April 2019

Research Data Management (RDM) Capabilities at the University of Ghana, Legon

Bright Kwaku Avuglah
University of Ghana, Legon, bkavuglah@gmail.com

Peter G. Underwood
Dissertation Supervisor, University of Pretoria, pgunderwood@wol.co.za

Research Data Management (RDM) Capabilities at the University of Ghana, Legon

Bright K. Avuglah
University of Ghana, Legon
ORCID: http://orcid.org/0000-0002-6244-2811

Peter G. Underwood
Dissertation Supervisor, University of Pretoria
ORCID: http://orcid.org/0000-0002-9618-5032

Abstract

The purpose of this study was to assess Research Data Management (RDM) capabilities at the University of Ghana (UG). The study focused on four key capability elements: policy framework, technological infrastructure, skills and knowledge, and support services. It explored the extent to which RDM is embedded in research practices at UG and provides insight into the preparedness of UG to develop RDM. A qualitative case study method was adopted for the study and data was gathered using semi-structured interviews and document analysis. The instrument for the assessment was informed by the Collaborative Assessment for Research Data Infrastructure and Objectives (CARDIO) Matrix tool and respondents were drawn from the Library, IT department, Research Office and senior researchers. The results of the study show that RDM at UG is currently underdeveloped but with immense potential for growth. Though there is no formal RDM infrastructure in place, RDM is considered an essential research integrity issue. Capabilities were generally found to be limited, uncoordinated and not officially instituted. The study recommends that a clear and comprehensive policy framework for RDM should be developed to articulate RDM aspirations and express management's commitment. It also recommends that research support staff should be supported to build their capacity for RDM promotion and support.

Keywords

Research data management, RDM capability, University of Ghana

Introduction

The growing influence of the digital revolution and its concomitant advancements in Information and Communication Technologies (ICT) are reshaping how research is practiced; so also, is the ethos of science changing. Research is increasingly becoming more computational, data intensive and collaborative over virtual and networked platforms (Wang, 2013) leading to the so-called fourth paradigm (Hey, Tansley & Tolle, 2009). These present new opportunities and challenges for Higher Education Institutions (HEIs), including the effective and sustainable management of the research data generated during research (Procter, Halfpenny & Voss, 2012). HEIs – mainly in the developed countries – have begun to develop capabilities to support this emerging research culture.

There is a growing body of literature on RDM, and the surge can be ascribed to the increasing awareness and recognition of the data deluge phenomenon and its implications, the prospects for data reuse and the need to maximise the return on investment for research (Wong, 2009: 125). HEIs and research institutions are also beginning to approach RDM strategically, considering the research data emanating from internally funded research as assets rather than by-products of research (Cox & Pinfield, 2014: 300; Lynch & Carleton, 2009: 236). But these discussions have mostly concentrated on developed countries and experiences from such countries have dominated the body of literature on RDM development at HEIs in particular.

Several case studies (Chiware & Becker, 2018; Chigwada, Chiparausha & Kasiroori, 2017; Jones *et al.*, 2015; Ball, 2013; Rice & Haywood, 2011; Takeda *et al.*, 2010) on institutional RDM implementation involves some institutional assessment. Whyte *et al.* (2014: 285) indicated that such assessments are important to establish what capabilities exist, their adequacy and how well they are being deployed to support RDM. These case studies also provided insight on the approaches and strategies adopted by institutions to implement RDM. The context of these studies are vastly different from Ghana and many developing countries in Africa. For instance, the legal and policy landscape of most developed countries provide the impetus for RDM uptake. Data mandates from government and funding organisations are a major driver for RDM initiatives in many HEIs (Henderson & Knott, 2015: 48). This is not the case in Ghana and most of Africa. There are no official government mandates, neither are the private and international organisations funding research on the continent strict on data management as they do elsewhere. Also, such an assessment of RDM capabilities in Ghana or even Africa is almost non-existent. This research, therefore, investigated an HEI in a developing country (Ghana) – the University of Ghana (UG). According to van Deventer and Pienaar (2015: 43) it is important to contribute to the RDM literature from developing countries like Ghana because this can provide some insightful perspectives even for experienced colleagues in the global north.

For UG, this study fits well into its vision and aspirations of becoming a world-class research-intensive university (UG, 2014). The issues, challenges and expectations of a world-class research-oriented HEI will include a strategic and systematic approach to supporting the management of the research data from research enterprises. Doing this opens such institutions up to opportunities for funding and partnerships (Hiom *et al.*, 2015: 491). But developing RDM must first start with an understanding of the current situation (Davidson *et al.*, 2014: 217; Jones, Pryor & Whyte, 2013: 6). The study was, therefore, conducted to assess existing RDM capabilities at the University of Ghana which can be harnessed for future RDM development and support.

Objectives

This paper reports on the results of a research study. On the basis of the problem outlined, the objectives of the study were to:

a. Identify what elements are necessary for assessing institutional RDM capabilities.
b. Assess what RDM capabilities currently exist at the University of Ghana.

Literature review

The literature shows that the responsibility for RDM development is most emphasised at the institutional level (Awre *et al.*, 2015; Jones, Pryor & Whyte, 2012), even though one can point to some national level infrastructure investments (e.g. Data Intensive Research Initiative of South Africa [DIRISA], Australian National Data Service [ANDS] and UK Data Archive) and capacity building support (e.g. the Digital Curation Centre [DCC]). Fortunately, HEIs who wish to develop RDM today have a second-mover advantage following several documented experiences from RDM pioneers (Henderson & Knott, 2015: 49). Earlier developments were exploratory, allowing for a variety of approaches (Hodson & Molloy, 2014: 208). Also, institutional context influenced the approach adopted (van Deventer & Pienaar, 2015: 43; Cox & Pinfield, 2014: 300). However, these previous and continuing experiences are enabling the development and refinement of best practices and transferable toolkits by which late adopters can be guided (Davidson *et al.*, 2014).

