Intersect Australia Ltd
PO Box H58, Australia Square
Sydney NSW 2000

T +61 2 8079 2500
F +61 2 9262 3040
www.intersect.org.au

ABN 67 131 752 657

National Collaborative Research Data Infrastructure and tools

NSW considerations on state/national data storage infrastructure.

1 March 2010

Version 3: "an emerging model"

markus.buchhorn@intersect.org.au

Table of Contents

This document is based on feedback to an initial set of questions and a broad-brush model, and will need to be further evolved and informed. It is also set within the context of a national planning framework that is itself evolving. This version of the paper is based on feedback on an initial summary of views, sets out more specific next steps, and seeks further and more specific feedback from the NSW research community.

1. Executive Summary

The need to effectively manage research data, during its collection, analysis, publication and eventual re-use is becoming increasingly evident to the sector. The federal and state governments offer funding at times to assist organisations, institutions and other agencies to tackle the issues. In NSW many groups have agreed that a strategic and collaborative approach will provide the greatest benefit to its participants. The federal funding through the 2009 SuperScience initiative of around $145m for data-related services, and more for networking and computing services, provides a catalyst to develop a plan for the next 5+ years and well into the future. This paper summarises the issues, requirements and opportunities, and proposes a three-pronged approach to reach a well-considered, appropriately-governed, and user-informed framework running on a trustworthy and robust infrastructure. It will be carefully connected into any federal and state planning processes, especially the EIF/SuperScience "Data Storage and Collaboration Tools" process currently emerging.

2. Introduction

Research data is becoming an ever more valuable asset for researchers in Australia, and the volume of data being collected is increasing dramatically every year. The way it is captured, managed, stored, accessed, presented, valued and preserved is becoming an issue for researchers, their institutions and agencies, and governments at all levels. Access to historical data sets, collected under earlier programs, is also becoming a more obvious issue.

There are many ongoing organisational, regional, and state-wide research aspects that will benefit from shared knowledge and a strategic approach to the management of research data. When dealing with the longer term there is a need to consider the ongoing responsibilities and costs which implies the support of enduring organisations.

This document sets out a range of issues, requirements and opportunities related to the management of primary research data across NSW research institutions and agencies. It has been informed by the research community and by the management of the organisations in which they work; the universities, research centres, government agencies and other bodies.

The major trigger for this document is the significant recent federal funding for research data infrastructure and services, and ensuring it is effectively allocated, within the context of a much longer-term strategy. This paper provides advice to the federal initiatives on behalf of the NSW research community, as well as articulating a process that NSW research organisations will immediately undertake to develop a state plan for the management and appropriate sharing and re-use of research data. This process will be closely connected with federal processes as appropriate, and managed under a formal project-management and governance framework.

It is important for the acceptance of any large-scale infrastructure investment strategy in this space to gain the confidence of the research community. It must lead to services that are seen as trustworthy, reliable, robust and long-term stable. These depend on a solid foundation, of people with skills and well-designed and -operated infrastructure, which themselves depend on a proper governance, planning and budgeting framework.

3. Background/Context

The federal government has invested significantly in infrastructure and services to support research data management and accessibility. The major channel recently has been the NCRIS Platforms for Collaboration capability, which manages the ARCS and ANDS programs, which between them have been funded with around $190m by NCRIS and more recently from the EIF. This provides the single largest per-capita investment in research data infrastructure in the world.

The design process outlined in this paper should build upon the work carried out by ANDS for the Australian Research Data Commons (ARDC – *ref. TADC paper*). It should also be informed by, and provide information to, the ARCS Technical Working Group on research data services and collaboration tools.

