

# Academic libraries' role in Research Data Management Services: a South African perspective

Elisha Chiware<sup>1</sup> and Zanele Mathe<sup>2</sup>

chiwaree@cput.ac.za ORCID: orcid.org/0000-0002-8375-9156

mathez@cput.ac.za ORCID: orcid.org/0000-0002-9996-2992

Received: 8 November 2015

Accepted: 13 January 2016

*Research Data Management (RDM) services are being implemented by academic and research libraries globally in support of university research activities. In South Africa, some libraries are beginning to provide frameworks for these services with some degree of success as policies are being formulated, infrastructure set up, library staff trained, and awareness and advocacy campaigns held with academic staff and researchers. Challenges being faced include availability of resources and infrastructures and limited data management skills among library staff. This paper reports on how the Library at the Cape Peninsula University of Technology is developing and integrating RDM services into institutional research workflows. The paper includes issues that are driving e-research at the institution and how requirements of researchers in the field of biomedical research have been used in a pilot e-research project. The report also details how the university library is using these user requirements to develop tools such as data management plans, electronic laboratory journals and systems for integration with institutional research workflows. The paper further outlines how an international collaborative approach has assisted the Library to participate in the development of an open source platform for the management of the full research lifecycle in support of RDM. It concludes with how further skills development within the Library is being undertaken to support data services and some of the likely challenges for further development of the services.*

**Keywords:** Academic libraries, e-research, e-science, research data management, South Africa

## 1 Introduction

Research funding agencies and national governments are increasingly demanding that grant recipients preserve and make their data publicly accessible with the provision for re-use for further research. The National Science Foundation in the United States of America (USA), the Australian National Data Service in Australia, and the e-Science Core Programme in the United Kingdom (UK) have all been involved in enforcing mandates and advocating for national legislative instruments on data retention and frameworks on responsible conduct of research. The national policies have been used as the guiding principles at various institutional levels to develop and enforce research data management (RDM) services. In South Africa, the National Research Foundation (NRF) has released a statement on open access for data retention to enforce the retention of research data for research that it has funded. Besides the NRF statement, there are many other national initiatives working towards formulating and implementing national frameworks on how academic and research institutions should develop and implement data management services. Lewis (2010) points out that the potential rewards of managing research data – which include sharing; minimising the need to repeat laboratory, field or library work; ensuring the longevity of data; multidisciplinary use of data and its potential to offer new insights; research integrity; and emergence of multidisciplinary research teams – are the driving forces behind the development of these services.

The Cape Peninsula University of Technology (CPUT) Library, which is part of a new division that focuses on Knowledge, Information and Technology Services (KITS), is taking a leading role in creating platforms, systems and processes for the management of research data. The KITS division is made up of Library Services, e-Learning and Educational Technology Services, Management Information Systems (MIS), Computer and Telecommunications Services (CTS), as well the Web Development and Innovation Office. Within this environment, the Library has better access to technology services and to various working committees created to drive knowledge, information, data and technology services within the university. The working committees within KITS are made up of members from all units within this division to ensure synergies are built and approved projects have the necessary institutional support and resources.

The development of RDM services is part of the Library's own e-strategic plan which is closely linked to the institutional goal of being "at the heart of technology education in Africa" (Cape Peninsula University of Technology 2015). The Library's e-strategy includes provision of, among other library technology-related services, e-research support through online research tools and systems that support research and encourage research collaboration within CPUT and beyond. The Library's digitisation, scholarly communication and open scholarship initiatives through its institutional

---

1. Elisha Chiware is Director: Cape Peninsula University of Technology (CPUT) Libraries

2. Zanele Mathe is Research Information Services Manager at CPUT Libraries

repository (which is called Digital Knowledge), are the result of the e-strategy implementation and are seen as an integral part of the research data services which enable research publications to be effortlessly linked to datasets.

The other key driver in terms of the development of research data services at CPUT is the institutional research strategy under the division of Research, Technology, Innovation and Partnerships (RTIP). The RTIP's research strategy, as stated in the Research, Technology Innovation and Partnerships blueprint (Cape Peninsula University of Technology 2013), outlines the Library's role in RDM support as:

Curation, dissemination and promotion of traditional outputs of research in terms of articles and theses, and curation of research data and innovation output, including enhanced research data management systems.

This direct integration and involvement of the Library's role in the university's research strategy has provided it with the avenues to work with the RTIP and other institutional stakeholders to mobilise institutional collaborations for RDM services within the university.

Over the last five years, CPUT has witnessed a growth in its research outputs. This rapid increase in research outputs at CPUT is a result of coordinated and sustained efforts to grow the institution's research profile following government's re-orientation of science and technology in South African higher education institutions through research, technology and innovation policies, strong research governance systems, increased research funding, recruitment of highly skilled staff, research infrastructure development, and the building of collaborations within the university's structures to provide the necessary support to researchers. The university's research strategy as outlined in the Research, Technology Innovation and Partnerships blueprint (Cape Peninsula University of Technology 2013) includes the establishment and appointment of several top-rated researchers as Research Chairs, participation in international research programmes, as well as the mentoring and support of young researchers and growing the number of postgraduate students and postdoctoral positions. These objectives have all contributed to a rich research environment within the institution. Since the merging and inception of the university in 2005, research has been growing in many areas, most notably in the Applied Sciences and Engineering faculties. The focus now is to re-direct the university's research activities towards multidisciplinary research areas. The growing research outputs, together with national and international collaborations in various research projects and the use of multidisciplinary approaches, have provided an opportunity for the university library to initiate data management services to the growing research activities.