A critical part of most RDM implementation initiatives is assessing institutional capacity or preparedness to implement a feasible RDM infrastructure. According to Jones, Pryor and Whyte (2012: 142), institutional RDM capabilities denote the ability of an institution to articulate and attain RDM objectives. One of the implications of the fourth paradigm of science has been the growing necessity for HEIs to develop capabilities to handle complex data-intensive science (Lyon *et al*., 2012: 9). Several models have been postulated in the literature for assessing data management capabilities and their maturity. Some are at the institutional level like the Australian National Data Service (ANDS) adaptation of the Capability Maturity Model for assessing institutions research data capability maturity level (ANDS, 2017), and those for assessing data capabilities at the project level (Sallans & Lake, 2014; Lyon *et al.*, 2012; Crowston & Qin, 2011). Using the maturity model to assess current levels of institutional data management capability enables institutions to identify pressure points that need to be enhanced. The ANDS model assesses five key capability elements (Policies and procedures; IT infrastructure; support services; managing metadata; managing research data) along five levels of maturity.

Another model, the Cornell Three-Legged Stool model, was originally developed to evaluate HEI's response to digital preservation along three dimensions (organisation, technology and resources). These have been adopted and adapted as elements essential for a workable and sustainable RDM effort, first, through the AIDA self-assessment tool and later the Collaborative Assessment for Research Data Infrastructure and Objectives (CARDIO) tool by the UK Digital Curation Centre (DCC) (Jones, 2014: slide 8; Whyte & Allard, 2014: 13; Pryor, 2013: 187). What is peculiar about the CARDIO tools is that, the model allows for local level adaptation: it specifically emphasises research data management and the three dimensions can be assessed to different degrees of granularity according to the level of engagement desired - project, departmental or institutional level (Whyte & Allard, 2014: 13). A variant of the tool (CARDIO RDM Matrix) has also been used to assess institutional readiness to comply with

Engineering and Physical Sciences Research Council (EPSRC) Policy Framework for RDM. The tool addresses three thematic issues essentially encapsulating nine-point expectations:

a. RDM policy, strategy, governance and sustainability
b. RDM support services and skills development
c. Technical infrastructure to facilitate storage, preservation and sharing of research data (Jones *et al.*, 2015).

These capability elements also represent the aspects that institutions need to consider when planning for an institutional RDM programme and must be developed in the light of adequate resources provision (financial and staffing), well defined roles and responsibilities and commitment from senior management (ANDS, 2017: 2; Whyte *et al.*, 2014: 285).

A few studies report on institutional RDM capability assessments in the literature. Takeda *et al.*, (2010) used the AIDA self-assessment tool to benchmark the level of data management capability at the University of Southampton. The findings of the assessment revealed among other things, limited RDM guidance and incoherent policy framework, a lack of formal training around data management, and limited support and guidance for researchers, varied capabilities across campus with pockets of best practices, and limited awareness about existing capabilities and resources.

Jones *et al.*, (2015) reported on how four institutions in the UK (University of East London, University of Edinburgh, University of Leeds, and University of St. Andrews) complied with the EPSRC mandate on RDM. Using the CARDIO Matrix framework, they report that three of the four universities adopted a policy-first approach, while University of St. Andrews started their RDM implementation with a strategy document (roadmap) instead and developed a policy later on. Overall, technical infrastructure was focused on storage solutions in the form of data repositories, but University of Edinburgh also had a high-performance computing (HPC) infrastructure in place. RDM was generally promoted through service offerings and relationship building. Support included guidance on writing data management plans (DMPs) and training, which were generally done by embedding RDM trainings into graduate programmes. On-demand trainings were also offered to faculty members and students. Guidance was also provided through library websites and links to relevant resources on the web such as the University of Edinburgh's online management training (MANTRA) resource.

Chigwada, Chiparausha and Kasiroori (2017) explored RDM practices across Zimbabwean HEIs and found that there was a general lack of policies and guidelines on RDM, limited financial and human resources, lack of robust and secure technological infrastructure and a general lack of support from the management of the institutions on the issue of RDM. They recommended that trusted data repositories be established to encourage best data practices among researchers.

Chiware and Becker (2018) conducted a study to determine the readiness of HEIs in Southern Africa to lead and participate in institutional RDM development. They found that most institutions were not fully ready to comprehensively support RDM in their institutions due to a number of resource, infrastructure and human capacity constraints. There is disparity among countries in terms of developing policies and guidelines for RDM, with most of the institutions having no policies at all. Institutional repositories were also not fully harnessed to manage datasets and their metadata. There was a skills gap, but some institutions are working at bridging the gap. They recommend training for librarians and organizational restructuring to align existing library research services to RDM.

These studies reveal that librarians play a critical role in developing RDM and it is absolutely important that investments are made into their capacity development to be effective. It is also evident that the level of development is disparate and environmental factors and organisational culture tend to shape institutional response and challenges. What is more, these studies help to understand which capability elements tend to be emphasised in such assessments. In this case, they are the policy framework, technological infrastructure, skills development and support services.

Methodology

In this study, the qualitative approach and case study strategy were adopted. According to Yin (2009), case studies are most appropriate for exploring contemporary issues within specific and bounded contexts. Creswell (2009: 177) also asserted that the case study strategy is appropriate for exploring processes and activities. Accordingly, this research is about exploring RDM capabilities at UG; it is a new area of engagement for academic institutions and is still evolving in terms of its practices and responsibilities in this part of the world. The study investigates a single case – the University of Ghana.

