5-2019

DATA SHARING PRACTICES AND ATTITUDES OF SCIENTISTS IN THE GULF COOPERATION COUNCIL (GCC) COUNTRIES

Jimmy R. Malone
University of Tennessee, jmalone4@vols.utk.edu

DATA SHARING PRACTICES AND ATTITUDES OF SCIENTISTS IN THE GULF COOPERATION COUNCIL (GCC) COUNTRIES

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Jimmy R. Malone
May 2019

ACKNOWLEDGEMENTS

Many thanks and much gratitude to each member of my doctoral committee for your advice, guidance and patience with me as I travelled my path to complete this dissertation. To my committee Chair, Dr. Suzie Allard, I would not have completed this journey without your persistence and perseverance in keeping me motivated, and I thank you and the other members of the committee so very much.

ABSTRACT

Advances in science and technology over the last century have brought dramatic changes to most societies of the world, with a parallel increase in the amounts of research data being produced. Scientific progress in the Middle East geographic region of the world has, in general, lagged far behind Western countries during this same time period. Several Gulf Cooperation Council (GCC) Persian Gulf countries of the Middle East have recently made huge investments in developing their educational and research capacities, with the goal of establishing a culture and practice of scientific innovation. Several recent studies looking at data sharing and re-use among scientists in North America and Europe have insisted that sharing data is central to the goals of scientific progress. Using the *Diffusion of Innovation Theory* as a framework, this research looked at the data sharing practices and perspectives of scientists in the Middle East, through the lens of the four main elements of this theory: the innovation, communication channels, time, and the social system. The analysis of this phenomenon may provide a clearer understanding of data sharing as part of the emerging practice of conducting scientific research and its importance to the region. A mixed-methods research approach, using semi-structured interviews and an online survey, was conducted with scientific researchers in GCC nations. A separate analysis was conducted for the country of Qatar.

Keywords: Data sharing, data reuse, research datasets, Middle East, scientists, Persian Gulf, diffusion of innovations, Big Data, GCC

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTON

Data is the information at the heart of all scientific research. The practice of enabling scientific data sharing and re-use is a central tenant of doing good science. In the past few decades, scientific research has largely transitioned from the historical practice of lone researchers producing the majority of new discoveries, to the present time where science has become a data-intensive and collaborative process. This transition has been enabled by the fast pace of development of new technologies that allow for the capture of large data sets via sophisticated computing power. The increasing complexity of conducting scientific research projects, and the resulting large data sets, require researchers with various types of expertise to work together, and thus share data (Tenopir et al., 2011). The accumulation of data is growing at an almost exponential rate, and methods to organize, preserve and share this data have been undertaken. Many are following the Organization for Economic Cooperation and Development (OECD) international guidelines in this matter and working to make these data available in various repositories (Darby et al., 2012).

Some of the identified benefits of data sharing and re-use include: facilitating additional research discoveries from the same data through the use of alternative methodologies by researchers in other disciplines; increasing the rate of scientific discovery by removing the need to duplicate already-existing data sets; validating existing data results through re-use and replication of studies; and reducing the cost of research by the re-use of existing data (Tenopir et al., 2011). Some of the negative results from withholding access to data can have profound impacts on the next generation of scientists. The lack of access to data can prevent some

beginning researchers from pursuing a promising line of research, resulting in the necessity for the research to be dropped (Vogeli et al., 2006). Understanding the data-sharing practices of scientists is essential to help facilitate the progress of science.

Study Focus

With the advent of nascent research cultures in the Middle East, there comes the question of how these scientists will share their scientific data. Scientific progress in the countries of the Middle Eastern geographic region of the world has significantly lagged behind the efforts of the Western countries (Kassir, 2006). In the past couple of decades, the countries of the Middle East have begun to use their vast natural resource wealth to begin the process of kick-starting the transition to knowledge economies, or at least to initiate scientific research projects and educational reforms aimed at bringing the individual nations, and the region, into the mainstream of modern scientific and societal progress. The establishment of well-funded science and technology parks and Western-style research universities are examples of the steps being taken by the governments of these developing countries to accomplish these goals (Adams, 2011). These efforts to increase the amount and level of scientific research will naturally result in increasing amounts of research data being created.

To understand the data-sharing practices of scientists, and its implications for scientific progress in general, this study looked at the various research institutions in these countries to determine the current attitudes of their scientists toward data sharing and how this is accomplished. It is hoped that this research will be meaningful, and the resulting data will be of use to the institutions at which scientists were surveyed, in order to inform future policy decisions aimed at improving the quality of scientific research through data sharing.

In the last few years of scientific development, the disparities in scientific progress between the Western countries and the developing regions of the Middle East continue to be vast, but the rapid growth and proliferation of communication and information technologies, and investment in research and education are catalysts for change (UNESCO, 2010).

Goldemberg (1998) points out that one of the misconceptions of developing countries is that they should emulate the highly developed nations, such as the United States, in pursuing large research programs in areas such as nuclear science and satellite technology. He insists that a wiser use of research resources would be to target areas of need that are in alignment with national needs and requirements. In addition to solving urgent needs, this approach would also bolster the educational system in the countries. This is echoed by Harris (2004), who insists that developing nations must make sincere efforts not only to transfer technology to their countries, but also to assist with the education, training, and information needs of the local scientists and researchers. Part of this education and training must include the importance and impact of data sharing on the development of young scientists (Vogeli et al., 2006).

Many recent studies looking at data sharing and re-use among scientists in general (mostly studies conducted using European and North American populations) have insisted that data is central to the goals of science. The benefits of data sharing have been explicated to include: (1) the fulfillment of the credo of science to make data and results available to the public (especially when the research has been funded by the public), (2) the facilitation and inclusion of new interpretations using alternative methodologies, (3) the assistance data sharing provides for the progress of scientific discovery, and (4) the validation of existing research through the re-use of the data (Tenopir, 2011; Borgman, 2010).

In formulating a theoretical or conceptual framework for studying how developing countries in the Middle East share data, the Diffusion of Innovation Theory (DOI) serves as a useful model. According to Rogers (2003, p. 5), "Diffusion is the process in which an innovation is communicated through certain communication channels over time among the members of a social system. It is a special type of communication, in that the messages are concerned with new ideas." This newness leads to a certain amount of uncertainty and the need to acquire additional information. Information would include any content that would help an information seeker to resolve the uncertainty. There are four basic elements to the DOI theory: (1) the innovation, (2) communication through various channels, (3) the time dimension, and (4) the context of a social system (Rogers, 2003). This can include the exchange of new scientific knowledge creation between scientists in the region of study.

An *innovation* can be any new idea, method, tool or technology with the potential to impact the status quo of a social system (Rogers, 2003). In the context of this study, the need for research data sharing brought on by the era of big data in science can be thought of as a new idea, enabled by new information and communication technologies (ICTs), and one that is having a definite impact on science and societies in general.

Communication is understood to be the method by which two individuals or agents share information through various channels (Rogers 2003). Communication happens in various ways, primarily through either mass media or interpersonal communication channels. Mass media channels are usually where knowledge about an innovation is first disseminated. The mere knowledge of an innovation alone is not enough to convince a potential adopter to accept and use the innovation. Most adopters will be influenced by change agents through interpersonal communications. Diffusion communications have the distinction of containing some degree of

heterophily, and a defining trait of diffusion communications. Heterophily refers to the degree to which two individuals or change agents are dissimilar in various traits, such as cultural beliefs, education level, or social standing, etc. Most human communication occurs between individuals who are homophilous, thus facilitating more effective communication. Homophily is the opposite of heterophily and describes the degree to which individuals or change agents share similar traits, such as beliefs, education level, or social standing. These heterophilic communications naturally lead to some difficulties in permitting change agents to affect potential adopters positively (Rogers, 2003). In this study, the various communication pathways used by scientists to share research data are an important element for understanding this phenomenon in the GCC.

Time has a role in diffusion research by differentiating several types of adopters and for categorizing the rate of adoption of the innovation. According to DOI, adopters can be grouped into **five categories: innovators, early adopters, early majority, late majority, and laggards (**Rogers, 2003). Innovators are the first to use an innovation, and if their experiences are positive, they can influence a group of early adopters to try the innovation. With the increased success of an innovation, a large segment of users, known as the early majority, will begin to use the innovation. When an innovation has become almost mainstream, the late majority will begin to use the innovation. Eventually, the remaining number of users that were reluctant to use a new innovation, the laggards, will adopt the innovation. This adoption pattern is graphed as an "S-curve," with the graph resembling an 'S' lying on its side. The rate of adoption can be thought of as the time frame it takes for a social system to adopt an innovation, from the first time knowledge is gained about the innovation, until its full adoption is confirmed (or rejected). The time element of the DOI theory was of particular interest in this study, because this study

established a "snapshot" view of the current prevalence of data sharing within the scientific communities of the Persian Gulf, at this point in time.

Social systems are defined as an interrelated group of individuals or agents within a system with common goals. These systems are generally defined by behavioral norms and have a tendency toward maintaining the status quo. In DOI Theory, those individuals or agents within a social system who can consistently influence others, either positively or negatively, to adopt a new innovation are considered to be gatekeepers or opinion leaders (Rogers, 2003). The social system in this study is represented by the scientists and researchers in the GCC countries, and they share commonalities relevant to this study: geographical proximity, history/culture, investment in Western-style higher education institutions, a recent focus on developing a culture of research, and their efforts to help their nations transition from oil/natural resource extraction economies to becoming knowledge societies (Gremm et al., 2017).

The principles of DOI Theory have been widely applied in research in various disciplines, such as communications, sociology, marketing, medicine and other fields. Kedia and Bhagat (1988, p.561), developed a model incorporating Diffusion of Innovation to help explain the effect of cultural influences on the transfer of technology across national borders. This model suggests that cultural-based differences in the individual societies can have a moderating effect on the adoption of new technologies or innovative ideas.

Although the emphasis of DOI Theory would appear to be on the communication aspects of the diffusion process, there is at the heart of these processes (the innovation and communication through channels) a central role for information. Data sharing is primarily a sharing of new information among scientists and other researchers. When scientists are first exposed to the concept of sharing data, this newness leads to a degree of uncertainty on the part

of a potential adopter and triggers the need to acquire additional information. Information would include any content that would help an information seeker to resolve the uncertainty. Information theory includes several well-documented variations on information-seeking behaviors. Uncertainty is a primary stage in the information-seeking process detailed by Kuhlthau (2004), among others. This information perspective of data sharing in developing countries certainly aligns with DOI Theory.

In light of the converging environmental, cultural and economic phenomena related to the practice of scientific research by the region's scientists, this study focused on the data dimension, guided by this over-arching research question:

RQ: How and why are research data being shared by scientists in the GCC countries of the Middle East/Persian Gulf?

This research project employed a mixed-method (qualitative and quantitative) research approach to explore the data sharing practices of scientists of the GCC. The two methods used in investigating this research focus were semi-structured interviews and an online survey. This combination of the use of qualitative (interviews) and quantitative (survey) methods provided a more balanced and holistic approach to the research. The data generated from these methods of conducting research, allowed for a depth and breadth of insight that could not be obtained if either approach were used alone.

CHAPTER TWO
LITERATURE REVIEW

Big Data

Before perusing the literature on data sharing, it is necessary to understand why the sharing of research data by scientists has become an important issue. This requires a quick dive into the topic of "Big Data." A search of the Web of Science (WoS) citation database for "big data" in the title of articles produced over 9,700 results in the time period of 1900-2018 (Figure 2.1). Although there were a few publications in the early 2000s, the publication numbers ramped up exponentially within a five year period, beginning with the year 2012. As will be noted in the following section on data sharing, these two topics share a very similar publication trajectory.

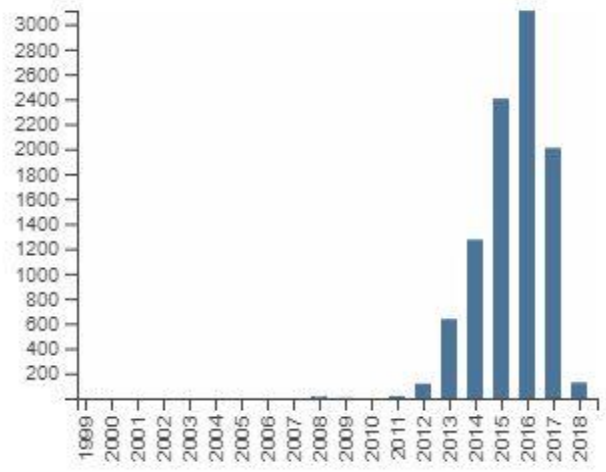

Figure 2.1: "Big Data" - Total Publications by Year (1900-2018)

In a highly influential talk on the topic of *e*Science at the National Research Council-Computer Science and Telecommunications Board (NRC-CSTB) meeting in 2007, Jim Gray outlined the basic components of what he referred to as "The Fourth Paradigm" of scientific exploration (Hey 2009). This refers to the historical development of scientific paradigms over the last several thousand years through four distinct phases:

- Empirical approach - individuals gathering empirical data on natural phenomena (thousands of years ago)
- Theoretical approach - theory development using models for generalizations (Last few centuries)
- Computational approach - allowing for simulation of phenomena (last few decades)
- Fourth Paradigm (eScience) is computationally intensive data capture by sensor instruments, processed by computer software, storage in a computer, and using software for data analysis (present day)

  (Hey 2009).

This new paradigm in science has only been possible because of tremendous changes in information and communication technologies over the last several decades. The output of data from this new way of conducting science has led to the era of Big Data. The definition of Big Data varies depending on which community of experts you ask. Initially thought of as any dataset or information exceeding the memory capacity computers used for processing (Mayer-Schonberger and Cukier 2014), this definition does not hold since more computing power is being packed into smaller machines over time. MIT scientist, Alex Pentland, describes big data as "the newly ubiquitous digital data now available about all aspects of human life," and includes

structured and unstructured data (Pentland 2014, p.8). This speaks to the increasing volume of data being created by corporate and personal online transactions, and geospatial data creation, and is a preoccupation of much academic and economic research. There is not one generally accepted definition of big data across technology practitioners, research scholars and data analysts. One of the simplest big data definitions may be "where data collection has grown tremendously and is beyond the ability of commonly used software tools to capture, manage, and process within a 'tolerable elapsed time'" (Wu et al 2014, p.97).

There are hopes that big data will assist with increasing efficiencies and insights in healthcare, government, manufacturing, and retail (Chen et al 2014). An example of the use of big data to assist with a public health crisis was during the H1N1 flu pandemic of 2009. Hospitals could not provide information to the CDC, and other disease prevention centers, on a timely basis to assist with determining the geographic spread of the disease, since most people could be infected for 1-2 weeks before presenting with symptoms. Google used specially developed algorithms to analyze flu related search terms by the public, scouring massive amounts of data from prior years to help predict where the flu may spread geographically, based on the location of people searching Google for related terms (Ginsberg 2009). There appears to be many opportunities to use big data for solving practical and theoretical questions in all research domains.

Some of the challenges for using big datasets involve the "4 Vs": volume, variety, velocity and veracity. The increasing amounts (volume) of datasets, with the various internal structures of homogenous and heterogeneous data (variety), produced more quickly with each passing year (velocity) are making it more difficulty to ensure the accuracy (veracity) of big datasets (McCulloch 2013). All these aspects of the data make extracting useful information and

knowledge a difficult task, but with continued innovation in IT capabilities there is reason to believe any challenges will be overcome.

The big data phenomenon has changed the way science is being done. The research infrastructure has begun to evolve over the last couple of decades and no longer is geared toward individual scientists producing data and resultant publications. Data intensive research now requires collaboration of multiple researchers and sometimes multiple research centers. The IT infrastructure has been evolving also, from centralized data centers, to a mix of centralized and distributed nodes of data and must continue on this path to support this new paradigm in science (Hey 2009, p.112).

Data Sharing

Publication Trends

This section of the literature review looks at the various aspects of data sharing and re-use in more detail. To begin this discussion, a look at the correlation between "big data" and "data sharing" in the published literature shows they have similar/merged publication trajectories after a specific point in time, but have divergent histories leading up to that point in time.
A search of the Web of Science (WoS) citation database to find literature related to the phrase "data sharing" in the title resulted in a total of 1,565 citations (Figure 2.2). After analyzing these for relevance to data sharing of research datasets, a smaller and more specific subset of over 300 items was produced. These include articles, books, proceedings and other items with this term listed in the title. Although the entire backfile (1900-2018) was searched, the first mentions of research data sharing appeared in the literature in the years 1971 (Kliemer) and 1981 (Williams) discussing energy datasets and seismic datasets, respectively. Non-relevant citations found in this search prior to the year 2000 tended to be focused on how electronic data was shared between

computer systems or databases. Relevant citations began to substantially increase in the year 2000 and subsequent years. There has been a corresponding increase in citations of this literature during the same time period. The selected approach was taken to target the most salient items.

It is of interest to note that "data sharing" began showing up more regularly and in greater numbers as a topic in the peer-reviewed literature around the year 2000, about a decade prior to the topic of "big data." This indicates that data sharing has been of concern to scientists prior to the advent of the big data phenomena. When big data articles began appearing in peer-reviewed publications, the numbers skyrocketed within a couple of years. At the exact same time, there was a corresponding leap in the number of publications on data sharing. This could indicate that the interest in big data is driving an increased interest in data sharing in general; and data sharing is seen as even more necessary due to the increasingly data intensive nature of research and the need for collaboration between researchers.

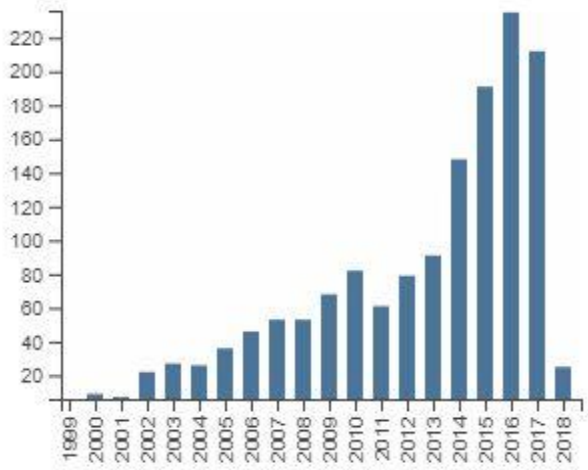

Figure 2.2: "Data Sharing" - Total Publications by Year (1900-2018)

Data Sharing Opportunities and Challenges

There are various aspects of data sharing and re-use detailing the opportunities and challenges of this endeavor. Following is a discussion of the status of data sharing in specific scientific disciplines, and drivers of data sharing – including policies of funding agencies and publishers.

