

Research Data Management in a Developing Country: A Personal Journey

Martie van Deventer
University of Pretoria and CSIR

Heila Pienaar
University of Pretoria

Abstract

This paper explores our own journey to get to grips with research data management (RDM). It also mentions the overlap between our own ‘journeys’ and that of the country. We share the lessons that we learnt along the way – the most important lesson being that you can learn many wonderful and valuable RDM lessons from the international trend setters, but in the end you need to get your hands dirty and get the work done yourself. You must, within the set parameters, implement the RDM practice that is both appropriate and acceptable for and to your own set of researchers – who may be conducting research in a context that may be very dissimilar to that of international peers.

Received 05 January 2015 | *Accepted* 10 February 2015

Correspondence should be addressed to Heila Pienaar, University of Pretoria, Lynnwood Road, Pretoria. Email: Heila.Pienaar@up.ac.za

An earlier version of this paper was presented at the 10th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Introduction

Our personal journey started with the South African Research Information Services (SARIS) project early in 2004. The SARIS project was funded by the Ford Foundation, and initially focussed on finding a solution to the extremely high costs to access electronic research literature (Page-Shipp, et al., 2005). From that investigation it was very soon clear that a new research paradigm (known as eResearch or eScience) was in the process of emerging and that we needed to widen the scope of our investigation to include open access as well as research data management (RDM).

Visits to various institutions in the UK brought it to our attention that eResearch could be seen as a composite of three main trends:

- The ability to transfer large volumes of data and to share computation capacity between remotely situated researchers as the basis of eScience as ‘faster, different, better research’ (T. Hey, personal communication, June 22, 2004);
- The need to make better use of expensively created databases by ‘the active management and appraisal of data over the life cycle of scholarly and scientific interest’ as the basis of a new field of endeavour called Digital Curation (P. Burnhill, personal communication, June 28, 2004);
- Scholarly discourse was taking place on the dual playing field of commercial publication and open access (Page-Shipp, et al., 2005). For published information this trend was already clearly visible but open data soon followed.

We also discovered, back in South Africa, that researchers (even at our own institutions) were initiating many disconnected, small projects to enable participation in the global research network. Much funding was being used, often ineffectively, which threatened sustainability because too many of these initiatives were isolated. In addition, valuable data and information were being transferred to international initiatives – with very little regard for the intellectual property that was being developed. It was understood that a ‘Team SA’ approach, with high level participation and commitment to the interests of all researchers, would be considerably more beneficial. The journey to get to that point though was much more difficult than what any one of us could have imagined. Root causes for the delay in fast tracking the South African initiative included a lack of knowledge and skills, limited commitment from very senior decision makers and a naiveté regarding the infrastructure required to make eResearch (and with that then RDM) a reality.

Today the research infrastructure to manage the exchange of large data sets and to manage data in general is substantially better understood. Fernihough’s (2011) research made it much easier for all of us to understand where to position RDM. Obviously the Square Kilometre Array (SKA) project, with its large data management requirements on the one hand and attempts by people like Lötter (2011) to raise general awareness regarding the ‘long tail’ of research data on the other, has provided the South African community with considerably more maturity than we experienced in 2008.

However, each of our organisations has followed its own path of development. The chronological development of RDM initiatives at the University of Pretoria started in 2007 with the development of an initial RDM policy for the University. Between then and the final draft policy, which is currently (January 2015) with the University

Executive, there has been an in-depth evaluation of practices and several pilot installations of fit-for-purpose RDM initiatives.

At the CSIR the RDM progress was considerably less marked. We did, however, appoint a data librarian during 2013 and it is now possible to make real visible progress. An extensive review of the CSIR's RDM practices was done early in 2014. This allowed the data librarian to collect factual evidence (Patterton, 2014) that CSIR research staff behave no different than any other organisation's researchers – similar to what was established through several international studies.

The national journey is also starting to gain momentum (Lötter, 2014). The most promising initiative is known as DIRISA (the Data Intensive Research Initiative of South Africa). Our community of practice, the Network of Data and Information Curation Community (NeDICC), as well as our institutions individually are actively participating in the DIRISA activities in an attempt to make RDM a reality for all.