Major funding opportunities for infrastructure are listed below; all have particular purposes and constraints that need to be balanced within the overall plan:

- NCRIS, both within individual capabilities that have prioritised certain discipline areas, as well as PfC which manages ARCS and ANDS on behalf of the entire research sector. NCRIS can fund operations as well as capital, but will finish during 2011/2012.
- EIF, through the 2009 federal budget SuperScience initiative, which has provided funding to ARCS and ANDS, but will only fund infrastructure and its development, not its enduring operation. The EIF funding from 2009 will cease in 2011 for ANDS and 2013 for ARCS.
- The various federal funding programs under the ARC and NH&MRC provide funding mostly for infrastructure and development but not significantly for any level of operations.
- Institutions and research agencies have recurrent budgets which could be applied to capital and operational costs for research data infrastructure and services, if they see it as a strategic priority i.e. they are able to identify the value of research data, but of course this would take away from other purposes.
- In NSW, the State Government provides the Science Leveraging Fund, which is targeted at attracting other funding to support NSW research initiatives.
- Commercial organisations may see benefit in providing in-kind support for the storage and management of research data.

While the various sources have particular constraints and priorities on what their funds can be used for, there are also wider contexts and imperatives which inform infrastructure planning. These include:

- The Australian Code for the Responsible Conduct of Research (ACRCR) sets out fairly broad requirements for the longer-term retention of and, to a lesser degree, access to research data. Not all research organisations though come under the Code.
- The ARC and NH&MRC are starting to adopt the emerging approaches from the OECD, US, Europe, and the UK that research data collected through public funding must be made (appropriately) publically accessible.
- Some international collaboration projects with an Australian presence are supported by agencies in the US, the UK and elsewhere, with data-sharing guidelines or policies.
- Some projects are constrained by ethics and privacy frameworks, which can be at odds with the rules of the funding sources unless suitably broad consent can be sought and obtained.
- Some journals require "open-access" to the research data that is evidence for a publication.
- Some datasets are acquired from primary data providers under commercial relationships and may have additional access or curation requirements.

4. Research Data Storage/Management Purposes

4.1. Purpose principles

When considering the development of a significant research data infrastructure, providing a rich set of services, it becomes easy to attempt 'everything'. Even with significant funding available, it will not fund 'everything', nor will it fund it in perpetuity. This pragmatic consideration alone means that plans must be carefully scoped and tied to particular purposes, without constraining options for future expansion.

There have been major efforts, within Australia, the UK and the US, to map the space of research data management needs and services, and then identify priority areas that can be funded and lead to operational services. ANDS and ARCS have guiding business plans, through both NCRIS and EIF funding, which identify major areas of development.

However these plans do not define boundaries around areas such as policy frameworks, content "ownership" and research priorities. This immediately leads to questions such as whose research data could/should be supported, at what stages in their lifecycle, under what guarantees, and even which organisations should be supported and how. Not all "research data" comes from "research agencies", e.g. government, industry and private datasets that are used to inform and support research. If there are limitations to be imposed these must be defined early, to set out the rules for participants. Some organisations may sit on the boundary of the research space, e.g. those that have collecting imperatives or are targets of donations of historical materials, such as cultural organisations, are often not fully funded to actively acquire, or provide ongoing storage and curation, of data assets but are expected to support research initiatives. This could lead to unintended cost-shifting, and will require stated boundaries of what data can be supported and what cannot.

It is widely agreed that all research-supportive data should be supported at some level within a state and national data management framework. Data itself will evolve through a lifecycle of processes, and there may be different stages at which data is stored on the proposed infrastructure, for many different reasons. Stages in data processes include

- Raw data collection (and storage), preceding any calibration or initial analysis process
- collaborative access to raw or calibrated data during a project,
- wider sharing and re-use beyond a project, becoming 'reference' data
- archiving for policy and publication purposes,

with ongoing management of the data and its metadata throughout all of these stages by appropriate authorities. These management processes become highly discipline-specific at some level, and move into a separate governance domain.

All of these processes also need to be linked, and ideally integrated, with the publication process. There are increasing imperatives for the evidence underpinning a publication to be made accessible.