The usefulness of data sharing can be seen even in the distant past, in the example of astronomer Nicholaus Copernicus, who developed his heliocentric universe theory, with the help of using data collected by others (Packer 2016). The ongoing and ever increasing importance of data sharing is evident in the scientific literature.

Once research data has been collected, curated, analyzed, and preserved, it can be made available for sharing or re-use. As previously mentioned, the benefits and advantages of data sharing have been stated to be part of the fulfillment of the central purpose of science and allow researchers to:

- make data and results available to the public (especially when research has been funded by public funds),
- allow for new interpretations using alternate methodologies,
- assist with the progress of scientific discovery, and
- validation of existing research through re-use of the data

(Tenopir 2011, Borgman 2010).

In the first survey of astrobiologists, Aydinoglu et al (2013), found the surveyed participants acknowledging the benefits of sharing data (including re-use of data for validation of results and avoiding redundant data collection) and some common reasons this did not happen, including

lack of resources (time and money) and lack of technical skills. The more interdisciplinary research becomes, the more complex the challenges of sharing data.

Sharing of ecological data by scientists involves many of the familiar challenges in other scientific disciplines. Although most ecologists and environmental scientists are willing to share their data, they generally do not have the skills or training necessary to do so (Michener, 2015). Agricultural data collection and storage is fragmented, with no real standards across the discipline; in addition, most data are collected using paper field notebooks, and may not be entered into a digital format for months or years, thus delaying analysis and any possibility of sharing of data (Diekmann, 2012).

Attitudes Towards Sharing Data

Ceci (1988) conducted one of the early surveys on the attitudes scientists have toward data sharing, showing that scientists indicate that they agree that data sharing is a hallmark of science and should be done, while also indicating that 59% said their colleagues had refused to share data. The most common reason scientist gave in this survey for not sharing data was the fear of another researcher publishing their results ahead of the original researcher. In addition, scientists in the biotechnology fields expressed fear of financial loss of possible patent rights or future funding if they did share data. These survey results have been echoed years later in additional studies (Campbell et al, 2002; Piwowar et al, 2007; Wicherts et al, 2006).

Significant follow-up research around scientists' perceptions and attitudes toward data sharing would not be conducted again till the 2000s, although smaller, discipline specific research did occur. In an international survey of 1,329 researchers, the main reasons given for not sharing data electronically were a perceived lack of time to share, and a lack of funding to support this activity (Tenopir et al 2011).

In the biomedical sciences, many researchers are concerned with sharing their data for various reasons, including: fear others will take their data and publish before they have a chance to do so; fear others will use their data to questions the original researchers' findings; and concerns over others misinterpreting their data (Federer et al, 2015).

Individual motivations toward data sharing were identified by Kim and Stanton (2015), and include positive associations of perceived career benefit and scholarly altruism; while perceptions of career risk and perceived effort (time) had a negative effect on data sharing by individual scientists.

The attitudes of scientists toward sharing their primary research data is similar in many respects across many disciplines, although there are specific reasons held by each discipline, with the most resistance to sharing held by scientists in the biomedical fields. The challenges of sharing data from a practical stand point are also similar, with most indicating more time and resources are needed to accomplish the tasks.

Data Sharing Policies

Funding agencies' policies governing the sharing of data produced during funded research are becoming common place. In 2011, the National Science Foundation (NSF) in the United States implemented their data sharing policy, and began requiring a data management plan from all funded researchers and researchers seeking funding, detailing how they would manage the study data including how they would share primary data with other researchers (NSF, 2018). Soon afterward, the National Institutes of Health (NIH), responded to a White House Office of Science and Technology Policy (OSTP) memorandum directing all federal agencies and offices to make research data resulting from federally funded research available to the public and the scientific community of researchers (NIH, 2018) by requiring all NIH funded

research by its' many sub-agencies be made available to the public. In response to the same memorandum, the National Aeronautics and Space Administration (NASA) put forth their own plan, detailing the requirements of researchers funded by that agency to share their research data with others as their research results were published or soon afterwards (NASA, 2014).

Outside of North America, other governmental bodies are also pursuing policies to promote the sharing of publicly funded data. The Australian government is coordinating with public and private industries to develop data sharing guidelines and legislation, to make research data accessible to the public and researchers (DPMC, 2018). This trend toward policies requiring data sharing is also seen in Europe, with the European Commission's funding through its Horizon 2020 project mandating open access to research data (EC, 2018) as well as resultant publications.

The effectiveness of journal publisher policies to increase data sharing is not yet quantified and the policy is not yet uniform across publications in each domain. In a recent study, only 21% of biomedical journals stated explicitly in their policies, that data sharing was required, while 32% made no mention of data sharing at all (Vasilevsky et al, 2017). In Savage and Vickers (2009), results indicated that only one out of ten researchers provided raw data sets when it was requested, even when the journal policy required the data to be shared. More recently, literature suggests there is a growing importance of journal publishers and their data sharing policies in ensuring researchers share their primary data along with their journal articles (Tal-Socher and Ziderman, 2017). The American Geophysical Union (AGU) is one of the largest Earth and space science publishers, with a data policy that states all available data in their publications must be made available if possible. In 2018, the AGU "along with a large number of other Earth and space science publishers endorsed a new Commitment Statement of the

Enabling Fair Data Project. This project includes a common set author guidelines and FAQ's around data, software, and samples (AGU 2018)."

Supporting broader data sharing practices has also been influenced by the creation of data journals where data can be published as a separate type of article describing a data set; and publisher funded repositories, where researchers can deposit their primary data sets, once they have published their research results (Ware and Mabe, 2015).

The peer-reviewed literature does not contain any mention of the Diffusion of Innovation (DOI) theory being applied to explain data sharing (of data sets) in the larger scientific community. This suggest the current research project is the first to do so. A search of the WOS database for the concurrent terms "diffusion of innovation" and "data sharing" did not produces any meaningful results (no limits on year range in the results). DOI has been used to explicate the way knowledge and information is shared in many areas: technology adoption, health care, business management, and social networks; and still can be used to explain new areas of scientific inquiry.

Cultural Context

According to Vickery (2000), the Arab countries of the Middle East have a long history of scholarship dating back to the seventh century, including the development of algebra. Over the centuries, centers of learning have grown and subsided in the region. In recent years, the vast oil and natural gas wealth of the modern countries in the Persian Gulf region forms the basis of the shared economic and political interests of the Gulf Cooperation Council (GCC) founded in 1981 (The World Bank, 2018). Currently, there are six members of this council: Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and United Arab Emirates (UAE).

These countries are using their vast wealth to strive to transition from oil economies to knowledge societies (Gremm et al 2017, p.164) via large investments in science and education to promote a culture of research. Qatar in particular is embracing this path, as laid out in the Qatar National Vision 2030 (GSDP 2008), the first pillar of this strategic plan is to build a world-class education system to meet the needs of their rapidly advancing society..

To take full advantage of this push for better education, especially higher education, it is important to understand the current situation in the education arena in the Middle East in general, and the GCC region specifically. Hassan Diab, Vice President of the American University of Beirut, addressed the future of higher education in the Middle East during his 2015 speech to the 50th Anniversary Meeting of the International Association of University Presidents, indicating the need for improvement in the entire K-12 as well as university education system (Diab 2016),. Diab states that it is common for high school graduates in Middle Eastern countries, and in particular the GCC countries, to require remedial courses in math and science when beginning their Freshman year university studies. Focusing on the education of the youth must be strong for the push for higher education to be successful. Many Arab countries in the Middle East and North Africa (MENA) region have decided on non-traditional ways of meeting the need for quality educational opportunities, through partnerships with transnational universities.

Beginning in the first decade of the twenty-first century, there were approximately 100 branch campuses world-wide operated by universities in countries outside their country of origin (host countries), and more than one-third were located in the MENA region (Miller-Idriss and Hanauer 2011). The UAE had a particularly high number of these foreign-operated educational institutions (n=35), and Qatar with eight. A general surge in the numbers of universities has been witnessed in the MENA region over the last several decades, with over two-thirds of the

existing universities in the region having been founded since the 1980s. The GCC countries were among the last to participate in this surge in higher education, but have now become major proponents of various higher education institutions. Saudi Arabia alone went from having eight universities to over 100 universities from 2003-2009, with an annual budget over $15 billion USD (Romani 2009). Currently, Saudi Arabia has is seeing one of the fastest growths of their education system among all the countries of the world (Abouammoh 2018, p.327)

This rapid growth of educational endeavors in the GCC has had an effect on the breadth of scientific research being conducted there, and is often driven by national priorities. Most GCC research institutions are participating in international collaborations with experienced researchers to promote a culture of research. Medical education is a priority for Saudi Arabia, UAE and Qatar, with each of these countries investing billions of dollars to build medical and other research education facilities and to train physicians and other scientists. This has led to an increase in research articles being produce from these countries, but medical research still lags behind other regions of the world (Meo et al 2015). Renewable energies research is also of great interest, with a high potential for solar, wind and ocean wave energy generation in the GCC region. These renewables are becoming more attractive to the governments in the region for various reasons: uncertainty of economic global markets and dealing with the effects of pollutants related to oil and gas exploration and production (Al-Maamary et al 2017). This focus on science, education and research in the region will inevitably lead to the growth of research data

In the field of Science and Technology Studies (STS) it is believed that scientific knowledge and its' products are made and not discovered, through the social activities and conditions in which scientists and engineers find themselves (Erickson 2005). Scientific

research is central to solving problems in society, such as risk mitigation, environmental issues, infectious disease tracking or sustainable development of resources; and the ability to do so is becoming more dependent on access to quality data produced by academic and other research communities. Researchers in many developing countries and emerging economies are generally lacking in data sharing policies and formal mechanisms for openly sharing of their data (National Research Council 2012). An understanding of the current data sharing landscape in the GCC is therefore needed to assist researchers and society in finding ways to overcome these hurdles, and further the scientific research culture that is rapidly developing in the region.

CHAPTER THREE

METHODS

This research project employed a mixed-method (qualitative and quantitative) research approach to understand the data sharing practices of scientists of the Middle East. The methods used to investigate this research focus were semi-structured interviews, online survey, and a case study. This combination of the use of qualitative (interviews) and quantitative (survey) methods, along with a case study, provided a more balanced and multi-pronged approach to the research. The data generated from these methods permitted the researcher to explore the research question effectively and attempt to explain them (Creswell, 2009). Overall, this approach allowed for a depth and breadth of insight that could not be obtained by using either method alone.

Population

The target population for this study included any researchers in member nations of the Gulf Cooperation Council (GCC), which includes the Arab states bordering the Persian Gulf (except Iraq), that are engaged in scientific research, at organizations that are oriented to a Western-style higher educational model. The countries that were included in this list are delineated in Figure 3.1 below. The specific countries of interest were: Qatar, United Arab Emirates, Saudi Arabia, Bahrain, Oman, and Kuwait. These six countries have each invested substantial amounts of resources into developing research capabilities and infrastructure to support scientific research that will meet international standards.

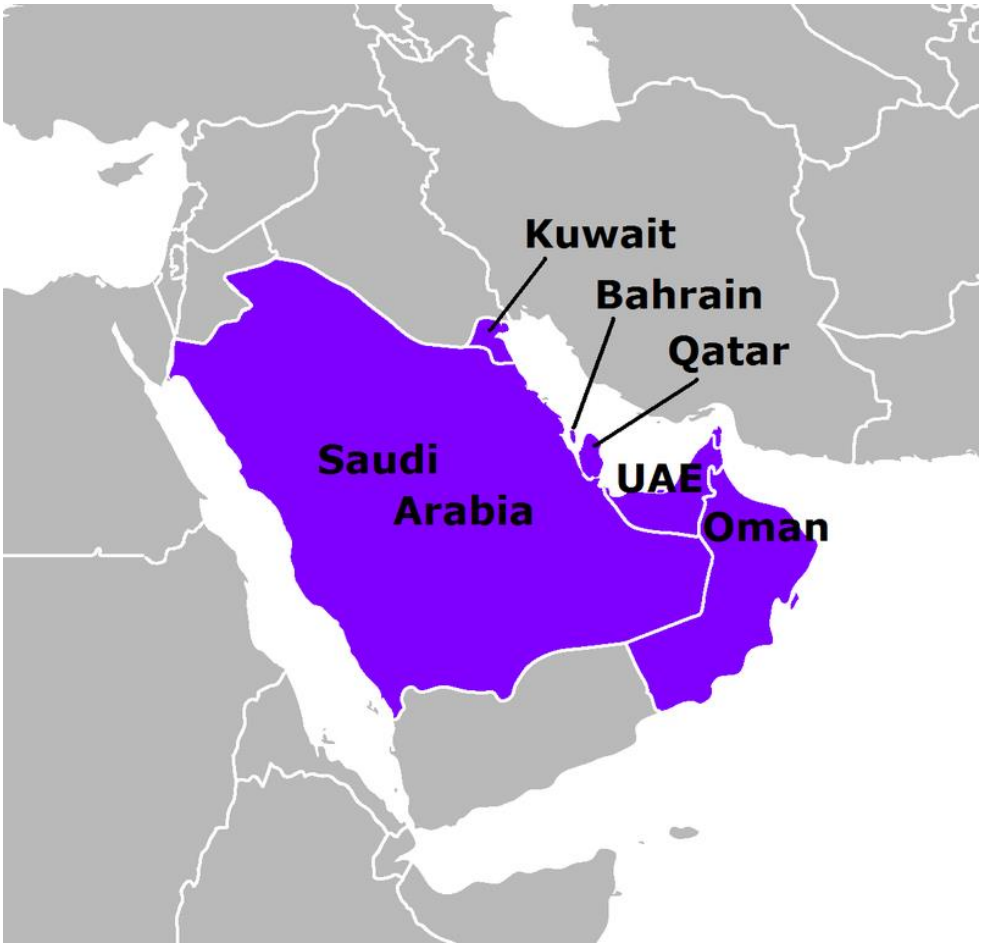

Figure 3.1: Gulf Cooperation Council (GCC) Countries (WorldAtlas, 2018)

The *UNESCO Science Report: Towards 2030* (UNESCO, 2015) gives an idea of the total number of researchers in each of the UNESCO Arab region countries, including most of the Persian Gulf countries which were the focus of this study: Qatar, Kuwait, Oman, and Bahrain (Table 3.1). The Saudi Arabia numbers are from the *UNESCO Science Report 2010* (UNESCO, 2010).

Sample and Sampling Method

The interview study's sample was drawn from the population defined above. The initial interview participants were identified via a convenience sample in the country of Qatar, using personal/professional contacts of the researcher, from locations where the researcher previously worked and has connections to the scientific communities there. A snowball method was used to identify additional participants after the researcher was in the field. The culture of this region of the world values personal connections, therefore using a snowball sampling technique that began with people that are personal/professional contacts of the researcher was a sensible and successful approach.

For the survey, the researcher obtained the contact information and email addresses of individual scientists and researchers that were identified using existing trusted scientific and academic networks. A comprehensive list was constructed via these scientific networks, and the survey was sent to all of these identified scientists with sufficient contact information. A list of the known source networks is provided in Table 3.2 below.

Table 3.1: GCC Number of Researchers

Country	Researchers per Million Population	Total Number of Researchers
Qatar	587	1,203
Kuwait	135	439
Oman	137	497
Bahrain	50	67
Saudi Arabia*	42	1,033
UAE*	n/a	n/a
	TOTAL POPULATION	3,239

*Data from the UNESCO Report 2010 [not available for the 2015 report]

Table 3.2: Scientific Networks in the Middle East

Organization	Description	Note
Arab Science and Technology Foundation (ASTF)	ASTF is an independent, non-profit non-governmental organization (NGO), which works regionally and Internationally to encourage investment in Science and Technology. Composed of an elite cluster of renowned Arab scientists who have outstanding scientific achievements known at the national and international levels.	
The World Academy of Sciences (TWAS)	Focused on the advancement of science in developing countries – works to support sustainable prosperity through research, education, policy and diplomacy. TWAS is a global science academy based in Trieste, Italy, with regional offices, including one in Alexandria, Egypt.	
Qatar Foundation (Education City)	The umbrella organization in Qatar that incorporates Education City and its universities, and QNRF – the research arm of QF.	Researcher has personal connections at most of the 8 major universities
KAUST University (Saudi Arabia)	The preeminent research university in Saudi Arabia, with a Western-style higher education structure.	Co-workers on my job have personal connections with faculty and arranged introductions. I also used other professional contacts via LinkedIn to tap into various Saudi universities.
Kuwait Universities		Professional connections in Kuwait provided assistance with connections
UAE Universities		Used personal contacts at a couple of the universities, and professional connections at others.
Oman		Professional connections via LinkedIn at Oman universities
Bahrain		Professional connections via LinkedIn at several Bahraini universities

Data Collection Instruments

The interviews and survey were both conducted using the English language. An IRB application was completed for this study and was approved by the University of Tennessee, Knoxville's Institutional Review Board (IRB) prior to the start of the interviews.

Semi-structured Interviews

A qualitative research study focuses on asking and answering research questions. Two types of questions are important: a central question which is broad in scope and explores the main concept or phenomenon of the study and supporting questions that hone in on specific dimensions of the phenomenon. In the current study, the central phenomenon is the data sharing practices of scientists in the GCC/Persian Gulf countries. Each central question was followed up by several sub-questions that helped to narrow the topic into more specific information but were still left sufficiently open-ended to allow for effective exploration of the topic (Creswell, 2009). For example, when an interviewee was asked to discuss their thoughts on sharing data with others and a simple "yes or no" was provided, a follow-up question to have the researcher to describe their process of sharing data or who they were willing to share data with would be asked. Concepts and categories that arose from the interview transcripts were identified; these concepts and categories were then used to better understand the data gathered in the survey, which was the second instrument to be used in this research project. Each interview was scheduled for a one-hour time period, with most interviews taking from thirty minutes to one hour to complete; a couple of interviews lasted slightly longer than one hour.