Each of these markers or signposts on our journey is discussed in some detail in the following sections. Our learning did not happen in isolation and it was fortunate for us that the international community was so willing to share the knowledge that it had already generated.

Learning Gained from International Colleagues

As mentioned previously, our journey started with a visit to the UK during June 2004. We were in awe when, at one of our first engagements we were informed that:

‘The focus area for JISC in future would be to effectively use management information systems to make e-Learning as well as e-Business much more efficient at higher education institutions. This would include making available both the technology (infrastructure and applications) as well as the skills to analyse the data harvested to ensure that each student and his/her progress is seen as a whole entity rather than small bits of a puzzle that needs to be fitted together’ (M. Reed, personal communication, June 22, 2004).

At that stage we also did not quite understand what was meant but as we were exposed to the thinking of senior staff at JISC, ContentComplete, University College London, ResearchResearch, the MIMAS data centre at the University of Manchester, the EDINA data centre at the University of Edinburgh, EEVL (The Internet Guide to Engineering, Mathematics and Computing) and the British Library we came to realise that we had stumbled onto something much bigger than what we anticipated and that we needed to refocus our research. This we did before other team members visited Brazil, the USA (Ohio), Canada (Toronto) and Australia. For that second phase of the study tour, we deliberately identified only institutions that were similar to either the University of Pretoria or the Council for Scientific and Industrial Research (CSIR) so that we could also investigate how they were making progress within this larger context. In hindsight, this was probably a mistake as we had, without realising it, started focussing on the operational side of getting things done rather than on the decision making that enables the process.

Although research funding is made available via our National Research Foundation (NRF), South Africa does not have a funding agency such as the UK's JISC. Most of the

research infrastructure investment is done via our Department of Science and Technology (DST). We naively thought that existing infrastructure and relationships would be sufficient – that our new knowledge could lead to services and products that could exist within an existing context. We now know that it may have been more appropriate to have focussed on actively cultivating a network of people, from a variety of disciplines, and to persuade the Department to include libraries when developing its cyberinfrastructure plans.

Nevertheless, our report and recommendations were well received by both the funder and all South African stakeholders. The research was subsequently published (Page-Shipp et al., 2005). Our Department of Science and Technology initiated a workshop in 2007 with the objective explore the application of the OECD Principles and Guidelines on Access to Research Data in South Africa and to identify a policy direction for the country in this regard, so that research data in the country may be more comprehensively used for the benefit of the research community, industry and society at large. However, shortly after the workshop several of the key stakeholders moved on and it was only when Fernihough (2011) expanded the eResearch model that we initially developed that we understood the entire eResearch framework better, were able to start focussing on the service layer of the model and could explain where our RDM activities and focus should be.

In the interim, the University of Pretoria Library was evaluated in 2010 by an international panel of experts. One of the evaluation panel experts was Dean Jim Mullins of the University of Purdue Library, USA and an international leader on the topic of RDM. He invited two managers to visit their library to gain first-hand experience of their RDM activities. In April 2014 Dr Heila Piennaar and Mr Isak van der Walt, a Library IT specialist, visited Purdue as did Mr Robert Moropa during November of the same year. Valuable information was gathered, especially about the role of the information specialist in RDM, the Purdue Research Repository (PURR) and the long-term preservation of data.

NeDICC's Role and Activities

Immediately upon the completion of our funded SARIS project, we initiated a process to establish a network of people involved in data and information curation activities (NeDICC¹), using the SARIS partners as the base community. Several meetings later, and after attending the first DCC conference, we received advice from Chris Rusbridge (personal communication, October 14, 2005) to kick-start our initiative with a conference that would pull in several international experts. This we did and the 1st African Digital Curation Conference and Workshop was held in Pretoria on 12th and 13th February 2008. At that conference it was decided, in principle, to formalise the NeDICC group. This decision was ratified at several events and the network was then formally established during 2010. Although initiated by the authors, the conference(s) and the establishment of NeDICC directly resulted from learning gained from international colleagues.

Conferences were subsequently held in 2009, when the title was changed to the African Digital Scholarship and Curation Conference (van Deventer and Piennaar, 2009), 2010 (Gaberone, Botswana), 2011 (Pretoria) and 2013 (Durban). After the 2013

¹ Network of Data and Information Curation Communities (NeDICC):
<http://nedicc.com/2012/08/23/nedicc/>

conference it was decided to rather join forces with the eResearch Africa conference² and to specifically develop a research data management track – our community is currently too small for two such conferences to exist side by side.