In line with the rationale for purposive sampling and to avoid unnecessary duplication of responses, seven respondents, comprising five respondents from research support units (University of Ghana Library System (UGLS) [two], University of Ghana Computing System (UGCS) [two] and Office of Research, Innovation and Development (ORID) [one]) and two senior researchers, were selected to participate in the study. This sample size is in consonance with the recommendation by Creswell (2013: 157) who asserted that a sample size in the region of five is appropriate for a single case study research.

Selection of respondents was done through "priori criteria sampling" (Pickard, 2008: 64). By this method, a set of criteria were set as baseline for including information-rich respondents. For the respondents from the research support units, the following selection criteria were applied: they must be senior members (this is a management/administrative level rank), and should have worked in that capacity for not less than three years. The researchers believe that this provides ample time for the respondents to have acquired rich information about the capabilities, programmes and policies of the university in their respective units. For the researcher-participants, they must be a senior researcher in UG, a previous recipient of research funding (internal and external), and must have extensive research experience with at least ten published scholarly works. Discussions were held with them on their expectations and perceptions about current capabilities, infrastructure and support for RDM at UG.

Data was collected using semi-structured interviews and document analysis. Pickard (2008:171) asserted that interviews are the most used data collection method in Library and Information Science (LIS) research and is the most appropriate technique for qualitative and in-depth studies such as case study research. The following institutional documents were also analysed and used to corroborate the primary data: "UG Strategic Plan 2014-2024", "UG Research Policy", "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management", "UG Institutional Repository Policy", "UG Research Ethics Policy", "UG Intellectual Property Policy" and "UG Library System Draft Strategic Plan 2014-2019". Combining data collection method and sources is also consistent with the practices for qualitative case study research in extant literature on qualitative inquiry (Creswell, 2013; Pickard, 2008). The researchers used the United Kingdom's Engineering and Physical Sciences Research Council's (EPSRC) CARDIO matrix capability elements as the criteria to assess RDM capabilities at UG. The CARDIO framework informed the questions for the assessment.

The focus was on four aspects: policy, technology, skills and knowledge of support staff and existing data services and support. The data was analysed by first transcribing all the seven interviews, individually analysing each transcript and institutional document to identify key terms and topics, grouping these topics into categories using colour codes and annotations, comparing the categories across the different transcripts and documents for patterns, and regrouping and condensing them to form themes which were then presented and discussed. The presentation and discussion were done by mixing data from both sources as well as corroborating data from the interviews with data from the institutional documents.

The study was approved by ethics committees from the University of Pretoria and the University of Ghana.

Results

Institutional Policies framework on RDM

The data from the interviews conducted and document analysis shows that UG has no explicit RDM policy. It is however captured as one of the key policy statements in the "UG Research Policy" (section 5.6). Four RDM issues are addressed in the Research Policy:

a. Recognition of RDM as a good research practice and integrity issue;

   "The credibility of research findings depend[s] on record keeping and good data management";

   "Data management is one of the essential areas of responsible conduct of research".

b. Institutional commitment to developing systems to support RDM;

   "the University will create a meta-database of research materials/data repositories"

c. Researchers' responsibility as the main steward of research data; and

   "Under normal circumstances the original materials and data sets will be held by the PI who undertook the research."

d. Mandate of researchers to keep datasets for not less than ten years after the completion of a research project.

   "The PI is expected to maintain this data set for a minimum of ten years after the final project close-out. In certain special circumstances, this minimum period may be extended."

Despite the lack of an explicit policy, there was a guideline for RDM - "UG Research Policy Guideline on Good Practices: Record Keeping and Data Management". This was the most pronounced institutional document on RDM that spells out in greater details a number of best practices for guidance in the management of research data for the research community. The guideline was developed by the Office of Research, Innovation and Development (ORID). It captures many important aspects of RDM such as data ownership, data collection and documentation, data storage and retention, data protection, data privacy, and data sharing and publication. For instance, the guideline outlines the minimum documentation or metadata that must be provided for every dataset.

"Record anything that seems relevant to the project, its data, and the standards of the project. At a minimum, records should include the following information: Date and time; Names and roles of any team members who worked with the data; Materials, instruments, and software used; Identification number(s) to indicate the subject and/or session; Data from the experiment and any pertinent observations from the data's collection. It may also be helpful to include a summary of the day's data collection activities and a task list for the next day."

"Data should be retained for a reasonable period of time to allow other researchers to check results or to use the data for other purposes. There is, however, no common definition of a reasonable period of time."

"Data should however not be shared without the permission of the University."

Probing further on the level of uptake and compliance with the guideline, respondents (designated as [R1] to [R7]) had the following to say:

...but a guideline is just to guide you, a policy reinforces or is binding, a guideline is not binding [R5].

I have no idea about institutional policy on RDM, but I know a research policy exist and then there is ethics policy. I am not particularly aware of any RDM guideline... [R7].

Some other policy documents of the institution also capture data related issues. For instance, the Institutional Repository (IR) policy stipulates that "datasets" are one of the acceptable content formats to be deposited. It also outlines clearly the metadata schema for describing items (including datasets), institutional services and support for deposited items as well as the standards of operation. The intellectual property policy also defines data as a tangible research property and addresses the terms of data ownership:

"Research data shall be jointly owned by the University and researcher(s) or determined on a case by case basis. Either party shall have a right to use the data for its research purposes...In spite of these provisions, entitlement to the ownership of primary data, software, and other products of research may vary, depending on the circumstances under which the research is conducted..."

The ethics policy as well addresses the risk and integrity issues about research and data collection. It highlights the issues of confidentiality and privacy as it pertains research subjects. In the policy, the ethics committee is tasked to:

Put in place procedures to ensure participants privacy, maintenance of confidential data, and adequate protection of vulnerable participants.