Survey

A cross-sectional survey was used to gather data for this research project. This type of instrument was chosen because it provided the ability to collect a large amount of data in a quick, economical and simple way, especially given the geographically-dispersed nature of the population, which included the Persian Gulf countries of the Middle East. The survey questions were structured to provide an understanding of participants' practices and attitudes regarding the constructs at the center of this survey: the respondents' attitudes toward scientific data sharing and how their data sharing is done. The instrument contained self-reported, closed-ended items, open-ended items and demographic questions.

The survey used for this research was the *Scientists Second Follow-Up-Champions* survey distributed by the DataONE project (DataONE 2018a). DataONE (Data Observation Network for Earth) is a collaborative cyberinfrastructure project between various academic, governmental and other member institutions that provides access to datasets in the earth and environmental sciences disciplines from member repositories (DataONE 2018b). This survey was chosen as the instrument of choice because it contained virtually all the relevant questions (24 total) of interest to the researcher and was an already created and vetted instrument. Using a proven instrument also allows for the possibility of comparing the results of this survey with those from scientists in other settings, as a foundation for subsequent research on the topic.

This survey was created and ready for distribution beginning the month of April 2018. After the University of Tennessee (UTK) IRB approval process was complete, the researcher began initial distribution of the survey April 10, 2018 via an email distribution list, and continued with follow up reminder emails for the next four weeks, ending in early May 2018.

Case Study

A case study approach was also used to provide further insight into the data from this mixed-methods research approach. Case studies allow for a researcher to provide an in-depth analysis of a unit of measure in the research design, that is bounded by specific criteria (Creswell 2014). The unit of measure for this case study focused on data at the country level, which provided an opportunity to look at one country in-depth. After the initial analysis of the survey and interviews was completed, it was noted that a large percentage of the data was a result of data collected from the country of Qatar. As a result, a case study analysis was performed for the country of Qatar; this provided a means to understand the data sharing practices and attitudes of scientists in this individual country.

Data Collection Procedures

The semi-structured interviews required that the researcher travel to the Middle East/Persian Gulf country of Qatar to have access to a sufficient number of participants. Face-to-face interviews were conducted with each interview participant. The need for primarily using face-to-face interviews was a result of needing to build trust with the interviewees, since the culture there is built on personal connections and in-person interactions. Several interviewees noted that they would not have agreed to be interviewed if it had not been in-person. Even in those instances where the researcher could not travel to a country included in the study, spending time in the Persian Gulf and having a temporary base there was very helpful in having a closer proximity to the participants. Respondents were more likely to participate with a "local" researcher than with someone thousands of miles away in the USA.

A digital (electronic recording device) copy was obtained for each interview. In the event travel to any of the countries of interest was not possible for any reason, telephone or Skype interviews was used to collect these data.

The survey was administered via the UTK online survey software Qualtrics. A link to the survey was provided and the requests for participation were distributed via email solicitation. The survey included twenty-four questions related to demographics, familiarity with the DataONE project, research data, and views on data use and reuse. Except for the questions regarding familiarity with DataONE, all other questions were relevant to the current study and used for analysis.

An attempt was first made to derive an email distribution list from a list of contacts gathered from the institutions that are physically located in the countries in the Middle East/Persian Gulf. These contacts were ultimately identified from a list of researchers and scientists in the ProQuest Pivot database of curated profiles, limited to the countries in the GCC region.

Pivot contains millions of research profiles from around the world and allows for the easy identification of scientists and researchers by country. For reasons unknown, Bahrain researchers were not included in this database. Also, the DataONE team distributed the survey globally using a separate distribution list, and the survey link could be shared by recipients with their colleagues. The researcher also posted a link to the survey on LinkedIn, to expose the survey opportunity to a wider audience, so it is not possible to determine the total number of researchers that received the survey in these latter snowball distribution channels.

Through this method, a list of 5,642 potential survey participants was compiled for the six countries of the GCC region. After the initial invitation to participate in the survey was sent out via email, there was a large amount of emails that were undelivered for various reasons: the email address was no longer active or the institutional email server rejected the email because it was considered spam. This resulted in an estimated 2,253 emails not being delivered. Two

follow up messages were sent to the email list at one week intervals to prompt additional responses. The number of emails that were presumed to have ultimately been delivered successfully equals 3,389.

Data Analysis

Interview data (recordings and transcripts) were stored on the researcher's password-protected laptop computer. A secure cloud service (Google Drive) was used to ensure automatic back-ups of the interview data files to an online server. This cloud data storage service is also password protected. Each completed interview was transcribed into a typed text transcript from the digital recordings. Transcripts were analyzed according to accepted qualitative methods, that identified emergent concepts and categories that were revealed across the set of interviews (Miles, Huberman & Saldana 2014). Although the interview questions were grouped by two general categories (how data is stored and why data is shared), the initial coding cycle of the transcripts was an open coding approach; taken to allow specific ideas or topics to emerge into initial codes that were descriptive of the what the interviewees were attempting to convey. In this way, codes were applied to sentences or groups of words. A secondary coding cycle, using axial coding, allowed the many disparate and initial codes to be pulled together into more inclusive or broader categories (Saldana 2016). These categories assisted with understanding how the interview data related to the main research question of data sharing. The data were coded manually, using Microsoft Excel software to help structure the text of the interviews into categories. In accordance with accepted qualitative research practice, interviews were conducted until a saturation point in the resulting data was reached, and no new additional concepts or categories were surfacing.

Data from the survey portion of the study was collected and stored via the UTK survey software server (Qualtrics). Qualtrics aggregates the data and allows for export to Excel, SPSS or other software for analysis. The UTK DataONE team were in control of the survey data collected with this survey instrument. Mention why? Access to the survey data was limited to the researchers only, who could access it via a password- protected account. Storage of the data set on the server ensured the existence of a reliable backup copy of the data, in accordance with good practices of data management. A copy of the final raw DataONE survey dataset was downloaded from the Qualtrics server, and emailed to the author of this document, where it was saved on the researcher's individual laptop, accessible via a password known only to the researcher, and a backup was saved to a cloud storage account (Google Drive).

The raw data set contained survey results from respondents world-wide, therefore it was necessary to clean the data and remove any responses not relevant to the GCC. Standard descriptive statistics were used to analyze the survey data and were conducted using the SPSS v.25 software provided by UTK. The analysis of the survey data focused on the identification of meaningful descriptive variables: types of data, format of data, reasons for not sharing data, and reasons for sharing data. The results and findings from the qualitative interviews, the quantitative survey data analysis, and case study are reported in Chapter 4.

Limitations and Delimitations

One of the limitations of this study is the low response rate on the survey. Although the response rate was ~3%, this is the only survey that has been targeted toward this population of researchers. Also, this population is very reluctant to participate in research surveys and interviews because of trust issues, further impacting the response rate negatively. Only participants currently residing and working in countries identified as part of the GCC Persian Gulf region were surveyed

or interviewed. Compiling an email distribution list from the various university and organizational websites for the GCC countries proved difficult, and in some instances contact emails were impossible to find.

Another limitation was the necessity of using a convenience sample for the interviews, as this may have resulted in the inclusion of some interview participants that are not typical of the population under study. The survey and interviews were conducted using the English language only, and this may have precluded some scientists from participating. However, English is the most commonly-used second language internationally in most scientific research disciplines and is used almost exclusively at the higher education/research institutions in this study, so this issue likely had only a limited impact on the study's results.

CHAPTER FOUR

RESULTS AND DISCUSSION

This study employed a survey of researchers from the Region 6 (Middle East and Africa) with a total of 118 respondents. Of these, 94 respondents were from the GCC countries specifically. This is less than a 3% response rate overall. Also, 18 individual interviews were conducted to gather data about the data sharing practices and attitudes of scientists and researchers in the GCC countries of the Persian Gulf. Additionally, since most of the survey respondents (55.6%) and 10 of the interviewees reported that they work in Qatar, a case study of data sharing in Qatar is conducted.

GCC Analysis

GCC survey results

Demographics (GCC)

Of the 94 relevant survey respondents, 92 were from academic institutions and two were from non-profit organizations (Table 4.1). Of these, over 27% were engaged in lab research, while 28% were field researchers. Relatively few respondents conducted modeling research (9%) or GIS-related research (3%). The largest segment in the "type of research" category was "other," and these were not defined by the respondents (Table 4.2).

The most prevalent subject disciplines of the respondents (Table 4.3) were Information/Computer Science and Engineering (both at 12.8%), and Physical Sciences at 6.4%. Various other disciplines in medicine, social sciences or humanities scored below 3% each, with most at only 1% of the responses. The largest proportion did not respond to this question (42.3%).

Table 4.1: Q2. Work Sector.
Which one of the following best describes your primary work sector?

	Frequency	Percent	Cumulative Percent
1 Academic	92	97.9	97.9
4 Non-Profit	2	2.1	100.0
Total	94	100.0	

Table 4.2: Q4. Primary Research.
Which of the following best describes your primary type of research activity?

	Frequency	Percent	Cumulative Percent
Field Research	27	28.7	28.7
Lab Research	26	27.7	56.4
Modeling	9	9.6	66.0
GIS	3	3.2	69.1
Other	29	30.9	100.0
Total	94	100.0	

Table 4.3: Q3. Subject Discipline.
Which one of the following best describes your primary subject discipline?

		Frequency	Percent	Cumulative Percent
	Agriculture and Natural Resources	3	3.2	3.2
	Biology	2	2.1	5.3
	Information/Computer science	12	12.8	18.1
	Environmental Science/Ecology	2	2.1	20.2
	Engineering	12	12.8	33.0
	Geology/Earth Science	2	2.1	35.1
	Law	1	1.1	36.2
	Physical sciences	6	6.4	42.6
	Psychology	1	1.1	43.6
	Other (No Response)	45	42.3	89.1
	Architecture	1	1.1	90.2
	Dentistry	3	3.2	94.7
	Health informatics and information management	1	1.1	95.7
	Humanities	1	1.1	96.8
	Nursing	1	1.1	97.9
	Sociology	1	1.1	98.9
	Virology and immunology	1	1.1	100.0
	Total	94	100.0	

Of the six countries of the GCC, survey respondents were primarily working in Qatar (40.4%) and Saudi Arabia (31.9%), with UAE and Kuwait both at 11.7%. Smaller levels of participation came from residents of Oman and Bahrain, at 2% each (Table 4.4). Also, three-fourths of respondents were male (74.5%), while 24.5% were female; one respondent preferred not to indicate a gender (Table 4.5).

The age groups represented in the survey results (Table 4.6) were a normal distribution with slightly more than 3% under age 29, with 24.5% in their thirties, 33% in their forties, 26.6% in their fifties, and 10.6% over the age of sixty. The year in which respondents completed their degrees ranged from 1982 to 2016, reflecting a representation of early career to late career researchers (Table 4.8, n=31). These last two tables (Tables 4.7 and 4.8) were the result of answers to Q7 "Year highest degree was attained," an open text field. It appears that the wording was not clear since about one third chose to answer with the exact year (numerical value), while the majority provided the actual level of their degree (PhD, Masters, etc.).

Table 4.4: Q5. Country.
Which of the following countries is your primary place of employment?

	Frequency	Percent	Cumulative Percent
Bahrain	2	2.1	2.1
Kuwait	11	11.7	13.8
Oman	2	2.1	16.0
Qatar	38	40.4	56.4
Saudi Arabia	30	31.9	88.3
United Arab Emirates (UAE)	11	11.7	100.0
Total	94	100.0	

Table 4.5: Q8. Gender.

	Frequency	Percent	Cumulative Percent
Male	70	74.5	74.5
Female	23	24.5	98.9
Prefer not to answer	1	1.1	100.0
Total	94	100.0	

Table 4.6: Q6. Age_Group.

	Frequency	Percent	Cumulative Percent
29-under	3	3.2	3.3
39-30	23	24.5	28.3
49-40	31	33.0	62.0
59-50	25	26.6	89.1
60-over	10	10.6	100.0
Total	92	97.9	
No Response	2	2.1	
Total	94	100.0	

Table 4.7: Q7. Highest Degree Attained.

	Frequency	Percent
DScD	1	1.6
Master	8	12.7
Other	1	1.6
PhD	52	82.5
PhD Student	1	1.6
Total	63	100.0

Table 4.8: Q7_1. Year Highest Degree Was Attained.

	Frequency	Percent
1982	1	3.2
1983	1	3.2
1984	1	3.2
1985	1	3.2
1988	1	3.2
1991	1	3.2
1992	2	6.5
1996	1	3.2
1998	1	3.2
1999	3	9.7
2001	1	3.2
2003	1	3.2
2005	3	9.7
2007	1	3.2
2009	2	6.5
2010	1	3.2
2011	1	3.2
2012	4	12.9
2013	1	3.2
2015	2	6.5
2016	1	3.2
Total	31	100.0

The following tables provide an overall view of the survey respondents' knowledge of their funding agencies, whether they are required by their funding agencies to provide a data management plan and their awareness of the DataONE project. Most researchers are not required to provide a data management plan as a condition of their research funding (55.3%), while about a quarter are required to provide one (27.7%), and 17% do not know if this is a requirement (Table 4.9). Respondents' research is predominantly funded by federal or national governments (35.1%) or by a researcher's own institution (38.3%), with relatively few researchers (6.3% to 2.1%) being funded by any of the various other sources (Table 4.10).

Table 4.9: Q9. Data Management Plan.
Does your primary funding agency require you to provide a data management plan?

	Frequency	Percent	Cumulative Percent
Yes	26	27.7	27.7
No	52	55.3	83.0
Don't Know	16	17.0	100.0
Total	94	100.0	

Table 4.10: Q11. Funding Agency.

	Frequency	Percent
Internal/my institution	36	38.30
Federal/national government	33	35.11
Local government	6	6.38
Private foundation	5	5.32
Corporation	4	4.26
Internal/my institution (self funded)	4	4.26
State/regional government	4	4.26
Other	2	2.13
Total	94	100

Due to the lack of awareness of DataONE and very low response to the section of questions in the survey (see Appendix A) pertaining specifically to DataONE, questions numbered Q1_1 through Q1_5 are not included in these results. These questions are not directly relevant to the research on data sharing in the GCC countries.

How Data are Stored and Shared (GCC)

Most researchers indicated they do not use any metadata standards to describe their data (63.1%), while some use the standard created within their lab (10.7%) or institution (7.8%); several other metadata standards are used to a much lesser extent (3.9% or less) (Table 4.11).

Researchers in the GCC store data in several locations (Table 4.12), with the greatest amount of storage (using the top two categories of "Most" or "All" of their data) on: their personal computers (69.1%), and thumb/external drives (36.2%). The least likely place for GCC researchers to store their data is in repositories (using the bottom two categories of "None" or "Some" of the data): a discipline-based repository (60.6%), publisher repository (67%), institutional repository (63.8%), and other data repository/national data center (65.9%). Other places in which researchers tend not to store data (using the bottom two categories "None" or "Some" of data) are: on paper in their offices (68.1%), or on departmental servers (68%).

As shown in Table 4.13, the majority of GCC researchers indicated they are mostly satisfied (using the top two positions "Agree somewhat" and "Agree strongly") with the process for storing data during the life of the project/short-term (74.5%), and with the process for storing their data beyond the life of the project/long-term (59.6%). Less definitive are the satisfaction levels (using the top two positions "Agree somewhat" and "Agree strongly) with tools for preparing metadata (40.4%), the ability to track and verify provenance information (43.6%), and ease of locating a suitable repository for the deposit of data (38.3%).

Table 4.11: Q10. Metadata Standards.
What metadata standards do you currently use to describe your data

	Frequency	Percent
DC (Dublin Core)	3	2.9
DwC (Darwin Core)	0	0
DIF (Directory Interchange Format)	2	1.9
EML (Ecological Metadata Language)	0	0
FGDC (Federal Geographic Data Committee) CSDGM (content standard)	0	0
ISO 19115.xx (Geographic Information – Metadata)	2	1.9
Other ISO metadata standard	4	3.9
OGIS (Open GIS)	2	1.9
ANZLIC metadata profile	1	1
Net CDF	1	1
Metadata standardized within my institution	8	7.8
Metadata standardized within my lab	11	10.7
None	65	63.1
Other	4	3.9
Total	103	100

Table 4.12: Q12. Data Storage Locations.

How much of your data do you currently store or deposit in the following locations? (For each location, choose only the one best answer.)

	None of my data	Some of my data	Most of my Data	All of my data	Not sure	No Response	Total %
On my institution's server	28.7	22.3	18.1	14.9	6.4	9.6	100
On the principal investigator's server	28.7	18.1	13.8	21.3	7.4	10.6	100
On a departmental server	45.7	22.3	4.3	6.4	6.4	14.9	100
On my personal computer	3.2	20.2	11.7	57.4	1.1	6.4	100
On paper in my office	24.5	43.6	10.6	7.4	1.1	12.8	100
Thumb/external drive	19.1	25.5	11.7	24.5	3.2	16	100
In a discipline-based repository, (e.g. NEON or LTER)	53.2	7.4	4.3	1.1	18.1	16	100
In a publisher or publisher-related repository (e.g., specific publisher or Dryad)	56.4	10.6	1.1	3.2	12.8	16	100
Other data repository or archive (e.g., national data center)	58.5	7.4	8.5	9.6	0	15	100
In my institution's repository	48.9	14.9	5.3	9.6	5.3	16	100
Cloud storage	29.8	23.4	13.8	11.7	7.4	13.8	100
Other	34	2.1	2.1	1.1	12.8	47.9	100

Table 4.13: Q14. Data Storage Satisfaction.
The following statements relate to how you store and manage your data. Tell us how much you agree with the following ways to complete this sentence: I am satisfied with the…

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Process for storing my data during the life of the project (short-term)	7.4	6.4	9.6	38.3	36.2	1.1	1.1	100
Process for storing my data beyond the life of the project (long-term).	7.4	13.8	11.7	28.7	30.9	4.3	3.2	100
Tools for preparing metadata	4.3	8.5	18.1	22.3	18.1	23.4	5.3	100
Ability to track & verify provenance information	4.3	10.6	16	21.3	23.4	20.2	4.3	100
Ease of locating a suitable repository for the deposit of data	4.3	12.8	18.1	21.3	23.4	14.9	5.3	100

Most GCC researchers indicate that their institution either does not have a formal process for managing data during the life of the project/short-term (36.2%), or they do not know if there is a formal process in place (33%) (Table 4.14). The majority of researchers also indicate their institution does not have a formal process in place for storing data beyond the life of the project/long-term (43.6%), or they do not know if such a formal process is in place (31.9%) (Table 4.15).

