International
Journal
Of Advanced Research in Engineering & Management (IJAREM)

Research Data Management with Context of Different Countries

Bilalkhan Raufkhan Pathan
Assistant Professor
Jamia Institute of Engineering & Management Studies, Akkalkuwa, Maharashtra, India

Abstract— The reuse and management of research data are becoming increasingly important as Australian research funders are increasingly following the global trend of lodging publications and data in public repositories.Research data are defined here asthe factual records (e.g. numerical and textual records,microarray, images and sounds, etc.) used asprimary sources for research, andthat are commonly accepted in theresearch community as necessary tovalidate research findings.
Publication and re-usability of research data bring great benefits such as research and researcher effectiveness, enhancing the reputation of researchers and institutions, meeting obligations to funders, and compliance with Open Access agendas. To achieve this, it is critical that research data is properly managed from the pre-research planning stages, through to post research completion. The US National Science Foundationdescribes digital data as both
the products of research and,increasingly, the starting pointfor new research. In the digitalenvironment, it is possible forresearchers to re-purpose data –to use them in innovative ways andcombinations not envisioned bythose who created them. It is alsocommon for researchers to usedata from one study to buildon and enhance the findings ofprevious research; and to undertakelongitudinal studies that comparedata from repeated observations ofthe same items over long periods oftime often many decades [1].
Keywords: longitudinal study,data management,open access, research, reuse

I. INTRODUCTION

Over the next five years, the worldwill produce more research data thanhas been created in all of human history [2].For the most part these data are borndigital, and stored and managedelectronically, making them easy toshare, replicate, and combine withother data. However, in order toshare and reuse data, they must becreated and maintained in a mannerconsistent with the goal of long-termpreservation. This involves activedata management throughout thelife-cycle of the data, beginning at thetime they are first envisioned. Research is becoming more data-intensive. The ability to make research data available for further use delivers a series of benefits including individual and institutional reputation (through data citation), better research- and data validation, improved cost-benefit ratios, and compliance with emerging Open Access agendas. Data sharing will further reduce duplication and will enable more targeted future research. It also supports complex, international research projects [3]. Right now in Canada, the vast majority of research data is being lost. For example, a study of Social Sciences and humanities Research Council funded research projects found that only 3 out of 110 studies had archived their data in a repository, and those 3 were all housed in the US [4]. Research data in Canada are not being systematically managed and therefore, valuable data are under-utilized. While certain disciplines and research projects have institutional, national or international support for data management, this support is available for a minority of researchers only. A crucial aspect to creating data with long-lasting usability is to ensure that the accompanying documentation is user-friendly, clear, and comprehensive. Ideally,metadata and documentation should be produced at the start of a research project and enhanced throughout the course of the data life-cycle. Planning and support from data professionsat the initial stages of the research project can significantly reduce the time and money needed to provide long-term access [5].

II. OBJECTIVE

Australian research funders are increasingly following the global trend to require research outputs, including data, to be lodged in public repositories. This is best achieved using standardized approaches to data capture, storage, attribution and metadata. Recent studies have shown that the benefits of freely available Public Sector Information (PSI), which includes publicly funded research data, outweigh the costs [6]. Effective research data management also plays a vital role in managing research risk. All research is subject to a range of data-related risks such as data loss or corruption, and privacy or copyright breaches. These risks come with significant, potentially catastrophic impacts. Effective research data management can go a long way towards preventing and

managing such risks. Perhaps the most compelling reason for effective research data management is that it directly benefits both researchers and institutions by minimising the risk of loss or deterioration of research findings including the data.

III. COMMON STEPS IN RESEARCH DATA MANAGEMENT

Below figure shows the common steps evolved in management of research data.

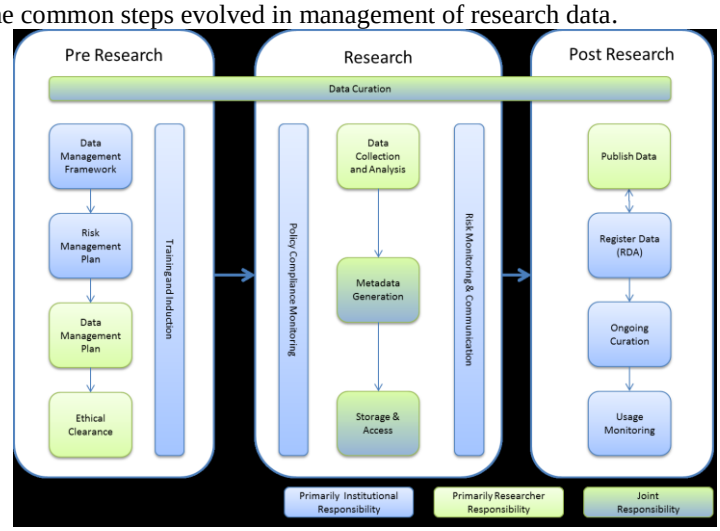

Figure 1Ssteps in Research Data Management

3.1 Research Data Curation

With traditional publication, most curation activities occur at the end of the research cycle. In contrast, digital curation of data is characterised by activities planned for from the outset and occurring throughout the data lifecycle. These will include, for example, the processes outlined as below,

