Communication

Judith Mavodza*

Research Data Management: A review of UAE academic library experience

https://doi.org/10.1515/opis-2022-0128
received October 23, 2019; accepted March 22, 2022.

Abstract: Purpose: This paper is a review of the status of Research Data Management (RDM) efforts in UAE public university libraries. Approach: The investigation is through examining available literature about the topic using region-specific articles when available, librarian comments, and the information provided by UAE university library websites. Existing lessons and policy documents are sought, and plans suggested for local solutions, suggesting avenues for progress. Findings: Though not a new concept, findings indicate that local RDM activities are emerging, but knowledge of their importance of exists. Research limitations: This review is limited to public university libraries though the results and experiences could be generally relevant to more research establishments. Practical implications: Taking advantage of the existing awareness to organize tangible RDM efforts can facilitate retrieval and availability of data relevant to the region. Value: An intricate range of activities involved in the organization of RDM services is revealed.

Keywords: Research Data Management (RDM); Big Data; Data Literacy; eFada; Open Access.

1 Introduction

This discussion covers the Research Data Management (RDM) efforts in national university libraries (i.e. those created by government with national aspirations) in the United Arab Emirates (UAE). These are predominantly moving from a teaching focus, becoming research-intensive. While Puri-Mirza (2020), in the Statista database, lists 15 public universities, and 79 private universities and colleges, several of the private ones are satellite campuses of Western-based universities. A list of licensed higher education institutions is also reflected on the UAE Ministry of Education (2018) website. Western-based institutes sometimes have already-existing RDM arrangements, e.g. the University of Wollongong Dubai, Paris Sorbonne University in Abu Dhabi, or New York University Abu Dhabi, have highly established digital data collections. In this essay, focus is on efforts in UAE public universities. However, it is noted that when it comes to research matters, academics identify more with the notion of "academic tribes" (Becher & Trowler, 2001; Trowler, Saunders, & Bamber, 2012) than with their parent institution. As such, local RDM-related services and advances are pointed out to demonstrate and affirm their contributions to the national and global research environment, but without positioning the extent of researcher support provided by librarians in the various universities.

As in many other parts of the world, an enhanced emphasis on research among UAE universities, transforming them from teaching-only establishments to a comprehensive agenda includes research, development, innovation, and entrepreneurship. Previously, research by Abouchedid and Abdelnour (2015) revealed that faculty research productivity in six Arab countries (UAE included) was very low. Thereafter, additional research institutes in UAE universities have been opened as illustrated on their websites. Johnson and Potluri (2020, p.1) are of the viewpoint that in the UAE, "research data service is slowly gearing up in many academic libraries". This slow but progressive development of the research agenda requires that resources and support be in place to enable researchers to seriously engage in research. Supporting elements of the research cycle that include research data services (RDS) or RDM is thus a major library activity. But then, it is not entirely the availability of and access to research resources and data that determines the

***Corresponding author: Judith Mavodza,** Zayed University, United Arab Emirates, E-mail: Judith.Mavodza@zu.ac.ae

success of research, innovation, and entrepreneurship. Other factors are at play too, for example, teaching workload matters. According to Kelsky (2018), top-ranked research universities have a standard 2-2 load (meaning an individual teaches two courses each semester of the academic year) in the humanities and most social sciences (and a 1-2 or 1-1 or 1-0 in STEM fields). Both Zayed University (ZU) and United Arab Emirates University (UAEU) have teaching workload formulas in faculty handbooks on their websites that reflect efforts to balance teaching and research productivity. These are examples of an approach geared for encouraging academic researchers to be productive. Other impactful factors are discussed in the next sections.

2 Research Data in Academic Libraries

The proliferation of huge amounts of data and information that academic scholars and researchers are confronted with drives the need for RDM services. Their development in various regions of the world have national implications and are based on universities identifying the existing need for them, developing plans of action, considering existing tools and service models that may be reused. Recording the long history of American RDM efforts starting from 1950s, Rice and Southall (2016, p. 10) point out that it "would be a mistake to think all these developments have been smooth and of the same kind". In comparison, RDM projects in libraries of UAE public universities are in their early stages but have the potential advantage of benefitting from the experiences and examples of established services both inside the country and outside. In fact, Johnson and Potluri (2020, p. 10) mention that the "UAE librarian's community is playing a vital role in changing research behavior by supporting the researchers in building the knowledge and skills required to retrieve and manage data by conducting hands-on workshops, seminars, and data skill development programs."

In academic research environments, librarians are sometimes expected to make research data as accessible and visible as possible. Where this is within the boundaries of their professional control, they make them retrievable by creating metadata schemes and taxonomies as well as designing standard retrieval methods. Ackers (2013) points out that in university environments, academic librarians are likely to encounter Small Data, therefore should focus on developing approaches to the documentation, organization, preservation and dissemination of small datasets that have no permanent home outside of the labs and offices in which they were born. They must also increasingly include data literacy in their instructional programmes. The Small Data sets are the result of research activities. While highly valuable, the data are also diverse and heterogeneous in nature (Wang, Xu, Chen, & Chen, 2016) and can be huge in a university setup where active research is going on. The data are useful for mining by other researchers per their needs, but that requires proper infrastructure and a policy framework. Librarians thus support by curating the data to make them more valuable and available for other future uses. Decisions therefore should be made about what to save and in which format. That leads to data archiving. Researchers usually store their data locally on their computer disks or in the cloud (Elsayed & Saleh, 2018; Tenopir, Birch, & Allard, 2012).