RDM knowledge and skills among service providers

Respondents were asked to comment on whether they consider research support staff possess adequate knowledge and skills to support RDM and these were some of their comments:

I have a feeling that yes, most people should have a fair idea, but again this is not a system that has been introduced fully in the University. Once that is fully introduced, staff will be trained on how to do things, but I have a feeling that people have a fair idea about how [research] data should be managed [R5].

I don't think so; I think we need to retool. Because in the first place, when we look at the library school from which many of us are trained we don't even talk about those things at all, so we really don't have what it takes, [but] I think we can learn. We don't have the right skills [and tools] now, but we are capable of doing it so long as we are retooled, I think there is so much out there that we can read, and learn, and maybe visit people who are doing it and be exposed to what is being done, we can come and replicate it here [R3].

I don't think currently, [we have the necessary skills to support data curation], but data curation is not nuclear science so I believe the staff concerned, once we are told to move in that direction within a short period of time people, we will be abreast with the skills to do that, but then also I believe there should be some formal training in that respect…Curation will have to come in, so I believe even if the person has some informal training I believe at the end of the day there should be some formal training before the person is mandated to do that task [R4].

I think librarians are already doing it. This Institutional Repository thing is something along that line [R2].

IT infrastructure and support services

The researchers enquired about existing IT infrastructure that can support researchers' data storage, preservation and backup, data processing and analysis, data sharing and data security needs. The results revealed a number of IT systems and applications that can be harnessed and extended to support data management. It is important to establish that these systems are not necessarily provisioned for RDM, but they represent potential for data storage and preservation, analysis, access management, publishing and sharing.

On infrastructure to support storage and backing up of data (active storage and long-term preservation), respondents stated the following:

As for storage we have about 120 Terabyte on the cloud infrastructure…we have what we call the HP Cloud matrix...and it comprises of Storage Area Network (SAN), Network Installation Management (NIM), optical devices and multiple storage to really provide the needs of other departments [R1].

Per the IR policy it is written that it permits the deposit of datasets but in actual fact we are not at the moment accepting datasets. The system itself can accept datasets but for now we are not accepting it, it is more of a future thing that we are thinking of doing, but for now we are more or less taking the end products of research, that's the PDFs and other formats. We are not accepting the raw data formats [R4].

We take backups of the kind of data that we need to take backup, so when there is an emergency we can restore them….but we have been smart to move students to the Google cloud which is virtually limitless storage, Google has given us gigabytes and we can always ask for more, in fact, we have that clause that we can negotiate for more [R1].

They also mentioned the capacity of the IR to harvest metadata about researchers' publications from external systems in addition to the manual approach.

We also struck this agreement with BioMed where all publication authored by faculty and researchers from University of Ghana are automatically pushed onto our platform together with the full text. It is something that we are planning to expand, for instance with Elsevier…DATAD is able to harvest these things onto their platform so we have that interoperability with other systems where we can harvest things from other people and other can also harvest from us [R4].

Respondents were also asked to comment on institutional support for collaborative research, data analysis and computational science and these were some of their comments:

There was recently some Italian collaboration they did with the department of Physics and Computer Science and we had to go in to provide what we call High Performance Computing infrastructure (HPC). We didn't have that infrastructure; we had to build some basic infrastructure to support the kind of lab they wanted to run that project…. Currently too we are also collaborating with Prof. Awandare and his team at Biochemistry department. They have had the need to build specialised […] HPC to run genome, I mean to run the kind of multiple algorithms to support the kind of chemical related or biomedical related research they are doing. And […] we have collaborated with them and IBM to provide such infrastructure. Even though we have an HPC, it's not so much unique for the rest of the community, so we are trying to build another one that could be leveraged by all other departments… like I said, it had to be done within [the] shortest possible time, maybe two or three months because the VC then and incoming VC were all interested and we were able to work with procurement to get some few servers, put together heavy processor-base and memory-base to be able to enable them [to] run that thing [R1].

Apart from the resources that we have, we also have this package for data analysis; the NVivo, for example, is for qualitative data analysis [R2].

We have a training unit in UGCS,…that does training for faculty for specific software to use for their research activity, SPSS and the like and then when faculty [members] have issues with how to handle such data […] they always come here and there are people who have been supporting them over the years [R1].

When asked if respondents had received any training on RDM, they replied with such phrases as "Never before" [R6] and "Not here, not in the USA" [R7] to show that they have never received any RDM related training. One respondent from the library however mentioned that in one of the Authoraid workshops faculty members received some training relating to data collection and publishing:

…they are taking through how to collect data and put all these things together and even find appropriate journals to publish…everything about authorship [R2].

We've also done a bit of training in using research software like reference managers, and then we have also collaborated with publishers and done author workshops for researchers just to enhance the research process. And this has brought out issues of where they could be publishing, how they can select journals for publishing and what is required of an author in publishing [R3].

The result also shows that another support offered to researchers at UG is guidance on funding application and best practices on data management. The RDM guidelines, among other things, advice researchers on the need to understand and comply with funder regulations. One respondent describing the research support role of the ORID emphasized that the Research Office provides information to researchers about funding opportunities and requirements:

> *…In terms of providing research grants to researchers to conduct research, ORID plays a vital role…We provide research advisory role, we send information to faculty members regarding funding opportunities through a platform we call 'research alert' just so faculty will know the information as far as research grant is concerned* [R5].

Data management planning was not captured in any of the institutional documents on RDM and the respondents corroborated this by stating that writing a data management plan (DMP) was not part of the research process or was not part of the requirements for applying for funding through the Research Office (ORID). These were some of their comment in relation to DMPs:

> *So far, I haven't seen that [requirement for a data management plan] in the internal funds that we give researchers... international donor I believe will require for data management plans but, as far as our internal grants are concerned, I have not seen that on our forms yet, but like I said that could be the next level* [R5].