1. All Stages
2. Rich metadata
3. Access Continuum:

I. Private
II. Shared
III. Public

Digital data curation requires the capturing of rich metadata, and depositing in appropriate formats into well-managed stores as defined in the Research Data Management Plan as shown in below figure.

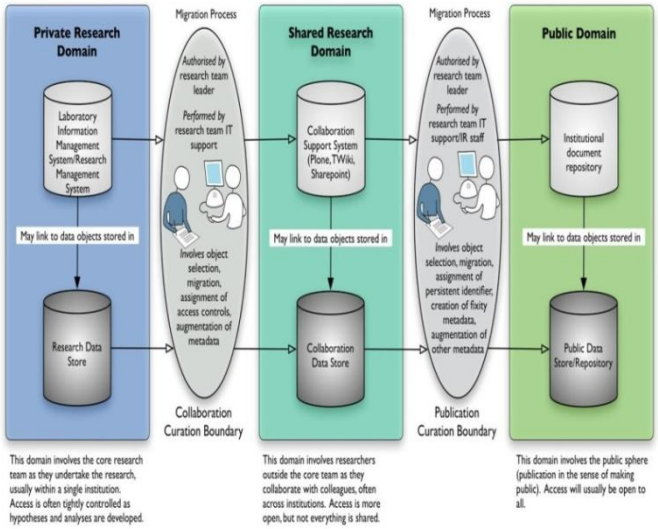

Figure 2 Curation Research Data

3.2 Pre Research

3.2.1 Research Data Management Framework:

The research institution is responsible for providing an adequate Research Data Management Framework, providing the basic elements required within an institutional context to support effective research data management. These elements comprise four categories:

- Institutional policy and procedures: these should be up to date, addressing data-related issues, and be publicized to all those who have a data creation and/or management role;
- IT Infrastructure: the hardware, software and other facilities which underpin data-related activities, as well as identity management and access control;
- Support services: people and other means of providing advice and support, such as web-pages;
- Metadata management: so that data records can be used for both internal and external purposes.

3.2.2 Risk Management Plan

Risk management is the process of identifying, assessing and responding to risks, and communicating the outcomes of these processes in a timely manner. Developing a risk management plan is important to avoid preventable losses associated with research data management. Risk can arise from e.g. data loss or corruption, under-utilisation of research outcomes, or breaches of privacy, confidentiality or copyright. 4 Research Data Management in Practice A risk management plan identifies and assesses risks based on likelihood of occurrence and respective impact. It identifies effective mitigation measures based on risk criticality, which are allocated to specific people, and monitored throughout the research project.

3.2.3 Research Data Management Plan

Development of a Research Data Management Plan is a critical aspect of the pre-research stage. It lays out what data will be created, what policies will apply to the data, which will own and have access to the data, what research data management practices will be used, what facilities and equipment will be required, and who will be responsible for each of these activities.

It will include activities such as:

- Data organisation and storage;
- Metadata standards and guidelines;
- Backups;
- Archiving for long-term preservation;
- Version control and derived data products;
- Data sharing or publishing intentions, including licensing through AusGOAL
- Ensuring security of confidential data;
- Data synchronisation; and
- Governance, roles and responsibilities.

The plan usually defines all research data management related activities during and subsequent to the research activity.

It also includes following phases;

I. Ethical Clearance
II. Training and Induction

3.2 During Research

3.3.1 Risk Monitoring and Communication

Continuous monitoring and review are vital components of an effective risk management process. In terms of the research project, review of risk would normally be incorporated into the review of the research data

management plan. The primary purpose of monitoring and review is to determine whether risks still exist, whether new risks have emerged, and to reassess the risk priorities.

3.3.2 Research Data Collection and Analysis

Throughout the research cycle data will be collected. Data analysis will generate derived data and in many cases data will be shared between researchers and institutions for collaboration purposes.Data at this point may be made available to collaborators, as specified in the research data management plan.