NeDICC currently functions as a community of practice. It's membership represents three large research universities and three science councils. The focus of our activities is on developing own expertise and on collectively raising the awareness of RDM as an essential component of good research practice. As a group the community is responsible for several training and development initiatives to encourage the active and reliable management of research data. The NeDICC:

- Provides a forum for discussing issues of mutual concern, providing support and working towards solutions for specific constraints;
- Exposes the community to new developments and trends, providing opportunities to engage with a wider audience, as well as to showcase work and initiatives;
- Develops the knowledge and skills of the members;
- Promotes awareness/best practices relating to digital preservation, dissemination and use the of research data;
- Collaborates on projects in support of shared objectives.

Collectively the community has:

- Investigated the role of the funder in encouraging good RDM practice;
- Gained a detailed understanding of data management planning;
- Experimented with *Bag-It* as a preservation technology;
- Received training in the management of large data;
- Investigated the integration of RDM with the ethics clearance process;
- Gained a detailed understanding of the management of human sciences data – across the RDM lifecycle;
- Investigated long term preservation issues;
- Experimented with the integration of RDM with the Records Management activities;
- Developed workflows that ensure that research data is surfaced and is traceable;
- Experimented with RDM within a virtual research environment context (van Deventer and Pienaar, 2012; Pienaar and van der Walt, 2014);
- Actively trained librarians in the theory (and limited practice) of RDM (Carnegie-funded continued professional development programme);
- Contributed towards the understanding of data management policy (van Deventer, 2013a)
- Participated in developing the principles of data citation and is actively promoting proper data citation (van Deventer, 2013b);

² eResearch Africa Conference: <http://www.eresearch.ac.za/>

- Investigated the use of persistent identifiers to add research data to existing repositories;
- Completed RDM situation analyses that enables cross institution comparisons (Pienaar, 2010; Patterton, 2014).

In the immediate future the community needs to formalise some of the work that has been done and to then also actively become part of the DIRISA national initiative, which is discussed in more detail below.

Linking NeDICC to Africa

One of the most difficult things to do is to keep track of colleagues in Africa, where professionals appear to be in a constant state of flux between jobs, institutions and even between countries. It is therefore not always clear who the right contacts would be at any given organisation. The formal structures that support networks just do not function as well as they appear to do in the developed world. Here too the question remains: do we build contacts with the operational staff – those who need to actively do and promote RDM – or should we be looking for the decision makers? Again the national initiative (DIRISA) would in all probability provide more clarity but in the meantime we have, through the African conference, our rookie RDM workshops, and through the Carnegie-funded training initiatives, been able to establish a network of librarians in Africa who at least are aware of RDM, even if they are not able to immediately start implementing good RDM practices. This is perhaps a very ‘bottom-up’ approach but we prefer to inform and influence practice rather than to, at this stage, pretend that we know what would work in ‘their’ environment.

Placing our RDM Activities within the Broader National Context

Although both the University of Pretoria (UP) and the CSIR are members of the NeDICC community, each of our organisations has followed its own path of development.

RDM at the University of Pretoria

The chronological development of RDM initiatives at the University of Pretoria started in 2007 with the development of a RDM policy for the University. Between then and the new/updated draft policy (which was sent to the Deputy Director for Research and Postgraduate Studies during September 2014) there has also been an in-depth evaluation of practices and several pilot installations of fit-for-purpose RDM initiatives.

