universities and elsewhere, which may have use outside the HE sector. | L |
| International organisations, e.g. RDA | A UK-wide Discovery Service and outputs from this work will be of interest internationally as a potential solution in other countries. | L/M |

### 1.7 Related Projects

The Discovery Service project is related to Research Data work undertaken by Jisc and the DCC. This includes the Research at Risk Jisc Co-design challenge, the Jisc CASRAI-UK pilot (in particular the Data Management Planning Working Group), Research Information Management projects and other work coordinating metadata profiles (e.g. RIOXX). Links with the equipment.data and data.ac.uk initiatives based at the University of Southampton can be facilitated via Jisc working with DCC as part of the strategic development of the Jisc data service infrastructure (however, Southampton is already involved with the discovery service work as an active partner in phase 1).

International links (in addition to ongoing interaction with ANDS), prioritised via Jisc research data infrastructure discussions, offer opportunities for access to developments such as the US-based SHARE initiative.

It is anticipated that discussions and progress on metadata in the project will be shared widely and forms parts of the wider Jisc Research at Risk metadata work.

### 1.8 Constraints

The following constraints will need to be managed in order for the project to successfully complete:

- Resources – ensuring the resources are available to deliver the project's outputs, including developer expertise, business case expertise, metadata expertise, engagement expertise, communications expertise, etc. Ensuring there is appropriate communication between all project partners including clear and open sharing of information to support the project's ability to progress efficiently and overall likelihood of success.
- Agreement – lightweight but formal agreements between Jisc and the HEIs and Data Centres will be required to ensure they are fully aware of their requirements to deliver their metadata to the project and work closely with the project throughout.
- Engagement – through the User Group to the HEIs and Data Centres is required. Project activity should also include engagement with the UK research community to ensure they are prepared for the Discovery Service and with international metadata development. ANDS and OKFN engagement is key to the evaluation activity that underpins the successful delivery of the Discovery Service.

### 1.9 Assumptions

Assumptions include the availability of key staff, in particular development staff, the willingness of HEIs and Data Centres to be fully involved with the project throughout its lifetime, the help and support of ANDS and OKFN with the evaluation of their software and assistance with service development, the technical infrastructure to operate a working service is available.

### 1.10 Risk Analysis

| Risk Description | Probability (P) | Severity (S) | Risk Score | Detail of action to be taken |
|---|---|---|---|---|

| | **1 – 5 (1 = low 5 = high)** | **1 – 5 (1 = low 5 = high)** | **(PxS)** | (mitigation / reduction / transfer / acceptance) |
|---|---|---|---|---|
| Delay to project start up | 5 | 3 | 15 | The pilot concluded spring/summer 2014, although final reports were delivered after summer workshops. It's important now to ensure the project is initiated as soon as possible to remove this risk. |
| Ensuring key developer expertise is available | 4 | 5 | 20 | Clear agreement on roles and resources between partners, including availability. |
| Lack of engagement from Data Centres | 2 | 4 | 8 | Data Centres have been engaged in the pilot project and have expressed a desire to continue their involvement in developing a research data discovery service. Clarifying the DCs involved and issuing grant letters early in the project will remove this risk. |
| Lack of engagement from HEIs | 2 | 4 | 8 | HEIs have been engaged in the pilot project and have expressed a desire to continue their involvement in developing a research data discovery service. Clarifying the HEIs involved and issuing grant letters early in the project will remove this risk. |
| Issues in continuing metadata schema or interchange format development | 2 | 3 | 6 | The pilot project developed a number of cross-walks to convert metadata into the ANDS format. Further ANDS and new CKAN cross-walks will be required. Metadata Advisory Group and Technical Advisory Group need to ensure this work is progressing and fully tested. |
| Lack of ANDS/OKFN engagement | 2 | 5 | 10 | Whichever solution is adopted engagement, in particular help and support, from these organisations is key to the delivery of a discovery service within the timeframe of the project. |

| | | | | |
|---|---|---|---|---|
| Inability to specify a sustainable business case | 2 | 4 | 8 | Ensuring up-front in the plan what the essential elements are and key resources are recognised. There may be a need to deliver the full business case following the project BUT progress should be made throughout so the options and information are available. |

### 1.11 Technical Development

The pilot phase of the project demonstrated that it is possible to redeploy the ANDS software to harvest UK-based research data repositories and data centres, and in order to do so, to create crosswalks to convert other metadata formats to RIF-CS. Lack of a standard metadata schema across all pilot participants means that crosswalks would be necessary for most participants regardless of the metadata format being used by the software.