For those GCC researchers who indicated their organization does have formal processes for short/long-term data storage, the following departments generally assist with this process: research support units (25.5%), information technology support units (14.9%), the library (13.8%), and colleagues in their own unit/department (12.8%) (Table 4.16).

Table 4.14: Q16_1. Managing Data (Short-term).
The following statements relate to how your organization is involved with managing and storing data. Tell us how much you agree with the following ways to complete this sentence: My organization has a formal process for: - ...managing data during the life of the project (short-term).

	Frequency	Percent	Cumulative Percent
Yes	26	27.7	27.7
No	34	36.2	63.9
Don't know	31	33.0	96.8
Total	91	96.8	100.0
No Response	3	3.2	
Total	94	100.0	

Table 4.15: Q16_2. Managing Data (Long-term).
The following statements relate to how your organization is involved with managing and storing data. Tell us how much you agree with the following ways to complete this sentence: My organization has a formal process for: - ...storing data beyond the life of the project (long-term).

	Frequency	Percent	Cumulative Percent
Yes	21	22.3	22.3
No	41	43.6	65.9
Don't know	30	31.9	97.9
Total	92	97.9	100.0
No Response	2	2.1	
Total	94	100.0	

Table 4.16: Q17. Formal Process for Storing Data.

You have expressed agreement that your organization or project has a formal process for managing or storing data during or beyond the life of the project (short-term or long-term). Which of the following are involved with this process? (Choose all that apply, multiple answer may apply)

	Frequency	Percent
Research support unit(s) (e.g. Office of Research, Office of Sponsored Programs and Contracts)	24	25.5
The Library	13	13.8
Information technology support unit(s) (e.g. Office of Information Technology, IT Support Center)	14	14.9
Administrative office(s) (e.g. Department Heads, Deans, Provosts, Program Offices, Research Offices, Divisions, Directorates / Directors, Managers)	9	9.6
Designated data manager(s)	4	4.3
Colleagues in my own unit / department	12	12.8
Other	0	0
Total	94	100.0

Why Data is Shared or Not Shared (GCC)

The main reasons GCC researchers indicated for not making their data available to others include: the need to publish first (59.6%), that they don't have the rights to make the data public (29.8%), and that their sponsors do not require it (29.8%) (Table 4.17). Other less prevalent reasons were that they had insufficient time to make the data available (25.5%), and that the researcher would lose control of the data (25.5%).

Researchers in the GCC are willing to share data, if certain conditions are met (Table 4.18). The majority would share if: a citation of the data is provided in all work making use of the data (83%), an acknowledgement is given of the data in all work making use of the data (78.7%), they had an opportunity to collaborate on a project using the data (72.3%), they received co-authorship on publications resulting from the use of the data (64.9%), legal permission to use the data is obtained from the researcher (54.3%), and research results based (at least in part) on the data must obtain researcher approval prior to dissemination (53.2%).

Organization-based training related to data management appears to have a low prevalence, or a low level of awareness among these respondents (Table 4.19). GCC researchers indicated the following levels of involvement by their organizations (using the top two positions "Agree strongly" and "Agree somewhat") for: training on best practices for data management (21.2%), training on how to cite datasets (29.8%), assistance with creating data management plans (26.6%), and assistance on creating metadata to describe data or datasets (18.1%). Similar percentages of researchers are not sure if their organization offers these trainings and assistance.

Table 4.17: Q13. Why Data Not Available.

If all or part of your data are not available to others, why not? (Choose all that apply.)

	Frequency	Percent
Lack of funding	24	25.5
Lack of standards	20	21.3
People don’t need them	18	19.1
There is insufficient time to make them available	24	25.5
There is no place to put them	16	17.0
Sponsor doesn't require it	28	29.8
Don't have the rights to make the data public	28	29.8
I would lose control of the data	24	25.5
I need to publish first	56	59.6
I have insufficient skills to make my data available	15	16.0
Other	3	3.2
Total	94	100.0

Table 4.18: Q15. Expectations if Others Use My Data.

The following statements relate to conditions for use of your data. Indicate whether you agree or disagree with each condition. For others to use my data, I would expect the following in exchange:

	Yes	No	Not Sure	No Response	Total %
Co-authorship on publications resulting from use of the data.	64.9	12.8	18.1	4.3	100
Acknowledgement of the data providers in all disseminated work making use of the data.	78.7	9.6	7.4	4.3	100
Citation of the data providers in all disseminated work making use of the data.	83	5.3	7.4	4.3	100
The opportunity to collaborate on a project using the data.	72.3	10.6	11.7	5.3	100
Results based (at least in part) on the data could not be disseminated in any format without the data provider's approval.	53.2	18.1	23.4	5.3	100
Results based (at least in part) on the data could not be disseminated without the data provider having the opportunity to review the results and make suggestions or comments, but approval not required.	40.4	26.6	27.7	5.3	100
Reprints of articles that make use of the data must be provided to the data provider.	50	27.7	16	6.4	100
The data provider is given a complete list of all products that make use of the data, including articles, presentations, educational materials, etc.	48.9	24.5	21.3	5.3	100
Legal permission for data use is obtained.	54.3	14.9	25.5	5.4	100
Mutual agreement on reciprocal sharing of data.	66	12.8	16	5.3	100

Table 4.19: Q18. My Organization or Project Provides:

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Training on best practices for data management.	23.4	14.9	14.9	13.8	7.4	21.3	4.3	100
Assistance on creating data management plans	19.1	17	13.8	21.3	5.3	20.2	3.2	100
Assistance on creating metadata to describe my data or datasets	24.5	12.8	16.0	13.8	4.3	24.5	4.3	100
Training on how to cite datasets	17	14.9	12.8	20.2	9.6	21.3	4.3	100

GCC researchers indicated an interest in sharing their data and re-using other researchers' data (Table 4.20), primarily in the following circumstances (using the top two positions "Agree somewhat" and "Agree strongly"): they were willing to share data across a broad group of researchers (73.4%), they would use other researchers' datasets if they were easily accessible (71.3%), and they would be willing to place at least some of their data into a central data repository with no restrictions (56.4%). The item researchers agreed upon the most (combining the top two positions "Agree somewhat" and "Agree strongly") was having their data cited when used by other researchers (83%).

Table 4.20: Q19. Sharing Scientific Data.

The following statements relate to sharing scientific data. Tell us how much you agree with each statement.

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
I would use other researchers' datasets if their datasets were easily accessible.	6.4	9.6	3.2	28.7	42.6	5.3	4.3	100
I would be willing to place at least **some** of my data into a central data repository with no restrictions.	6.4	9.6	12.8	26.6	29.8	10.6	4.3	100
I would be willing to place **all** of my data into a central data repository with no restrictions.	16.0	21.3	17.0	16.0	16.0	9.6	4.3	100
I would be more likely to make my data available if I could place conditions on access.	4.3	12.8	12.8	33.0	28.7	4.3	4.3	100
I am satisfied with my ability to integrate data from disparate sources to address research questions.	3.2	7.4	20.2	34.0	20.2	10.6	4.3	100
I would be willing to share data across a broad group of researchers.	6.4	3.2	6.4	40.4	33.0	6.4	4.3	100
It is important that my data are cited when used by other researchers.	3.2	1.1	5.3	14.9	68.1	3.2	4.3	100
It is appropriate to create new datasets from shared data.	4.3	3.2	12.8	27.7	36.2	11.7	4.3	100

When asked about the access to and comfort level with using others' research data (Table 4.21), GCC researchers indicated (combining the top two positions "Agree somewhat" and "Agree strongly") that data may be misinterpreted due to poor quality of the data (77.6%) or the complexity of the data (76.6%). Lack of access to data generated by others is considered (using the top two positions "Agree somewhat" and "Agree strongly") an impediment to the progress in science (70.2%) and to the ability to answer scientific questions (54.3%).

Researchers in the GCC indicated their confidence in re-using data collected by others (Table 4.22), primarily if certain conditions were in place. Respondents endorsed data re-use (based on the top two positions "Agree somewhat" and "Agree strongly") if the data were accompanied by written details about collection and quality assurance methods (66%), or if the metadata standards utilized were explicitly stated with the data (50%).

When GCC researchers need data to answer a research question, they primarily approach the situation in these four ways (based on the top two positions "Agree somewhat" and "Agree strongly"): they make a plan to generate or collect the data they need within their research team (72.3%), they search for data to use for analysis (66%), they generate or collect the data themselves (74.5%), or they ask colleagues if they know of data that can be used for analysis (52.1%). Respondents were somewhat more likely to turn to a librarian (38.3%) than their data manager (27.7%) in this circumstance, although these options garnered far less support than the others (Table 4.23).

Half of the researchers in the GCC said that they either occasionally or frequently conduct research with data collected by someone other than a member of their immediate research team (50%), while 43.6% indicated that they seldom or never use other researchers' data to conduct research (Table 4.24).

Table 4.21: Q20. Use of Scientific Research Data.

The following statements relate to your views on the use of scientific research data. Tell us how much you agree with each statement.

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Lack of access to data generated by other researchers or institutions is a major impediment to progress in science	6.4	3.2	12.8	35.1	35.1	4.3	3.2	100
Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.	6.4	11.7	19.1	30.9	23.4	5.3	3.2	100
Data may be misinterpreted due to **complexity** of the data.	1.1	5.3	9.6	44.7	31.9	4.3	3.2	100
Data may be misinterpreted due to **poor quality** of the data.	0	3.2	9.6	34.0	43.6	5.3	4.3	100
Data may be used in other ways than intended.	1.1	4.3	10.6	36.2	35.1	9.6	3.2	100

Table 4.22: Q21. Re-use of Scientific Research Data.

The following statements relate to your views on the reuse of scientific research data. Tell us how much you agree with the following ways to complete this sentence: I would have increased confidence in re-using data collected by others if…

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
The metadata standard(s) utilized were explicitly stated with the data	2.1	2.1	20.2	28.7	21.3	19.1	6.4	100
The data were accompanied by written details about collection and quality assurance methods	2.1	5.3	9.6	36.2	29.8	11.7	5.3	100
A recorded workflow from a standard workflow system (Kepler, VisTrails, Taverna, etc) were also available with the data	2.1	4.3	19.1	24.5	17.0	27.7	5.3	100
Detailed information about the provenance (data lineage, chain of custody) were available with data	2.1	2.1	14.9	25.5	21.3	26.6	7.4	100
Other	0	2.1	9.6	7.4	8.5	24.5	47.9	100

Table 4.23: Q22. When I need Data to Answer a Research Question.
Tell us how much you agree with the following ways to complete this sentence: When I need data to answer a research question, I...

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Make a plan to generate or collect the data I need **myself.**	1.1	3.2	12.8	20.2	54.3	3.2	5.3	100
Make a plan to generate or collect the data I need **within my research team.**	2.1	5.3	9.6	23.4	48.9	4.3	6.4	100
Ask colleagues if they **have data** I can use for analysis.	13.8	10.6	18.1	30.9	18.1	3.2	5.3	100
Search for data to use for analysis	7.4	4.3	10.6	33.0	33.0	6.4	5.3	100
Ask colleagues if they **know of data** I can use for analysis.	11.7	6.4	18.1	31.9	20.2	6.4	5.3	100
Talk to a librarian about my data needs.	18.1	9.6	20.2	22.3	16.0	8.5	5.3	100
Consult my data manager.	23.4	5.3	20.2	16.0	11.7	16.0	7.4	100

Table 4.24: Q23. Data Collected by Someone Else.
How often do you conduct research in which some or all of the data analyzed was collected by someone besides yourself or members of your immediate research team? (Choose the one best answer.)

	Frequency	Percent	Cumulative Percent
Never	18	19.1	20.0
Seldom	23	24.5	45.6
Occasionally	29	30.9	77.8
Frequently	18	19.1	97.8
Always	2	2.1	100.0
Total	90	95.7	
No response	4	4.3	
Total	94	100.0	

Those GCC researchers who indicated they did use data collected by others besides themselves or their immediate research team, gave the following reasons for doing so (based on the top two positions "Agree somewhat" and "Agree strongly"): it is easier than having to collect all their own data (71.3%), it saves time (69.1%), it is efficient (63.8%), and it helps answer their research questions (66%). If data from others was not used, the main reason given was that it requires too much trust in others' methods (62.7%, based on the top two positions "Agree somewhat" and "Agree strongly") (Table 4.25).

Table 4.25: Q24. Research with Data Collected by Others.
Tell us how much you agree with the following ways to complete this sentence: Conducting research in which some or all of the data analyzed was collected by others besides myself or members of my immediate research team...

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
…saves time	3.2	2.1	13.8	28.7	40.4	4.3	7.4	100
…is efficient	5.3	6.4	10.6	31.9	31.9	4.3	9.6	100
...is easier than having to collect all my own data for analysis	4.3	4.3	9.6	28.7	42.6	4.3	6.4	100
...is hard to explain in methods section	6.4	8.5	19.1	33.0	14.9	11.7	6.4	100
...requires too much trust in others' methods	2.1	9.6	14.9	28.7	34.0	4.3	6.4	100
...improves my results	4.3	5.3	27.7	36.2	13.8	6.4	6.4	100
...helps me answer my research questions	3.2	3.2	14.9	42.6	23.4	6.4	6.4	100
...is **harder** than conducting research using only my own data	7.4	20.2	19.1	26.6	11.7	8.5	6.4	100
...**takes longer** than conducting research with only my own data	8.5	21.3	16.0	23.4	14.9	9.6	6.4	100

GCC Interview results

Demographics

A total of 18 interviews were conducted either in person or via phone/Skype/online calls during the months of April and May 2018. The basic demographics of this group is presented in Table 4.26. Most of the interviews were in person during the time the researcher travelled to Qatar (n=10). The remaining interviews were via phone/Skype and WhatsApp remotely, with participants based in Kuwait (n=2), Saudi Arabia (n=3) and UAE (n=3).

The range of primary subject disciplines of the interview participants was similar to those reported by the survey respondents: Information/Library Studies (n=2), Physical Sciences (n=2), Engineering (n=3), Health/Medical (n=4), Life Sciences (n=1), Social Sciences (n=2), Computer Science (n=3), and Energy (n=1). Almost all of the researchers were engaged in work at primarily academic institutions (n=17), while one was located at a government research facility.

All interviewees spent some portion of their time on research activities, ranging from 10% to 100%, depending on their additional teaching and administrative duties. The most common amount of time spent on research was 30% to 40%. The experience level of the researchers ranged from early career (<5 years) to late career (>15 years) researchers, with the majority being mid-career (6 to14 years) in the research field.

A majority of the interview subjects were male (72%, n=13), while about 28% were female (n=5). Their ages ranged from 36 to 68, with the majority being in their forties (n=9).

Table 4.26: GCC_1. Demographics of Interview Subjects by Country.

	Primary Subject Discipline(s)	Work Sector	% Time spent on Research	Years in current position	Terminal Degree	Gender	Age range
Kuwait (n=2)	Health Informatics Information Studies	Academic	10-40%	4-13yrs	All PhD	1 Female 1 Male	49-52
Qatar (n=10)	Epidemiology Electrical Engineering Sociology Microbiology and Immunology Information Library Studies Physics Chemical Engineering Family Medicine Energy Archaeology	All Academic, except one Government research department	20-100%, most commonly around 50%	1-12yrs, most common 4-8yrs	8 PhD 1 EdD 1 MD	2 Female 8 Male	41-68, majority in 40s
Saudi Arabia (n=3)	Dentistry, High Performance Computing Computer Science and Electrical Engineering	Academic	20-50 %	1-10yrs, average of 5yrs	2 PhD 1 Master	1 Female 2 Male	46, and two ages not given
UAE (n=3)	Ecology-Remote Sensing Computer Science-Modeling Academic Writing	Academic	30-70%, with 30% most common	1-7yrs, average of 4yrs	3 PhD	1 Female 2 Male	36-45, mostly in 30s

Most researchers obtained their terminal degrees in either the USA or Europe (n=15), while two studied in GCC countries, and one in the Russian Federation. All of the researchers indicated that they use English as their primary research collaboration language (n=18), with an occasional researcher (n=4) also using additional languages (Arabic or French) for collaboration work if needed. All interviewees (n=18) stated that English was used exclusively in their daily work and teaching duties; and it was indicated that English was mandated as the primary language by their institutions.

How Data is Stored and Managed (GCC)

The types of datasets created and used by the interviewed researchers varied, because of the range of data being collected in their various disciplines (Table 4.27). These data are stored in a variety of file formats: structured/unstructured, free text/spreadsheet, and specialized software data formats. Most do not use any metadata standard to describe their data (n=15), while some were not sure if a metadata standard was used (n=3).

Table 4.27: GCC_2. Type of Dataset.

	Examples
Abiotic	Chemical properties of nanomaterials, ecological measurements
Biotic	Wet lab samples and observations
Experimental	Experimental physics
Observational	Data sheets (paper)
Social science	Interviews, surveys/questionnaires, focus groups, ethnographic observations
Other	Simulation code (engineering), primary archival documents and recordings, archaeological artifacts, epidemiology data

These datasets can range in size from a few megabytes up to the terabyte and petabyte scale. Most of these researchers create small data sets of less than 500 MB (n=11), while some create medium-size data sets of 500 MB to 1 TB (n=3), and others make large datasets >1 TB (n=6). Some of the large datasets can be in the hundreds of terabytes, and generally result from computationally-intensive data gathering. These size designations were indicated by the interviewees, and as the size increases, so does the complexity of the data. When asked if their datasets would be considered "big data," most said no (n=10), a few said yes (n=5), and some were not sure (n=3). Some researchers indicated they were not sure of the definition of big data, with one researcher stating, "I don't know how many people actually understand big data. I think it's one of those clichés that people put in research proposals because it's a sexy topic. To me big data would be if you took lots of broad data to determine causalities (QA02)."