The chronological steps in more detail:

- 2007: Policy for the preservation and retention of research data (Rt 306/07)
- 2010: Survey of RDM practices at UP (October 2009 – March 2010) during which the following recommendations were identified:

- A RDM policy should be developed for UP;
 - A central RDM office should be established;
 - Formal RDM practice should be initiated in each of the Faculties;
 - The Impact of RDM on workload and time of researchers and students should be taken into account;
 - A data repository should be established for UP;
 - The necessary IT infrastructure, that can accommodate small and big data sets, as well as high performance computing should be investigated;
 - A timeframe for the roll-out of a RDM system should be developed for UP (Pienaar and van der Walt, 2014)
- 2013: Assignment of a full time staff member to be responsible for RDM
 - 5 Feb 2013: Meeting with Deputy Dean Research of the Health Sciences Faculty to discuss the possibility of pilot project in the Faculty
 - 11 April 2013: Meeting with Professor Michael Pepper – Director of Institute for Cellular and Molecular Medicine (ICMM) – to discuss the possibility of a pilot project with his students
 - 10 May 2013: Discussion session with ICMM students to determine their needs
 - 24 May: Demo session to Prof Pepper, where we showed an example of one of the completed doctoral student’s data on the Moodle software. Data organised according to the research cycle (van Deventer and Pienaar, 2012)
 - 27 May and 6 June 2013: Demo sessions to the Library Executive and the UP Research IT committee
 - 19 July 2013: Demo session to ICMM students using Alfresco software
 - August – September 2013: Training of Prof Pepper, Laboratory manager and students on the Alfresco system
 - August – November 2013: At the request of the Vice Principal, research interviews were held with the Deputy Deans of Research from all the Faculties to determine “Essential research data that the University should manage”
 - December 2013: Feedback on the interviews to the UP Library Advisory Committee
 - December 2013: Second pilot project – Neuro-Physiotherapy (word of mouth)
 - January – February 2014: Training of students of Neuro-Physiotherapy and their Information Specialist on the Alfresco system
 - 22 January 2014: Feedback on the interviews with the Deputy Deans of Research to the UP Research IT committee
 - 17 March 2014: High Level Report compiled on RDM internationally, nationally and at UP for the Vice Principal of Research
 - April 2014: Visit by Deputy Director for Innovation and Technology and Library IT Specialist (UP Library Services) to Purdue University in USA to investigate

Purdue's Research Data Repository (PURR) and long-term preservation processes as possibility for replication at UP

- June 2014: Assistant Director for RDM attended CODATA International Training Workshop in Big Data for Science for Researchers from Emerging and Developing Countries, in Beijing, China
- July 2014: High Level Report on RDM sent through to University Executive for review
- July 2014: Implementation of more pilot projects, including Potato Pathology Programme, Powdery Scab, and Psychiatry Dissociation
- August 2014: Proposed new University policy on RDM sent through to University Executive for review
- January 2015: Task Team (Library and IT) to investigate infrastructure needed for RDM across the University.

RDM at the Council for Scientific and Industrial Research (CSIR)

The CSIR is a research institute and not a research funder – as is the practice elsewhere. Although we receive public funding, our research is often funded by private institutions and our data could therefore not be made available in open access in all instances. We like to be acknowledged as an institution with academic standing but it is often a real challenge to understand the complexity of RDM within the CSIR context. What sets the CSIR's RDM apart from that of its NeDICC colleagues is that we have only just started our activities formally and we are positioning our RDM to be part of our records management initiative. Research data sets are included in our file plan and are defined as 'data that were accumulated in the process of due diligent research in accordance with a signed research contract'.

Our first attempt to get involved in RDM started in 2010 when our COGIS³ pilot project was launched. We soon realized that the bigger challenge is to first provide researchers with the necessary infrastructure to conduct data science. It is only once this infrastructure functions well that the researchers would be persuaded to follow RDM guidelines. The pilot project did ensure that the infrastructure was put in place to:

- Provide access to research output and associated geospatial data,
- Promote the use of geo-information in research,
- Facilitate access to geospatial data,
- Ensure compliance to legislation pertaining to geospatial data,
- Contribute to an increase in the quality of research output, and to
- Facilitate collaboration between CSIR staff members.

A workflow was developed to surface geospatial data so that the data set could be linked to all subsequent publications. The workflow did not ensure that RDM principles

³ Cooperative Geographical Information System (COGIS): <http://gsdi.geoportal.csir.co.za/>

were implemented across all projects but it has given researchers a means to ensure that their data (whether geospatial or from any other discipline) could follow a mandated process to land in our institutional repository. Although we were not the instigators of the project we gained enormous insight into the challenges associated with RDM in one research discipline. We also came to realise the importance of context-giving documentation, such as research contracts, as well as all the subsequent publications that make use of a specific data set. All of these are regarded as research records and hence the decision was taken to then also include data in the file plan.