This paper reports on the initiatives at CPUT that are aimed at providing researchers with dedicated RDM services. The paper also outlines the steps that have been followed to date and concludes with the successes and challenges faced and what should be done in the future to ensure sustainable services.

## **2 What is Research Data Management?**

RDM services refer to the storage, access and preservation of data produced in particular investigations or research projects. These services support the full data lifecycle including data management planning, digital curation and metadata creation and conversion (Tenopir et al. 2012). A further aim of the RDM services is to ensure research integrity and enable the use of existing data for future research endeavours. As Pinfield, Cox and Smith (2014) assert "research data management is a complex issue involving multiple activities carried out by various actors addressing a range of drivers and influenced by a large set of factors". According to Davidson et al. (2014), researchers, librarians, administrators, ethics advisors and IT professionals all have a vital contribution to make in ensuring that research data and related information is available, visible, understandable and usable over the mid- to long-term. These services are provided in complex environments and require systematic and holistic approaches from many of the key players involved in research support services.

According to Carlson and Garritano (2010), the changes brought about through e-science offer academic libraries opportunities to be involved in the development of cyberinfrastructures and to address the issues and challenges of e-science. University and research libraries have thus been responding by taking on the responsibility and leadership of managing research data. Recent surveys in the USA and Canada (Tenopir, Birch & Allard 2012) and the UK (Davidson et al. 2014) showed that the emergence of data-intensive science and the establishment of data mandates are motivating academic and research librarians to develop research data services for faculty and students. These responses are prompting library directors to plan for additional research data services to be offered by their libraries. At the same time, many librarians are looking for opportunities to develop their research data services-related skills (Tenopir et al. 2014). On the question of whether it is the responsibility of academic and research libraries to manage research data, Lewis (2010: 145) responds both "yes" and "no": "yes", in the sense that data from research represents an integral part of the global knowledge base and their management is an extension of the current library's role; "no", in the sense that the scale of the challenge in terms of infrastructure, skills and culture change requires action by a range of stakeholders and not just university libraries (Lewis 2010). The response, then, has been that of caution on the part of libraries to ensure that they do not lose their credibility along the way.

## **3 Changing nature of scientific research**

Global scientific research outputs have been growing, prompted by the economic and social benefits derived from research. They have also been growing due to the growing collaborative work among scientists across borders and disciplines and due to increased funding from national governments and funding agencies. Science is said to have

entered a “fourth paradigm” which is more collaborative, more computational and more data-intensive than the previous experimental, theoretical, and computational paradigms (Tenopir et al. 2014). This changing nature of scientific research has prompted research and academic institutions to respond by investing in systems to leverage the research data being generated at an exponential rate.

RDM services are being established in response to the changing scientific research landscape and being supported both by the existence of cyberinfrastructures and data-sharing mandates by funding agencies and researchers who are committed to open science, which advocates for open verification and reproduction of research data. Tammaro and Casarosa (2014: 138) further emphasise that an RDM service

encompasses practices such as open publishing of research, campaigning for open access, encouraging scientists to practice open notebook science, and generally making it easier to publish and communicate scientific knowledge.

The growing amount of data being generated due to the new technologies available offers potential for collecting both digital data and research objects (Kruse & Thestrup 2014).

#### **4 The RDM services landscape in South Africa**

In South Africa, some academic and research libraries and research councils have introduced programmes towards the realisation of RDM services in their institutions. The University of Cape Town (UCT) for example has established the UCT eResearch Centre to work and partner with researchers to find appropriate IT solutions for their research activities. The University of South Africa (UNISA) has completed an investigation into RDM practices at UNISA as part of a project to establish data management services (Macanda, Rammutloa & Bezuidenhout 2015). The Universities of Pretoria (van Wyk & van der Walt 2014; van Deventer & Pienaar 2015), Stellenbosch and Witwatersrand (Wits) are all at different planning and implementation stages. The Council for Scientific and Industrial Research (CSIR) and the Data Intensive Research Initiative of South Africa (DIRISA), as well as the National Integrated Cyberinfrastructure System (NICIS) are putting systems in place towards the realisation of technical infrastructures and services for data, processing and connectivity required to enable data exchange, collaborative work and remote access (van Deventer & Pienaar 2015). The IT infrastructures and digitisation activities that exist in many South African universities today, together with national initiatives and communities of practice like the Network of Data and Information Curation Communities (NeDICC) (van Deventer & Pienaar 2015) will provide the basis for the growth and development of RDM services in research and academic institutions in South Africa. The analysis of the South African RDM landscape by Kahn et al. (2014) shows that there are high levels of awareness of RDM in universities and that issues of leadership, policy and skills require the most urgent attention.

The NRF, as the leading government research funding agency, has been involved in many initiatives to allow the sharing of research outputs and datasets. The NRF maintains the South African Data Archive (SADA) which

promotes and facilitates the sharing of research data and related documentation of computerised raw quantitative data of large scale regional, national and international research projects mainly in the humanities and social sciences (National Research Foundation 2015a).

In addition, the NRF maintains a number of research support and knowledge networking databases which contribute to knowledge generation for the support and promotion of research development. The databases facilitate access to research results, technology and innovation in order to allow their utilisation and include: the Current and Completed Research Projects Database; the National Electronic Theses and Dissertations Portal; the NRF Funded Projects Database; the Forthcoming Conferences Database; the South African Professional Associations Database; the Research Information Management Systems (RIMS); Data Resources in Africa; and the Digital Initiatives Register and Heritage Repository (National Research Foundation 2015a).