Many researchers indicated they used Excel or Access files to structure basic data or text, and used statistical software such as SPSS, Stata, MATLAB or SAS for survey and observational data. Additional software tools, such as Origin and TensorFlow, were used for computationally-intensive data (big data sets in the hundreds of gigabytes or greater).

Most researchers (n=15) are satisfied with their current process of collecting and analyzing their data, and some were somewhat satisfied (n=3). Of those not completely satisfied, there were suggestions on how to improve the process, including obtaining help with data analysis, help with entering data into a database, and needing more access to other datasets. One interviewee said, "[I] think it could be better, but have not looked into how to make it better (SA01)." Another interviewee summed up his experience with data management by saying, "I don't think you will find an academic who is ever truly satisfied. Just the sheer volume of data,

somethings takes days to move data from supercomputer to archives. I think we are still in the dawn of managing data (SA03)."

Storage of data is accomplished in several ways, and many researchers use a mix of storage and backup methods. The most common locations for short-term storage of research data is on the researcher's work laptop/desktop (n=13), an external hard drive (n=9), an institutional network server (n=6), in cloud storage (n=5), or in paper files (n=3). Some mentioned occasionally saving to a personal laptop at home temporarily, if they work there (n=?). Dropbox is the most common choice for cloud storage (not an institutional network drive). One interviewee stores data on magnetic tapes, indicating "The amount is huge, the archival business is huge. You would be amazed at the amount of tapes that are being produced. This is not just in Saudi Arabia, also in the USA (SA03)."

When backing up data, the same types of storage methods are employed, but in different amounts. Most use cloud storage (n=10), an institutional server/network drive (n=3), an external hard drive (n=6), a home laptop (n=2), paper files as a backup (n=1), and magnetic tapes (n=1). The popularity of cloud storage can partially be attributed to the ease of access to data from any location, and as one participant noted, "I can think about it at home at night when family is 'nerding' on their devices (QA04)." Keeping data safe via cloud/network storage is also a concern, "In case of fire or disaster, store in separate cloud location, on separate server, geographically diverse locations (SA02)." The frequency of data backups is a bit mixed, with many (n=10) not sure how often their data is backed up. Of those who have a regularly scheduled time interval for data backup, the most common are: instantly/immediately to a server/cloud storage (n=2), daily (n=1), weekly/every two weeks (n=3), and monthly (n=1). One interviewee indicated a specific schedule was not set, saying "Backup to university network

drive, when a reminder is sent out twice a year to researchers to backup any local data to the university server (UAE03)."

When asked if a data management plan is in place to guide the short and long-term storage and disposition of their data, most researchers say "no," (n=14), while a few do not know (n=3). Only one researcher indicated that some type of data management plan was in place, and this was determined by the individual university departments for which data was archived. Several researchers (n=4) indicated that the only data management guidelines they have are those included as part of their institutional review board (IRB) document and are specific to each research proposal. As far as funding agency requirements, most researchers either don't know (n=7) or specifically indicated that no policies were in place (n=8) to require a data management plan. One researcher knew that the funding agency did require a data management plan, stating, "Yes, absolutely, they are very strict in managing data and reports (QA08)."

Why Data is Shared (GCC)

Most researchers are willing to share their data (n=13), while with others, it depends on the situation/details of sharing (n=4). One researcher does not share. When determining who to share with, most will share with anyone (n=14). A few are more selective in sharing (n=3), limiting it to members of their organization/team, or to those researchers who have the skill set to properly use the data.

The interviewees mentioned several **reasons why they would not share** their datasets:

- Fear of data being stolen prior to publication
- Lack of a data sharing culture
- Data has commercial or patent potential

- Data may be taken out of context
- Qualitative data are not conducive to sharing (others are interested only in resultant publications)
- Collaborators may not agree to share
- If requestor has a questionable track record/reputation
- Cannot share before IP (intellectual property) agreement is signed
- No time to clean data/remove errors
- Sensitive data (personal or political related data, culturally sensitive data)
- IRB limits sharing

These are selected quotes regarding reasons not to share data, that were mentioned by the participants:

1. "Part of the reason I never have is because my IRB policy prohibits this (QA03)."
2. "Only original data/research counts toward promotion, so researchers in this culture do not share. Maybe with time, if promotion policies change, it will encourage data sharing (KU02)."
3. "People in the Middle East region are horribly protective of data sets produced locally, resulting in much money spent to collect, and never published (QA01)."
4. "In Qatar research is still young and they are still figuring it out (QA02)."
5. "Afraid the notes would be taken out of context of my original research (QA03)."
6. "If data can potentially become an invention, it could be reason not to publish or share data (QA04)."
7. "Sensitive data, such as video of drivers (females) may not share, because of cultural issues. Have to seek permission if want to share (SA02)."

8. “I think many researchers are not comfortable with their capabilities, and will not share to allow others to verify their findings (KU01).”

Researchers identified the following reasons/**incentives why they would share** their data:

- Data were cited in publications using that data
- Collaborations with others, built trust with Middle East researchers
- To change culture to a data sharing culture
- Scientific compassion – data could make a difference in lives/preventative care of patients
- Networking with other researchers

This quote offered an adamant reason to support sharing data:

- “very selfish act of a scientist to not share data. Even if I lose intellectual property, sharing knowledge is essential (QA04)."

Benefits to sharing data with others as mentioned by the participants:

- Validation of data
- Increase in research efficiencies (re-use data)
- Collaboration
- Benefits to society
- Getting advice from reviewers/referees about the data during the pre-publication phase
- Further regional research (in Middle East)
- Helps humanity/society

- Contribute to my scientific field

These quotes from participants support the benefits of sharing data:

1. “if you are a scientist you are not in this field for money, what drives me is a fascination with science, from a hobby when I was younger , and now into a profession. Science data can have impact on society (QA04)”
2. “here (in Middle East) it will benefit everyone to further water pretreatment research (QA07)”

Barriers to sharing data, noted by the interviewees:

- Cultural tendency to NOT share
- Collaborations require an MOU (memorandum of understanding)
- Data is unstructured
- IRB limits access to and sharing sensitive data
- Researchers not familiar with data repositories
- No mechanism in place to share data
- Competition in research and development (R&D) areas
- Time and resources are too scare to prepare data for sharing

These quotes from the participants describe barriers to sharing data:

1. “Also in Qatar, there is a huge drive to protect data if it can lead to commercialization (QA01)”
2. “Marketing of repositories needs to be done to make researchers familiar with them. Name recognition needed with repositories (QA05)”

Most researchers agree that sharing data is important for the progress of science. All researchers (n=18) stated that sharing data is important for their institutions and their local regions of the globe They indicated that it is not only a possibility for researchers to share, but a responsibility.

These are quotes from the participants about data sharing's relationship to the **progress of science**:

1. "When data becomes available for others to use, it has an exponential impact on scientific progress (QA01)."
2. "Absolutely! I think it is important in general to benefit society, especially if funded by the government. We could achieve more efficiency if we all shared data (QA02)."
3. "Definitely, progress in general, and especially when you talk about nationally or regionally. Funding resources will be optimized if data is shared, and money is not spent on redundant research (QA07)."
4. "Very important to reach the fastest progress in R&D area (QA09)."
5. "I believe it is important to society and to science, to economic impact, for people to understand the issues around sharing data. Data sharing is important (SA02)."
6. "Absolutely, no question about that. This is the best enabler to the progress of science (UAE02)."
7. "Data is a photograph of society at a specific time, and we are obligated to share it on that level (UAE03)."

Most researchers (n=15) indicate they are not required to share their data by their funding agencies , while others do not know if this is a requirement or not (n=2). One researcher was

self-funded and could make their own determination whether to share. Most researchers indicated that there were no specific policies in place regulating data sharing and they could make this determination themselves (n=14), while few (n=2) had to get prior institutional approval to share , and an equal number (n=2) said it depended on the sensitivity of the data

Qatar: A Case Study in data sharing

Survey Results (Qatar)

Most of the data for this research focusing on the GCC region showed a disproportionate representation from the country of Qatar; with 55.6 % of the interviews and 40.4% of the survey respondents situated in Qatar. Therefore, it was of interest to analyze the data for Qatar as a separate case study to provide understanding of the variation in applying data sharing in GCC and also to reflect on the trend in data sharing practices in Qatar, in addition to the combined analysis for the GCC region.

Demographics (Qatar)

All (n=38) of the Qatar respondents indicated that they were employed in the Academic sector (Table 4.28). None indicated that they were employed in government, non-profits, or commercial sectors.

Table 4.28: QATAR_Q2. Work Sector.
Which one of the following best describes your primary work sector?

	Frequency	Percent	Cumulative Percent
Academic	38	100.0	100.0

Most of the Qatar researcher respondents did not provide their primary subject discipline (n=22). Of those who identified their primary subject disciplines, Information/Computer Science and Physical Sciences (13.2%) had equal amounts, with Engineering (10.5%) and Humanities and Sociology both reporting 2.6% each (Table 4.29).

Most Qatar researchers reported (Table 4.30) that they were engaged in either field research (34.2%) or lab research (28.9%). Several respondents did not indicate which type of research they are involved in (26.3%).

Table 4.29: QATAR_Q3. Subject Discipline.
Which one of the following best describes your primary subject discipline?

	Frequency	Percent	Cumulative Percent
Information/Computer science	5	13.2	13.2
Engineering	4	10.5	23.7
Physical sciences	5	13.2	36.8
Humanities	1	2.6	39.4
Sociology	1	2.6	42.0
Missing	22	58.0	100.0
Total	38	100.0	

Table 4.30: QATAR_Q4. Primary Research
Which of the following best describes your primary type of research activity

	Frequency	Percent	Cumulative Percent
Field Research	13	34.2	34.2
Lab Research	11	28.9	63.2
Modeling	3	7.9	71.1
Social Science	1	2.6	73.7
Other	10	26.3	100.0
Total	38	100.0	

The majority of respondents were male (73.7%), with fewer than a quarter of respondents being female (23.7%). One respondent chose not to indicate their gender (Table 4.31).

The majority of Qatar survey respondents (n=38) were in the 40-49 year old age group (n=16, 42.1%), with 31.5% (n=12) in the 50 or older category. The smallest age group was represented by the 39 and under group, at a combined 21% (n=8), with some respondents not answering the question (n=2, 5.3%) (Table 4.32).

Most researchers indicated (Table 4.33) they had completed a Ph.D. degree (55.3%), while another 42.1% (n=22) did not indicate their terminal degree. This appears to have resulted from a confusion on what information was being requested in this open text field, because the missing (n=22) respondents provided the year they received their terminal degree. The years ranged from 1985 through 2015, reflecting the various age groups indicated earlier (Table 4.34).

Table 4.31: QATAR_Q8. Gender.

		Frequency	Percent	Cumulative Percent
	Male	28	73.7	73.7
	Female	9	23.7	97.4
	Prefer not to answer	1	2.6	100.0
	Total	38	100.0	

Table 4.32: QATAR_Age_Group

		Frequency	Percent	Cumulative Percent
	1 29-under	1	2.6	2.8
	2 39-30	7	18.4	22.2
	3 49-40	16	42.1	66.7
	4 59-50	11	28.9	97.2
	5 60-over	1	2.6	100.0
	Total	36	94.7	
	Missing	2	5.3	
Total		38	100.0	

Table 4.33: QATAR_Q7. Highest degree attained.

	Frequency	Percent	Cumulative Percent
MBA	1	2.6	2.6
Ph.D.	21	55.3	57.9
Missing	16	42.1	100.0
Total	38	100.0	

Table 4.34: QATAR_Q7_1. Year highest degree was attained:

	Frequency	Percent	Cumulative Percent
1985	1	2.6	2.6
1992	2	5.3	7.9
1998	1	2.6	10.5
1999	1	2.6	13.2
2003	1	2.6	15.8
2005	2	5.3	21.1
2007	1	2.6	23.7
2009	1	2.6	26.3
2011	1	2.6	28.9
2012	3	7.9	36.8
2013	1	2.6	39.5
2015	1	2.6	42.1
Missing	22	57.9	100.0
Total	38	100.0	

Most funding agencies in Qatar do not require researchers to provide a data management plan as part of their research proposals (50%). While 26.3% of respondents indicated that a plan was required, almost the same amount (23.7%) did not know if a data management plan was required (Table 4.35).

The Qatar national government is the largest funder of research in the country, funding 42.1% of all research done by these respondents, followed by 34.2% that is funded internally by the researchers' institutions (Table 4.36). Since most academic institutions are also funded by the national government, these two amounts could be looked at as one unit, indicating a 76.3% funding level by the Qatar national government. Much smaller amounts of respondents' research are funded by corporations (2.6%) and private foundations (5.3%). Of interest is the amount of self-funding indicated (5.3%) by two of the researchers.

Table 4.35: QATAR_Q9. Data Management Plan. Does your primary funding agency require you to provide a data management plan?

	Frequency	Percent	Cumulative Percent
Yes	10	26.3	26.3
No	19	50.0	76.3
Don't Know	9	23.7	100.0
Total	38	100.0	

Table 4.36: QATAR_Q11. Funding Agency.
Which of the following best describes the primary funding agency for your research?

	Frequency	Percent	Cumulative Percent
Federal/national government	16	42.1	42.1
State/regional government	2	5.3	47.4
Local government	2	5.3	52.6
Corporation	1	2.6	55.3
Private foundation	2	5.3	60.5
Internal/my institution	13	34.2	94.7
No funding sources / self-funded	2	5.3	100.0
Total	38	100.0	

How Data is Stored and Shared (Qatar)

The majority of Qatar researchers do not use any metadata standards to describe their data (68.4%) (Table 4.37). Only a small portion indicated that they use any metadata standardized within their lab or their institution (7.9% and 5.3% respectively).

Researchers in Qatar store their data in many locations (Table 4.38), with Most or All data (top two positions) being stored primarily on the researcher's personal computer (65.8%). The locations least favored for storing data, with either None or Some data stored on or in them (the bottom two positions on the response scale), are departmental servers (79%), institutional servers (55.2%), the principal investigator's server (52.7%), and paper files in the researcher's office (71%). Repositories were the least likely place for researchers to store their data, with None or Some data being stored on a discipline-based repository (60.5%), publisher's repository (71.1%), institutional repository (73.7%), or other types of repositories/national data centers (79%).

Table 4.37: QATAR_Q10. Metadata Standards.
What metadata standards do you currently use to describe your data?

	Frequency	Percent
DC (Dublin Core)	1	2.6
DwC (Darwin Core)	0	0
.6IF (Directory Interchange Format)	0	0
EML (Ecological Metadata Language)	0	0
FGDC (Federal Geographic Data Committee) CSDGM (content standard)	0	0
ISO 19115.xx (Geographic Information – Metadata)	1	2.6
Other ISO metadata standard	2	5.3
OGIS (Open GIS)	0	0
ANZLIC metadata profile	0	0
Net CDF	0	0
Metadata standardized within my institution	2	5.3
Metadata standardized within my lab	3	7.9
None	26	68.4
Other	3	7.9
Total	38	100.0

Table 4.38: QATAR_Q12. Data Storage Locations.
How much of your data do you currently store or deposit in the following locations? (For each location, choose only the one best answer.)

	None of my data	Some of my data	Most of my Data	All of my data	Not sure	No Response	Total %
On my institution's server	26.3	28.9	13.2	15.8	7.9	7.9	100
On the principal investigator's server	31.6	21.1	15.8	13.2	10.5	7.9	100
On a departmental server	55.3	23.7	0	2.6	2.6	15.8	100
On my personal computer	0	28.9	5.3	60.5	0	5.3	100
On paper in my office	18.4	52.6	13.2	5.3	0	10.5	100
Thumb/external drive	21.1	21.1	15.8	26.3	2.6	13.2	100
In a discipline-based repository, (e.g. NEON or LTER)	50.0	10.5	2.6	2.6	18.4	15.8	100
In a publisher or publisher-related repository (e.g., specific publisher or Dryad)	65.8	5.3	0	2.6	10.5	15.8	100
Other data repository or archive (e.g., national data center)	73.7	5.3	2.6	0	2.6	15.8	100
In my institution's repository	57.9	15.8	2.6	5.3	2.6	15.8	100
Cloud storage	23.7	28.9	15.8	13.2	5.3	13.2	100
Other	34.2	2.6	0	0	13.2	50.0	100

Most Qatar researchers seem to be satisfied with their short-term data storage methods (Table 4.39), with 71.1% agreeing either somewhat or agreeing strongly with this statement. Slightly fewer researchers agree somewhat or strongly with their long-term data storage methods (60.6%). Researchers have mixed opinions on whether they are satisfied with the tools they have to prepare metadata; 34.2% agree either somewhat or strongly, and an equal portion (34.2%) are not sure.

About one-third of the researchers (31.6%) indicated their organizations don't have a formal process in place for managing data during the life of a research project (short-term), and another 34.2% do not know if their organization has such a process in place (Table 4.40). The situation is very similar for long-term storage of data (Table 4.41), beyond the life of a project: 36.8% indicate their institutions does not have a formal process, and 31.6% do not know if there is a process for long-term storage.

Of those researchers who indicated that their institution has a formal process for managing or storing data during or beyond the life of the project (short-term or long-term) (Table 4.42), they are most likely to be assisted by: research support units (34.2%), information technology support units (23.7%), the library (15.8%), and administrative offices (13.2%).

Table 4.39: QATAR_Q14. Data Storage Satisfaction.