We appointed a data librarian in 2014 and as a first step we attempted to identify and understand existing researcher behaviour when it comes to RDM. We wanted to surface any existing good standard operating procedures/practice if any were identifiable. Our approach was a survey-based investigation (personal interviews) between 1st September 2013 and 10th December 2013 with a representative sample group of 36 Research Group Leaders (RGLs) from all nine research units. The majority work with long tail data but we also included large data generators in the sample.

The survey included 23 open-ended questions. Patterton (2014) provided a detailed report on the study she conducted at the CSIR. Based on the results, she was able to get a better idea of the data held by CSIR scientists, research data practices and trends present in research groups, challenges faced as well as the research data service requirements of researchers.

Results were very similar to those established in several surveys completed by international colleagues (Martinez-Uribe, 2008; Pienaar, 2010; Tenopir et al., 2011; Keralis et al., 2012; Beile, 2014 and several others). It was found that research data held in the CSIR was generally regarded as confidential in nature, varied in size, and are mostly stored as spreadsheets, text (pdf and MS Word) and images (jpeg and tiff). There are, however, also data generated by and stored in proprietary systems.

The findings were used to complete a full CARDIO-model evaluation in conjunction with our ICT department. The outcome of that was that ICT too now has a better understanding of the challenges associated with RDM. Going forward, our immediate focus is on getting an RDM policy accepted by the organisation. The policy would require each new project (as of April 2015) to submit a detailed data management plan. Plans generated during 2015 would then be used to motivate for curation infrastructure and also for developing selection criteria that would apply to those data sets earmarked for long term preservation or possible transfer to national infrastructure. Simultaneously we have to:

- Develop and implement an awareness programme as part of a change management process – to create enthusiasm and to obtain support and buy-in for the drive from all our stakeholders;
- Align the RDM activities with the CSIR's enterprise information management activities, as future storage solutions would need to fit into the architecture and make use of standardised workflow solutions; and to
- Share our experiences and continue our involvement with the growing NeDICC community.

It is extremely important for us to remain in contact with the national RDM initiative (DIRISA). The CSIR is one of the larger generators of natural science data and as such is able to contribute to the development of the skills and knowledge necessary for the management of data in these domains.

DIRISA

The national journey has been frustratingly slow as the country needed to first develop the necessary national infrastructure to allow eResearch. Lötter (2014) provided much detail regarding the current state of RDM in South Africa. Various members of our community do represent the country in international bodies and initiatives coordinated through, for example ICSU – CODATA and the Belmont Forum. The country is signatory to the OECD declaration and subsequent principles on publicly funded research (OECD, 2004; 2007) and several individuals are members of the Research Data Alliance (RDA) and the Alliance for Permanent Access to Records of Science in Europe Network (APARSEN).

Although our CoP NeDICC is based in Gauteng and represents seven of the larger research data generators, there are also start-up activities in other institutions dotted around the country. For example, there is a trusted data repository based at the University of Cape Town (UCT) and significant progress is being made with the implementation of RDM practices at the Cape Peninsula University of Technology (CPUT).

The most significant development to date is the framework document for the formation of the NICIS (National Integrated Cyber-Infrastructure System), which is owned by our Department of Science and Technology (NICIS, 2013). The document was developed by an international panel with the collaboration of three national task teams. Direct participation in task team activities and the writing of the task team recommendations was an enlightening experience, as we were again brought in contact with international learning!

The NICIS framework allows for the following initiatives to be collectively managed as a sustainable, national system:

- High Performance Computing (with CHPC-Rosebank as a central node)
- South African National Research Network (SANReN)
- South African GRID (SA-GRID)
- Data Intensive Research Infrastructure for SA (DIRISA)

These initiatives are currently not all at the same level of maturity. Significant effort is being put into DIRISA to ensure that the management of research data is acknowledged as a country asset and to ensure that the essential skills are developed. To that effect the newly established Sol Plaatje University⁴ has just announced the first programme to develop the skills to manage large data generated by the Square Kilometre Array (SKA) project (Mothibi, 2015).