3.3.3 Metadata Generation

Data discovery and access is dependent on the availability of rich metadata. Metadata is collected at both collection and object level and can be stored separately, or embedded in the data collection. Collection level metadata is generated by the researcher as part of the research process and supplemented by object level metadata for data publication, in most cases, by librarians or other data 6 Research Data Management in Practice Good metadata creation can be supported by tools designed to simplify metadata input and to enhance interoperability. Metadata is also useful in tracking the history of derived data products.

3.3.4 Storage and Access

The choices made regarding the approach to data storage have implications for cost, security, and future access. It is an institutional responsibility to ensure that adequate and appropriate storage facilities are available. The goals of "reusing and sharing data more often" are met by storage solutions which make data discoverable and accessible over the long term, which means the tendency, should be towards more metadata rich, curated stores with a wide community scope.

3.4 Post Research

There is an increasing expectation that the outputs of publicly funded research, including the data, will be made available for others to use. That means published data should be well-described (metadata), citable, discoverable and re-usable wherever possible. Potential re-users of research data need to have clear guidance about what they can and cannot do with the data: this is normally achieved via a licence. ANDS supports the Australian Governments Open Access and Licensing Framework which provides support and guidance to facilitate open access to publicly funded information. Research data can be published in the form of collection descriptions, citable and online accessible data elements, or citable other objects such as web services, APIs, or concept definitions. Sometimes just the existence of data collections is published; this occurs when data cannot be accessed or accessed under strict conditions. Online data publication for download or web-service access is desirable for those kinds of data which are not restricted.

IV. CURRENT CONTEXT

Approaches to the management of research data vary significantly according to discipline. Some fields, such asgenomics, proteomics, high-energy physics, and astronomy have long-standing traditions of data archiving andsharing. Others, such as chemistry and the humanities and social sciences have less established traditions. Researchersin some fields oppose sharing their data for a number of reasons, such as: ownership/intellectual property concerns;lack of awareness of the value of data sharing; or, lack of knowledge or time to prepare data for dissemination.

International

There are a growing number of large discipline-based data archives like PubChem, GenBank, Protein Data Bank,Digital Sky Survey, World Data Centers (solar, geophysical and related environmental data), Global BiodiversityInformation Facility, International Virtual Observatory Alliance, and the Interuniversity Consortium for Political and SocialResearch (ICPSR), and so on. These archives are international in scope and aim to collect data from around the world.They also provide broad access to the data they collect in order to further research and knowledge creation. Thevast majority of these archives are funded through national government departments and/ or funding agencies inthe country which they are housed- although there are also a few international data archives maintained by privateindustry. In addition to these international, discipline-based archives, many countries maintain national data archivesthat have a mandate to collect valuable national data in a given subject area. For example most countries collect andarchive population data, climate data, health data and so on.In 2004, 34 countries, including Canada, signed the OECD “Declaration on Access to Research Data from PublicFunding” [7]. The premise of the declaration is that publicly funded research data should be openly available tothe maximum extent possible. Following on this, a number of countries are investigating how they can moresystematically exploit the data created through publicly funded research

United States

In the US, the National Science Foundation (NSF) announced a $100 million funding program called"Sustainable Digital Data Preservation and Access Network Partners (DataNet)" [8]. The program isdeveloping "new methods, management structures and technologies to manage the diversity, size, andcomplexity of current and future data sets and data streams by creating a set of exemplar national andglobal data research infrastructure organizations".Also in the US, an Interagency Working Group on Digital Data was formed in December 2006. Nearly30 government agencies, offices, and councils were named as members or participants, reflectingthe broad range of interests in digital scientific data. The group's final report proposes the followingstrategy for the US: "Create a comprehensive framework of transparent, evolvable, extensible policiesand management and organizational structures that provide reliable, effective access to the fullspectrum of public digital scientific data. Such a framework will serve as a driving force for Americanleadership in science and in a competitive, global information society" [9].

United Kingdom

The UK already has a very robust infrastructure of data repositories for collecting research data,managed by several RCUK funding agencies. To build on this, the Consortium of Research Libraries inthe UK and Ireland, and the IT Directors Group undertook a study to assess the feasibility and costs ofdeveloping and maintaining a national shared digital research data service for the UK Higher Educationsector. The study concluded that considerable efforts are being made in other countries to surface andexploit data on a national or regional scale and that there is a need for a UK-wide approach to researchdata management to ensure that it does not fall behind [10].