The following statements relate to how you store and manage your data. Tell us how much you agree with the following ways to complete this sentence: I am satisfied with the…

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Storing my data during the life of the project (short-term)	7.9	5.3	13.2	23.7	47.4	2.6	0	100
Process for storing my data beyond the life of the project (long-term).	5.3	15.8	15.8	21.1	39.5	2.6	0	100
Tools for preparing metadata	2.6	0	23.7	18.4	15.8	34.2	5.3	100
Ability to track & verify provenance information	0	7.9	13.2	10.5	28.9	34.2	5.3	100
Ease of locating a suitable repository for the deposit of data	2.6	10.5	18.4	18.4	26.3	18.4	5.3	100

Table 4.40: QATAR_Q16_1. Managing Data (Short-term).

The following statements relate to how your organization is involved with managing and storing data. Tell us how much you agree with the following ways to complete this sentence: My organization has a formal process for: ...managing data during the life of the project (short-term)

	Frequency	Percent	Cumulative Percent
Yes	12	31.6	32.4
No	12	31.6	64.9
Don't know	13	34.2	100.0
Total	37	97.4	
Missing	1	2.6	
Total	38	100.0	

Table 4.41: QATAR_Q16_2. Managing Data (Long-term)
The following statements relate to how your organization is involved with managing and storing data. Tell us how much you agree with the following ways to complete this sentence: My organization has a formal process for: - ...storing data beyond the life of the project (long-term).

	Frequency	Percent	Cumulative Percent
Yes	12	31.6	31.6
No	14	36.8	68.4
Don't know	12	31.6	100.0
Total	38	100.0	

Table 4.42: QATAR_Q17. Formal Process for Storing Data.
You have expressed agreement that your organization or project has a formal process for managing or storing data during or beyond the life of the project (short-term or long-term). Which of the following are involved with this process? (Choose all that apply, multiple answer may apply)

	Frequency	Percent
Research support unit(s) (e.g. Office of Research, Office of Sponsored Programs and Contracts)	13	34.2
The Library	6	15.8
Information technology support unit(s) (e.g. Office of Information Technology, IT Support Center)	9	23.7
Administrative office(s) (e.g. Department Heads, Deans, Provosts, Program Offices, Research Offices, Divisions, Directorates / Directors, Managers)	5	13.2
Designated data manager(s)	2	5.3
Colleagues in my own unit / department	3	7.9
Other	0	0
Total	38	100.0

Why Data is Shared or Not Shared (Qatar)

The most common reason given for not sharing research data with others (Table 4.43) is the need to publish first (57.9%). Secondary reasons include: having insufficient time to make the data available (28.9%), the sponsor does not require the researcher to share (26.3%), the researcher feels they would lose control of their data (26.3%), and researchers think people do not need the data (26.3%).

Researchers generally expect a lot of control and reciprocation in return for the use of their data by others (Table 4.44). In particular, they indicated that they would expect co-authorship on publications resulting from the use of their data (68.4%), acknowledgement of the data providers in all disseminated works making use of their data (78.9%), citation of the data providers in all disseminated works making use of the data (92.1%), the opportunity to collaborate on a project using the data (65.8%), and a mutual agreement on reciprocal data sharing (60.5%).

The majority of Qatar researchers are unsure if their organization provides training on best practices for data management (31.6%), assistance with creating metadata to describe data/data sets (36.8%), or training on how to cite data sets (34.2%) (Table 4.45). Researchers agree either somewhat or strongly that their organization provides assistance on creating data management plans (34.2%), but a similar number (31.6%) do not know if assistance is given to create a data management plan.

Table 4.43: QATAR_Q13. Why Data Not Available.
If all or part of your data are not available to others, why not? (Choose all that apply.)

	Frequency	Percent
Lack of funding	8	21.1
Lack of standards	7	18.4
People don’t need them	10	26.3
There is insufficient time to make them available	11	28.9
There is no place to put them	7	18.4
Sponsor doesn't require it	10	26.3
Don't have the rights to make the data public	10	26.3
I would lose control of the data	10	26.3
I need to publish first	22	57.9
I have insufficient skills to make my data available	6	15.8
Total	38	100.0

Table 4.44: QATAR_Q15. Expectations if Others Use My Data.
The following statements relate to conditions for use of your data. Indicate whether you agree or disagree with each condition. For others to use my data, I would expect the following in exchange:

	Yes	No	Not Sure	No Response	Total %
Co-authorship on publications resulting from use of the data.	68.4	13.2	15.8	2.6	100
Acknowledgement of the data providers in all disseminated work making use of the data.	78.9	13.2	5.3	2.6	100
Citation of the data providers in all disseminated work making use of the data.	92.1	2.6	2.6	2.6	100
The opportunity to collaborate on a project using the data.	65.8	21.1	10.5	2.6	100
Results based (at least in part) on the data could not be disseminated in any format without the data provider's approval.	47.4	26.3	23.7	2.6	100
Results based (at least in part) on the data could not be disseminated without the data provider having the opportunity to review the results and make suggestions or comments, but approval not required.	44.7	31.6	21.1	2.6	100
Reprints of articles that make use of the data must be provided to the data provider.	50.0	34.2	13.2	2.6	100
The data provider is given a complete list of all products that make use of the data, including articles, presentations, educational materials, etc.	55.3	23.7	18.4	2.6	100
Legal permission for data use is obtained.	57.9	18.4	21.1	2.6	100
Mutual agreement on reciprocal sharing of data.	60.5	21.1	15.8	2.6	100
The data provider is given and agrees to a statement of uses to which the data will be put.	55.3	18.4	23.7	2.6	100

Table 4.45: QATAR_Q18. My organization or project provides:

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Training on best practices for data management.	15.8	7.9	15.8	21.1	7.9	31.6	0	100
Assistance on creating data management plans	10.5	13.2	10.5	26.3	7.9	31.6	0	100
Assistance on creating metadata to describe my data or datasets	10.5	13.2	15.8	18.4	5.3	36.8	0	100
Training on how to cite datasets	13.2	18.4	7.9	21.1	5.3	34.2	0	100

Qatar researchers agree somewhat or strongly on most of the items listed in the table below (Table 4.46). It is important for their data to be cited when used by other researchers (86.9%). They are willing to share data across a broad group of researchers (73.7%), and would be willing to place at least some of their data into a central data repository with no restrictions (81%). They would use other researchers' data sets if they were easily accessible (81.5%), and feel it is appropriate to create new datasets from shared data (65.8%). Qatar researchers are satisfied with their ability to integrate data from disparate sources to address research questions (63.1%).

Table 4.47 shows the main reasons indicated by Qatar researchers for not sharing data (using the top two points Agree somewhat and Agree strongly) are that data may be misinterpreted due to the complexity of the data (78.9%), that data may be misinterpreted due to the poor quality of the data (73.7%), and that data may be used in ways other than they were intended (71.1%). Impediments to Qatar researchers' work related to data access (using the top two positions Agree somewhat and Agree strongly) are that a lack of access to data generated by other researchers is a major impediment to progress in science (76.3%), and that a lack of access to other researchers' data has restricted the ability to answer scientific questions (55.3%).

Table 4.46: QATAR_Q19. Sharing Scientific Data.

The following statements relate to sharing scientific data. Tell us how much you agree with each statement.

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
I would use other researchers' datasets if their datasets were easily accessible.	0	13.2	2.6	28.9	52.6	2.6	0	100
I would be willing to place at least **some** of my data into a central data repository with no restrictions.	5.3	7.9	10.5	34.2	36.8	5.3	0	100
I would be willing to place **all** of my data into a central data repository with no restrictions.	10.5	18.4	13.2	26.3	21.1	10.5	0	100
I would be more likely to make my data available if I could place conditions on access.	5.3	15.8	18.4	26.3	28.9	5.3	0	100
I am satisfied with my ability to integrate data from disparate sources to address research questions.	0	7.9	18.4	34.2	28.9	10.5	0	100
I would be willing to share data across a broad group of researchers.	7.9	5.3	5.3	31.6	42.1	7.9	0	100
It is important that my data are cited when used by other researchers.	2.6	0	7.9	21.1	65.8	2.6	0	100
It is appropriate to create new datasets from shared data.	5.3	5.3	13.2	21.1	44.7	10.5	0	100

Table 4.47: QATAR_Q20. Use of Scientific Research Data.
The following statements relate to your views on the use of scientific research data. Tell us how much you agree with each statement.

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Lack of access to data generated by other researchers or institutions is a major impediment to progress in science	7.9	2.6	7.9	39.5	36.8	5.3	0	100
Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.	10.5	7.9	21.1	34.2	21.1	5.3	0	100
Data may be misinterpreted due to **complexity** of the data.	2.6	2.6	10.5	50.0	28.9	5.3	0	100
Data may be misinterpreted due to **poor quality** of the data.	0	2.6	10.5	39.5	34.2	10.5	2.6	100
Data may be used in other ways than intended.	2.6	5.3	5.3	31.6	39.5	15.8	0	100

Table 4.48 indicates that many Qatar researchers feel (per the top two points Agree somewhat and Agree strongly) that they would be more confident in re-using data collected by others under certain conditions. These include if the data were accompanied by written details about collection and quality assurance methods (71.1%), if the metadata standard utilized was explicitly stated with the data (65.7%), and if detailed information about the provenance were available with the data (50%).

When Qatar researchers need data to answer a research question (Table 4.49) they primarily (based on the combination of Agree somewhat and Agree strongly): make a plan to generate or collect the data within their research team (76.9%), generate or collect the data themselves (76.3%), search for data to use for analysis (71.1%), ask colleagues if they know of data they can use for analysis (58.2%), or talk to a librarian about their data needs (44.7%). Many would not consult their data manager (42.1%).

Table 4.48: QATAR_Q21. Re-use of Scientific Research Data.

The following statements relate to your views on the reuse of scientific research data. Tell us how much you agree with the following ways to complete this sentence: I would have increased confidence in re-using data collected by others if…

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
The metadata standard(s) utilized were explicitly stated with the data	0	5.3	7.9	36.8	28.9	15.8	5.3	100
The data were accompanied by written details about collection and quality assurance methods	2.6	2.6	5.3	31.6	39.5	13.2	5.3	100
A recorded workflow from a standard workflow system (Kepler, VisTrails, Taverna, etc) were also available with the data	2.6	2.6	15.8	23.7	13.2	36.8	5.3	100
Detailed information about the provenance (data lineage, chain of custody) were available with data	2.6	2.6	5.3	23.7	26.3	31.6	7.9	100
Other	0	0	2.6	5.3	7.9	23.7	60.5	100

Table 4.49: QATAR_Q22. When I Need Data to Answer a Research Question.

Tell us how much you agree with the following ways to complete this sentence: When I need data to answer a research question, I...

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
Make a plan to generate or collect the data I need **myself**.	0	0	15.8	15.8	60.5	2.6	5.3	100
Make a plan to generate or collect the data I need within my **research team.**	2.6	2.6	10.5	26.3	50.0	2.6	5.3	100
Ask colleagues if they **have data** I can use for analysis.	18.4	7.9	23.7	23.7	18.4	2.6	5.3	100
Search for data to use for analysis	7.9	2.6	7.9	39.5	31.6	5.3	5.3	100
Ask colleagues if they **know of data** I can use for analysis.	10.5	7.9	15.8	39.5	18.4	2.6	5.3	100
Talk to a librarian about my data needs.	21.1	13.2	13.2	26.3	18.4	2.6	5.3	100
Consult my data manager.	36.8	5.3	15.8	13.2	10.5	10.5	7.9	100

Researchers in Qatar occasionally or frequently conduct research with data collected by someone other than a member of their immediate research team (57.9%), while 39.5% indicate they seldom/never use other's data to conduct research (Table 4.50).

Those researchers who indicated they did use data collected by others besides themselves or their immediate research team (Table 4.51), gave the following reasons for doing so (using Agree somewhat and Agree strongly combined). It is easier than having to collect all their own data (78.9%), it saves time (71%), and it helps answer their research questions (68.4%). If data from others was not used, the main reasons given by those researchers were (using Agree somewhat and Agree strongly) that data re-use requires too much trust in others' methods (68.4%), and that it is hard to explain in the methods section (57.9%).

Table 4.50: QATAR_Q23. Data Collected by Someone Else. How often do you conduct research in which some or all of the data analyzed was collected by someone besides yourself or members of your immediate research team? (Choose the one best answer.)

	Frequency	Percent	Cumulative Percent
Never	8	21.1	21.1
Seldom	7	18.4	39.5
Occasionally	14	36.8	76.3
Frequently	8	21.1	97.4
Always	1	2.6	100.0
Total	38	100.0	

Table 4.51: QATAR_Q24. Research With Data Collected by Someone Else.
Tell us how much you agree with the following ways to complete this sentence: Conducting research in which some or all of the data analyzed was collected by others besides myself or members of my immediate research team...

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure	No response	Total %
…saves time	0	2.6	15.8	28.9	42.1	7.9	2.6	100
…is efficient	2.6	7.9	18.4	28.9	34.2	5.3	2.6	100
...is easier than having to collect all my own data for analysis	0	2.6	10.5	36.8	42.1	7.9	0	100
...is hard to explain in methods section	13.2	5.3	15.8	44.7	13.2	7.9	0	100
...requires too much trust in others' methods	2.6	13.2	10.5	36.8	31.6	5.3	0	100
...improves my results	0	7.9	31.6	36.8	15.8	7.9	0	100
...helps me answer my research questions	2.6	2.6	18.4	36.8	31.6	7.9	0	100
...is **harder** than conducting research using only my own data	15.8	18.4	15.8	26.3	15.8	7.9	0	100
...**takes longer** than conducting research with only my own data	18.4	15.8	23.7	21.1	10.5	10.5	0	100

Interview results (Qatar)

Demographics (Qatar)

A total of 10 interviews were conducted in Qatar during the month of April 2018. The basic demographics of this group is presented in Table 4.52 below. Most of the interviews were conducted in person during the time that the researcher travelled to Qatar (n=9), with one interview conducted via phone.

The primary subject disciplines of the Qatar interviewees was similar to those reported in the overall GCC interview data: Medicine (n=3), Engineering (n=3), Physics (n=1), Social Sciences (n=2), and Archaeology (n=1). Most of the researchers were engaged in work at academic institutions (n=9), while one was located at a government research facility. All interviewees spent some of their time on research activities, ranging from 10 to 100%, depending on their additional teaching and administrative duties. The most common proportion of time spent on research was 50 to 70% (a significantly larger percentage than for the GCC researchers as a whole, at 30 to 40%). The Qatar researchers have spent from one to 12 years in their current positions, with an equal distribution of those with less than five years' experience in their position (n=5), and those with greater than 5 years (n=5).

A majority of the Qatar interview subjects were male (n=8), with two being female. Their ages ranged from 41 to 68, with an average of 51 years old. The largest proportion were in their forties (n=6), with the remaining in their fifties or sixties (n=4).

Table 4.52: INQ_1. Demographics of Interview Subjects (Qatar).

	Primary Subject Discipline(s)	Work Sector	% Time spent on Research	Years in current position	Terminal Degree	Gender	Age range
Qatar (n=10)	Epidemiology, Electrical Engineering, Sociology, Microbiology and Immunology, Information Library Studies, Physics, Chemical Engineering, Family Medicine, Energy, Archaeology	Academic (n=1), Government (n=1)	10-100%, with most commonly 50- 70%	1-12yrs, most common 4-8yrs	8 PhD, 1 EdD, 1 MD	2 Female, 8 Male	41-68, majority in 40s

Most researchers obtained their terminal degrees in either the USA or Europe (n=9), plus one who studied in Russia. None were obtained in Qatar. Although these researchers currently work and live in Qatar, none are originally from Qatar (n=0). Instead, several came from countries in the Middle East/North Africa region (n=4). Most came from other countries of origin (n=6), primarily from Western nations. This is a small number of interviews, obtained by a convenience method, but this finding still highlights the current situation that Qatar faces. Qatar has an emerging scientific research endeavor, fueled by experienced researchers primarily from North America/ /Europe, and Arabic-speaking MENA region countries.

All of the researchers indicated that they use English as their primary research collaboration language (n=10), with only a couple of researchers (n=2) occasionally using additional languages (Arabic or French) for collaboration work if needed. All interviewees stated that English was used almost exclusively in their daily work and teaching duties (n=10), with the rare use of Arabic or French (n=6). It was indicated that English was mandated as the primary language by their institutions.

How Data is Stored and Managed (Qatar)

The types of datasets created and used by the Qatari researchers varied, because of the range of data being collected in various disciplines: mapping/GIS data, data models, qualitative interviews/focus group data, ethnographic observational data, audio/video recordings, archival documents, wet lab observational data, experimental data, and instrumentation data.

These datasets can range in size from a few megabytes (MB) up to a few terabytes (TB). When asked if their datasets would be considered "big data," most said no (n=9), with one researcher indicating that some datasets required computationally-complex analysis, but not all. Some researchers indicated they were not sure about the definition of "big data," with one researcher stating, "I don't know how many people actually understand big data. I think it's one of those clichés that people put in research proposals because it's a sexy topic. To me, big data would be if you took lots of broad data to determine causalities (QA02)."

These data are stored in a variety of file formats: structured/unstructured, free text/spreadsheets, and specialized software data format. Many researchers indicated that they used Excel or Access files to structure basic data or text and relational database files, and one used ASCII text files for raw data. Various other file formats were used, depending on the tools used to collect the data, including audio/video files, computational data models, and observational lab data. Commonly-used statistical software includes SPSS, Stata, MATLAB or SAS for survey and observational data. Additional software tools were used for computationally-intensive raw data generated by instrumentation (for big data sets in the hundreds of gigabytes or greater), such as Origin and TensorFlow. Most do not use any metadata standard to describe their data (n=9), while one researcher indicated basic metadata was captured but a standard was not used.

All researchers in Qatar agreed that they are satisfied with their current process of collecting and analyzing their data (n=10), with some indicating they could use help with entering paper data sheets into a database, and obtaining more access to other datasets.

Storage of data is accomplished in several ways, and many researchers use multiple storage and backup methods simultaneously. The most common locations for short-term storage of research data are on the researcher's work laptops/desktops (n= 8), on external hard drives (n=7), on institutional network servers (n= 4), using a cloud storage service (n=2), and paper files (n=2). Some mentioned occasionally saving to a personal laptop at home temporarily, if they happen to be working here. Dropbox is the most commonly-used cloud storage (that is not an institutional network drive).