> *Not really, it's more interested in how you are going to execute, analyse the data, the outcomes and how you are going to disseminate the [outcome]. Normally that's what I have seen, but there may be a question "how long you will keep the data?"* [R7].

> *I have not had the need to write an RDM plan* [R6].

The respondents from the UGCS who provided insight into the technological capabilities appeared confident about the robustness and resilience of the network and security infrastructure:

> *We have reasonable bandwidth…We have a first line of security, the network is interfaced with a firewall and also at the end of the users we have also put other security devices like Universal Threat Management, which is Sophos UTM, we have also put another device that also actually checks within the network to see whether there is any harmful thing coming in – intrusion detection – which has been deployed by the IT Security Unit of the UGCS and that warns us to watch out as the network people provide the necessary remedial solutions…* [R1].

Discussions and Conclusions

According to Whyte *et al.* (2014: 285) the capabilities required by institutions to successfully develop RDM range from human and technical capabilities to adequate financial resources needed to deploy them. They add that how well these capabilities are deployed and their adequacy to enable institutions to implement their own policies or comply with external funders policies is a question information professionals may be required to answer by conducting some form of institutional assessment. In this study, the researchers conducted an institutional scoping to identify existing capabilities for RDM by benchmarking with established criteria and best practice. The literature review shows that the elements of institutional data

management capabilities range from issues relating to policy framework, strategy, governance, technical infrastructure (including ICTs), human resource capabilities (expertise), support services, metadata management and research data. These criteria are derivatives of a number of capability models on RDM in the literature (ANDS, 2017; Jones, 2014; Lyon *et al.*, 2012; Crowston & Qin, 2011) and represent the aspects that institutions need to consider when implementing RDM. More importantly, developing such capabilities towards a matured RDM service opens up academic institutions for funding and partnerships (Hiom *et al.*, 2015: 491), something that resonates strongly with the aspirations of UG and most research-focused institutions. In this study, four of the elements received attention, namely: policy, technical infrastructure, skills and knowledge, and support services.

The data shows that there is currently no explicit policy on RDM at UG. However, it is captured very briefly in the institution's research policy and more elaborately in the policy guideline for data management. Aspects of RDM are also loosely captured in the "UG Institutional Repository Policy", "UG Research Ethics Policy" and "UG Intellectual Property Policy". While this is not unusual, – in fact, guidelines are normally part of the institutional policy frameworks (Ball, 2013; Jones *et al.*, 2012), and may sometimes be the only institutional document available to articulate RDM aspirations – the challenge is that guidelines, in most cases, are dynamic, non-mandatory controls which typically evolve as services develop (Searle *et al.*, 2015: 444). This is confirmed by the data which show that the guidelines were not necessarily binding on researchers, thus, the guidelines do not represent a "public statement of intent and an expression of the commitment of management" as a policy is expected to denote (Pryor, 2014: 16-17). In the opinion of Pryor (2014) and Rans and Jones (2013), a clear policy is essential if an institution wants to develop or adopt RDM because the policy outlines the key principles, expectations and roles for every specific data management activities, and more importantly, it provides the mandate which drives a cultural change.

These results are consistent with the situation in some institutions in the developed world. Searle *et al.* (2015) in their study at Griffith University, Australia, found that rather than an explicit policy, RDM was covered briefly together with other important research integrity issues in the "Griffith University Code for the Responsible Conduct of Research" and the "Best practice guidelines for researchers: managing research data and primary materials" at Griffith University provided a highly-detailed and practical response to evolving cultural change for RDM. Also, with respect to the coverage of RDM issues, the study by Higman and Pinfield (2015) on RDM policy development among UK HEIs found that many institutional policies lacked detail and specificity – failing to address some critical aspects such as funding. The authors also asserted that while such lack of details is to be expected especially at the early stages in RDM policy development, it is likely to limit how much influence these policies can have in shaping practices within new and emerging RDM systems.

Despite efforts by the ORID to publicise the guideline on RDM, like other policy documents, mainly through its website and circulated booklets, the researchers interviewed were oblivious to its existence. They admitted it might exist, but this was unknown to them. While this cannot be generalised because of the number of researchers who were invited to participate, this revelation is still instructive especially coming from senior researchers with an average of twenty years of research experience in the institution. It signals the need for the University to take a look at how it communicates and creates awareness about institutional policies and guidelines and how effective those approaches are to reaching the local research community. Compliance monitoring could be an effective way of measuring this. Khokhar *et al.* (2017) posit that monitoring RDM policy compliance can be challenging; nonetheless, non-compliance (particularly with funder policies) could attract sanctions. The situation with institutional

policies is less stringent as they have been used largely as lobbying and advocacy tools (Higman & Pinfield, 2015: 374). Yet, policy adherence can help avert catastrophic research risks. The Australian National Data Service's *Research Data Management in Practice* manual emphasises the need to demonstrate compliance through effective audit frameworks that can help institutions to respond to compliance breaches in a more systematic, rather that ad hoc, manner (ANDS, 2013: 5). Yet, the data does not reveal a clear institutional framework for monitoring compliance with existing policies and guideline provisions on RDM. As far as internal systems are concerned, satisfying the requirements for internal or external funding application seems to be how researchers' appreciation of, or adherence to, the policies, guidelines, and regulations are gauged. While the review process of internal funding applications may pass for an audit framework, there is no certainty that compliance with local and funder data mandates are key concerns in this process especially as the data show that writing a DMP is not yet a requirement in research funding application at UG.