Australia

In September 2008, representatives from the major Australian research community endorsed theBrisbane declaration, which states among other things that "Every citizen should have free open accessto publicly funded research, data and knowledge" and "Every Australian university should have access toa digital repository"[11].The Australian government recently announced the creation of the Australian National Data Service(ANDS). ANDS began in September 2008 and was established as part of the National CollaborativeResearch Infrastructure Strategy. Among other stated goals, ANDS aims "to ensure that Australianresearch data are well managed, made available for access and discoverable" [12].

Canada

In October 2000, the Social Sciences and Humanities Research Council and the National Archivist of Canadaestablished a Working Group of research and archival experts and asked them to assess the need for a nationalresearch data access, preservation and management system. The Working Group found large gaps in theinfrastructure for collecting and preserving research data. The final report, published in 2002, recommended thecreation of a new national research data archival service [13]. In January 2004, Canada was a signatory of the OECD Declaration on Access to Research Data from Public Funding.But, unlike some other national signatories, Canada has not responded with concrete actions to give its agreement tothe MOU substance.

Canadian GroupsConcerned with DataManagement:

Canadian Research DataStrategy Working Group

On January 25, 2008, the Canada Institutefor Scientific and Technical Information(CISTI) hosted a meeting with keystakeholders to discuss how to beginimplementing the National Consultationon Access to Scientific and ResearchData outcomes [14].

Canadian NationalCommittee for CODATA(CNC/CODATA)

This committee is the Canadian voiceof CODATA (Committee on Data forScience and Technology) [15]. CODATA arean interdisciplinary Scientific Committeeof the International Council for Science(ICSU), which was established 40 yearsago. CNC/CODATA is sponsored byCISTI (Canada Institute for Scientificand Technical Information), part of theNational Research Council of Canada.CODATA have no formal governmentalstatus, but works through its nationalmembers and task groups to "improvethe quality, reliability, management andaccessibility of data of importance to allfields of science and technology".

Different other Data Policy:

McGill University:

"Data must be organized in a manner that allows ready verification….Subject to exceptionsbased on a duty of confidentiality and the laws respecting intellectual property andaccess to information, after data are published, they must be made available to any partypresenting a reasonable request to examine them. In cases where there is a disagreementbetween the researcher and the person requesting the data, the matter shall be referred tothe Office of the Vice Principal Research for resolution; (a) all original data must be retainedfor a reasonable length of time. (b) A period of at least five years from the date of publication isrecommended" [16].

Fisheries and Oceans

Canada: "Management Policy for Scientific Data" states, "It is the responsibility of Science andOceans managers to ensure that data collectors under their management submit theirdata as well as data collected under contract to or partnership with other agencies, tothe appropriate data centre in a timely fashion. This is important to ensure that data arequickly migrated into a 'managed' environment where they are properly backed up andsecured from accidental or circumstantial loss, and where the supporting metadata areintegrated with the data to preserve the long-term usefulness of a data set"[17].

International Polar Year(IPY**)**

The "IPY Data Policy" states, "in order to maximize the benefit of data gathered under theauspices of the IPY, the IPY Joint Committee requires that IPY data, including operational
data delivered in real time, are made available fully, freely, openly, and on the shortestfeasible timescale."[18] In addition, "All IPY data must be archived in their simplest, usefulform and be accompanied by a complete metadata description." [19]

Federal Government:Personal InformationProtection and ElectronicDocuments Act (PIPEDA):

Personal information shall not be used or disclosed for purposes other than those forwhich it was collected, except with the consent of the individual or as required by law.Personal information shall be retained only as long as necessary for the fulfillment ofthose purposes. Personal information that is no longer required to fulfill the identifiedpurposes should be destroyed, erased, or made anonymous. Organizations shalldevelop guidelines and implement procedures to govern the destruction of personalinformation. [20]

V. ADVANTAGES OF RESEARCH DATA MANAGEMENT

Research Data has its own importance in innovation field some of them are listed below;

- Accelerates scientific progress. Thesound management of researchdata will allow researchers to accessand understand others' data andre-use them for their own scientificpurposes, thereby speeding up therate of new discoveries.
- Avoids duplication of research.When a dataset is publicly availableit is much less likely to be recreated,avoiding expensive and needlessdata collection/production activities.
- Enables replication and verification of research results. When data arearchived and shared, results are repeatable and data can be used for reanalysis,backing up original research findings. They may also be used toexpose errors or inconsistencies with original data analysis.
- Enhances collaboration: Publicly available data enable researchers tocollaborate with each other by sharing data sets, research environments andtools.
- Ensures compliance with fundingagency policies. A growing numberof funding agencies demand thatresearchers and host institutionsretain, manage and share their dataupon completion of a researchproject. Thus, universities andresearchers have legal and ethicalobligations to provide a legacy ofresearch data. Some publishers alsorequire that the data connected totheir publications are preserved.
- Increases the visibility and impactof research. Data made visiblethrough a data repository candramatically increase the impact ofthat research. Sharing research datahas been associated with increasedcitation rates. For example, a studyof citation rates for cancer clinicaltrials publications found that clinicaltrials that shared their data werecited about 70% more frequentlythan clinical trials that did not.[21]