The technical development in this phase will be based on either an implementation of the software used by ANDS for their research data discovery service[8], the CKAN open source data portal[9], or another solution that might be identified through engagement with the community. If there are other solutions available it's important to justify why they have, or haven't, been selected for further evaluation in the final business case. Assuming the choice is between ANDS and CKAN, it will require continued ANDS evaluation and the start of an evaluation of an alternative, CKAN. The core ANDS registry component[10] is available on Github. Additionally, their OAI-PMH Harvester component[11] is a Java application. CKAN[12] is available on GitHub and documented at http://docs.ckan.org. The CKAN JSON API guide is available at http://docs.ckan.org/en/latest/api/index.html.

In the pilot, the ANDS components were deployed using Microsoft's Azure cloud platform and made available via the DCC website. A sustainable solution to hosting the Discovery Service, where Jisc provides the cloud infrastructure, will be explored as part of this project. The option for institutions to filter the contents of the Discovery Service, so it appears to be a local catalogue, should be included and tested.

The Technical Advisory Group will provide oversight to the technical development of the project.

### 1.12 Standards

| Name of standard or specification | Version | Notes |
|---|---|---|
| RIF-CS <http://services.ands.org.au/documentation/rifcs/guidelines/rif-cs.html> | version 1.4 | Registry Interchange Format - Collections and Services (RIF-CS) |

[8] http://researchdata.ands.org.au
[9] http://ckan.org/
[10] https://github.com/au-research/ANDS-Registry-Core
[11] https://github.com/au-research/ANDS-Harvester
[12] http://github.com/ckan

| | | |
|---|---|---|
| | | Schema used internally by ANDS software |
| GEMINI <http://www.agi.org.uk/uk-gemini/> | version 2.1 | Used by NERC data centres |
| DDI <http://www.ddialliance.org/> | version 2.5 | Used by UKDA |
| Dublin Core <http://dublincore.org/> | version dated 2012-06-14 | Used extensively as a common denominator in most metadata applications |
| CERIF <http://www.eurocris.org> | 1.5 | |
| DataCite <https://www.datacite.org/> | 3.0 | Used by ADS and some other repositories registering DOIs for their datasets |
| JSON | | JavaScript Object Notation is a lightweight data-interchange format used by CKAN's API |
| EPrints <http://www.eprints.org/> | | Used by HEI repositories implementing ReCollect plugin for EPrints |

### 1.13 Intellectual Property Rights

The project will follow Jisc guidelines: IP created by the project will reside with the creator, all outputs will be licensed as openly as appropriate and possible (e.g. CC-BY) except where the reuse of existing IP dictates otherwise.

## 2 Project Resources

### 2.1 Project Partners

| Partner | Contact | Role |
|---|---|---|
| Jisc | Rachel Bruce / Catherine Grout | Senior Responsible Owner |
| Jisc | Christopher Brown | Project Manager for the holistic set of activities |
| DCC | Kevin Ashley | Senior Project Oversight, DCC |
| DCC | Alex Ball | Metadata development, DCC |
| DCC | Laura Molloy | DCC Project Manager & HEI engagement |
| UKDA | Veerle Van den Eynden | Data Centre engagement |
| DCC | David Wilson | Technical evaluation and implementation |
| HEIs | User Group | Project participation through grant funding |
| Data centres | User Group | Project participation through grant funding |

The project team will have access to relevant sources of information and expertise across other Jisc services for appropriate deployment in the development of project outputs.

The project will identify and consider the options for an appropriate destination for the continued delivery of any service developed, which may include Jisc services such as Mimas. Service transfer issues, resilience and infrastructure and guides (code, etc.) may require clarification and agreement to ensure a successful handover of the service for ongoing delivery. Resource should be allocated specifically to these activities in the project

### 2.2 Governance Structure

It is important that members of selected HEIs and Data Centres are represented in, and engaged through, a user group. Due to the technical and metadata development of the project, including software development and metadata updates, it's equally important that the developers working in the project and the developers of the potential solutions (ANDS and CKAN) are represented in a technical group. Similarly, metadata expertise is valuable and should be represented in a metadata advisory group. The provision of a separate advisory group for these areas allows a nimble approach and more focused input, but it is recognised that the technical and metadata advisory groups might need to work in parallel on specific activities. It will be a priority of the relevant project team staff members to maintain good communication between all elements of the governance structure.