There was consensus among respondents that there was a skills and knowledge gap in the institution to adequately support the RDM needs of researchers. This finding is consistent with what persists in most developing countries (Chiware & Becker, 2018; Ssebulime, van Deventer & Pienaar, 2018; Chigwada, Chiparausha & Kasiroori, 2017; Kahn *et al.*, 2014) and reflects the fact that the need for RDM is new to them. Support staff who will be taking up the new role of supporting researchers' data management needs must be trained. The roles and perception of research support units (e.g. Library, IT Services and Research Office) with regards to research and RDM are varied. IT Service providers tend to think about their role in terms of providing technological (software and hardware) and technical support, Research Officers think of it in terms of research administration and funding support, and librarians perceive their role in terms of collection management and curatorial support (managing data repositories), training, guidance and consultancy on writing DMPs (Cox & Verbaan, 2016). From the data, the UGCS has demonstrated the capacity to develop specialised systems to facilitate collaborative and computational research, the UGLS (Library) currently manages the Institutional Repository (IR) and provides data analysis software for small scale projects, and ORID has developed polices for research conduct and a guideline for RDM. While these may not be enough to operate effectively, they are still relevant and need to be considered in meeting the training needs of each of these units. For libraries, how to develop such professional capacities of librarians (Conrad *et al.*, 2017), and the competencies required to take up this new and additional responsibility (Cox & Pinfield, 2014; Brewerton, 2012; Cox, Verbaan & Sen, 2012) have been well researched and documented in the literature.

According to Beitz *et al.* (2014: 174) technical infrastructure for RDM can enable researchers to generate new knowledge and achieve a greater reach for their work, preserve and protect their research data, and facilitate collaborative research, data reuse and outcome validation. Ideally, the technical infrastructure must support every aspect of the data lifecycle and facilitate collaboration, enable integration of datasets as well as sharing data across dispersed teams (Whyte, 2012: 209-210; Wilson *et al.*, 2011: 275-276). The data shows that UG has invested and continues to improve institution-wide technology capability to support different aspects of institutional process, activities and services. Of particular relevance to the purpose of this study, a number of systems and tools were identified that can be harnessed and extended to support data management. These systems and tools range from an HPC facility, a private cloud facility (HP Cloud Matrix), an institutional repository developed on the DSpace platform, a robust network and security infrastructure and specific data analysis applications. It is important to establish that these systems are not necessarily provisioned for RDM, but represent potential for storage and preservation, analysis, access management, publishing and sharing.

The only notable data-related support revealed by the data are: guidance for best practices on RDM which the ORID has provided in the form of RDM guideline, which is available on its website, and support for data analysis (provision of and training on the use of data analysis software – SPSS and NVivo) offered by the UGCS and the Library. There is also limited support for collaborative and computational research, through the provision of higher performance computing (HPC) and high storage capacity infrastructure. The issue is that these services are disjointed and not formally instituted as RDM services. This finding is consistent with the literature. The studies by Chiware and Becker (2018), Chigwada, Chiparausha and Kasiroori (2017) and Takeda *et al.*, (2010) all reported limited RDM support. This is in part because RDM is new and underdeveloped and there is a lack of appropriate skill and adequate knowledge about RDM.

Taking all together, this study reveals that the situation at UG is similar to what persists in some institutions across Africa and even the developed world. First of all, RDM is a new concept in Africa. Institutions sometimes articulate their RDM aspirations through non-mandatory guidelines rather than a binding policy. RDM is undeveloped, but there is potential for growth. There is a considerable RDM skills and knowledge gap and technological capabilities were generally found to be small-scale, uncoordinated and not necessarily provisioned for RDM. Though there is no formal RDM infrastructure or programmes in place at UG, it is still considered an essential research integrity concern by the University's management. This also suggests that RDM issues in themselves are a natural part of research activities and even when they are not formally instituted as a service, HEIs and research institutions are conscious about proper handling of data, which is the basic ingredient for scientific inquiry, knowledge production and validating research findings (Jones, Pryor & Whyte, 2013: 1-2; Lynch & Carleton, 2009: 236). To answer the question "How to systematically and effectively do this?", is the opportunity for research stakeholders like the library to provide leadership.

The outcomes of this study have practical implications for future RDM development at UG. It provides some pointers to the University management and stakeholders with regards to which RDM capabilities currently exist, which ones need to be developed and a snapshot of the general institutional preparedness. To the broader academic and research community, this study makes a significant contribution to the body of knowledge on institutional RDM development from the perspective of a developing country, specifically, Ghana, by showing how institutions that are without any formal RDM programmes could respond to RDM as a critical part of an emerging research process.

While this study reveals that the state of RDM development at UG is similar to most institutions across Africa, further research would be required to validate the findings in this study. More researchers and other critical RDM stakeholders (e.g. University management, legal department, archives, Government etc.) would have to be engaged to provide a stronger and more accurate picture. The recognition of RDM as a research integrity issue and the response by the institution is worth emulating by other HEIs in Ghana and Africa who are yet to take any step.

Recommendations

On the basis of the findings of the study, the following recommendations are proposed for developing RDM at UG and other HEIs on the African continent:

1. The development of a clear and comprehensive institutional policy framework for RDM. This must be practicable, harnessing all the current capabilities identified and covering all

essential aspects including an explicit retention policy, data management planning and a framework for monitoring compliance. This is a key requirement for developing a coherent RDM programme and must be done through coordinated and collaborated efforts of the Library, Research Office and the local academic community.

2. A conscious and gradual programme to embed data management planning (DMP) into the research practices. For instance, at UG this can be done by including DMPs as part of the current ethics approval process. This will help to identify potential uses and risks for research data very early in the research process, so that the institution can identify which data are worth retaining or discarding or what actions must be taken to address data risks.