In March 2015, the NRF released a statement on open access for data retention, mandating that their funded research publications and supporting data be deposited in an accredited open access repository, “with the provision of a Digital Object Identifier for future citation and referencing” (National Research Foundation 2015b). These requirements will certainly change the face of research and drive data curation services in all academic and research institutions benefiting from NRF funding.

Van Deventer and Pienaar (2015: 34) found that “in South Africa, researchers were involved in many disconnected, small projects to enable their participation in collaborative global research projects”. They also noted that “much funding is being used, ineffectively, and that this posed a threat to sustainability because too many of these initiatives were isolated”. Van Deventer and Pienaar (2015) further asserted that “valuable data and information are being transferred to international initiatives with very little regard to intellectual property rights that were being developed”. They suggest that a South African team approach, involving high level participation and commitment to the interests of all researchers, would be considerably more beneficial to the country as a whole.

#### **5 CPUT and the role of its Library in leading RDM services**

One of the important elements of shaping the framework for providing RDM at CPUT was the issue of policy development. A policy framework is important in such an environment as the issues of managing researchers’ data are

very sensitive and contested. It is important in such a policy development process to indicate clearly the roles and responsibilities of the different institutional stakeholders and, more especially, of the library, which is often seen as a mere provider of information resources. Higman and Pinfield (2015) have noted that, in the UK's higher education sector, RDM policy formulation and services development have created complex sets of networks within and beyond institutions, involving different role players with varying priorities. All differing interests have to be addressed in order for the institutional policy to be accepted and for it to achieve the desired results.

The approach at CPUT was to set up an institutional RDM Working Group made up of the Library, Research Office, faculty research representatives, Information and Communication Technology staff, Records and Archives services, the institutional Quality Management unit, the institutional ethics committee chair, research chairs, heads of research units and centres, and the Centre for Postgraduate Studies. The main objectives of the RDM Working Group were to develop an RDM services roadmap for CPUT through the following:

- analysis of RDM practices at CPUT and its contribution to the management of research through its lifecycle;
- use the findings of the analysis to develop user requirements that will inform the development of a university strategy for RDM;
- use the CPUT requirements document as an input for the development of an infrastructure for RDM;
- use findings and infrastructure capabilities to identify relevant RDM services that the library can offer to support researchers; and
- develop models for close and continual communication with researchers and stakeholders in the university research process regarding RDM.

Within the Library, further groups were created to support the work of the institutional RDM Working Group. These groups were:

- Policy analysis and development;
- Requirements gathering and gap analysis;
- Technology and infrastructure; and
- RDM service development.

The RDM Working Group, through meetings and workshops, provided guidance in shaping the CPUT Research Data Management Policy. This policy, together with the institutional Open Access Policy, has now been used to develop the framework for RDM services. The working group also identified potential research groups within the institution that could be used in the pilot phase of the RDM project.

At a national level, the NRF's statement of March 2015, *Open Access to Research Publications from the NRF-Funded Research*, stated that "researchers should deposit final peer-reviewed manuscripts that have been accepted by the journals, to the administering institutional repository with an embargo period of no more than twelve months" (National Research Foundation 2015b). The statement has added further impetus to the RDM policy enforcement at CPUT. The NRF, as a government agent, has an important role to play in setting the framework for a national research data policy and its March 2015 statement will contribute towards this realisation.

Through discussions between the Library and the CPUT Research Office, the NRF requirements have now been incorporated into the NRF and other grant application processes in line with the institutional RDM Policy. The CPUT proposal writing process now requires a description of the types of data to be created or used, the standards used for the storage and preservation of the data, guidelines for ensuring future access to the data, and under which conditions such access will be provided.

## **6 Development of the e-Research Information and Communication (eRIC) Infrastructure at CPUT**

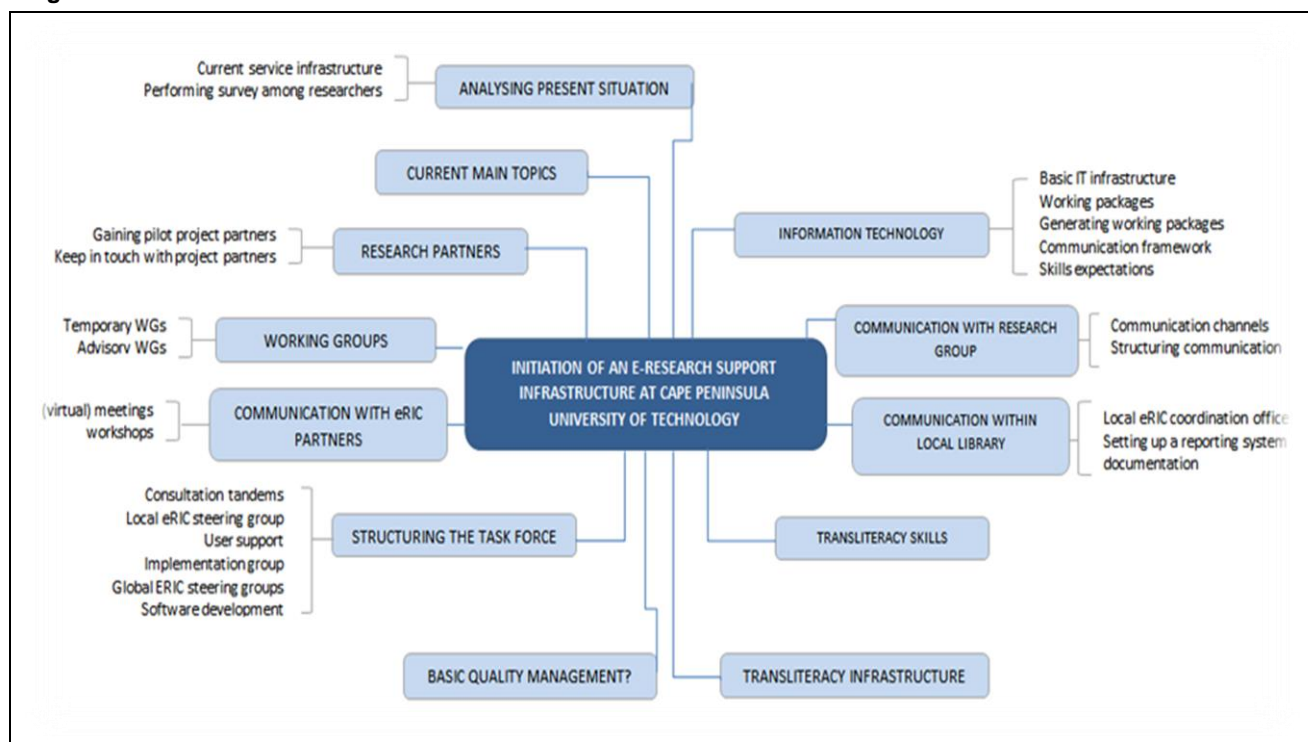
RDM services at CPUT are being developed on the premise that, within an e-research environment, several components exist including: infrastructure development; information flow and management; communication with researchers; development of tools related to the full research lifecycle and the means to store, curate and retrieve data for further use; and the training of researchers.