1) Oversight Group – representatives from Jisc and partner organisations to lead the direction of the project.
2) User Group – not researchers but people sharing catalogues and submitting data.
3) Technical & Metadata Advisory Group
    a. Looking at the service from a technical standpoint. Scope includes consideration of issues such as handling duplicates, deletions, choice of crosswalks for support, QA of crosswalks, transport mechanisms (e.g. OAI-PMH), and other relevant issues. Comprised of developers and architects within the project, plus developers from ANDS and CKAN and relevant technical experts in participating data centres and HEIs.
    b. Advising on the development of the metadata schema, including the necessary and desirable metadata elements to achieve discovery functionality and which conventions should be adopted when using these and other relevant issues. Comprised of metadata experts from within the project and relevant metadata experts in participating data centres and HEIs.
4) User Group – researchers. As the overall aim of the project is production of a service to provide improved discoverability of research data for reuse in research, it is critical that we provide a mechanism for researchers to interact with and feedback on the development of the service. This may be achieved by representative bodies and / or nomination of researchers by project partner institutions.

### 2.3 Project Management

The management of the project will be undertaken by Jisc (see Project Roles below). The Project Manager is responsible for the day to day running of the project and ensuring outputs are delivered by the end of the project.

The DCC and Jisc will work together as a project team where the Jisc Project Manager and the DCC Project Manager have regular contact and updates on project progress and issues.

The project will generate reports and documentation as required by the main partners and the Advisory and Oversight Groups.

The project will necessarily evolve over time and there should be regular check points every 12 weeks on progress and direction.

### 2.4 Project Roles

| Team Member Name | Role | Contact Details | Days per week to be spent on the project |
|---|---|---|---|
| Rachel Bruce | Deputy Chief Innovation Officer | r.bruce@jisc.ac.uk | 0.5 |
| Catherine Grout | Project Director | c.grout@jisc.ac.uk | 0.5 |
| Christopher Brown | Project Manager | c.brown@jisc.ac.uk | 3 |
| Kevin Ashley, DCC | DCC Director, Oversight of DCC input & advice to the project | kevin.ashley@ed.ac.uk | 0.5 |
| David Wilson, DCC | Registry Developer | David.Wilson@glasgow.ac.uk | 3.5 |
| Alex Ball, DCC/University of Bath | Metadata Coordinator | a.ball@ukoln.ac.uk | 3.5 |
| Veerle Van den Eynden, UKDA (+colleague) | Data Centre Coordinator | veerle@essex.ac.uk | 1(+1) |
| Laura Molloy, DCC/University of Glasgow | HEI Coordinator/DCC Project Manager | Laura.Molloy@glasgow.ac.uk | 4.5 |
| Mike Jones, Jisc Manchester | Longer term hosting/infrastructure for service provision | mike.jones@jisc.ac.uk | TBC |
| Mike Webb, Jisc | Longer term hosting/infrastructure for service provision | mike.webb@jisc.ac.uk | TBC |

### 2.5 Project Support

Project support is primarily undertaken within Jisc. Where required, support will be obtained from partners and members of the oversight group.

## 3 Detailed Project Planning

### 3.1 Evaluation Plan

| Timing | Factor to Evaluate | Questions to Address | Method(s) | Measure of Success |
|---|---|---|---|---|
| Ongoing | Project management | Is project on target? | Check progress against plan | Each stage of the project delivers on time |
| Dec 2014 | Confirm ANDS and CKAN commitment to supporting the project | Are both organisations committed to help and support project? | Engage with organisations and formalise | Organisations commit to offer help and support throughout lifetime of project |