3. The University management should set up a high-level working group comprising senior officials from the Research Office, Library, IT department, Legal department and any relevant stakeholders to champion RDM within the institution and also ensure a deeper cooperation between researchers and the RDM working group for broader consultations and assessment of institutional capabilities and RDM needs.

4. Invest in building capacity for research support staff. This is crucial for a competent and reliable professional expertise on RDM within the institution. Research support staff should be supported to attend trainings, short courses, conferences and workshops about research data management. This will retool current staff to take up data management support roles and also build their capacity as trainers.

5. The skills and knowledge gap also have implication for curricula development. It is recommended that the School for Library studies update its graduate curricula to reflect the new and emerging roles of librarians in the emerging data-driven research environment.

6. This study is also an awakening for Academic and Research libraries on the African continent about the emerging innovations that is revolutionizing the operations of the library. It is, therefore, recommended that librarians should be proactive by innovatively developing specialized research support services for their researchers whether there is a mandate or not.

Limitations

As is the case with qualitative studies in general, generalisation of results can be challenging. The transferability of the findings therefore is dependent on contextual proximity. A thick description of the institutional context can be found in the original research (Avuglah, 2016: 2-4). Also, given the nature of the study and the sample size used, some of the findings may not necessarily reflect the general state of affairs. The researchers used triangulation of data sources as a strategy to reduce this challenge by verifying some of the findings from the primary data with secondary data which were in the institutional documents. This in no way diminishes the veracity of the research findings which are still instructive, providing some useful pointers worth considering.

Acknowledgements

References

ANDS. 2013. *Research data management in practice*. Caulfield East, Victoria, Australia: Australian National Data Service. Available: http://www.ands.org.au/__data/assets/pdf_file/0009/394056/research-data-management-in-practice.pdf [accessed 9 April 2018].

ANDS. 2017. *Creating a data management framework: ANDS Guide*. Caulfield East, Victoria, Australia: Australian National Data Service. Available: http://www.ands.org.au/__data/assets/pdf_file/0005/737276/Creating-a-data-management-framework.pdf [accessed 9 April 2018].

Australian National Data Service *see* ANDS.

Avuglah, B.K. 2016. Developing an implementation plan for Research Data Management (RDM) at the University of Ghana. M.IT dissertation. University of Pretoria. Available: http://hdl.handle.net/2263/62100 [accessed 12 December 2018].

Awre, C., Baxter, J., Clifford, B., Colclough, J., Cox, A., Dods, N., Drummond, P., Fox, Y., *et al*. 2015. Research data management as a "wicked problem". *Library Review*, 64(4/5): 356-371.

Ball, J. 2013. Research data management for libraries: getting started. *Insights,* 26(3): 256-260.

Beitz, A., Groenewegen, D., Harboe-Ree, C., Macmillan, W. and Searle, S. 2014. Case study 3: Monash University, a strategic approach. In: Pryor G., Jones S. & Whyte A. (eds.) *Delivering research data management services: fundamentals of good practice.* London: Facet Publishing, 163-190.

Brewerton, A. 2012. Re-skilling for research: investigating the needs of researchers and how library staff can best support them. *New Review of Academic Librarianship*, 18(1): 96-110.

Chigwada, J., Chiparausha, B. and Kasiroori, J. 2017. Research data management in research institutions in Zimbabwe. *Data Science Journal*, 16 (31): 1-9. Available: https://doi.org/10.5334/dsj-2017-031 [Accessed 9 April 2018].

Chiware, E.R. and Becker, D. A. 2018. Research data management services in southern Africa: a readiness survey of academic and research libraries. *African Journal of Library Archives and Information Science*, 28(1): 1-16.

Conrad, S., Shorish, Y., Whitmire, A. L. and Hswe, P. 2017. Building professional development opportunities in data services for academic librarians. *IFLA journal*, 43(1): 65-80.

Cox, A. M. and Pinfield, S. 2014. Research data management and libraries: current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4): 299-316.

Cox, A., Verbaan, E. and Sen, B. 2012. Upskilling liaison librarians for research data management. *Ariadne*, 70. Available: http://www.ariadne.ac.uk/issue70/cox-et-al [accessed 9 April 2018].

Cox, A.M. and Verbaan, E., 2016. How academic librarians, IT staff, and research administrators perceive and relate to research. *Library & Information Science Research*, 38(4): 319-326.

Creswell, J. W. 2009. *Research design: qualitative, quantitative, and mixed methods approaches*. 3rd ed. London: Sage.

Creswell, J. W. 2013. *Qualitative inquiry and research design: choosing among five approaches*. 3rd ed. Los Angeles, CA: Sage.

Crowston, K. and Qin, J. 2011. A capability maturity model for scientific data management: evidence from the literature. *Proceedings of the Association for Information Science and Technology*, 48(1): 1-9. Available: doi:10.1002/meet.2011.14504801036 [Accessed 9 April 2018].

Davidson, J., Jones, S., Molloy, L. and Kejser, U.B. 2014. Emerging good practice in managing research data and research information within UK universities. *Procedia Computer Science*, 33: 215-222.

Henderson, M. E. and Knott, T. L. 2015. Starting a research data management program based in a university library. *Medical Reference Services Quarterly,* 34(1): 47-59.

Hey, T., Tansley, S. and Tolle, K.M. (ed.). 2009. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

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.

Hiom, D., Fripp, D., Gray, S., Snow, K. and Steer, D. 2015. Research data management at the University of Bristol: charting a course from project to service. *Program: Electronic Library and Information Systems*, 49(4): 475-493.

Hodson, S. and Molloy, L. 2014. Case study 5: Development of institutional RDM services by projects in the JISC managing research data programmes. In: Pryor, G., Jones, S. and Whyte, A. (eds.) *Delivering Research Data Management Services*. London: Facet Publishing, 205-237.