The RDM services at CPUT are part of a joint project called e-Research Infrastructure and Communication (eRIC, accessible at <http://eric-project.org/>). eRIC aims "to develop an integrated communication and data management infrastructure for accompanying the complete lifecycle of knowledge generation and transfer" (Mitscherling 2014). The project is a collaborative initiative among information specialists in academic and research libraries all over the world. Academic and research libraries in Germany, South Africa and Thailand are working in collaboration with systems developers, Information Technology (IT) experts, and researchers in various fields of expertise to develop software platforms and services to support RDM services (Chiwere 2015). Each institution is in control of its own environment and only the development of the platform, the exchange of ideas through working committees and groups, relevant skills transfer and problem solving are shared.

The end product is that CPUT Libraries has created a new model that serves as a roadmap for developing an RDM service infrastructure that will provide data management support services to the various research groups within the university. Figure 1 shows that at the centre of the eRIC project at CPUT are various elements that are considered during

the development process which make up the service model. At the core of the model is building partnerships with researchers and accompanying them through the full research lifecycle. The other important elements are: basic IT infrastructure; communication channels with research groups; communication channels within the Library; information literacy programmes; quality assurance measures; mechanisms for building working groups within the Library; the ability and capacity to identify potential researchers with whom to work and to be able to analyse their research topics; and to carry out needs assessments for data management services. The model ensures that there is constant communication with eRIC project partners in other countries for efficient project management and shared development of technical components.

**Figure 1 e-Research Service Model at CPUT**



## 6.1 Software development process

The approach at CPUT Libraries has been to work collaboratively with other universities internationally to develop an open source platform by engaging researchers about their requirements; the open source platform will be further enhanced by embedding different tools to enable the capture and storage of research data through their lifecycle. The software infrastructure being developed is open source – free for reuse by other interested libraries globally. All project results are documented and published as open source. The development process is collaborative in nature with researchers, librarians and IT experts providing input. The platform aims to provide modular solutions as all eRIC tools and services are part of a flexible multi-tier software and service architecture. They are also customer-oriented and all customer requirements are documented, analysed, and implemented from the bottom up (Chiwere 2015).

## 6.2 Pilot project partners and requirements assessment

CPUT Libraries decided to run a pilot project with one or two research groups before rolling out the full service infrastructure to the university. The development of RDM services should be based on the needs of researchers, the identification and integration of the necessary technologies as well as the expertise to manage the platforms and workflows. Through the CPUT RDM Working Group, researchers were invited to workshops on data services. The purpose of the workshops was to introduce the initiative, provide a platform for the institutional policy development processes, gather requirements from a larger population of researchers, and gain pilot project partners willing to commit to the initiative.

Subsequently, the Institute of Biomedical and Microbial Biotechnology (IBMB), one of the specialised research units within CPUT, agreed to work closely with the Library in the development of the services based on their requirements. A requirements-gathering audit with researchers within the IBMB was conducted, based on structured interviews that were meant to establish current research data creation and management practices, the type of services the Library could provide to improve data management, and the patterns of communication within the research environments and how these can be incorporated into future data services.

The RDM lifecycle concept was used to structure the interview questions that were used to guide the audit. The lifecycle concept has helped focus attention on issues of data quality and documentation at the time of creation, critical to data-driven research as well as for successful data preservation and sharing (Ray 2014). The practice of using the data

lifecycle concept for audit purposes has also been realised through the Curation Lifecycle Model developed by the Digital Curation Centre (DCC 2015). The model, according to the DCC (2015), provides a “graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt through the iterative curation cycle”. The DCC suggests that the model can be used to plan activities within the organisation or consortium “to ensure that all of the necessary steps in the curation lifecycle are covered”. It is for this reason therefore that the Library selected the Curation Lifecycle Model as a suitable model to guide the process of gathering the requirements and developing the service infrastructure for RDM at CPUT.