| | | | | |
|---|---|---|---|---|
| | | | project participation | |
| Dec 2014/Jan 2015 | HEI/Data Centres' agreements | Have agreements been reached and understood? | Grant letters signed | All grant letters signed and organisations engaged with project |
| April/May 2015 | Workshop | Are the aims of the project and roles fully understood? Are all partners, ANDS, OKFN, HEIs and Data Centres fully engaged? | Workshop report. Assessment by Advisory Groups | Stakeholders engage. Requirement agreed. Use cases defined. Software Evaluation started. |
| As required | Virtual Workshops | Is the project progressing to plan? Are there any issues? | Advisory Group feedback and workshop update | Project is progressing to plan |
| Mar 2016 | Workshop | At end of implementation phase for HEIs and Data Centres to discuss project progress, lessons learned, service implementation. Is the project progressing to plan? Are there any issues? | Advisory Group feedback and workshop update | Project is progressing to plan |
| Mar 2015 | ANDS/CKAN evaluation (including any alternatives) | Is there enough information available to decide on which one to choose? Is either viable as a discovery service? | Evaluation report and decision by oversight and advisory groups | Selection is made for use in developing Discovery Service |
| Dec 2015 | Crosswalks | Have crosswalks been developed to populate the discovery service with HEIs / Data Centres metadata? Are they fit for a production service? | Report summarising agreement around technical and metadata processes (metadata, exchange format, crosswalks etc). | Crosswalks used for production service once fully tested |
| Dec 2015 | HEIs participation | Has participation been a success? Can their metadata be | Report to advisory group on progress, | HEIs metadata exists in Discovery Service and can be harvested |

| | | | | |
|---|---|---|---|---|
| | | successfully harvested? | issues and challenges of harvesting metadata records from UK HEIs. | |
| Dec 2015 | Data Centres participation | Has participation been a success? Can their metadata be successfully harvested? | Report to advisory group on progress, issues and challenges of harvesting metadata records from UK Data Centres. | Data Centres metadata exists in Discovery Service and can be harvested |
| Post-July 2016 | Service implementation | Has the Discovery Service been deployed as a production service? | Review and evaluation of project and service | Discovery Service operational and in use |

### 3.2 Quality Assurance

Quality assurance of deliverables of the project will be implemented using the principles of peer review. All deliverables will be reviewed by at least two members of the project team who were not involved in the creation of the deliverable. Major deliverables, including reports, will be reviewed by the Oversight Group. The Jisc project manager is responsible for drawing up the quality assurance approach. Implementation is overseen by the Jisc/DCC project managers with the support of the project team.

| **Output / Outcome Name** | Research Data Discovery Service: production ready Discovery Service, comprising i) metadata records from UK data centres harvested by the discovery service; and ii) metadata records from UK universities' data repositories or catalogues, harvested by the discovery service. | |
|---|---|---|
| **When will QA be carried out?** | **Who will carry out the QA work?** | **What QA methods / measures will be used?** |
| July 2016 | Partners/stakeholder groups | Full deployment and testing from March 2016 to ensure pilot can run as a production service |
| July 2016 | Partners | Post-project maintenance and analysis of production service will be required |

| **Output / Outcome Name** | Software Implementation: Implementation and testing of ANDS and CKAN software, assuming these are the preferred options. | |
|---|---|---|
| **When will QA be carried out?** | **Who will carry out the QA work?** | **What QA methods / measures will be used?** |
| Mar 2015 | Development team | Checking that software instance complies with relevant standards |
| Mar 2015 | Development team | Checking that software instance complies with Jisc Open Source policy |

| | | |
|---|---|---|
| Mar 2015 | Stakeholder groups | Group will feed back on software implementation including usability and viability for use as production service |

| **Output / Outcome Name** | Expert engagement: metadata and technical processes | |
|---|---|---|
| **When will QA be carried out?** | **Who will carry out the QA work?** | **What QA methods / measures will be used?** |
| Ongoing – Dec 2015 | Metadata Advisory Group | Check metadata schema used complies with community standards |

| **Output / Outcome Name** | Reports on HEIs and Data Centres pilot implementation | |
|---|---|---|
| **When will QA be carried out?** | **Who will carry out the QA work?** | **What QA methods / measures will be used?** |
| Ongoing – Dec 2015 | Oversight Group and User Group | Check metadata harvesting and population of the discovery service is tested and signed off by each organisation |