Librarians get involved in attempts to systematically archive that data, meaning that they arrange, describe, document, and preserve the data to enhance and make it easily retrievable, and in the process facilitating new research. But then, there is frequently a data organization challenge due to the lack of subject-specific and interdisciplinary standards (Petersen, et.al, 2020; Wang, Xu, Chen, & Chen, 2016), and costly nature of the high-performing computer servers required to process such data (Wang, Xu, Chen, & Chen, 2016). Nonetheless, it is their role to curate and preserve the datasets, thus help researchers understand their meta-descriptions. Promoting and supporting the use of common standards (e.g. ISO/IEC JTC 1/SC 32 - Data management and interchange – specifically the ISO/IEC 11179 (Martin-Rodilla & Gonzalez-Perez, 2019, p. 31) family of standards) is to address the specification and standardisation of metadata schemas); and tools (e.g. research instruments such as questionnaires and/or scales, for implementing the research strategy) among researchers. This is how it is possible for researchers to relate to the research data lifecycle, data analysis, tools, statistics (Chiware & Mathe, 2015). In other words, data librarians do their best to maintain knowledge bases about data. They also champion the establishment of uniform data citation standards.

The topic of library initiatives organizing research data attracts attention in many parts of the world and has been discussed widely. Consensus is that when large networks and repositories are created, they provide support for effective data management and access. Peters (2018, p.2) mentions that there are "relative strengths in research data management across the institutions and working together...". HathiTrust (2017) is an example of such a co-operative network effort of more than 100 university libraries. It is administered by Michigan and Indiana universities and is a

collaborative repository of digital content from research libraries. This is not the only example, but one of several. In fact, "we would be mistaken to assume the feasibility of a one-ring-to-rule them all solution. We need many rings" (Padilla, 2018, p.298). Essentially, a smart policy framework is necessary because personal research data is being used in new ways, giving rise to many information sharing issues. Policy makers must therefore create proper guidelines for an information sharing environment and researcher privacy. Guidelines must define the extent and nature of use that can be considered fair dealing. But then, per Salo (2010):

> significant mismatches exist between research data and library digital warehouses, as well as the processes and procedures librarians typically use to fill those warehouses. Repurposing warehouses and staff for research data is therefore neither straightforward nor simple; in some cases, it may even prove impossible.

The same point is repeated by Wang, Xu, Chen, and Chen (2016) who suggest that this work could be done by data scientists rather than librarians, and by Rice and Southall (2016). This reality means that libraries planning to organize data should either train their personnel or hire already trained ones to fulfil the universities' research data management requirements. They are needed in providing researchers with advice on levels of control relating to their research data. Funding for existing and continued maintenance of data repositories is a far-reaching investment.

There is also a special challenge in libraries concerning the use of the various Content Management System (CMS) platform options whose control of the underlying technology is not necessarily controlled by librarians, therefore including the RDM option is not always an original component. The University of Exeter, for example, uses SWORD and Globus plug-ins to bridge the gap between DSpace and Big Data capture (Taylor, 2013). A similar arrangement is described in a study by Jurik, Blekinge, Ferneke-nielsen, and Møldrup-dalum (2015) who, in the context of the Danish State and University Library, refer to integrating Hadoop applications or Map Reduce with traditional repository systems such as Fedora. Furthermore, Motta, Puccinelli, Reggiani, and Saccone (2016) suggest an integrated system for the management and analysis of grey literature contents and metadata. Attention is required towards the effectiveness of query processing strategies in Big Data applications, such as Hadoop, that are intended to facilitate its adequate exploitation. But then, Salo (2010) expresses concern about the lack or shortage of library personnel expertise in creating and managing the data capture and control of the platforms. A similar point is repeated by Itani and Östlundh (2021) who confirm the challenge of staff who lack qualifications in required skillsets in their UAE University open scholarship research. However, Kafel, Creamer, and Martin (2014) note the increasing interest in library science training programmes to include data management in the curriculum, furnishing trainee librarians with data management, data mining, machine learning, deep learning capabilities, equipping them with data stewardship skills.

3 Research Data Types and Formats

There is a tendency for different universities and institutes to define research data according to how it matters to them. Cox and Verbaan (2018) list an array of definitions from universities and research organizations, but what they have in common is that it is all data in varying formats and types. Data types include results from experiments or laboratories, generated from test models (climate, mathematical or economic models), social media data such as tweets, logs of web server traffic, derivations or compilations (text and data mining, databases, 3D models), collections of peer reviewed datasets published (chemical structures, gene sequence databanks, statistical datasets), and datasets from national censuses (for demographic, occupational, and housing information to answer fundamental questions in social economic history), and more.

Research data formats include raw data files, artefacts, documents (PDF, Microsoft Word, Text), notebooks, questionnaires and transcripts, specimen, samples, photographs, and more. According to Essayed and Saleh (2018, p. 2), "research data includes every piece of data acquired and generated during the research process, and may comprise, among others, text, spreadsheets, questionnaires, photographs, films, test responses, slides, laboratory notes, statistics, observations, results of experiments, measurements, samples, algorithms, scripts, and workflows". Like Big Data, research data is characterized by its huge volume, its variety and the velocity with which it is being created, its veracity, i.e. reliability, and the value attributed to as well as resulting from it. RDM is important in facilitating the

making of informed data-reliant decisions. The role of university libraries as enablers of research is central to data capture, management, and access by researchers, and this is not always a smooth activity, as literature reveals.

4 Research Data Repositories

There are data repository directories, e.g. Re3Data which is a database of data repositories; Fairsharing.org, a catalogue of databases and related resources; DataCite, a database of datasets and repositories; the European Union Open Data Portal which is a catalogue of open datasets; Data Citation Index (DCI) which is a database of datasets; and Google Dataset Search. Beyond the directories are general data repositories, e.g. Dryad Digital Repository, Figshare, Harvard Dataverse, Open Science Framework, Zenodo, and Mendeley Data. The advantages and strengths of using one or the other, or several at the same time are determined by institutional requirements and preferences. Additionally, subject specific repositories are also the choice of many disciplinary scholars and researchers, e.g. in biology and life sciences, chemistry, computer science, earth and environmental science, geoscience, health sciences, humanities, physics, astrophysics and astronomy, and social sciences. Among these registries, there is evidence that research data from or about the UAE exists in several repositories. An example is that a search for the name "Emirates" (by no means an exhaustive search term) in March 2022 in PANGAEA reveals 18 datasets; DRYAD reveals 11, and ZENODO reveals 2739 datasets. A search for "United Arab Emirates" in Mendeley Research Data in March 2022 reveals 27420 datasets of various types from several data repositories across the research world. It is in light of this vast array of possibilities that the article, *So you want to be a data scientist?* by Catherine de Lange (2013) becomes useful in explaining the increasing demand for the position of Data Librarian, investment in RDM projects, including the hiring of copyright and data experts.