4.2. Classes of storage services

The research community has a diverse range of needs for the storage and management of research data. It is useful to broadly categorise the services under particular purposes, or classes.

Storage classes make different kinds of promises to users. For example: "Scratch" or "Drop-box" storage can imply short-term bulk storage for active projects, possibly high-performance for reading/writing but with limited reliability guarantees. "Collaboration" storage can imply medium-term storage for active projects, and likely to have significant access control requirements, but certainly higher levels of trustworthiness. "Archival" storage implies long-term retention and accessibility, hence firm operational support, and may have various levels of performance of access e.g. it may be stored offline. "Mirror" storage implies copies of authoritative information kept elsewhere, typically brought closer for performance issues or institutional requirements.

Any planning process for infrastructure and services must catalogue the storage classes, their functionality, performance and promises, and be informed by the needs of the research community and existing practices. Where a service-level statement is expected there must be a process to demonstrate performance and accredit that performance.

5. Generic and Specialised services

There is a continuum of services that can support data-related activities, ranging from simple concepts such as 'store' and 'access' through to more complex issues such as 'manage', 'curate' and 'discover', up to discipline-specific services related to e.g. 'merging', 'analysis' and business-specific services such as 'preservation' for the ACRCR requirements and 'reporting' (with 'access') for eventual ERA/HERDC requirements. Some of these may be common across all disciplines and all organisations, with minor tweaks for context, through to highly context-specific such as an astronomical cone-search.

It must be noted that 'generic' does not imply a sense of simplicity or low quality, but the concept of an underpinning infrastructure layer that is common across diverse communities, e.g. across disciplines or across institutions.

Funding sources (e.g. ARC/NHMRC grants, NCRIS) for 'services' may target certain discipline opportunities, but are unlikely to provide ongoing operational funding. Some university administrators have suggested that institutions may set research priorities and underwrite operational services for research activities at some level (maybe at the 'generic' level), or may choose to focus on where it adds strategic value ('specialised' level) in the expectation that generic services are funded from broader frameworks. Others have suggested that there is no strong boundary between them, or that services may too easily evolve from specialised to generic, or that the mesh of dependencies between them make the distinction meaningless.

The need for operational funding requires consideration of this concept. Given that discipline needs are set within institutional priorities, it suggests that generic ongoing services should be supported through institutional operational funding, with specific targeted funding for specialist services that are consistent with the University's strategic objectives.

It is clear though that data must be visible to both classes of services, allowing for discoverability and re-use in other contexts.

6. Use case gathering

To inform the design of the infrastructure, its scale and performance, and the classification and prioritisation of services, there is a requirement to gather the needs of the research community. These include researchers in universities, government agencies, and other organisations. The more deeply that gathering process is carried out, the better-informed the planning will be, and at the same time the better the community engagement will be.

There is broad support for a multi-stage requirements gathering process:

1. A working party will be formed from interested parties across the NSW research organisations. It will shepherd the process.

2. It will start by drawing on existing surveys and overviews that have been developed over the last two years, and develop an initial summary of requirements and priorities. This will naturally cluster around well-known (to infrastructure programs) research initiatives, but forms a baseline for discussion.

3. The initial summary will be followed by a simple online survey to gain a wider coverage of the research community, especially on matters of scale and performance.

4. Where necessary there will be a follow-up engagement of particular groups or projects.

The collection should capture not just basic needs (such as data volumes, file formats and particular processing tools) but also information related to issues around rights and other terms of use, as well as discipline-related specifics. It should capture a sense of relative priorities, to inform the deployment process. The process should take several perspectives, including that of the researcher in relation to

their workflows, of the data in relation to its curation and lifecycle, and of the institutions and other organisations that will be expected to provide ongoing support and have administrative process built on research outputs.

Intersect is well placed, and trusted, to provide support for this process.

7. Infrastructure and Services Design

There are widely-understood benefits in a coordinated strategy for storage infrastructure, including economies of scale, strategic capability building across the sector, improved support for collaboration, and simplification of security issues for cross-institutional projects.