NeDICC, as a group, and our institutions individually are actively participating in the development of the DIRISA business plan and activities in an attempt to make RDM a reality for all – not only for big data initiatives. However, it is quite obvious that the SKA project will be driving many of the most significant big data challenges of the coming decade. Taylor (2014) reported that data rates to researchers are already 103-104 times larger than typical only a few years ago and that these rates will continue to continue to grow exponentially. Scientific leadership on the pathway to the SKA requires not simply access to facilities for processing and storage of data but innovation in computational approaches and algorithms, technologies for visualization and visual

⁴ Sol Plaatje University: <http://www.spu.ac.za/>

analytics of big data, and e-science tools for access and collaborative research around big data by regionally or globally distributed teams of researchers.

Taylor excluded the skills required to preserve and curate the data – to ensure that the data remain accessible in the longer term. This in itself remains one of our biggest challenges. Both researchers and funders acknowledge the need to gain access to data science skills – those skills that will allow for the manipulation and interpretation of data collected. However, they are not yet ready to see the need for a different set of skills (as well as infrastructure) that will allow for long term access to the data. It is not understood that these skills need to be developed simultaneously or the data will be lost for future generations.

DIRISA itself has stated that it would primarily fulfil enabling, supporting and facilitation roles. It would therefore:

- Coordinate, not prescribe, data science capacity development;
- Fund capacity development – limited to DIRISA’s remit;
- Promote and support priority research but with caveats of data stewardship planning and capacity building;
- Guide research strategy and funding;
- Provide services and research environments but would not be not a domain research funder;
- Promote, not enforce, data contribution and adoption of open standards and open data, where feasible; and
- Support data stewardship in a federated context (Vahed, 2014).

The programme manager is completing a business plan with a three year implementation horizon – making ample provision for community and stakeholder engagement. We are convinced that the next three years would allow for much growth and fast tracking related to RDM.

Lessons Learnt

The most important lesson is perhaps to acknowledge that you can learn many wonderful and valuable RDM lessons from the international trend setters but in the end you need to get your hands dirty and get the work done yourself. You must implement the RDM practice that is both appropriate and acceptable for and to your own set of researchers – who may be conducting research in a context very unfamiliar to that of their international peers.

While you are in the process of getting your hands dirty it is sometimes encouraging to realise that developing country initiatives could at times also cause ‘aha’ moments for more experienced colleagues – as was seen with the initiative to explicitly incorporate the entire research process (also the RDM activities) in the planning of a virtual research environment. Looking at it from the other side, it is also about incorporating the entire research process when designing the RDM system – not to plan RDM as a function in isolation.

The downside of a bottom-up approach is that the individual is often able to unintentionally move outside their personal sphere of influence. When this happens it is

easy to create expectations (for oneself as well as within the sphere of influence) that cannot be realised – without knowing that this is the case at the time.

Therefore, a very fundamental and very valid lesson is that commitment from very senior decision makers is essential. Large financial investments are required and therefore that commitment can only come once the decision makers truly understand the e-Research context and with that then the role that RDM plays.

The role that funders play in developing country initiatives is often debated and in our case it remains the cause of much ambivalence. We received exposure to a changed research world before the country was ready and able to acknowledge that, which we now know is the new, often open, way of doing research. Many of our attempts to learn and develop functionality and services for our organisations were ahead of time and we could therefore associate with those of the opinion that when funding is acquired too early in an uncoordinated way (without holistic insight) it is often frustrating for the participants and wasteful for the funder!

On the flip side of the coin, though we have also learnt that one should not give up when you know you are busy with the right ‘things’ – even when it means you need to tread water for a while.

In South Africa it is not clear whether the training of RDM practitioners should be positioned within library schools or rather be transferred as an ICT programme. It does appear though that the cards are stacked towards the latter. There are several reasons for this but in our opinion it is necessary, even essential, that we gain clarity on roles and activities for ourselves. The value of each of these roles, also of the curation role, has to be clearly understood by all stakeholders.

Conclusions

The root causes for the delay in making the South African initiative eResearch initiative a reality have been addressed. Knowledge and skills have expanded, we now do have commitment from very senior decision makers in the country and the infrastructure required making eResearch and RDM a reality has been put in place. It is now up to the South African research community to make use of the opportunity to participate fully in the international research agenda. It does remain an important task for the South African library community to understand what it is that research librarians could and should do to be part of this changed research paradigm (van Deventer, 2014). The right skills still need to be recruited and developed, and that would only be possible if librarians truly understand the niche/specialisation area and when they participate in the management of research data as experts in their own right. Being professionals who could assist in making data accessible and who could be trusted to take care of the data in the longer term.