### 3.3 Communication and Dissemination Plan

| Timing | Dissemination Activity | Audience | Purpose | Key Message |
|---|---|---|---|---|
| Ongoing | Workshops | Partners / HEIs / Data Centres | Group meetings to discuss plan, progress and deliverables | Project progress and groups have opportunity to provide input to project |
| Ongoing | Mailing list | Partners / HEIs / Data Centres | Forum for discussion and feedback on project deliverables. | The HEIs and Data Centres are fully engaged with the project |
| Ongoing | Blog/social media | Partners / Groups / Research Data Community | Raise awareness, engage community in input/feedback | Project progress, engaging with community, preparing community for production service |
| Ongoing | Articles / presentations / events | Research data community | Raise awareness, engage community in input/feedback | Project progress, engaging with community, preparing community for production service |

### 3.4 Exit and Embedding Plans

| Project Outputs/Outcomes | Action for Take-up & Embedding | Action for Exit |
|---|---|---|
| Discovery Service | Production service running on Jisc cloud service and supporting local views and workflows | If ANDS/CKAN software solutions are not suitable following evaluation and no alternative has been selected, collate all reports and feedback and close project. |

### 3.5 Sustainability Plans

| Project Outputs | Why Sustainable | Scenarios for Taking Forward | Issues to Address |
|---|---|---|---|
| Discovery Service | Running as Jisc service as part of Digital Resources & most likely on Janet cloud service | Continued support, maintenance and testing is required to run as a production service. Incorporate other UK organisations metadata as required | Support and maintenance costs. Inputting other organisations metadata into live service. |

## Appendices

### Appendix A – DRAFT Budget

See “Jisc UK Research Data Discovery Service Project Budget” document for budget details

**Appendix B – Work packages**

See "Jisc Research Data Discovery Workpackages" document for details on work packages and project deliverables.

The work packages are summarised below:

| Work package | Description |
|---|---|
| WP1: Project Management | Managing the project (led by Jisc) through its entire lifetime to ensure successful delivery of outputs, managing project partners, engaging with and reporting to Advisory Groups and coordinating the work packages within the project plan. |
| WP2: Engagement | Engagement with advisory groups/HEIs/DCs throughout the project via groups, mailing lists, virtual and face-to-face meetings, etc. Includes face-to-face meetings of the Advisory Groups and project participants to discuss progress, outputs, technical evaluation and recommendations. Specific topics for discussion depend on timing of the workshop. |
| WP3: Requirements Gathering | Gathering user requirements from the HEIs, data centres and wider sector, defining the criteria for evaluating potential solutions. Include researcher requirements, research managers, etc. Requirements should include those for the software and the crosswalks. Define the evaluation criteria for evaluating potential solutions. Develop key use cases for the use of the registry. |
| WP4: Software Evaluation | Evaluation of CKAN, ANDS and any other solution proposed from the requirements gathering. As ANDS was evaluated in phase 1, whereas alternatives have not been, this WP is split into the following strands:<br>It is important to ensure each potential solution is evaluated against the same criteria* to ensure a robust decision can be made as to which one is taken forward as a solution. It is inevitable that development work will be undertaken within this evaluation and a prototype approach should be taken. In order to respond to evolving user requirements and metadata developments, the evaluation should take an agile, iterative approach to development.<br>*(*The evaluation criteria action is an output from the Requirements Gathering WP)*<br>**ANDS Evaluation** - To build on the lessons learnt from the first phase of the pilot and investigate the suitability for use as the discovery service and work required to incorporate the latest updates, develop associated components (in particular the Harvester). Further develop and test crosswalks.<br>**CKAN Evaluation** - An evaluation of the suitability of CKAN including comparison against ANDS solution, |