This means there ultimately needs to be a technical framework that takes into account every level of service provision from the desktop through the institutional, state-wide and discipline services, and how they connect into national frameworks. It needs to address the higher-level 'services' view and map that onto a physical infrastructure. It also needs to consider the need for support services, which enhance the capability of the research community to make use of the infrastructure.

For the research community, the two crucial elements are that firstly, the services are robust, trustworthy and effective, and secondly that they provide user-friendly and useful interfaces. The back-end technology is important, but fundamentally most researchers will not particularly care where or how the data is stored as long as the interfaces provide the required services. Regardless of the implementation it should be transparent to users, no matter what model is employed – the back-end must be informed but transparent to a researchers who says 'store this'.

This design process needs to start quickly and be able to evolve. The consensus is to start with a simple set of services, and to add functionality over time as maturity and trust are established and the operational model is tested. To evolve the framework there must be a process to publish any proposed extensions or changes as early as possible to allow local groups to engage with and plan around future developments.

Some infrastructure already exists at institutions and other organisations, and is funded, or is about to be funded. That investment needs to be recognised, not just for financial accounting of 'leverage' but also the institutional strategic priorities implicit in such investment. As part of the technical design, there will be a need to link with institutional infrastructure; some level of abstraction and well-defined interfaces will simplify the communication, and enhance the engagement between the state, national and local infrastructure.

The process to develop the design will follow an agile and structured approach:

1. A technical-design working party will be established with representatives from interested research and research-support organisations. As part of the design process input will also be actively sought from any other relevant organisations.

2. An initial design will be derived from the state-wide ARC/LIEF 2010 proposal for a large scale collaborative storage infrastructure. This identified the need for robust, trustworthy and well-

connected central storage, with a range of performance characteristics (fast storage for processing support, bulk storage for large-scale data sharing, fully replicated and backed up storage for archiving needs), together with a significant level of institutionally-based storage that straddled institutional firewalls and also acted as caches as necessary.

3. This initial design will be distributed widely across the sector for comment, as well as to the State and Federal governments, and the national infrastructure programs.

4. As the use-case gathering (Section 6) and governance design (Section 8) is completed, the technical design will be fine-tuned, taking into functionality requirements as well as operating cost models, and distributed again accordingly.

The design will need to specify the capital cost requirements for a spectrum of functionality to allow for the uncertain funding being proposed. It will also need to specify what services will be initially available, what network infrastructure may be required, and the design of the institutional 'caches'. The funding model for various elements will be left to the governance and management design process (Section 8) but the design will need to provide information to that process.

Intersect will provide the coordination for this activity. It can run in parallel to the planning of the governance and management framework, taking into account the evolving context in the ARCS/EIF process.

8. Governance and Management

8.1. Oversight and coordination, at the state level

A strategically planned infrastructure with significant federal, state and institutional investment needs a solid, transparent, appropriate and well-understood model for oversight and coordination, with a high level of trust and assurance.

At a minimum, there must be a way for stakeholders to participate actively in the governance of the infrastructure. There needs to be a formal core structure for governance, with advisory groups on policy and technological issues. Some committee-based structure for governance, with representation from participants, is required, however it should avoid creating yet another organisational entity if possible, rather it should build off existing organisations. Such a governance structure must also take into account that there are different classes of stakeholders, such as those related to infrastructure/service providers, or related to particular content providers.

As with the technical infrastructure design, it may be beneficial to start quickly with an interim model, and allow it to evolve, again taking into account the evolving national context. But governance and management must be planned for the longer term, beyond any initial investment of capital, and acknowledging that interim models can become entrenched and hard to change.