When backing up data, the same types of storage methods are employed. Most use cloud storage services (n=4), an external hard drive (n=6), an institutional network server (n= 3), their laptops/desktops (n= 1), or paper files (n=1). The popularity of cloud storage can partially be attributed to the ease of access it provides to the stored data from any location. As one interviewee noted, "I can think about it at home at night when family is 'nerding' on their devices (QA04)." The frequency of data backup is not defined for most Qatar interviewees, with many (n=6) not sure how often their data is backed up. Of those who have a regularly-scheduled time interval for data backup, the most common are: instantly/immediately auto-saving to a server or cloud storage (n=1), daily backups (n=1), every two to three weeks (n=1), and monthly (n=1).

When asked if a data management plan (DMP) is in place to guide the short/long-term storage and disposition of their data, most Qatar researchers say that there are no DMP

requirements or that they simply do not know (n=5). Others indicate that the only DMP requirements they have are indicated in their IRB application, and are specific to each proposal (n=4). Only one researcher indicated that some type of data management plan was in place, and this was determined by the individual university department for which the data was archived.

As far as funding agency requirements for data management, most researchers either don't know or specifically indicated that no policies were in place (n=8) to require a data management plan, if funded by a Qatar university or agency. One researcher indicated that the funding agency did require a data management plan, stating, "Yes, absolutely, they are very strict in managing data and reports (QA08)," particularly if the funding was from an external source outside Qatar. One interviewee thought that their Qatar institution did require a DMP.

Why Data is Shared (Qatar)

Most Qatar researchers are willing to share their data (n=9), depending on the situation or details of sharing, but one researcher does not share. When determining who to share with, most will share with anyone (n=7), while a few are more selective in sharing (n=2), by limiting sharing to members of their organization/team, or to those researchers who have the skill set to properly use the data.

The Qatar interviewees mention several **reasons why they would not share** their datasets:

- Fear of their data being stolen prior to publication
- Lack of a data sharing culture

    - "People in the Middle East region are horribly protective of data sets produced locally, resulting in much money spent to collect, and never published (QA01)."
    - "In Qatar research is still young and they are still figuring it out (QA02)."
- Data has commercial / patent potential
    - "If data can potentially become an invention, it could be reason not to publish or share data (QA04)."
- Data may be taken out of context
    - "Afraid the notes would be taken out of context of my original research (QA03)."
- If the requestor has a questionable track record or reputation
- Cannot share before IP agreement is signed
- No time to clean data/remove errors
- IRB limits sharing
    - "Part of the reason I never have [shared] is because my IRB policy prohibits this (QA03)."

Researchers in Qatar identified the following reasons or **incentives for why they would share** their data:

- Data were cited in publications using the data
- Collaborations with others, built trust with Middle East researchers
- To change culture to a data sharing culture
- Scientific compassion – data could make a difference in lives/preventative care of patients

    - "[It is a] very selfish act of a scientist to not share data. Even if I lose intellectual property, sharing knowledge is essential (QA04)."
- Networking with other researchers

Qatar researchers indicated these **benefits to sharing data** with others:

- Validation of data
- Increased research efficiencies (re-use of data)
- Collaboration
- Get advice from reviewers/referees on data during the pre-publication phase
- Benefits society/humanity
- Further regional research (in the Middle East)
    - "Here [in the Middle East] it will benefit everyone to further water pretreatment research (QA07)."
- Contribute my scientific field
    - "If you are a scientist you are not in this field for money, what drives me is a fascination with science, from a hobby when I was younger, and now into a profession. Science data can have an impact on society (QA04)."

These **barriers** to sharing data were identified in Qatar:

- Cultural tendency to <u>not</u> share
- Collaborations require an MOU (memorandum of understanding)
    - "…a MOU has to be in place, even if the different organizations are funded by the same source (Qatar government) - these take time (QA02)."
- Data file structure is difficult for others to use

- IRB restricts sharing sensitive data
- Researchers not familiar with data repositories
    - "Marketing of repositories needs to be done to make researchers familiar with them. Name recognition needed with repositories (QA05)."
- No mechanism in place to share data
- Competition in R&D areas
    - "Also in Qatar, there is a huge drive to protect data if it can lead to commercialization (QA01)."
- Time and resources are scarce for preparing data for sharing

Most researchers agree that sharing data is important for the progress of science; most researchers stated that it is important for their institutions and their local regions of the globe (n=9). They indicated that it is not only a possibility for researchers to share data, but a responsibility, as exemplified in this quote, "Yes, the institutional process of sharing data should be facilitated. Research cannot develop without sharing data (QA09)." Other supporting quotes from Qatar interviewees include:

- "When data becomes available for others to use, it has an exponential impact on scientific progress (QA01)."
- "Absolutely! I think it is important in general to benefit society, especially if funded by the government. We could achieve more efficiency if we all shared data (QA02)."
- "Definitely, progress in general, and especially when you talk about nationally or regionally. Funding resources will be optimized if data is shared, and money is not spent on redundant research (QA07)".
- "Very important to reach the fastest progress in R&D area (QA09)."

Most researchers indicated they are not required to share their data by their funding agencies (n=9), while one researcher did not know (n=1), "Not sure, have not been exposed to this information. I think everyone is doing their own rules (QA07)." The same number of researchers indicated there were no specific policies in place regulating data sharing, and they could make the determination themselves if they wished to share data (n=9), while some had to get prior institutional approval to share (n=1).

When reflecting on additional factors that could have an effect on whether research data were shared or not, one researcher indicated the important role of the government in Qatar, since most research is government funded:

> "This region really needs to push for open access to data. When governments and societies share data, they are more accountable…This applies to all types of data sets. … I think data sharing could definitely have an impact on the progress of science in the GCC and Middle East region, absolutely (QA01)."

Discussion

The main goal of this research has been to understand the people, the scientists, who are engaging in research, and to understand their thoughts and attitudes towards data sharing, and how this affects their current practices surrounding data sharing. Taking a dual approach of analyzing the data for the collective GCC countries and conducting a separate analysis for the country of Qatar, which accounts for the majority of the data from the survey and interviews, allowed for a more nuanced look at the findings. The data for the survey and interviews is segmented into three areas: demographics, how data is stored and managed, and how and why data is being shared in Qatar and the GCC region.

The demographic information helps to characterize and provide some context to understand the researchers at the heart of the study. Most of the survey and interview participants were primarily male, at approximately 75%, with female participants comprising close to 25%. The age of the participants was spread in a normal distribution, ranging from those in their twenties to those over age sixty; with the bulk in their forties.

The majority of participants work in an academic sector (over 90%), and the remainder work for governmental agencies or non-profits. Most survey participants did not define their type of research, but of those that did, approximately equal numbers were engaged in lab and field research (about a fourth for each type), with some also doing modeling research. The primary subject disciplines of the participants ran the gamut of the STEM fields, with engineering and information/computer science making up the largest segments, at just over 12% each. Most participants had obtained a Ph.D. or M.D. (over 82%), with the remainder holding a master's degree.

The majority of survey respondents were from Qatar (~40%) and Saudi Arabia (~32%). There were survey respondents from all six of the GCC countries. Interview participants were from only four of the GCC countries (none were from Bahrain or Oman), with ten out of the 18 interviews with Qatar researchers.

Data Management and Storage

The "how" part of the research question has helped uncover answers to questions regarding data management and storage in the GCC and Qatar. For the GCC, the survey results indicated the majority of researchers (~ 72%) are not required to have a data management plan as part of their research projects, or don't know if one is required; this is echoed by similar results in Qatar

of ~74%. With most funding reported as being supplied by national government agencies in direct ways, or secondarily via the government financing of most universities or research organizations in the GCC (~70%), there is an opportunity for the national governments/funding agencies to effect change in the area of data management policies requirements. Interview participants across the GCC and Qatar overwhelmingly reported the lack of policies governing data management.

Most of the GCC researchers surveyed here store their research data in multiple locations, with most or all of their data being stored on personal computers (~69%) and thumb/external drives (~36%). Most do not seem to prefer data repositories, with over 60% storing none or only some of their data there, and over 68% of researchers preferring not to store their data on departmental servers or paper copies in their office. This survey data is reinforced with interview data from researchers, with many of them facing hectic work schedules and resorting to storing their data in the most convenient locations for daily use. Although survey results do not show a large usage of cloud storage in the GCC (~25% store some or all of their data there), interview data indicates the growing popularity of cloud storage for research data, because of the convenience of accessing it from any Internet-connected location, and the ease of collaboration when sharing with other researchers via this medium.

Most researchers surveyed indicated they are somewhat or strongly satisfied with their processes for storing data during the life of their research projects/short-term (~74%), and ~60% are satisfied with their storage processes after their projects are complete/long-term. Despite these high levels of satisfaction, approximately 69% of researchers indicated their institutions do not have a formal process for managing data during the life of a project/short-term, with slightly high numbers (~75%) reporting their institutions do not have a formal process in place for

storing data beyond the life of the project/long-term. This may indicate the researchers are happy/satisfied with the lack of formal processes surrounding long/short-term data storage. Researchers report that when institutions do assist with data storage (long or short-term), the most likely departments to be involved are research support units (25.5%), information technology support units (14.9%), the library (13.8%), and colleagues in their own unit/department (12.8%). This may present an opportunity for libraries to step in and fill a gap for researchers, and assist them with data storage and data management needs.

spacing

Attitudes Toward Data Sharing

Researchers reported strong opinions and attitudes about some aspects surrounding data sharing of their own data and their re-using other researchers' data sets. When GCC researchers do not make their data available to others, the main reasons they reported for this were: the need to publish prior to sharing the data (~60%), that they don't have the rights to make the data public (~30%), and that their sponsors/funding agencies do not require it (~30%). A lack of time to make the data available, and the researchers' concern with losing control of their data were both reported at a rate of ~25%.

When GCC researchers were willing to share their data, they would expect one or more conditions to be met. Their primary expectation was that attribution must be given in some form: a citation or acknowledgment of the data should be included in all work making use of the data (83% and ~79%, respectively). Researchers also would insist on co-authorship of publications resulting from the use of their data (~65%), and felt that legal permission must be obtained from the researcher prior to use of the data (~54%).

Most GCC researchers are willing to share data with other researchers and they agree somewhat or strongly) on the following points. They are willing to share data across a broad group of researchers (73.4%), they would use other researchers' datasets if they were easily accessible (71.3%), and they would be willing to place at least some of their data into a central data repository with no restrictions (56.4%). GCC researchers are concerned about the re-use of data by others, with 77.6% expressing concern that data may be misinterpreted due to the poor quality of the data or the complexity of the data (76.6%).

Researchers are open to using others' datasets but have some concerns and would want these items addressed, among others (based on those who agree somewhat or agree strongly): data should be accompanied by written details about collection and quality assurance methods (66%), and metadata standards utilized should be explicitly stated with the data (50%).

When researchers need data to answer a research question, they tend to (using agree somewhat and agree strongly) collect the data they need within their own research team (~72%) or collect the data themselves (~74%). They also search for data for analysis (66%), or ask colleagues if they know of data that can be used for analysis (~52%). Only about 50% of the time do GCC researchers conduct research with data collected by someone other than a member of their immediate research team, while ~44% indicate they seldom or never use others' data to conduct research.

GCC researchers do understand that using data collected by others has several benefits, with the following being the primary reported items (per the top two points agree somewhat and agree strongly). Re-using data is easier than having to collect all their own data (71.3%), it saves time (69.1%), is efficient (63.8%), and helps to answer their research questions (66%).

GCC vs. Qatar

A look at the survey data for the GCC region and for Qatar specifically, shows minor differences in most reported data storage, data management and data sharing and re-use practices and attitudes. The research question that drove this dissertation project reflects an exploratory type of research project, and the author hopes the findings sheds some light on the data management and sharing practices of scientists in the region.

Diffusion of Innovation Theory and Data Sharing

Using the framework of the Diffusion of Innovation (DOI) Theory has provided insight into the phenomenon of data sharing in the GCC. The results of this current research can be viewed in context of the four basic elements of the DOI Theory: (1) the innovation, (2) communication through various channels, (3) through time, and (4) in a social system (Rogers 2003).

The innovation can be stated as the practice of exchanging or sharing of new scientific knowledge/information between scientists in the GCC region. This new knowledge/information takes the form of research data or datasets. In this context, the innovation is the act of sharing research datasets. The survey and interview data clearly indicate that researchers in the GCC/Qatar are engaged in the sharing of data.

Communication between various researchers and scientists and funding agencies regarding data sharing seems to be in the beginning stages. Funding agencies in the GCC are just beginning to discuss data management and data sharing polices, which are needed to be part of the structure that governs the sharing of data. Also, scientists around the world are beginning to discuss data sharing as part of efforts to collaborate with other researchers. Although collaboration in science is not new, the increasing amount of data being created, and the

complexity of the data, requires teams of specialists to collaborate to form a functioning team. GCC researchers do not exist in a bubble, and they do communicate regularly with researchers worldwide. During interviews for this research project, scientists have stated that in their communications with researchers outside the GCC and Middle East region, they recognize that data sharing is more advanced than it is in the GCC.

Over the last few years, the research infrastructure of the GCC has advanced dramatically. This is a driver for the further creation of datasets in the region and the need for access to datasets for analysis and to answer fundamental scientific questions. Over time scientists in the region see the need for more dialogue with each other, their funding agencies, and scientific publishers on data management and data sharing. According to DOI Theory, those who adopt a new technology or idea, fall into five categories: innovators, early adopters, early majority, late majority, and laggards. This researcher has observed that those scientists willing to participate in this research project could be considered innovators or early adopters of the data sharing practice. This is likely true based on their mostly positive support of the concept of data sharing, and their desire to see policies and tools put in place to facilitate data management and data sharing.

This diffusion of innovation in the area of data sharing has some unique cultural challenges. The social system of research scientists is part of a larger social system of the GCC economic bloc countries regionally and the Middle East more broadly. The GCC countries do not have a cultural norm of sharing research data, so more communication must be done to overcome this challenge. It was common for those interviewed for this research to note this cultural challenge for the GCC, as it moves toward a more mature understanding and practice of sharing research data. The adoption of the idea of sharing data will necessarily need to be

approached from the top down (policies and research management decisions) and from the bottom up, with scientists connecting and discussing their best practices and concerns for data management and sharing.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The topic of research data sharing has seen an explosion of interest over the last couple of decades, and especially in the most recent decade. This can be seen in the amount of peer-reviewed literature published on the topic of "data sharing" and "data re-use." In tandem with the interest in data sharing, the topic of "big data" has also seen a similar rise. Articles on data sharing of scientific datasets have been published in peer-reviewed journals for decades, but they experienced an exponential growth in publication counts after the topic of big data arrived on the scene in the late 2000s. An exponential growth in research has a corresponding effect on publications and the creation of datasets; this is happening in all developed countries around the world.

The GCC, Qatar and the wider Middle East region is undergoing a dramatic evolution in the capacity to conduct scientific research, with the amount of research and the complexity of the research increasing as the research infrastructure matures. The number of researchers being trained and imported to support the research and societal goals of the individual countries is impressive. This increased scientific workforce will need supportive policies and infrastructure, as well as continued funding.

Although indications are that the GCC and Qatar are just beginning to think about the details of data sharing, this process can be moved ahead by the numerous immigrant workers in the academic and research fields. Many have studied and worked abroad in the USA, Europe and other Western countries and have become acclimatized to more advanced data management concepts and practices during their time at those organizations. This familiarity with best

practices can be influential on the GCC research community. In addition, most GCC researchers have the ability to leverage professional organizations in their area of expertise and to use conference attendance to network with researchers around the world who are willing to assist with improving the data management practices of the region.

This research project has provided a snapshot in time of the data management and data sharing practices of the GCC and Qatar. It is a small study, but the combination of the surveys and interviews has sufficient depth and breadth of findings that can be utilized to inform policies and practices in the region.

Recommendations

From the analysis of the survey data and interview data, several recommendations can be made concerning data sharing in the GCC. The most pertinent are listed here for further consideration:

Data management policies need to be implemented to ensure research data are managed from the inception of a project, through active collection/use/analysis of the data, to the post-publication phase, when it needs to be stored for the long term after a project is completed and made available for re-use. Currently, only about 28% of researchers' organizations have formal processes in place for managing data in the short-term (during the life of a project), and only 22% have formal processes in place for long-term data management (beyond the life of the project). This is slightly higher in Qatar, with approximately 31% for both short and long-term data management

Training in use of metadata standards to describe research data and datasets is needed. This would facilitate proper organization of the data and improve its usefulness for others when it is

re-used or shared. Currently, in the GCC, more than 65% of researchers do not use a metadata standard to describe their datasets. Enforcement of metadata standards usage by funding agencies would improve the use of these standards.

Training on best practices for data management is need. Currently, training is provided by only 21% (based on agree somewhat plus strongly agree) of organizations to their researchers.

Training on the use of data repositories (institutional, publisher and discipline-based) is urgently needed. Researchers in the GCC generally are not aware of the existence of these repositories, nor how to use them for long-term storage of their datasets. Currently less than 15% of researchers use institutional repositories, and the situation is direr with publisher repositories (only a 4% usage rate) and discipline-based repositories (5% usage) to store or deposit datasets. Increased usage of data repositories would ensure curation of datasets for long-term storage and re-use. Polices requiring data to be stored in appropriate repositories need to be implemented as part of a good data management plan to ensure long-term preservation of datasets.

Tools and training for preparing metadata could use some improvement, with only 18% of survey respondents agreeing somewhat or strongly agreeing their organizations provide assistance with creating metadata.

Tools for tracking provenance of data also could use some improvement, with only 45% of survey respondents reporting that they are satisfied with these tools.

Data storage (long-term) processes are not well known by researchers, with 76% of survey respondents indicating that they do not have such processes in place or do not know if these processes exist. Libraries have shown they can assist with this, with 14% of respondents saying they currently work with libraries for long-term storage.

Institutional departments need to get involved with assisting researchers with short/long term storage of their data. Currently, surveyed researchers indicated that only 15% of IT, 10% of administrative offices, 4% of managers, and 13% of colleagues assist in the formal process for storing data.