| | |
|---|---|
| | user feedback on its suitability and its suitability for use as the Discovery Service. Identify requirements to develop and test crosswalks.<br>**Other(s) solution** - as with CKAN evaluation, for any proposed alternative solution. |
| WP5: Metadata Development | The adoption of a metadata standard for use with phase 1 of the pilot was determined by the selection of ANDS ORCA software. This phase will either continue RIF-CS standard adoption or work on adopting an alternative based on the choice of solution. The decision should be determined by the data providers. The work items depend on the selection from WP4. |
| WP6: Data Centre Pilots Implementation | Engagement of Data Centres to participate in the pilot project (phase 2), develop crosswalks for metadata ingest and contribute their metadata to the discovery service. Includes those Data Centres that contributed to Phase 1 and those identified in the final report willing to be involved in Phase 2. |
| WP7: HEI Pilots Implementation | Engagement of HEIs to participate in the pilot project (phase 2), develop crosswalks for metadata ingest and contribute their metadata to the discovery service. Includes those HEIs that contributed to Phase 1 and those identified in the final report willing to be involved in Phase 2. |
| WP8: Service Definition and Design | To take the selected solution deploy as a test service. Develop a business case and operational design presenting options and costs for running a sustainable service, including evidence based market research and cost-benefit analysis. This WP will need to be revised following requirements gathering and evaluation phases. |
| WP9: Dissemination | Disseminate outputs to the community via blog posts, websites, community engagement, presentations, events, etc. Communicate progress, promote the project nationally and internationally and prepare the community for the new service. |
| WP10: Institutional Implementation | It is feasible that for some universities a discovery service will be able to fulfil their requirement for a local research data registry/catalogue. In addition to the sustainable service implemented in WP8, this work package will design, develop and test such an additional service with a localised view to their data in the registry. |

## Appendix C – HEI and Data Centres Selected

There is approximately £400K allocated to 3rd Party Grants/University & Data Centre engagements, to support universities and data centres making their catalogues/data available. These need to be made interoperable with the central service.

The selection includes those HEIs and Data Centres that contributed (voluntarily) to Phase 1 of the pilot. As well as these research pioneers/early adopters a number of HEIs have shown an interest in being part of the next phase of the pilot.

### HEIs actively involved in phase 1 of the pilot project

University of Edinburgh
University of Glasgow
University of Hull
University of Leeds
University of Lincoln
University of Oxford
Oxford Brookes University
University of Southampton
University of St Andrews
Note: The above list are those institutions who volunteered to actively participate in the harvesting activity in phase 1. Other institutions were involved in other ways, such as being on conference calls, receiving regular updates and being generally part of the engagement work.

### Further HEIs who may participate in future activity

The work of phase 1 of the pilot project was successful in attracting interest from HEIs across the UK and abroad, as well as other types of potential stakeholder. At the time of writing, six UK HEIs have expressed interest in active participation in a second phase of activity, and a further six HEIs, including a prestigious US-based institution, have asked to be kept up to date with project progress.

### Data centres involved in phase 1 of the pilot project

UK Data Archive[13] (one of the partners in the pilot project)
Archaeology Data Centre[14]
Various NERC data centres, via the NERC Data Catalogue Service[15] (DCS)

### Further Data Centres who may participate in future activity

*All have shown an interest in being involved in future registry activities.*
Visual Arts Data Service (VADS)[16]
ISIS ICAT data catalogue[17]
Cambridge Crystallographic Data Centre[18] (CCDC)
STFC Energy Research Unit[19]

### Others

[13] http://www.data-archive.ac.uk/
[14] http://archaeologydataservice.ac.uk/
[15] http://data-search.nerc.ac.uk/
[16] http://www.vads.ac.uk/
[17] http://www.isis.stfc.ac.uk/groups/computing/data/icat11680.html
[18] http://www.ccdc.cam.ac.uk
[19] http://www.technology.stfc.ac.uk/ERU/

Other interested bodies and research data-holding institutions can be included as active partners or added to the list of interested organisations as appropriate, depending upon their ability to actively participate and their desired involvement. Support work with EMBL-EBI will be included through UKDA liaison work.

DCC liaison activity has produced a broad base of engaged partners during phase 1. Jisc will put the relationship with HEIs and disciplinary archives into more formal agreement through grant funding, for those institutions who have already contributed to phase 1 and are willing to participate in phase 2. The project will continue, through the liaison workpackages, to foster positive relationships with interested organisations, aided by input from the oversight group.

It is expected that participating institutions will receive approximately £30K for their contribution to the project. The exact conditions and deliverables will be detailed in the grant funding letter sent to these organisations. It is currently anticipated that around 9 HEIs and 5 data centres will be funded. This equates to 14 * 30K = £419K available to their involvement. It is expected that other funds will be made available to support a wider group of HEIs.