8.2. Relationship between national and regional frameworks

There are 'infrastructure' and 'service' providers at multiple levels with Australia, from the institutional to the national. All of them have their own governance and management mechanisms, such as the joint ventures underpinning ARCS and ANDS, and incorporated entities such as Intersect. Any state-

wide governance structure has to interface with those frameworks, and has to rely on those other ventures interfacing effectively amongst themselves

To ensure that the relationship between any NSW program and federal programs is optimal and effective, a group such as Intersect is crucial, to provide a single point for the state where the various programs can be combined and aligned.

8.3. Managing Risks

The governance and management framework must deal with risk; an appropriate register of risks at all levels must be established, maintained and monitored, regardless of how the services are ultimately delivered.

8.4. Implementation

An initial and extensible model for the governance and management structure of this infrastructure organisation is summarised as follows

8.4.1. The Board

The role of the board is to oversight the overall management and operations of the organisation and its infrastructure. It reviews and approves business plans, development and operating plans and performance against those plans. Its membership should be broad, including a mix of independent and representative members, but whose role on the Board is to serve in the best interests of the organisation.

8.4.2. The Strategy and Services Committee

The role of the Strategy and Services Committee includes

- Providing advice to the board on a range of relevant topics,
- Reviewing the portfolio of services being offered, their performance, and their engagement with the user community,
- Reviewing technical affiliations i.e. connectivity of this state infrastructure with other infrastructure service at the national and institutional level, and
- Developing and evolving a Resource Allocation Policy

To ensure that the services being offered, and how they are offered, are aligned with the needs of the community the membership of this committee should be representative of all stakeholders.

8.4.3. Resource Allocation Committee

The role of the Resource Allocation Committee is to review applications, and apply the allocation policy of the organisation. To ensure that it is seen as working fairly across institutions, and is professionally informed in its decisions, its membership must be representative of the user-base, including senior academics and discipline experts.

8.4.4. Other sub-committees

As required, the Board and its committees may establish additional sub-committees or working parties to advise the organisation on issues as they arise.

8.4.5. Moving ahead

The structure outlined here is one that has already been developed by the universities of NSW when they established Intersect, and the effectiveness of these arrangements has been demonstrated. Intersect's mission encompasses the management of state-wide research infrastructure on behalf of its members, and it is also deeply connected with the various national framework programs for research infrastructure. Intersect provides a natural organisation to manage the infrastructure for the NSW research community starting immediately, and can evolve as required to meet the requirements of the sector for this infrastructure.

9. Planning

Having arrived at an overall architecture, there need to be plans for at least three distinct aspects: the overarching/longer-term strategy, the deployment phase(s), and the ongoing operations and refresh. These will necessarily overlap. They will also need to evolve over time, as maturity and expectations grow.

9.1. Strategic Plan

The Strategic Plan must include the business model and indicative budget, participation expectations, and a broad outline of the services framework over time. It needs to be reviewed annually, and set out at least a five-year timeframe commensurate with and beyond the initial NCRIS/EIF funding. It must also identify the broad principles for sustainable ongoing funding, support and evolution beyond that timeframe.

To keep the infrastructure relevant means ongoing review of the services framework, of the market offerings, and refresh of the underlying technological platforms. There must be an asset replacement schedule in the budget, as well as maintenance planning. These will ensure that there are sufficient operational funds, underpinning the operational plan, to support the infrastructure and to provide for timely upgrades.

The business model must identify sustainability mechanisms, including consideration of subscription models and fee-for-service models. Participation should entail a commitment for up to the timeframe of the strategic plan initially (i.e. 5 years) and an understanding of longer-term responsibilities.

9.2. Deployment and Commissioning plan

The demand for research data infrastructure is already very high, and the expectations have been raised by programs such as EIF, the ARC2009 NSW grant for HPC and storage, and the unsuccessful ARC2010 proposal for large-scale state-wide storage. There is wide agreement that researchers need operations to start as soon as possible, implying a deployment plan that rolls out basic functionality very quickly. That plan should outline a professional project management approach which includes detailed time lines, milestones, deliverables, regular reports, and audits. This process is likely to be strongly influenced by the processes for the ARCS/EIF-infrastructure plans.