5 The Context of Research Data Management

The management of research data is increasingly recognized from the perspectives of different professions because of "the deluge of data arising from new types of science, a crisis in confidence in research integrity in certain fields and the general movement for open data…" (Cox & Verbaan, 2018, p. 5). Research data is relevant to all subject areas and contributes to the bigger Big Data landscape used in and for policy making. Though this is not a new proposition, it is true that university libraries, regardless of geographical location, have a special role to play in supporting academic researchers with some of the data-related tools they need to pursue their disciplinary activities. Literature reviews and library website scrutiny indicate that RDM activities in selected UAE university libraries are developing, and knowledge about their importance exists. Some of the expressions of progress at UAEU were through a research article by Itani and Östlundh (2021) with the title "The role of institutional repositories in advancing open scholarship: A case study from the United Arab Emirates University", and at Zayed University during the *UAE Innovates Week* at the Expo 2020 during February 2022 in a presentation with the title: "Supporting Innovation with Open Access Publications and Data - ZU Scholars Institutional Repository" (Lappalainen, 2022).

Flores *et al.* (2015) propose that a needs assessment exercise and policy development must be put in place before confirming RDM projects. Bryant, Lavoie and Malpas (2018, p. 18) suggest that the motives for organizing a RDM service are:

> compliance with mandates or policies that establish formal requirements for documenting research data management plans or for demonstrating progress toward open science goals; evolving scholarly norms that influence disciplinary perspectives on what constitutes good scientific practice, including expectations of reproducibility and transparency in documenting protocols, methods, and data sources; institutional strategies that are aided by more rigorous and systematic attention to monitoring research productivity and performance and improving (or maintaining) institutional reputation in data-intensive research areas; direct or derived demand from researchers with unmet (or imperfectly satisfied) data management needs; for example, evidence that university researchers are turning to external services to meet data storage, management, or sharing needs that could be met by the university.

Once RDM programmes are in place, advocacy programmes, awareness campaigns, training sessions and advisory services, data repository development, helpdesk services, and data management plans (DMP) must be organized. It is notable that a RDM skills course is now integrated in the curricula of many library schools and available through professional development resources, including those run through the Research Data Management Librarian Academy that subsequently posts the training and publications files on GitHub (https://rdmla.github.io/) that librarians in the UAE are engaging with; or the Elsevier Researcher Academy webinars, and so on. The attributes of those courses cater to the "research data scientist" or "institutional intelligence officer" or "data librarian" type positions mentioned above that require computational and statistical expertise.

6 Questions About RDM Efforts

A Google search for "research data management" jobs in UAE libraries does not yield results. However, there are a few instances of "research data scientist", or "institutional intelligence officer" positions advertised, unrelated to libraries. In only a single instance was the position of "data librarian" advertised in 2020. But then, the fact that activities or terms of this nature are not reflected on university websites does not conclusively suggest that they are not happening. One needs to comment on possible reasons for delayed adoption of the efforts, and how libraries are involved. In so doing, the following questions arise:

a) Are academic libraries taking leadership in coordinating RDM efforts in their universities?
b) Does an enabling environment for those efforts exist?
c) What are the possible explanations for delays or impediments in RDM activities?

A close look was made at the status of RDM efforts in libraries of the three UAE national universities, using information about those libraries that is publicly available on their websites, unsolicited comments from librarians, and from literature review. Interest in the selected universities was due to their UAE national origins and efforts while they operate on an international education and research arena.

7 Commenting on RDM at UAE Public Universities

Efforts specific to the public national university libraries are at varying stages, mostly being works in progress. For example, in some libraries, positions exist in the Systems and Digital Library Services Section of a Digitization Technician, and the position of Librarian for Research, Data Management and Instruction. These repositories contain some research data, while others' efforts are developing. Meanwhile, after searching the library websites of the 15 public universities, there is reference to thesis repositories, but with a limited number of formalized RDM projects. However, the United Arab Emirates University (UAEU), for example, holds a Mars project-related dataset in its repository; Zayed University has links to seven datasets ranging from 2017 to 2021. From this information, one can see that the university library has taken leadership in making the resources known through being posted on the websites. The results of a United Nations e-Government survey (2014) indicated the UAE as one of the countries with a score of more than 66.6% in data publishing, suggesting an inherent enabling environment for research data harnessing. For the sake of propping up the UAE research environment, researchers must have support in finding information on where this data is readily accessible.

One must bear in mind that the strides being made by universities exist in a context where language, research scope, and resources (human and infrastructural) sometimes lead to the undervaluation of academic research output and in many instances limits the ability of regional scholars to contribute to international science. To overcome this shortcoming, an eFada initiative (Khalifa University, 2013) using the Ankabut platform, a UAE Advanced National Research and Education Network (NREN) was created. One of its goals was to provide high performance computing services and "data visualization services" (Khalifa University, 2013). As at 2022, eFada has accomplished some of its goals, but the RDM services plans are yet to be fulfilled. Meanwhile, regional research output in the Arabic language is increasingly becoming available through databases such as Al Manhal, Banipal, Dar AlMandumah, E-Marefa, eBook

Arabic Collection (EbscoHost), Kotob Arabia, to name a few. But then, if the research data emanating from the works housed in those databases were harnessed for access, they could contribute to global research, adding to the pool of original research pursued in Arabic language. Abalkhail (2018) highlights the challenge of translating qualitative data from Arabic to English for purposes of writing articles in English. This challenge is suggestive of the need for capturing and using research data in the original language. A proposal by Larabi Marie-Sainte, et.al (2019) suggests a New Arabic Dataset (NADA) for text categorization purposes. The same is echoed by Hammoud (2021) in reference to the importance of having an Arabic medical dataset for diseases classification. Resulting from infrastructural deficiencies, an additional stumbling block for scholars and researchers in low/middle-income countries (LMICs) is expressed by Bezuidenhout and Chakauya (2018, p. 39) to be the fact that "low-resourced environments shape data sharing activities but are rarely examined within Open Data discourse", meaning that they too suffer exclusion. Nonetheless, libraries are counted among the essential support entities available to all scholars.