Jones, S., Duke, M., Rans, J. and Welgert, V. 2015. Meeting the requirements of the EPSRC research data policy. *JISC*. Available: https://www.jisc.ac.uk/guides/meeting-the-requirements-of-the-EPSRC-research-data-policy [accessed 29 April 2018].

Jones, S., Pryor, G. and Whyte, A. 2012. Developing research data management capability: the view from a national support service. In: *Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES)*. Paper presented at iPRES, Toronto, 1-5 October, p. 142-149. Available: https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf [accessed 29 April 2018].

Jones, S., Pryor, G. and Whyte, A. 2013. How to develop research data management services - a guide for HEIs. DCC how-to guides. *Edinburgh: Digital Curation Centre*. Available: http://www.dcc.ac.uk/sites/default/files/documents/publications/How-to-develop-RDM-services_finalMay2013rev.pdf [accessed 29 April 2018].

Jones. S. 2014. CARDIO: Collaborative Assessment of Research Data Infrastructure and Objectives. *SlideShare*. Available: http://www.slideshare.net/sjDCC/cardio-34772337 [accessed 29 April 2018].

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.

Khokhar, M., Schwamm, H., Krug, J. and Albin-Clark, A. 2017. Data Management Administration Online (DMAOnline). *Procedia Computer Science*, 106: 291-298.

Lynch, C. and Carleton, D.E. 2009. Lecture: impact of digital scholarship on research libraries. *Journal of Library Administration,* 49(3): 227-244.

Lyon, L. Ball, A., Duke, M. and Day, M. 2012. Developing a community capability model framework for data-intensive research. In: *Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES)*, Toronto, 1-5 October, p. 9-16. Available: https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf [accessed 29 April 2018].

Pickard, A. J. 2008. *Research methods in information*. London: Facet.

Procter, R., Halfpenny, P.and Voss, A. 2012. Research data management: opportunities and challenges for HEIs. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet: 135-150.

Pryor, G. 2013. A maturing process of engagement: raising data capabilities in UK higher education. *International Journal of Digital Curation*, 8(2): 181-193.

Pryor, G. 2014. A patchwork of change. In: Pryor G., Jones S. and Whyte A. (eds.) *Delivering research data management services: fundamentals of good practice.* London: Facet, 1-19.

Rans, J. and Jones, S. 2013. *RDM strategy: moving from plans to action. DCC RDM Services case studies.* Edinburgh: Digital Curation Centre. Available: http://www.dcc.ac.uk/resources/developing-rdm-services [accessed 12 April 2018].

Rice, R. and Haywood, J. 2011. Research data management initiatives at University of Edinburgh. *International Journal of Digital Curation*, 6(2): 232-244.

Sallans, A. and Lake, S. 2014. Data management assessment and planning tools. In: Ray, J. M. (ed.) *Research data management: practical strategies for information professionals*. Series: Charleston Insights in Library, Information, and Archival Sciences. West Lafayette, IN: Purdue University Press, 87-108.

Searle, S., Wolski, M., Simons, N. and Richardson, J. 2015. Librarians as partners in research data service development at Griffith University. *Program: electronic library and information systems*, 49(4): 440-460.

Ssebulime, J. B, van Deventer, M. and Pienaar, H. 2018. The role academic libraries could play in developing research data management services: a case of Makerere University (Preprint). Available: https://bit.ly/2AL8qEw [accessed 15 December 2018].

Takeda, K., Brown, M., Coles, S., Carr, L., Earl, G., Frey, J., Hancock, P., White, W. *et al.* 2010. *Data management for all: the institutional data management blueprint project.* Presentation at the 6th International Digital Curation Conference. Available: http://eprints.soton.ac.uk/169533/1/6th_international_digital_curation_conference__idm b_final_paper_revised.pdf [accessed 29 April 2018].

UG. 2014. *University of Ghana Strategic Plan, 2014 – 2024*, University of Ghana. Available: http://www.ug.edu.gh/sites/default/files/documents/UG%20Strategic%20Plan.pdf [assessed 3 August 2018].

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.

Wang, M. 2013. Academic Library, e-Science/e-Research, and data services in a broader context. In: *Proceedings of the ACRL 2013 Conference in Indianapolis,* Indiana, 10-13 April, 2013. Association of College and Research Libraries. Available: http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/201 3/papers/Wang_AcademicLibrary.pdf [accessed 29 April 2018].

Whyte, A. 2012. Emerging infrastructure and services for research data management and curation in the UK and Europe. In: Pryor, G. (ed.), *Managing Research Data.* London: Facet, 205-234.

Whyte, A. and Allard, S. (eds.) 2014. *How to discover requirements for research data management services. DCC How-to Guides.* Edinburgh: Digital Curation Centre. Available: http://www.dcc.ac.uk/how-discover-requirements [accessed 29 April 2018].

Whyte, A., Molloy, L., Beagrie, N. and Houghton, J. 2014. What to measure? Towards metrics for research data management. In: Ray, J. M. (ed.) *Research data management: practical strategies for information professionals.* Series: Charleston Insights in Library, Information, and Archival Sciences. West Lafayette, IN: Purdue University Press, 275-302.

Wilson, J. A., Martinez-Uribe, L., Fraser, M. A. and Jeffreys, P. 2011. An institutional approach to developing research data management infrastructure. *International Journal of Digital Curation*, 6(2): 274-287.

Wong, G. K. 2009. Exploring research data hosting at the HKUST institutional repository. *Serials Review*, 35(3): 125-132.

Yin, R.K. 2009. *Case study research: Design and methods*. 4th ed. Series: Applied Social Research Methods vol. 5. Los Angeles, CA: Sage.