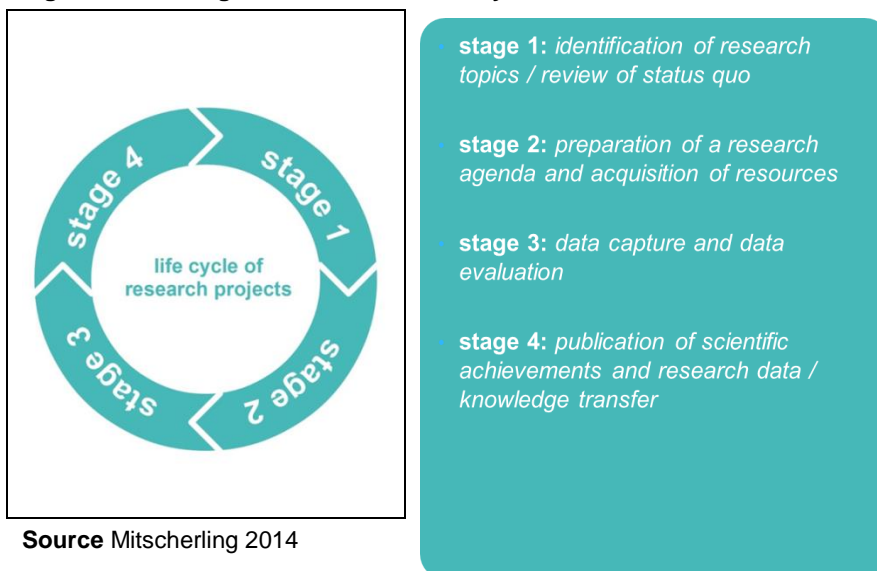
The results of the data management requirements survey collected during the workshops and those collected specifically from the IBMB group proved to the Library that there was a great need for structured services and tools for research groups within the institution. The current methods and tools used to manage data are not organised in any systematic way and do not provide for access by researchers outside the research group, unless special arrangements have been made. In many cases, researchers within these environments have different approaches to managing their data and, in the end, expose themselves to possibilities of losing critical data. Furthermore, the analysis of the IBMB requirements revealed that researchers within this group have a need for other specialised library services like digitisation of research reports and laboratory journals to ensure their accessibility to the public and for long-term preservation.

The IBMB presented CPUT Libraries with an opportunity to pilot the RDM project by developing services closely accompanying the data lifecycle during the development process. This research group, formerly part of the Medical Research Council, has joined CPUT to create a new research hub within the medical sciences. Led by a Research Chair, the group of experienced researchers, postdoctoral students and postgraduate students was seen as an ideal source of various data sources that can be used for experimentation on the platform being developed.

### 6.3 Data lifecycle and generic service layer

After the analysis of the requirements gathered from the IBMB group and workshop attendees, the next phase was to extract the interdisciplinary aspects of the requirements gathered and aggregate them to map out an ideal research lifecycle process that supports data services. The four-stage lifecycle was revealed as a generic software layer that consists of: identification of research topics that are being investigated and reviewing the status quo; development of a research plan and acquisition of resources; data capture and data evaluation; and publication of scientific achievements and research data or knowledge transfer. Through these stages, data management solutions that meet the interdisciplinary needs of the researchers are identified which aid in identifying and accounting for the roles, responsibilities, milestones, and other key components that must be considered in developing a data management service infrastructure to support the university community. Figure 2 presents the four stages of an ideal research lifecycle that supports interdisciplinary management of research data. As indicated earlier, the development of the software and service infrastructure is based on user requirements and its bottom-up, incremental solutions are prioritised as per requirements analysis results. The stages are outlined in the sections that follow and detail the progress that has been made thus far.

**Figure 2 Four stages of the research lifecycle**



#### 6.3.1 Stage 1: Identification of research topics or reviewing the status quo

The requirements from the group revealed that there is an interest in literature sources that have detailed publications of datasets that are specialised and subject-specific. In response to this requirement, solutions responsive to identification of research topics and reviewing the status quo were identified. Therefore in Stage 1, the research group is supported

through the Library's traditional services of providing diverse literature sources in different formats. During this stage, new services aimed at supporting researchers in identifying sets of discipline-specific data sources relevant to the research group have been identified. These services include the identification and recommendation of relevant discipline-specific data repositories for reviewing secondary data sources, a list of data journals both generic and subject-specific, and much more. The Library has thus set up alerts to receive updates on new and upcoming publications which are accompanied by data sources and which can be added to the Library's collection. At this stage, the aim is to integrate traditional library services related to the review of literature and new services that are specifically related to the review of data sources onto one platform and have them operate together seamlessly. The CPUT systems developer is working in close collaboration with developers from other project partners in this integration.

### **6.3.2 Stage 2: Development of a research plan and acquisition of resources**

At this stage, the objective is to support researchers in creating a project structure, planning individual work stages, using resources, and documenting entire research projects. Project planning includes management tools to support efficient resource planning including human resources, budget planning and scheduling (Mitscherling 2014). Thus far, there has been a successful integration of the project management and data management tools within the eRIC platform. At this point, CPUT is working on the implementation of the Lightweight Directory Access Protocol (LDAP) and Shibboleth user authentication standards to synchronise user accounts across multiple applications.

### **6.3.3 Stage 3: Data capture and evaluation**

This stage provides for the collection and analysis of primary data, manually or electronically. As there are varying multidisciplinary requirements in the collection of field or laboratory data, provision is made within the systems to cater for discipline-specific data capturing and evaluation requirements. The requirements gathered from the IBMB group revealed that research data is generated and captured using diverse data capturing techniques and tools which include the use of:

- questionnaires on human studies and focus groups, conducting experiments with gene expression using real-time polymerase chain reaction (PCR) machines for molecular procedures;
- different machines to conduct experiments on proteomics (the large-scale study of proteins); and
- DNA imager to visualise images and do image analysis.