The deployment plan must take into account existing institutional infrastructure, and allow for the deployment of perhaps centrally-managed or -coordinated equipment at institutions.

While the use-case gathering outlined above will provide more detail, some commonly identified priorities include:

- Storage for large data sets.
- More general data storage mechanisms to protect/rescue data from researchers' desktops and make it more easily available as appropriate.
- Archival storage for data that supports publications and needs to be maintained and available for set periods of time.
- Support for multidisciplinary or multi-institutional projects to collaborate around data.

9.3. Operating Plan

Once the infrastructure starts to be deployed, without infinite capacity and resourcing up front, there will be operational pressures, including managing income streams, managing service levels and managing allocation of resources to certain activities. These include staffing for support services. The operating plan must include the budget and resource allocation processes, metrics and auditing mechanisms, implementation and prioritisation processes for the service offerings, and how all of these will be able to evolve within the wider Strategic Plan.

10. Investment approach

As noted earlier, there are a diverse range of funding sources, each with their own funding priorities and particular constraints. The technical design process together with the governance and planning mechanisms will identify the best ways to seek access to those funds and target them effectively and optimally.

There must be clear indications of:

- What level of co-investment (subscriptions, fee-for-service) participating institutions will commit to, what purposes they may be tied to, and what benefits will result.
- The significant cost of expertise and infrastructure required to comply with various policy frameworks, such as recommended retention periods.
- The timeframes, and timing, for expenditure on infrastructure, service development and operations.

The generally preferred model is one where the initial funding is used to create the infrastructure, but subsequently the services become "just another service", funded by the institutions.

There is clearly a need to refresh the infrastructure over time. A massive up-front deployment may lead to a massive refresh being needed in 3-5 years, and may deliver more capacity than can be initially used. A phased introduction will allow for a steady uptake and capability building, and will stagger the refresh pressure as well.

11. Next steps and time line

This paper has evolved towards a broadly agreed position, and this version sets out more specific elements. There are three major themes of activity that can now be undertaken in parallel: the use-case gathering, the infrastructure design and the governance design. The development of the planning documents will follow naturally from these.

To ensure the process continues to move, and keeps up with the EIF discussions, feedback on this paper should be provided to the author by **Friday 19 March 2010**. Comments are sought from institutions and organisational perspectives as well as the wider user community. This includes agencies outside of the universities.

The three themes of activity above already have broad agreement, and can start immediately. Intersect is prepared to develop the underlying processes and documentation. **Institutions are asked to identify appropriate contacts in their organisations for each of the themes by the same deadline (19 March 2010),** to allow the working parties to be formed and discussions to begin.

Intersect will work to ensure that all three are carefully aligned with the processes being managed by DIISR for the ARCS/EIF planning, as well as the efforts being undertaken by ANDS.

12. Appendix 1: Glossary

The benefit of a (carefully-scoped) national e-research acronym directory is becoming obvious. As an aside the community could collaboratively establish a single online location and include it by reference. Obvious locations could be at www.pfc.org.au or www.eresearch.edu.au. Intersect can assist this process.

13. Appendix 2: References and other useful sites

The references here come directly from the feedback provided to-date, and will be structured and referenced more appropriately in future drafts.