Regardless of where RDM activities take place, libraries lower barriers that researchers sometimes face, making institutional scholarly works and research data available. Depending on the available infrastructure, librarians can integrate data catalogues with regular library retrieval services such as Koha, Alma, WorldShare Management Services (WMS), and Sierra among others, making the search for data hassle-free. But then, in the said UAE academic libraries, involvement of the requisite data specialists is an evolving effort. The required skills include creating common meta-description schemas and common citation practices so that data sets have persistent identifiers and can be crosslinked between publications and datasets. Awareness about the impact of datasets is also necessary for the data creators. For that reason, librarians must guide them on bibliometrics (for example, impact factor, h-index) and altmetrics. That is facilitated with the use of tools such as Mendeley, Scopus author management, ORCID identifier, and DMP Tool (Chiware & Mathe, 2015). Data librarians are also essential to higher education research projects because they provide advice on conditions under which data can be re-used. It is important for them to guide researchers on discipline specific best practices in data creation and intellectual property rights. They also offer RDM support, including DMPs for grant applications, intellectual property rights advice and the integration of data management into the curriculum.

The responsibilities mentioned above demonstrate that the work of librarians keeps evolving due to continuously changing demands and expectations of the academe. The ACRL (2019) summarizes those responsibilities to include:

> helping researchers to deposit data in institutional and disciplinary repositories, assisting with data management plans, and consulting with research teams. Librarians also serve the research community by creating workshops, webinars, and tutorials. ... develop a robust understanding of tools and support mechanisms available on campus... collaboration with campus computing or statistical support services...providing background information as well as advanced information about RDM...through, newsletters, or webpages is also important...have an RDM presence online ...provide information related to creating data management plans, data documentation, metadata standards, storage and preservation.

A similar description is provided by the Association of European Research Libraries (2018) and repeated in their 2018-2022 vision for the research landscape, suggesting the recognition of data management librarians. Sharing experiences in the context of South Africa, Chiware and Mathe (2015) concur with both the Association of European Research Libraries and ACRL on the need to employ data librarians in research-heavy data-reliant environments. It is the responsibility of the data librarian to connect with useful RDM procedures. The foregoing account affirms the importance of data librarian skills to scholarly research as it takes centre stage in UAE higher education.

Reference to increased use of research data in due to its exponential growth. But then many libraries must create Data Librarian posts, and this is not unique to the UAE. The development of professional staff skills for data librarianship has become essential so that the library can actively participate in institutional research data policy development, including resource plans, and adopt open data policies where appropriate in the research data life cycle. On a wide scale, it is important to partner with researchers, research groups, data archives and data centres to foster an interoperable infrastructure for data access, discovery and data sharing. This is where the Dataverse/ Bayanatona approach mentioned by a few UAE university libraries and, in conjunction with eFada would fit. Other RDM options that can be investigated too, including the Comprehensive Knowledge Archive Network (CKAN), DSpace, and more as listed by Lewis (2014), depending on the functions required and defined by the university.

It appears that university libraries need champions and advocacy especially when they do not yet possess data repositories or immediate plans for RDM services. A study of the UAE library websites suggests that requisite data experts are rare, causing delays in initiating any RDM plans. It is worth noting that there is value in learning from

university libraries with established RDM services in the UAE regardless of their status, to accurately articulate the advantages of having systems in place, the types of data and objects for inclusion, intellectual property (IP), and licensing implications. Harnessing knowledge about awareness, access, and use of these resources is key for the benefit of researchers.

Comments about RDM from librarians at the universities studied reflect that they view local initiatives as still developing. However, they admit that the researchers they support produce research data from their projects but organizing the same data for re-use, possibly by other scholars, could be enhanced. They are generally aware that for RDM programmes to be initiated, certain infrastructure must be in place, and various constituents involved as these are major initiatives. Examples of required participants include research offices, faculty evaluation processes, the library, information technology, legal counsel for intellectual property advice, and more. It is also understood that file storage strategies are necessary because storage formats are liable to fail, and file formats eventually become obsolete, therefore storage strategies help to minimize the risk of loss or destruction of data. Some conversations have suggested a national repository to centrally formalize RDM processes in the UAE, echoing the intentions of the eFada initiative. One can make the comment that this arrangement would enable the harvesting, management, and centralized regularized access to research data. Talib et.al. (2015), at the realization of a huge demand for a unified database for social science research in the UAE, advocate that UAE social science researchers upload their work to the Dataverse Network (coined with the name Bayanatona). Dataverse is "an open source web application to share, preserve, cite, explore, and analyse research data" (Dataverse Project, 2020). The name appears on some university websites, but the effort appears to require continuity and champions to keep it vibrant. Yoon and Schultz (2017) suggest improving aspects of library RDM webpages to include service development and thoroughness of information offered. The RDM topic has also been addressed at conferences and workshops in the UAE, e.g. INDEPTH Data Management Programme Training Workshop of 2017 in Dubai.