In response to the data capturing and evaluation requirements identified from the IBMB group, it is envisioned that the main platform will make provision for interoperability with other data processing systems such as SPSS, Matlab, Origin, and many other statistical processing packages, while simultaneously providing interoperability with related data-capturing proprietary software used by the research group.

### **6.3.4 Stage 4: Publication of research results and knowledge transfer**

eRIC aims to provide tools and services to support researchers in creating both electronic and printed publications (Mitscherling 2014). In response to the requirement made by the IBMB group on digitisation of research reports and laboratory journals to ensure their accessibility to the public at the appropriate time and ensure their long term preservation, the Library, with the assistance from the research group members, has started identifying laboratory journals that should be prepared for digitisation and thereafter stored on Digital Knowledge (the institutional repository) with links to the relevant research outputs. Workflows on submission of these resources to the Library are being discussed and designed. This service has become generic, requested by other laboratories within the institution. The integration and implementation of the electronic laboratory journal will eliminate this exercise in future as the tool will be accessible within the platform.

## **6.4 Specialised tool layer of the service infrastructure**

The requirements analysis of the IBMB group revealed that there is a need to accommodate subject-specific requirements which cannot be reduced to interdisciplinary features that have already been identified. Subsequently, a specialised tool layer that will meet special subject-related needs that the group has required has been created. Figure 3 shows the layered software and service architecture and how specialised tools work within the framework to support a data collection, analysis and retention stages.

## **6.5 Working groups**

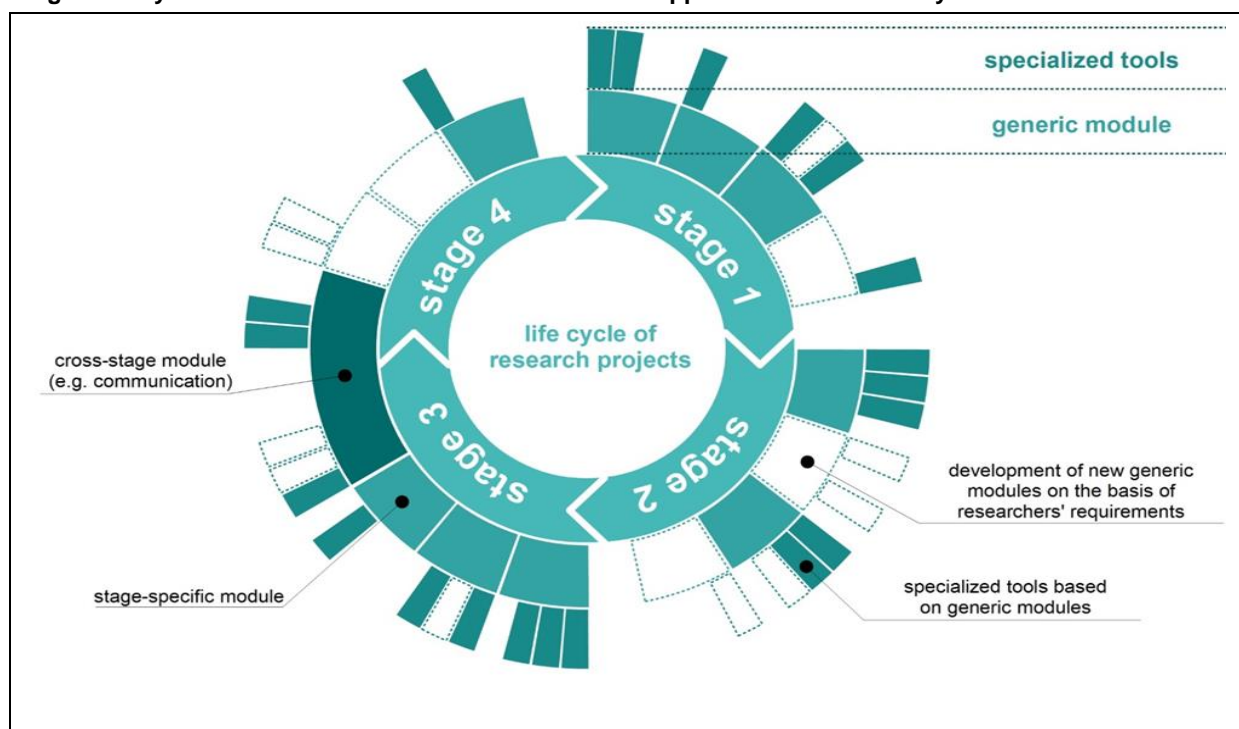
Working Groups have been set up both at collaborative and institutional levels to focus on the development of various tools and systems on the shared platform with the aim of subsequent internal and external implementation within the RDM platform, MediaTum. The Working Groups are:

- Work Group on Communication Tools and Researcher Profiles, tasked with the development of a set of specifications for the integration of communication tools, as well as tools for organisation and storage of ideas in the data management system and tools for the preparation and publication of researchers' profiles;

- Work Group on the Reference Management System and Visualisation of Data, responsible for developing specifications for the reference management system and their integration into the data management system to enable visualisation of scientific data;
- Work Group on the Research Plan and Electronic Project Journal, which aims at developing a set of detailed specifications for the integration of research plans and electronic project journals within the data management system framework; and
- Work Group on Search Functionalities, Semantic Methods, and Search Engine Optimisation, which aims to develop sets of specifications for new search functionalities within the data management system. In addition, semantic methods are evaluated and (if potentially useful) implemented. Methods for search engine optimisation should be gathered and implemented to optimise the visibility of digital objects in the data management system.