We believe that the curation infrastructure, which needs to be put in place to proactively manage research data, would in all probability become an extension of the existing library infrastructure but instead of the library being the only decision maker about the extent and the requirements of the infrastructure needed, it would need to be a shared responsibility between the library, ICT and various research disciplines.

We foresee that there is definitely a need for library professionals to guide and assist in developing standard practices, policies, procedures and workflows to ensure longevity in the life of research data being curated. Obviously, one should allow for experimentation – but this should be done wisely and with care (it is so easy to re-invent

the wheel). New practices should only be developed when it is absolutely certain that they do not already exist somewhere else – even in another discipline.

It is extremely important for both librarians and researchers to understand that librarians already have skills (and the inclination) to effectively contribute to the management of data. For this to materialise, librarians would need to advise data generators (researchers) and data scientists at the start of research projects as equals and not as support staff at the end of a long chain of more valuable activities. They also need to learn to collaborate and proactively take responsibility for very specific tasks, rather than to be reactive and wait for project closure before intervening.

Finally, and in conclusion, the questions remain: To what extent are our activities contributing to the library profession and the national debate regarding RDM in South Africa? Did our journey make a difference or does it not really matter? If we had to do this again – would we? The answers to these questions are uncertain but we believe that the journey has made a difference to our lives and we would most certainly advise anyone in a developing country context to get involved, to learn and to participate in their national debate – the journey is certainly worth the effort!

Acknowledgements

We wish to acknowledge the contributions that our NeDICC partners in developing our own understanding of RDM, digital curation and long term preservation.

References

- Beile, P. (2014). *The UCF research data management survey: A report of faculty practices and needs*. Retrieved from http://www.ist.ucf.edu/hpc/rcd/Beile_rcd2014.pdf
- DST. (2007). *Proceedings of the national access to research data workshop: A policy workshop on research data access in South Africa*. Pretoria, South Africa.
- Fernihough, S. (2011). *e-Research: An implementation framework for South African organizations* (Unpublished master's research report, University of South Africa). Retrieved from <http://hdl.handle.net/10500/4474>
- Keralis, S. D. C., Stark, S., Halbert, M., & Moen, W.E. (2012). *DataRes project primary survey* [Data set]. Retrieved from <http://digital.library.unt.edu/ark:/67531/metadc228265/>
- Lötter, L. (2011). *What can take the dark out of the long tail? Efforts to address the challenges of 'small science'*. Presentation given at the Centre for High Performance Computing Conference 2011, Cape Town, South Africa. Retrieved from http://www.chpccconf.co.za/files/2011/2011_08_L%20Lotter.pdf
- Lötter, L. (2013). *Data curation implementation at the Human Sciences Research Council: Case study*. Presentation given at the eResearch Africa Conference 2013, Cape Town, South Africa. Retrieved from http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/HSRC%20Case%20study.pdf