From a broader perspective, RDM activities in the UAE and the GCC region (Saudi Arabia, Qatar, Bahrain, Kuwait, Oman, UAE) reflect tendencies found in many other universities across the globe. A few have existing institutional repositories (IRs), but not necessarily active RDM/ RDS arrangements. The same holds for the role of librarians as expressed in a research data services study (RDS) by Cox, Kennan, Lyon, Pinfield, and Sbaffi (2019, p. 1432) which confirms "a picture of the spread of RDS, especially advisory ones. However, future ambitions do not seem to have seen much evolution. There is limited evidence of organisational change and skills shortages remain". In an ACRL White Paper, Tenopir, Birch, and Allard (2012) found that a quarter to a third of all academic libraries in the United States and Canada were planning to offer some research data services within the next two years. But then, another investigation by Tenopir, Sandusky, Allard, and Birch (2014) revealed not much difference two years later. Chigwada, Chiparausha, and Kasiroori (2017) make similar observations from their study focussing on Zimbabwe, where they conclude that RDM is "a relatively new concept in Zimbabwe's research institutions as compared to other institutions in the developed countries". Similarly, a research by Chiware and Becker (2018), covering Southern Africa (Botswana, Lesotho, Malawi, Namibia, South Africa, Swaziland, Zambia, and Zimbabwe) concludes that there is "need to advocate for awareness of research data management within institutions with academic and research libraries taking a leading role in spearheading data management and providing training and the technical support needed to store and retrieve research output and data sets". This too is the conclusion from a study of RDM practices in Australia by Krahe, et. al. (2019). Additionally, Tenopir, et.al. (2019, p. 36) found in a follow-up investigation of the USA and Canadian studies that the "provision of RDS is not yet widespread among academic librarians, although there is a group for which these services are integral to their jobs and another group who offer some of the services occasionally". Regardless, the perspective of Zhou (2018) on RDM/RDS is that services in the United States, Britain, Australia, and some universities in China such as Peking University, Fudan University, Wuhan University and Xiamen University have become more developed than elsewhere.

There are no specific studies either about the UAE or the GCC that are equivalent to the USA/Canada cases by Tenopir, or those from Southern Africa, but the findings of the said studies perhaps explain why the progress of RDM services via eFada in libraries of UAE public universities have not taken shape. In fact, efforts at Western-based universities, or non-federal ones, or private ones tend to be more advanced than those at public ones, e.g. New York University Abu Dhabi has a Social Sciences and Data Librarian position; the Abu Dhabi science-focused Khalifa University of Science, Technology and Research (KUSTAR) has well-established RDM guidance supported by the library team; Paris Sorbonne has the well-staffed SUAD Institutional Repository; the University of Wollongong Dubai has a service that one can

request through the library. RDM services are, however, also in place in other regional libraries, e.g. in the Kingdom of Saudi Arabia at King Abdullah University of Science and Technology (KAUST); or in Qatar at the Carnegie Mellon University Doha, and at University College London Doha, and so on. Omani and Kuwaiti library websites generally have no mention of data services, but that does not rule out that RDM efforts may be in existence in other forms.

8 Concerning Policy for RDM Efforts

According to the *UNESCO science report: Towards 2030* report by Zou'bi, *et al* (2015), it is important to put in place data policy framework for academic libraries, and data protection legislation. This defines the plan detailing what data to collect, documentation and metadata, ethics and legal compliance, selection and preservation procedures, storage and backup, data sharing provisions, and what resources are used, and allocation of responsibilities (Cox & Verbaan, 2018). In the event of RDM being outsourced, or where there is an independent data publishing model, it is necessary to set clear parameters for doing so. The independent data publishing model involves using spaces where research data from various departments, localities and units are integrated to make full use of international research data resources, developing series of data sets and products, building an intelligent and networked RDM and sharing service system (Xu, 2019). There must also be user training, and professional development opportunities relevant for librarians with limited data management knowledge or experience, raise awareness about research data availability and access, its advantages, and ethical use in and by university researchers. That reduces negativity and suspicion while assuring the researcher community of the range of support capabilities the libraries offer. There is merit in libraries becoming members of professionally supportive organizations, e.g. Research Data Alliance (RDA). The use of data handling tools also helps, e.g. DMPTool, for creating data management plans that meet institutional and funder requirements; the Spatial Data Access Tool (SDAT); and the UN StaTact toolkit for data governance and data utilization for policy-making purposes.

To enhance research data benefits, there is need to address data literacy skills among academic librarians and potential data consumers. This ranges from the ability to recognize the existence and need for data, to applying recognized ways of citing it. Elsayed and Saleh (2018) highlight the fact that when potential beneficiaries of research data do not possess data analysis capabilities to derive meaningful information from the data, that creates a gap. When there is no meaning derived from existing research data, it stands to reason that there may also be reluctance or inability to store it in retrievable spaces, because of not appreciating the point of doing so. Thus, the lack of clarity about the advantages of data sharing demands researcher guidance on RDM. A study on RDM and sharing among researchers in Egypt, Jordan, and Saudi Arabia by Elsayed and Saleh (2018, p. 2) reveals that "97% of researchers were responsible for their research data, and 64.4% of researchers shared their data". Among those unwilling to share, their inhibition was confirmed as being due to concerns about confidentiality and data misuse. This is not unique to the academics in the countries where the study was done, or the UAE. Rather, it is a concern shared by internet users worldwide, and inclusive of researchers and academics, and that is consistent with an earlier research by Tenopir, Birch and Allard (2012).

A 2017 multi-country survey by the European Conference of Information Literacy (ECIL), intended for assessing RDM appreciation among university researchers that was sent to several UAE participants, attracted comments expressing doubts, confusion, and suspicions about the need for RDM, as well as the mandate and expertise of their librarians. Additionally, a point that is relevant to the UAE, raised by Elsayed and Saleh (2018), is that when university libraries are introducing research data plans in a multi-lingual environment, sometimes researchers need more clarification about and familiarization with meanings of terms, e.g. research data, data management plan, data documentation, data repository, digital archive, etc. Thus, the addition of Arabic language on platforms for research is a practical solution. This is highlighted in the Science, Technology and Innovation Policy of the UAE Government (2015, p. 22) as necessary, stating that:

> The opportunity for "Arabic apps" in the UAE and for the broader Arabic speaking world is already significant, due to the presence of the suitable environment and infrastructure, and the rapid grow in its social, commercial, industrial, medical, government, and entertainment uses.