VI. CONCLUSION

Research Data is an important and vital recourse to preserve. For preserving and managing an individual as well as organization, Universities, Research Libraries may play an important role in order to store, manipulate and manage the research data. Some points are mentioned below;

Researchers:

- Commit to sharing research data as openly as possible.
- Develop data management plans before the beginning of a research project.
- Understand and comply with funding agency data management policies.
- Make use of the data professionals on campus to assist in collecting and managing research data.
- Identify and use international standards for data management.

University Administrators

- Develop policies that support sound data management activities.
- Support the implementation of data repositories at the institution.
- Provide education for researchers about data management practices.
- Provide support for researchers by hiring qualified data scientists or librarians and make theseprofessionals known to (and part of) the appropriate research teams.
- Recognize data sharing contributions in hiring, and promotion and tenure decisions.

Research Libraries

- Develop and manage data repositories at the institution.
- Support training for librarians in the area of data stewardship.
- Provide support for researchers by hiring qualified data librarians and make these professionalsavailable to the research community.
- Provide education for researchers about data management practices.

REFERENCES

[1]. "Community-based Data InteroperabilityNetworks". 2009. National ScienceFoundation. http://www.nsf.gov/pubs/2007/nsf07565/nsf07565.html
[2]. Beagrie, N. "Digital Preservation: Setting theCourse for a Decade of Change". 2007. www.abd-bvd.be/documents/60/BDA-1107_NB.ppt
[3]. Data Management Practice Guide
[4]. Humphrey, C. "Preserving research data: Atime for action". In Preservation of electronicrecords: new knowledge and decisionmaking.2003. Canadian ConservationInstitute: Ottawa. pg. 83-89.
[5]. An awareness toolkit commissioned by Canadian association of research libraries
[6]. Web Page http://ands.org.au/resourse/cost-benefit.html
[7]. "Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy atMinisterial Level, 29-30 January 2004 - Final Communiqué". 2004. Organization for Economic Cooperation and Development. http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html
[8]. Community-based Data Interoperability Networks". 2009. National Science Foundation. www.nsf.gov/pubs/2007/nsf07565/nsf07565.htm
[9]. Harnessing the Power of Digital Data for Science and Society". Report of the Interagency Working Group on Digital Data to the Committee onScience of the National Science and Technology Council. January 2009. pg.2 www.nitrd.gov/about/Harnessing_Power_Web.pdf
[10]. UK Research Data Service". 2009. www.ukrds.ac.uk
[11]. Open Access and Research Conference 2008". 2008 Queensland University of Technology. www.oar2008.qut.edu.au/
[12]. Australian National Data Service". 2009. Monash University. http://ands.org.au/
[13]. Final Report: National Data Archive Consultation". June 2002. Social Sciences and Humanities Research Council.www.sshrc.ca/site/about-rsh/publications/da_finalreport_e.pdf
[14]. Research Data Canada". 2009. CanadaInstitute for Scientific and TechnicalInformation. http://data-donnees.gc.ca/
[15]. Canadian National Committee for CODATA".2009. www.codata.org/canada
[16]. "Policy on Research Ethics". April 7, 2008. McGill University. www.mcgill.ca/files/secretariat/Research-Ethics-Policy-on.pdf

[17]. Management Policy of DFO Scientific Data". June 12, 2001. Fisheries and Oceans Canada. http://www.osl.gc.ca/sgdo/en/info-donnees/acces-donnees.html
[18]. International Polar Year 2007-2008 Data Policy". April 2008. http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf
[19]. The Personal Information Protection and Electronic Documents Act". 2009. Department of Justice Canada. http://laws.justice.gc.ca/en/P-8.6/FullText.html
[20]. Piwowar, HA, Day, RS, Fridsma, DB. Sharing Detailed Research Data Is Associated with IncreasedCitation Rate. 2007. PLoS ONE 2(3): http://dx.doi.org/doi:10.1371/journal.pone.00003