- Lötter, L. (2014). *Reflections on the RDM position in South Africa*. Presentation given at the LIASA WCHELIG/HELIG/DCC Workshop, Cape Town, South Africa.
- Martinez-Uribe, L. (2008). *Findings of the scoping study interviews and the research data management workshop*. Retrieved from <http://ora.ox.ac.uk/objects/uuid/%3A4e2b7e64-d941-4237-a17f-659fe8a12eb5/datastreams/ATTACHMENT02>
- NICIS. (2013). *A framework for the establishment and maintenance of a sustainable NICIS: Report of the International Committee for the Development of South Africa's National Integrated Cyber-Infrastructure System appointed by the Department of Science and Technology*. Retrieved from http://www.dst.gov.za/images/NICIS_Framework_Report_December_2013_26_May_2014.pdf
- OECD. (2004). *OECD declaration on access to research data from public funding*. Retrieved from <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157>
- OECD. (2007). *OECD principles and guidelines for access to research data from public funding*. Retrieved from <http://www.oecd.org/sti/sci-tech/38500813.pdf>
- Page-Shipp, R., Hammes, M.M.P., Pienaar, H., Reagon, F., Thomas, G., van Deventer, M.J., & Veldsman, S. (2005). eResearch: The challenge for South Africa. *SA Journal of Information Management*, 7(4). doi:10.4102/sajim.v7i4.287
- Patterton, L. (2014). *Research data management at the CSIR: An exploratory survey*. Unpublished internal report, Council for Scientific and Industrial Research, South Africa.
- Peters, D. (2013). *A roadmap for South African research data infrastructures*. Presentation given at eResearch Africa Conference 2013, Cape Town, South Africa. Retrieved from http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/DIRISA_Roadmap.pdf
- Pienaar, H. (2010). *Survey of research data management practices at the University of Pretoria, South Africa: October 2009 – March 2010*. Presentation given at the 22nd International CODATA Conference, Cape Town, South Africa. Retrieved from <http://www.slideshare.net/heila1/survey-of-research-data-management-practices-up2010>
- Pienaar, H. (2014). *Developing an institutional research data management plan: Guidelines*. Paper presented at the 2014 LIS Research Symposium, Unisa Science Campus, South Africa, and at the eResearch Africa Conference 2014, Cape Town, South Africa. Retrieved from http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/Developing%20%28a%20framework%20for%29%20an%20institutionalUnisa2014.pdf
- Pienaar, H., & Van der Walt, I. (2014). *Shepherding data on the way to long term stewardship: Research data management at the University of Pretoria: A case study*. Paper presented at the 12th Southern African Online Information Meeting, Pretoria, South Africa, and at the Unisa Library Technology showcase. Retrieved from <https://saoug.files.wordpress.com/2014/06/rdmsaoug-5-jun-2014.pdf>

- Mothibi, M. (2015, January 12). Sol Plaatje university first to offer data science in Africa. *SABC News*. Retrieved from <http://www.sabc.co.za/news/a/58112f8046e8fdb78d9eed9cbd593836/Sol-Plaatje-university-first-to-offer-data-science-in-Africa-20151201>
- Taylor, R. (2014). *The SKA and the rise of big data in radio astronomy*. Presentation given at the eResearch Africa Conference 2014, Cape Town, South Africa. Abstract retrieved from <http://www.eresearch.ac.za/presentations-2014-day-three>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, 6(6), e21101. doi:10.1371/journal.pone.0021101
- Vahed, A. (2014). *Data intensive research initiative for South Africa (DIRISA): A reinterpreted vision*. Presentation given at the eResearch Africa Conference 2014, Cape Town, South Africa. Retrieved from http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/DIRISA_busPlan_eResConf.pdf
- Van der Merwe, A., Patterton, L., & Van Deventer, M. (2014). *From process to practice: Establishing a research data management function in a resource-constrained environment*. Presentation given at the eResearch Africa Conference 2014, Cape Town, South Africa. Retrieved from http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/ErsearchAfricaFinal.pdf
- Van Deventer, M., & Pienaar, H. (2009). Report on the 2nd African Digital Scholarship and Curation Conference. *D-Lib Magazine*, 15(7/8). doi:10.1045/july2009-vandeventer
- Van Deventer, M., & Pienaar, H. (2012). *Employing the Moodle platform as a VRE tool while teaching eResearch skills to masters and PhD students*. Paper presented at the 4th International Conference on Education Technology and Computer (ICETC 2012), Cape Town, South Africa. Abstract retrieved from <http://hdl.handle.net/10204/6848>
- Van Deventer, M. (2013a). *Data policy*. Presentation given at the workshop 'Governance questions in respect of multi-national data centres', 5th African Conference for Digital Scholarship and Curation, Durban, South Africa.
- Van Deventer, M. (2013b). *Data citation (Out of cite, out of mind: Current state of play)*. Presentation given at the eResearch Africa Conference 2013, Cape Town, South Africa. Retrieved from http://www.eresearch.ac.za/sites/default/files/image_tool/images/140/Data%20Citation%20Oct%202013%20MvD%20Final.pdf
- Van Deventer, M. (2014). *Emerging trends in library support for research data management services*. Paper presented at the 2014 LIS Research Symposium on Research Data Management and Teaching Future Librarians, Florida, South Africa.