Other obstacles to RDM efforts include compromised reliability of some research data, e.g. when original data sets are not stored, remaining with only the aggregated results from one data gathering exercise to the next, it becomes complicated to have longitudinal studies of patterns. Furthermore, coding may not always be consistent over prolonged periods of time, resulting in many inconsistencies. When there are spelling mistakes, name variants, and specific language preferences, data collection and use requires unified effort as poor and loosely connected data representation happens easily. If there are no data capture and storage plans, with absence of qualified staff to create and manage the data management projects, accompanied by difficulty in creating RDM specialist job positions, lack of clarity about data capture, ownership, and sharing implications, e.g. data lifecycle, copyright, fair use, intellectual property, the challenges continue to be a hinderance.

9 RDM and Open Access (OA)

Several factors impact research data use. Lack of access to it, for example, is mentioned by Scaramozzino, Ramírez, and McGaughey (2012) as a major challenge particularly if there are no Open Access (OA) options. This can be a result of data repository restrictions (as is sometimes applicable with private funder specifications, concerns about researcher anonymity, and where there are anomalies to do with legal or ethical matters), not knowing of data existence, connectivity and technical infrastructure issues (Bezuidenhout & Chakauya, 2018), the state of the prevailing research culture, and sometimes the unavailability of organized relevant data sets. OA means that the results and data from research are made openly available. The implication is that there must be openness during the research process (Cox & Verbaan, 2018; Elsayed & Saleh, 2018), thus enabling readers to not only use journal articles and books but also be able to re-use or re-analyse the data they see (sometimes resulting in conclusions that differ from those of the originator of the data). That is the essence of academic tribal/ disciplinary dialogue.

For RDM, continuous close collaboration between the library and university leadership, research support units, researchers, and funders (including the applicable government offices) is essential to facilitate decision-making, e.g. the benefits of RDM efforts, funding, deciding which data management platforms to use, and organizing clear privacy policies and notices. Collaboration is practical for shaping decisions on basic plans, the life cycle of data for the long-term, and regularly assessing the RDM policy and practice. Because no single university can have all the data required by researchers, there is need to have interoperability and enhanced access among repositories, with appropriate permissions and limitations embedded. That is based on the OA concept that librarians and researchers must familiarize themselves with, a role that Boufarss and Harviainen (2021, p. 3) suggest as transitioning from "gatekeepers to gate-openers" who actively support best practices among researchers and scholars.

For visibility, data management spaces can be registered in the Registry of Open Access Repositories, and in OpenDOAR. In fact, there is increasingly a tendency for repositories to be ranked according to visibility and impact, e.g. by Institutional Repositories by Google Scholar (webometrics). The Repository Analytics & Metrics Portal (RAMP) is an example of a tool to evaluate the accuracy of repository analytics. Besides these mentioned impact measures, a library must regularly review the performance of its own RDM service through surveys, interviewing researchers, studying periodic data downloads and other relevant parameters or benchmarks it aspires to accomplish. As far as UAE repository visibility goes, in March 2022, OpenDOAR listed five of them including the Corepaedia University of Dubai, DSpace at American University of Sharjah (AUS), Paris Sorbonne University Abu Dhabi, The British University in Dubai (BUiD) Digital Repository, and ZU Scholars. The Registry of Research Data Repositories (re3data.org) which is an Open Science tool, indicates that as at March 2022, no UAE repository is registered in it. In addition to the Re3data.org registry, there are more types that do not reflect the efforts of UAE libraries, but the Registry of Open Access Repositories (ROAR) reflects ZU Scholars and Paris Sorbonne University Abu Dhabi; while OpenAIRE, and the Confederation of Open Access Repositories (COAR) which are necessary to reflect visibility and impact do not.

Established international research institutes, whether they are affiliated to universities or not, are increasingly making their research data available via OA because most of them are sponsored by public funds, therefore expected to allow the public to view and use them. The European OpenAIRE2020 project is an example of one such initiative. It is notable that as at 2022, there was no mandatory OA policy in the UAE for publicly funded research, with no policies listed in ROARMAP (Boufarss & Laakso, 2020). Thereafter, Boufarss and Harviainen (2021, p. 8) re-emphasize the same

point, suggesting that "…local OA policies and mandates are sporadic." This affirms the point about lack of policies being an existing shortcoming as mentioned in the earlier study by Elsayed and Saleh (2018) concerning research data in Arab universities. The same point is also made by Marlina and Purwandari (2019, p. 972) who found that "due to the absence of national or institutional policies, data is generally managed by researchers with limited access". However, the OpenAIRE website mentions that its project evolved to help implement and monitor open scholarship and improve the discoverability and reusability of research publications and data. There are also disciplinary repositories, e.g. Inter-University Consortium for Political and Social Research (ICPSR), incorporating OpenICPSR, DRYAD, PANGAEA (Data Publisher for Earth and Environmental Science), ONEMercury (geosciences and biology data), NIH Data Sharing Repositories (listing medical and biomedical data repositories), VADS (Visual Arts Data Service), British Library Sound Archive, and GIS datasets.

Because UAE academic libraries are predominantly staffed by librarians whose training is as varied as their countries of origin, they must depend on the universities' visions and goals in line with those of the country's Ministry of Higher Education and Scientific Research. This affirms the importance of national policy on RDM. Examples of data management beyond universities include the efforts of the Association of European Research Libraries (LIBER) as elaborated on by Tenopir, et. al. (2017), and those of China as explained by Xu (2019). Elsayed and Saleh (2018, p. 293) suggest that:

> The transition towards a culture of data management and sharing among researchers at Arab universities should start with training in data management and sharing practices; encouraging funded research projects to deposit research data in the funding agencies' repositories at least; providing infrastructures, including repositories, policies, guidelines and best practices, and tools supporting backup and accessibility; and finally, rewarding data sharing within and beyond universities

Likewise, with RDM matters, librarians must have or acquire expertise and objectively understand the options they are proposing or introducing to their patrons.