The groups work closely with feedback from the user requirements analysis and help shape workflows in the RDM services. All developments by the groups are done in close collaboration with researcher teams and potential users of the systems.

**Figure 3 Layered software and service architecture to support the research lifecycle**



Source Mitscherling 2014

## 7 Skills development

One of the biggest challenges facing academic and research libraries attempting to offer RDM services are the limited skillsets of librarians as well as the lack of understanding of the diversity of 'research data' itself within the context of different disciplinary and sub-disciplinary cultures and varying data practices (Cox, Verbaan & Sen 2012). Tenopir et al. (2012) have suggested that, to overcome the challenges of skillsets in RDM, the best approach is to reassign the existing staff to new roles with appropriate training. Cox, Verbaan and Sen (2012) list many other challenges that libraries face in staffing RDM activities and these include how to balance existing roles with new RDM roles; the lack of understanding by librarians of the motives and practices of researchers and the diversity of research data; the lack of domain-specific knowledge, using current experiences for creation of research data contexts; and the problem of engaging researchers who view librarians' role as that of supporting teaching. However, regardless of these views and challenges it is the responsibility of the librarians to change perceptions and convince researchers that they can undertake the role.

At CPUT Libraries, a special skills development plan has been established for a three-year period to enable librarians and especially subject/faculty librarians' participation in RDM. The Library has also started creating new roles dedicated specifically to managing and developing e-research platforms. One such position within CPUT Libraries is the position of an e-Research Systems Developer who will be responsible for developing all the systems and tools for RDM services.

The special skills development programme for librarians within CPUT Libraries includes the following areas:

- subject-specific orientation;
- understanding of institutional research processes and policies;



- knowledge of relevant tools (such as Mendeley, Scopus author management, Data Management Plan Tool, , ORCID identifier);
- communication and collaboration (video, web-conferencing technologies);
- current awareness – alerts and RSS services;
- research methodologies (for example, research data lifecycle, data analysis, tools, statistics);
- bibliometrics (for example, impact factor, h-index) and altmetrics;
- Intellectual Property rights;
- publication processes and requirements;
- awareness of DOAG.org and Sherpa Romeo; and
- academic networking.

The training programme is run both by internal staff and external service providers depending on the skills to be offered. Joint international training workshops within the eRIC partnership have also been held to further familiarise staff and build capacity in understanding approaches in RDM services.

In South Africa, it is encouraging to note that some library and information science schools have started to recognise the need for skills in this area and are beginning to offer a number of programmes dedicated to data management and curation. UCT's Library and Information Studies Centre offers a short course in RDM and an MPhil in Digital Curation. In other disciplines, the Wits School of Public Health offers a Master of Science in Epidemiology in Research Data Management. Some funding agencies and research councils like the NRF and CSIR have been organising awareness and information-sharing workshops to build capacity among library professionals. These exercises will go a long way to addressing the much-needed skillset in this emerging service area within academic and research libraries. Tamaro and Casarosa (2014) concluded in their investigation about the need for RDM in the curriculum to be an interdisciplinary approach, that data curation should move to be "beyond providing technology for storage and preservation, to embrace changes in scholarly production that emerging technologies have brought, to make sense of the digital curator as a whole".

## 8 Conclusion

Many lessons have been learnt from the RDM services initiatives within CPUT Libraries. These lessons have helped the Library to position itself well concerning engagements with researchers, approaches to requirements analysis, identification of gaps within research groups' data management practices, and the development of tools and systems for data services. Lastly, the mobilisation of institutional support through the CPUT RDM Working Group on RDM policy formulation and the start of a well-integrated institution-wide RDM service have also been important lessons to CPUT Libraries. The initiative has proven that collaboration between the Research Office, Library, researchers, IT and many other divisions can result in the effective and efficient delivery of RDM services. Furthermore, the initiative has created awareness of the RDM services within the different university structures including senior university management and management committees. International collaborative lessons have been learnt through the eRIC infrastructure development project and has provided library staff dedicated to the RDM project an opportunity to work at the international level. These approaches at CPUT could also be used by other universities seeking to develop and implement RDM services.

The pilot project has generated more interest within various research groups within the institution; the challenge now is to prepare more librarians to be at the forefront of the planned services. The challenge that the Library faces is the adaptation of librarians' workload and changing of mind-set to accommodate this kind of new service provision. It is believed that the skills development programme that has been put in place will play a significant role in preparing and improving librarians' skills in managing the research data needs of respective faculties. What also remains a challenge is the expansion of the current infrastructure to ensure that a university-wide e-Research Centre is built in order to support all researchers with dedicated services. At the national level, the South African government and the departments of Higher Education and Science and Technology need to provide clear national guidelines and legislative tools to guide universities, research councils and other entities in the development of research data services. Analysis from research on RDM services in other countries shows that a number of key elements for RDM services should be in place and these should include strategies, policies, guidelines, processes, technologies and services (Pinfield, Cox & Smith 2014). The proposed National Innovation Information Portal, the DIRISA initiatives, the High Computing Performance Centre within the CSIR, the NRF, the National Digital Library initiative and NeDICC will all eventually converge towards the realisation of a national e-research infrastructure that will enable the preservation and storage of research data and other national intellectual outputs.