13.1. Background for the purposes and processes

- Towards an Australian Data Commons: http://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf
- The Australian Social Science Data Archive (ASSDA – assda.anu.edu.au).
- The US National Science Foundation (NSF)and their review of the needs for "cyberinfrastructure" http://www.communitytechnology.org/nsf_ci_report/report.pdf
- the International Federation of Data Organisations for the Social Sciences (http://www.ifdo.org/)
- EDINA - the UK national academic data centre based at the University of Edinburgh. http://edina.ac.uk/
- The findings of the UK's Office of Science and Innovation (OSI) e-Infrastructure Working Group (http://www.nesc.ac.uk/documents/OSI/)
- Oxford EIDSCR http://eidcsr.blogspot.com/ ;
- NSF DataNet http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm
- APSR http://www.apsr.edu.au/publications/index.htm
- ANDS EIF: http://www.ands.org.au/infrastructure.html
- ARROW http://www.arrow.edu.au/docs/
- Robin Rice Data Sharing Continuum http://www.disc-uk.org/docs/data_sharing_continuum.pdf
- "Culture is not a Department: The Role of Governance in National Cultural Institutions". http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewFile/1393/1682
- http://www.naa.gov.au/about-us/director-general/2009-05-21.aspx Last accessed: 21/10/2009 See also the scope for the National Digital Heritage Archive: http://www.natlib.govt.nz/about-us/current-initiatives/ndha Last accessed: 21/10/2009

13.2. Use cases

Researcher perspective:

- The JISC SCARP has done full-immersion analysis of researcher's use of data and released four of the ten reports so far: http://www.dcc.ac.uk/scarp/
- Polydoratou, Panayiota. (2007). Use of digital repositories by chemistry researchers: results of a survey. Program: electronic library and information systems, Vol. 41 (4), pp386-399.

- Coggins, John (2009). A Researcher's Perspective: the Value and Challenge of Data. Paper presented at UKRDS Conference 26 February 2009; http://www.ukoln.ac.uk/events/ukrds-2009/presentations/j-coggins.ppt
- Dinkelmann, Karl; Edwards, Michelle; Fry, Jane; Humphrey, Chuck; Nakao, Ron; & Thomas, Wendy. (2009). Work flows - Data Discovery and Dissemination: User Perspective. DDI
- Van de Sompel, Herbert; Payette, Sandy; Erickson, John; Lagoze, Carl & Warner, Simeon. (2004). Rethinking Scholarly Communication, Building the System that Scholars Deserve. D-Lib Magazine, Volume 10 Number 9, September 2004; http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html
- McKay, Dana. (2007) Researcher data practices at Swinburne: results of a survey. Melbourne, Australia: Swinburne University of Technology, Information Resources

Data perspective:

- DCC Curation Lifecycle Model (2009). http://www.dcc.ac.uk/lifecycle-model/
- The Interagency Working Group for Digital Data (IWGDD) data lifecycle
- Dinkelmann, Karl; Edwards, Michelle; Fry, Jane; Humphrey, Chuck; Nakao, Ron; & Thomas, Wendy. (2009). Work flows - Data Discovery and Dissemination: User Perspective. DDI
- Lyon L. (2007). Dealing with data: roles, responsibilities and relationships, Consultancy Report. June,
- 2007, Bath: UKOLN. http://www.jisc.ac.uk/publications/publications/dealingwithdatareportfinal.aspx
- Green, Ann, Macdonald, Stuart, & Rice, Robin. (2009). Policy-making for Research Data in Repositories: A Guide. Version 1.2. JISC.
- Research Information Network. (2007). Stewardship of digital research data: a framework of principles and guidelines. Available at http://www.rin.ac.uk/dataprinciples.

Information gathering

- UNSW MemRE http://membranes.edu.au/wiki/index.php/Main_Page
- Intersect survey
- EUAsiaGrid survey http://www.surveymonkey.com/s.aspx?sm=ZtoA5sG3n0RuLVQ5ggBsdQ_3d_3d

Document Identifier			
Document Author	Markus Buchhorn		
Version No.	3	**Version Date**	1 March 2010
Document Name and File Location			

Revision History

Version No.	Revision Date	Summary of Changes	Revised By
1	**30/9/09**	**First version. Asks the questions.**	**MB/IG**
2	**30/10/09**	**Second version. Aggregates feedback from multiple institutions and sets out models**	**MB**
3	**1/3/10**	**Third version. Further feedback included, and specific steps for three themes set out**	**MB**