The MBRSG (2015) suggestions directly impact upon and empower university libraries to address the need for RDM services, and at the same time explain delays experienced as follows:

- Databases need to be set up to house research that is conducted by and for government institutions. This database should be open and readily available to all other government entities.
- The discrepancies in statistical data present in social science research in the UAE need to be remedied through the research institutions becoming more transparent in their research methods. They should open them up for independent verification and to be shared with others.
- There should be unification in accepted rigorous research standards, terminologies and methods among researchers in government and academia to ensure that research standards are maintained and are not questioned.
- Government must act as a driver and allocate more support to research that is in its priority areas.
- Public universities that are funded by government must be required to produce a certain specified number of research outputs. Funding for public universities must be connected to research KPIs.
- Academics need to be incentivized to conduct policy evaluations that help government to assess the impact of its policies

Once a university starts being research-heavy, the results show and rankings agencies, e.g. Times Higher Education World University Rankings (THE), the QS World University Rankings, and the Academic Ranking of World Universities (ARWU) reflect that. The ranking agencies are, in fact, examples of instances where existing data is used to provide visible results. Kelsky (2018) confirms that research-heavy universities attract researchers depending on the research emphasis demonstrated. That incentivizes academics to join successful and visible universities.

10 Conclusion

UAE university libraries that are national are in developing RDM programmes, and progress is slowly happening. Some university library websites already reflect strides to acknowledge and harness resources for research data to be made available to researchers. The eFada initiative reflects awareness though not necessarily fast action. What needs

developing is systematic access to data resources that includes datasets from and about the UAE beyond national census data. Literature indicates that RDM concerns are shared regardless of where, geographically, the efforts are being made. Where local RDM projects exist, the libraries must take leadership in running and coordinating efforts in their universities.

The MBRSG Policy Council newsletter suggests a healthy enabling policy and infrastructural environment for RDM efforts. What needs to be organized are openly available procedures to follow for successful RDM efforts that are locally appropriate and practical. The fact that the world wide web is available to most parts of the world puts all researchers and innovators in the position of being able to share as well as use research data sets in multi and interdisciplinary dialogues. Therefore, it stands to reason that UAE university libraries must play their part in data capture, management, and encourage data literacy. This essay concludes that the fast transformation of academic libraries, whose use of resources that are global in nature and access, means that librarians need to be prepared to provide the requisite expertise to support users, and notes UAE applicability whenever possible. Regular repository self-study is also a necessary activity. Lessons obtained from experiences of other university libraries, e.g. those from Western-based institutes, together with local ones such as KUSTAR, ZU, UAEU, and international ones, can be put in place for local practical solutions that promote additional RDM services in the UAE.

References

ACRL (2019). *Keeping up with …research data management*. Retrieved from http://www.ala.org/acrl/publications/keeping_up_with/rdm

ACRL (2015). *Scholarship as conversation*. Retrieved from https://sandbox.acrl.org/literacy-concepts/scholarship-conversation

Association of European Research Libraries (2018). *Data librarian, expert on research data management, description, archiving and dissemination*. Retrieved from https://libereurope.eu/blog/2018/05/14/data-librarian-expert-on-research-data-management/

Becher, T. & Trowler, P. R. (2001). *Academic tribes and territories*. McGraw-Hill Education (UK). Retrieved from https://www.mheducation.co.uk/openup/chapters/0335206271.pdf

Bezuidenhout, L., & Chakauya, E. (2018). Hidden concerns of sharing research data by low/middle-income country scientists. *Global Bioethics, 29*(1), 39-54. https://doi.org/10.1080/11287462.2018.1441780

Boufarss, M., & Harviainen, J. T. (2021). Librarians as gate-openers in open access publishing: A case study in the United Arab Emirates. *The Journal of Academic Librarianship, 47*(5), 102425. https://doi.org/10.1016/j.acalib.2021.102425

Boufarss, M., & Laakso, M. (2020). Open Sesame? Open access priorities, incentives, and policies among higher education institutions in the United Arab Emirates. *Scientometrics 124,* 1553–1577. https://doi.org/10.1007/s11192-020-03529-y

Bryant, R., Lavoie, B., & Malpas, C. (2018). *Incentives for building university RDM services. The realities of research data management, Part Three*. OCLC Research Report, OCLC Online Computer Library Center, Inc.

Chigwada, J., Chiparausha, B., & Kasiroori, J. (2017). Research Data Management in Research Institutions in Zimbabwe. *Data Science Journal, 16*,.31. http://doi.org/10.5334/dsj-2017-031

Chiware, E.R.T. & Mathe, Z. (2015). Academic libraries' role in research data management services: A South African perspective. *South African Journal of Libraries and Information Science, 81*(2), 1-10. https://doi.org/10.7553/81-2-1563

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

Cox, A. M. & Verbaan, E. (2018). *Exploring research data management*. London: Facet Publishing.

Cox, A.M., Kennan, M.A., Lyon, L., Pinfield, S. & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation, Vol. 75* (6), 1432-1462. https://doi.org/10.1108/JD-12-2018-0211

Dataverse Project. (2019). *About the project*. Retrieved from https://dataverse.org/about

Elsayed, A. M. & Saleh, E. I. (2018). Research data management and sharing among researchers in Arab universities: An exploratory study. *IFLA Journal, 44*(4), 281–299. https://doi.org/10.1177/0340035218785196

Hammoud, J., Vatian, A., Dobrenko, N., Vedernikov, N., Shalyto, A., & Gusarova, N. (2021, November). New Arabic Medical Dataset for Diseases Classification. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 196-203). Springer, Cham.

HathiTrust (2017). *Welcome to HathiTrust*. Retrieved 18 March 2022 from https://www.hathitrust.org/about

Itani, A. & Östlundh, L. (2021). The role of institutional repositories in advancing open scholarship: A case study from the United Arab Emirates University. *The Journal of Electronic Theses and Dissertations, 1*(1), 3.

Johnson, S., & Potluri, R. M. (2020). Are the UAE academic libraries ready to support research 2.0? *Library Philosophy and Practice*, 2020, 1-18.