## References

- Cape Peninsula University of Technology. 2013. *Research, Technology Innovation and Partnerships: ten year blueprint*. [Online]. [https://www.cput.ac.za/storage/research/RTI\\_Blueprint\\_2013\\_WEB.pdf](https://www.cput.ac.za/storage/research/RTI_Blueprint_2013_WEB.pdf) (5 November 2015).
- Cape Peninsula University of Technology. 2015. *Vision and Mission*. [Online]. <http://www.cput.ac.za/about/vision> (5 November 2015).

- Carlson, J.R. and Garritano, J.R. 2010. E-science, cyberinfrastructure and the changing face of scholarship: organizing for new models of research support at Purdue University Libraries. *Libraries Research Publications*, Paper 137. [Online]. [http://docs.lib.purdue.edu/lib\\_research/137](http://docs.lib.purdue.edu/lib_research/137) (5 November 2015).
- Chiware, E.R.T. 2015. Research Data Management Services at Cape Peninsula University of Technology. *Proceedings of the COAR-SPARC Conference on Connecting Research Results, Bridging Communities and Open Scholarship*. 15-16 April 2015. Porto, Portugal. [Online]. [http://coar-repositories.org/files/3\\_Chiware-COAR-SPARC-Presentation.pdf](http://coar-repositories.org/files/3_Chiware-COAR-SPARC-Presentation.pdf) (3 November 2015).
- Cox, A., Verbaan, E. and Sen, B. 2012. Upskilling liaison librarians for research data management. *Ariadne*, 70. [Online]. <http://www.ariadne.ac.uk/issue70/cox-et-al> (2 November 2015).
- Davidson, J., Jones, S., Molloy, L. and Kejsler, U.B. 2014. Emerging good practice in managing research data and research information in UK Universities. *Procedia Computer Science*, 33: 215-222.
- Digital Curation Centre. 2015. *DCC Lifecycle Model*. [Online]. <http://www.dcc.ac.uk/resources/curation-lifecycle-model> (8 November 2015).
- Higman, R. and Pinfield, S. 2015. Research data management and openness: the role of data sharing in developing institutional policies and practices. *Program: electronic library and information systems*, 49(4): 364-381.
- Kahn, M., Higgs, R., Davidson, J., and Jones, S. 2014. Research data management in South Africa: how we shape up. *Australian Academic & Research Libraries*, 45(4): 296-308. DOI:10.1080/00048623.2014.951910.
- Kruse, F. and Thestrup, J. B. 2014. Research libraries' new role in research data management, current trends and visions in Denmark. *LIBER Quarterly*, 23(4): 310- 334.
- Lewis, M. J. 2010. Libraries and the management of research data. In *Envisioning future academic library services*. S. McKnight, Ed. London: Facet. 145-168.
- Macanda, M., Rammutloa, M. and Bezuidenhout, R. 2015. Research Data Management at UNISA. [Online]. <http://hdl.handle.net/10500/13907> (20 December 2015).
- Mitscherling, C. 2014. Integrated communication and service infrastructure for libraries. *Proceedings of the IATUL Conferences*. 2- 5 June 2014. Aalto University, Espoo, Finland: Purdue. Paper 2. [Online] <http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2014&context=iatul> (5 November 2015).
- National Research Foundation. 2015a. *Databases*. [Online]. <http://www.nrf.ac.za/information-resources/databases> (20 December 2015).
- National Research Foundation. 2015b. *Statement on Open Access to Research Publications from the National Research Foundation (NRF)-Funded Research*. [Online]. <http://www.nrf.ac.za/media-room/news/statement-open-access-research-publications-national-research-foundation-nrf-funded> (5 November 2015).
- Pinfield, S., Cox, A.M. and Smith, J. 2014. Research data management and libraries: relationships, activities, drivers and influences. *PLoS ONE*, 9(12): e114734. DOI:10.1371/journal.pone.0114734.
- Ray J. 2014. Introduction to research data management. In *Research data management: practical strategies for information professionals*. J. Ray, Ed. West Lafayette, IN: Purdue. 1- 23.
- Tamaro, A.M. and Casarosa, V. 2014. Research data management in the curriculum: an interdisciplinary approach. *Procedia Computer Science*, 38: 138-142.
- Tenopir, C., Birch, B. and Allard, S. 2012. Academic libraries and research data services: current practices and plans for the future: an ACRL white paper. Chicago: Association of College and Research Libraries. [Online]. [http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir\\_Birch\\_Allard.pdf](http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf) (20 December 2015).
- Tenopir, C., Sandusky, R.J., Allard, S. and Birch, B. 2012. Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, 39(1): 70-78.
- Tenopir, C., Sandusky, R.J., Allard, S. and Birch, B. 2014. Research data management services in academic libraries and perceptions of librarians. *Library & Information Science Research*, 36: 84-90.
- Van Deventer, M. and Pienaar, H. 2015. Research data management in a developing country: a personal journey. *International Journal of Digital Curation*, 10(2): 33-47 [Online]. <http://www.ijdc.net/index.php/ijdc/article/viewFile/10.2.33/406> (20 December 2015) DOI: 10.2218/ijdc.v10i2.380.
- Van Wyk, J. and Van der Walt, I. 2014. Going full circle: research data management @ University of Pretoria. *Proceedings of the eResearch Africa Conference*. 23-27 November 2014. University of Cape Town, Cape Town: eResearch Africa. [Online]. [http://www.eresearch.ac.za/sites/default/files/image\\_tool/images/140/Going\\_Full-Circle\\_RDM\\_UP\\_VanWyk\\_VanderWalt\\_26Nov\\_2014.pdf](http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/Going_Full-Circle_RDM_UP_VanWyk_VanderWalt_26Nov_2014.pdf) (20 December 2015).