Khalifa University (2013). *Masdar Institute Joins Consortium of Libraries under UAE's 'Ankabut' Umbrella*. Retrieved 17 March 2022 from https://www.ku.ac.ae/masdar-institute-joins-consortium-of-libraries-under-uae-s-ankabut-umbrella/

Kelsky, K. (2018). The professor is in: Research first or teaching?. *The Chronicle of Higher Education*, January 29, 2018. Retrieved from https://www.chronicle.com/article/The-Professor-Is-In-Research/242371

Krahe, M. A., Toohey, J., Wolski, M., Scuffham, P. A., & Reilly, S. (2019). Research data management in practice: Results from a cross-sectional survey of health and medical researchers from an academic institution in Australia. *Health Information Management Journal*, *49*, (2-3), 108-116. https://doi.org/10.1177/1833358319831318

Lappalainen, Y. (2022). Supporting innovation with Open Access publications and data - ZU Scholars Institutional Repository [unpublished presentation]. Presentation at the *Dubai Expo 2020 in the Festival Garden*, February 15.

Larabi Marie-Sainte, S., Alalyani, N., Alotaibi, S., Ghouzali, S., & Abunadi, I. (2019). Arabic natural language processing and machine learning-based systems. *IEEE Access, 7,* 7011-7020. https://doi.org/10.14569/IJACSA.2018.090928

Lewis, J.A. (2014). *Research data management technical infrastructure: A review of options for development at the University of Sheffield*. Sheffield: University of Sheffield Library.

Marlina, E., & Purwandari, B. (2019). Strategy for research data management services in Indonesia. *Procedia Computer Science, 161*, 788-796. https://doi.org/10.1016/j.procs.2019.11.184

MBRSG Policy Council. (2015). *Research in the UAE: Present challenges and future opportunities*. Dubai: MBRSG Policy Council Series, Session No. 3. Retrieved 11 March 2022 from https://www.mbrsg.ae/getattachment/bbe59d16-6027-4c99-a785-dc09c47ae2b4/Research-in-UAE-Present-Challenges-and-Future-Oppo.aspx

Padilla, T. (2018). Collections as data: Implications for enclosure. *College & Research Libraries News, 79*(6), 296. https://doi.org/10.5860/crln.79.6.296

Peters, J. (2018). Shrinking horizons – or pushing boundaries? *Insights the UKSG Journal, 31*(1), 1C-9C. https://doi.org/10.1629/uksg.397

Petersen, M., Pramann, B., Toepfer, R., Neumann, J., Enke, H., Hoffmann, J., & Mauer, R. (2020). Research Data Management-Current status and future challenges for German non-university research institutions. *Research Ideas and Outcomes, 6*, e55141. https://doi.org/10.3897/rio.6.e55141

Potterbusch, M. (2018). What is Open Science, and how can radical collaboration facilitate it? *Research Library Issues*, *296*, 44–48. Doi: https://doi.org/10.29242/rli.296.6.

Puri-Mirza, A. (2020, November 3). *Number of higher education institutes in the United Arab Emirates (UAE) in the academic year 2013-2014 and 2016-2017*. Statista. https://www.statista.com/

ROARMAP. (2020). *Data visualisations*. Retrieved from: http://roarmap.eprints.org/dataviz2.html.

Scaramozzino, J. M., Ramírez, M. L., & McGaughey, K. J. (2012). A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries*, *73,* 4, 349-365. https://doi.org/10.5860/crl-255

Talib, M.A., AbuOdeh, M., Almansoori, A. & AlNauimi, A. (2015). Enhancing social science research in the UAE: An open source software solution. University of Sharjah (UOS) Case Study. *JCS, 11*(1), 98-108. https://zuscholars.zu.ac.ae/works/1505

Tenopir, C., Birch, B. & Allard, S. (2012) *Academic Libraries and Research Data Services. Current Practices and Plans for the Future.* Association of College and Research Libraries, Chicago, IL.

Tenopir, C., Allard, S., Baird, L., Sandusky, R. J., Lundeen, A., Hughes, D., & Pollock, D. (2019). Academic Librarians and Research Data Services: Attitudes and Practices. *IT Lib: Information Technology and Libraries Journal*, *1*, 1-37. https://trace.tennessee.edu/utk_infosciepubs/99

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research, 36*(2), 84-90.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Baird, L., Sandusky, R. & Allard, S. (2017). Research data services in European academic research libraries. *Liber Quarterly, 27*(1), 24-44. http://doi.org/10.18352/lq.10180.

UAE Government. (2015). Science, Technology and Innovation Policy in the United Arab Emirates. Abu Dhabi, UAE. Retrieved 9 March 2022 from https://government.ae/-/media/Science-Technology-and-Innovation-Policy.ashx

UAE Ministry of Education. (2018). List of licensed institutions in UAE. Retrieved 18 March 2022 from https://www.moe.gov.ae/En/MediaCenter/News/Pages/licensedInstitutions.aspx

United Nations. (2014). *Government for the future we want: United Nations E-Government survey 2014*. Retrieved 11 March 2022 from https://www.ica.gov.ae/userfiles/assets/BwvhnbUy.pdf

Yoon, A., & Schultz, T. (2017). Research Data Management Services in Academic Libraries in the US: A Content Analysis of Libraries' Websites. *College & Research Libraries, 78*(7), 920. https://doi.org/10.5860/crl.78.7.920

Xu, H. (2019). Challenges and Opportunities for Research Data Management in the Chinese Library Community. In *Proceedings of the 16th IFLA ILDS conference: Beyond the paywall - Resource sharing in a disruptive ecosystem*, held at National Library of Technology in Prague, Czech Republic, October 9-11, 2019

Zhou, Q. (2018). Academic Libraries in Research Data Management Service: Perceptions and Practices. *Open Access Library Journal, 5*(6), 1. https://doi.org/10.4236/oalib.1104693

Zou'bi, M.R., Mohamed-Nour, S., El-Kharraz, J. & Hassan, N. (2015). The Arab States. In *UNESCO Science Report: Towards 2030*, 430-469. Retrieved from http://unesdoc.unesco.org/images/0023/002354/235406e.pdf.