REFERENCES

Abouammoh, A.M. (2018). Chapter 19: "The Regeneration Aspects for Higher Education Research in the Kingdom of Saudi Arabia" in *Researching Higher Education in Asia History, Development and Future,* edited by Jung jisun, Hugo Horta, and Akiyoshi Yonezawa. DOI 10.1007/978-981-10-4989-7. Accessed at: https://www.researchgate.net/profile/Jisun_Jung5/publication/322171975_Researching_Higher_Education_in_Asia_History_Development_and_Future/links/5c077e10a6fdcc315f9deb27/Researching-Higher-Education-in-Asia-History-Development-and-Future.pdf#page=323

Adams, J. (2011). *Middle East: Exploring the Changing Landscape of Arabian, Persian and Turkish Research*. Evidence.

Al-Maamary, H. M.S., Kazem, H.A., and Chaichan, M.T. (2017). Renewable energy and GCC States energy challenges in the 21st century: A review. *International Journal of Computation and Applied Sciences (*IJOCAAS) Vol. 2, Issue 1, FEBRUARY 2017.

American Geophysical Union (2018). AGU100: Advancing Earth and Space Science. Data Policy FAQ. Accessed at https://publications.agu.org/author-resource-center/publication-policies/data-policy/data-policy-faq/

Arab human development report 2003: Building a knowledge society (AHDR) (2003). N.Y.: UNDP.

Aubert, J.E. & Reiffers, J.L. (2004). *Knowledge Economies in the Middle East and North Africa.* World Bank Report. DOI: 10.1596/978-0-8213-5701-9.

ASTF. (2013). Retrieved from http://www.astf.net/.

Bilsel, A., & Oral, O. (1995). Role of education, science and technology in developing countries. *Frontiers in Education Conference, 1995. Proceedings, 1995* (Vol. 2, pp. 4c4.11-4c4.14

vol.2). Presented at the Frontiers in Education Conference, 1995. Proceedings, 1995, IEEE. doi:10.1109/FIE.1995.483223

Borgman, C. L. (2010). Research data: Who will share what, with whom, when, and why? China-North America Library Conference, Beijing. Retrieved from http://works.bepress.com/borgman/238

Brown, J. R. (1986). Thought experiments since the scientific revolution. *International Studies in the Philosophy of Science*, *1*(1), 1–15.

Buchanan, E. A. (1999). An overview of information ethics issues in a world-wide context. *Ethics and Information Technology*, *1*(3), 193–201.

Campbell, E.G., Clarridge, B.R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N.A., Blumenthal, D. (2002). Data withholding in academic genetics: evidence from a national survey. *JAMA*, 287(4): 473–480.

Cochrane, G., & Atherton, P. (1980). The cultural appraisal of efforts to alleviate information inequity. *Journal of the American Society for Information Science, 31*(4), 283–292.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*, 3rd ed. Sage Publications: Thousand Oaks, California.

Darby, R., Lambert, S., Matthews, B., Wilson, M., Gitmans, K., Dallmeier-Tiessen, S., … Suhonen, J. (2012). Enabling scientific data sharing and re-use. In *2012 IEEE 8th International Conference on E-Science (e-Science)* (pp. 1–8). Presented at the 2012 IEEE 8th International Conference on E-Science (e-Science). doi:10.1109/eScience.2012.6404476

DataONE (2018a). *Scientists Second Follow-Up-Champions* survey. Provided by the University of Tennessee, Knoxville DataONE team, prior to distribution of the survey.

DataONE (2018b). Data Observation Network for Earth. Accessed December 12, 2018 at https://www.dataone.org/.

Department of the Prime Minister and Cabinet: Australian Government (2018). New Australian Government Data Sharing and Release Legislation: Issues Paper for Consultation. Retrieved from: https://www.pmc.gov.au/sites/default/files/publications/australian-government-data-sharing-release-legislation_issues-paper.pdf.

Diab, H. (2016). "The Future of Higher Education in the Middle East," in the Higher Education as a Bridge to the Future: Proceedings of the 50th Anniversary Meeting of the International Association of University Presidents, with Reflections on the Future of Higher Education. Edited by Jason A. Scorza. Rowman & Littlefield.

Diekmann, F. (2012). Data Practices of Agricultural Scientists: Results from an Exploratory Study. *Journal of Agricultural & Food Information*, 13:1, 14-31. DOI: 10.1080/10496505.2012.636005.

Dosa, M. L. (1985). Information transfer as technical assistance for development. *Journal of the American Society for Information Science*, *36*(3), 146-152. doi:10.1002/asi.4630360304

Erickson, M. (2005). Science, Culture and Society: Understanding Science in the Twenty-First Century. Polity Press, Malden, MA, USA.

European Commission (2018). Goals of research and innovation policy: Open Access. Accessed at http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess.

Federer LM, Lu Y-L, Joubert DJ, Welsh J, Brandys B (2015). Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff. *PLoS ONE*, 10(6): e0129506. doi:10.1371/journal.pone.0129506.

Goldemberg (1998). What is the role of science in developing countries? *Science*, Vol. 279, Issue 5354, pp. 1140-1141. DOI: 10.1126/science.279.5354.1140.

Gremm, J., Barth, J., Fietkiewicz, K. J., & Stock, W. G. (2018). *Transitioning Towards a Knowledge Society: Qatar as a Case Study*. Springer International Publishing. Retrieved from //www.springer.com/us/book/9783319711942

Gulf Cooperation Council (2018). Secretariat General of the Gulf Cooperation Council:The Cooperation Council for the Arab States of the Gulf (GCC). Accessed at http://www.gcc-sg.org/en-us/Pages/default.aspx.

Harris, E. (2004). Building scientific capacity in developing countries. *EMBO Reports*, *5*(1), 7-11. doi:10.1038/sj.embor.7400058

Hill, C. E., Loch, K. D., Straub, D. W., & El-Sheshai, K. (1998). A qualitative assessment of Arab culture and information technology transfer. *Journal of Global Information Management*, *6*, 29–38.

Inkeles, A. (1969). Making men modern: On the causes and consequences of individual change in six developing countries. *American Journal of Sociology*, 208–225.

InterAcademy Council. (2004). *Inventing a Better Future*. Amsterdam, Netherlands: InterAcademy Council. Retrieved from http://www.interacademycouncil.net/File.aspx?id=27016

Kuran, T. (2004). Why the Middle East is economically underdeveloped: historical mechanisms of institutional stagnation. *The Journal of Economic Perspectives*, *18*(3), 71–90.

Kedia, B. L., & Bhagat, R. S. (1988). Cultural constraints on transfer of technology across nations: Implications for research in international and comparative management. *The Academy of Management Review*, *13*(4), 559–571.

Kuhlthau, C. (2004). *Seeking meaning: A process approach to library and information services.* 2nd ed. Westport, CT: Libraries Unlimited.

Lewis, B. (2003). *What went wrong?: The clash between Islam and modernity in the Middle East*. Harper perennial.

McCracken, G. (1998). *The long interview*. California: Sage Publications.

Meo, S.A., Hassan, A., Aqil, M. and Usmani, A.M. (2015). *BMC Medical Education*, 15:8. https://doi.org/10.1186/s12909-015-0293-6 Accessed at: https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-015-0293-6.

Michener, W.K. (2015). Ecological data sharing. *Ecological Informatics*, 29: 33–44.

Miles, M.B, Huberman, A.M. and Saldana, J. (2014) Qualitative Data Analysis: A Methods Sourcebook. Thousand Oaks, Californinia : SAGE Publications, Inc.

Miller-Idriss, C. and Hanauer, E. (2011). Transnational higher education: offshore campuses in the Middle East. *Comparative Education*, Vol. 47, No. 2, May 2011, 181–207.

National Aeronautics and Space Administration (2014). NASA Plan for Increasing Access to the Results of Scientific Research. Retrieved from https://www.nasa.gov/sites/default/files/atoms/files/206985_2015_nasa_plan-for-web.pdf.

National Institutes of Health (2015). National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research. Retrieved from https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf.

National Research Council. (2012). The case for international sharing of scientific data: A focus on developing countries. Washington, D.C.: National Academies Press.

National Science Foundation (2018). Dissemination and Sharing of Research Results: NSF Data Sharing Policy. Accessed at: https://www.nsf.gov/bfa/dias/policy/dmp.jsp.

Packer, M. (Sept. 14, 2016). Data sharing: lessons from Copernicus and Kepler. *BMJ*, 354:i4911 doi: 10.1136/bmj.i4911.

Piwowar, H.A., Day, R.S., Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3): e308.

Rogers, E. M., & Rogers, E. (2003). *Diffusion of Innovations, 5th Edition* (5th ed.). Free Press.

Romani, V. (2009). The Politics of Higher Education in the Middle East: Problems and Prospects. *Middle East Brief*, No. 36. Accessed at: http://www.brandeis.edu/globalbrandeis/documents/MEB36.pdf

Saldana, Johnny (2016). The Coding Manual for Qualitative Researchers. SAGE Publications, Ltd., Thousand oaks, California.

Savage, C.J. and Vickers, A.J. (2009). Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. Published: September 18, 2009. Accessed at https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007078. DOI: https://doi.org/10.1371/journal.pone.0007078.

Tal-Socher, M. and Ziderman, A (2017). Data Sharing Policies in Scholarly Publications: Interdisciplinary Comparisons. Committee on Publishing Ethics (COPE), ^Bar-Ilan University, Israel. Retrieved from https://publicationethics.org/files/u7140/Data%20sharing%20poster_2017.pdf.

Tenopir, C., Allard, S., Douglass, K. L., Aydinoglu, A. U., Wu, L., Read, E., … Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS One*, *6*(6). Retrieved from http://works.bepress.com/kimberly_douglass/2/

TWAS. (2013). Retrieved from www.twas.org

UNESCO Science Report 2010: The current status of science around the world. (2010). (2nd ed.). Paris, France: United Nations Educational, Scientific and Cultural Organization. Retrieved from http://unesdoc.unesco.org/images/0018/001899/189958e.pdf

Vasilevsky, N.A., Minnier, J., Haendel, M.A., and Champieux, R.E. (2017). Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ*, April 25, 2017. Accessed at https://peerj.com/articles/3208/.

Vaughan, L. (2001). *Statistical Methods for the Information Professional: A Practical, Painless Approach to Understanding, Using, and Interpreting Statistics*. Information Today, Inc.

Vickery, B. C. (2000). *Scientific communication in history*. Scarecrow Press.

Vogeli C, Yucel R, Bendavid E *et al.* (2006). Data withholding and the next generation of scientists: results of a national survey. *Acad. Med.*81(2),128–136. Retrieved April 8, 2013, from http://journals.lww.com/academicmedicine/Fulltext/2006/02000/Data_Withholding_and_the_Next_Generation_of.5.aspx

Ware, M. and Mabe, M, (2015). "The STM Report: An overview of scientific and scholarly journal publishing" (2015). International Association of Scientific, Technical and Medical Publishers, The Hague, The Netherlands. Accessed at http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1008&context=scholcom.

Weiler, H. N., Guri-Rosenblit, S., & Sawyerr, A. (2006). Universities as Centres of Research and Knowledge Creation: An Endangered Species? Summary Report, Colloquium on Research and Higher Education. UNESCO: Paris, 29 November – 1 December 2006 Retrieved from

http://portal.unesco.org/education/en/files/52646/11736997525Finalreport_EN.pdf/Finalreport_EN.pdf

Wicherts, J.M., Borsboom, D., Kats, J., Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7): 726–728.

World Atlas (2018). Gulf Cooperation Council (GCC) Countries [map]. Retrieved from https://www.worldatlas.com/articles/gulf-cooperation-council-gcc-countries.html.

The World Bank (2018). The World Bank in the Gulf Cooperation Council: Overview. Accessed at http://www.worldbank.org/en/country/gcc/overview.

World Economic Forum (2012). *Arab world competitiveness report 2011-12*. Geneva: World Economic Forum.

Zewail, A. (2006). We Arabs must wage a new form of jihad. *The Independent*, in "Independent Voices", Retrieved from http://www.independent.co.uk/voices/commentators/ahmed-zewail-we-arabs-must-wage-a-new-form-of-jihad-413101.html

APPENDICES

Appendix A – Survey Instrument

Title: Scientists Second Follow-Up-Champions

The link that could be distributed: https://utk.co1.qualtrics.com/jfe/form/SV_77Xl6aXhjhx4PLD

Scientists and research data: Continuing to build an understanding of your data needs
You are invited to participate in an NSF-sponsored research study, in which the DataONE (Data Observation Network for Earth, www.dataone.org) organization is investigating how scientists work. Your responses will help us better understand how scientists manage their data, which will then allow DataONE to better serve their data management needs. The questionnaire should take about 20 minutes to complete. In addition to demographic information, other questions relate to the data management practices of scientists, the data education practices of scientists who are also educators, and finally how your organization and how designated data managers are involved with your research data. As such, no sensitive items are included in our survey, and your participation poses no foreseeable risks other than those one would encounter in everyday life. Also, your responses will be recorded anonymously so that no one can link your responses to you personally. Additionally, these responses will be openly shared via the Internet, as part of recent Office of Science and Technology Policy (OSTP) White House Open Data Policies (2013). This means once responses are anonymized, the data will be openly shared, but only after all possible steps are taken to increase anonymity. Your participation in this research is voluntary, and you may decline to participate without risk. While it is useful to be complete in your responses to the survey, you may skip any questions, and you are free to withdraw from the study at any time. If you have any questions about the study or procedures, please contact Dr. Carol Tenopir (ctenopir@utk.edu) or Dr. Suzie Allard (sallard@utk.edu) of the University of Tennessee. If you have questions about your rights as a participant, contact the Office of the Research Compliance Officer (865) 974-7697. If you would like to keep a copy of this consent statement, you can save or print this page. **By proceeding to the survey I acknowledge that I have read the above statements, I am 18 years old or older, and I agree to participate.**

First, we would like to ask you a few questions about yourself.

Q2 Which one of the following best describes your primary work sector?

- Academic
- Government
- Commercial
- Non-Profit
- Other

Q2_5_Text If other, please specify:

_____

Q3 Which one of the following best describes your primary subject discipline?

- Agriculture and Natural Resources
- Atmospheric science
- Biology
- Business
- Computer science
- Ecology
- Education
- Engineering
- Environmental science
- Geology
- Hydrology
- Information science
- Law
- Medicine
- Physical sciences
- Psychology
- Social sciences
- Other

Q3_Other_text If you selected other, please specify

_____

Q4 Which of the following best describes your primary type of research activity:

- Field Research
- Lab Research
- Modeling
- GIS
- Survey Research
- Secondary data analysis
- Other (specify)_____

Q5 Which of the following countries is your primary place of employment?

▼ Afghanistan ... Zimbabwe

Q6 Year of birth – Please select from the list below.

▼ 1905 ... 1998

Q7 Year highest degree was attained:

_____

Q8 Gender
- Male
- Female
- Other
- Prefer not to answer

The next group of questions asks about your familiarity with DataONE

Q_D1-1 Before this survey, had you heard of the NSF-funded DataONE project?
- Yes **If yes, skip to Q_D1-3 "Which of the following …"**
- No
- Don't Recall

Q_D1-2
How much do you know about DataONE on the following scale from 0 (nothing) to 5 (a great deal)?

	0 (Nothing)	1 (Very little)	2	3 (Some)	4	5 (A great deal)
How much do you know about DataONE?	○	○	○	○	○	○

Q_D1-3 Which of the following DataONE products and/or services have you used or participated in? (Choose all that apply.)
- Educational materials **(if selected, skip to Q_D1-3-1R "Educational materials")**
- Data search functions **(if selected, skip to Q_D1-3-2R "Data search functions")**
- General information on the web site **(If selected, skip to Q_D1-3-3R "general information")**
- Data deposit **(If selected, skip to Q_D1-3-4R "data deposit")**
- Data analysis tools **(If selected, skip to Q_D1-3-5R "data analysis tools")**
- Data description tools **(If selected, skip to Q_D1-3-6R "data description tools")**
- Data Management Plan Tool (DMP Tool) **(If selected, skip to Q_D1-3-7R "data management plan tool")**
- Data citation information **(If selected, skip to Q_D1-3-8R "data citation information")**

Q_D1-3-1R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Educational materials	○	○	○	○	○

Q_D1-3-2R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data search functions	○	○	○	○	○

Q_D1-3-3R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
General information on the web site	○	○	○	○	○

Q_D1-3-4R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data deposit	○	○	○	○	○

Q_D1-3-5R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data analysis tools	○	○	○	○	○

Q_D1-3-6R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data description tools	○	○	○	○	○

Q_D1-3-7R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data Management Plan Tool (DMP Tool)	○	○	○	○	○

D_D1-3-8R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data citation information	○	○	○	○	○

Q_D1-4
Which of the following DataONE products and/or services have you used or participated in? (Choose all that apply.)

- o General information on the web site (About DataONE, Working Groups, Member Nodes) **(If selected, skip to Q_D1-4-1R "general information on the website")**
- o DataONE educational materials (webinars, education modules, best practices, workshop/training) **(If selected, skip to Q_D1-4-2R "DataONE educational materials")**
- o DataONE Search **(If selected, skip to Q_D1-4-3R "DataONE search")**
- o DataONE MatLab Client **(If selected, skip to Q_D1-4-4R "DataONE MatLab client")**
- o DataONE R Client **(If selected, skip to Q_D1-4-5R "DataONE R client")**
- o Data Management Plan Tool (DMP Tool) **(If selected, skip to Q_D1-4-6R "Data management plan tool")**

D1-4-1R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
General information on the website	○	○	○	○	○

D1-4-2R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
DataONE educational materials	○	○	○	○	○

D1-4-3R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
DataONE Search	○	○	○	○	○

D1-4-4R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
DataONE MatLab Client	○	○	○	○	○

D1-4-5R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
DataONE R Client	○	○	○	○	○

D1-4-6R How useful was this service on a scale of 1 (poor) to 5 (excellent)?

	1 (Poor)	2	3	4	5 (Excellent)
Data Management Plan Tool (DMP Tool)	○	○	○	○	○

QD1-5 Are any of your data available through DataONE or through a member node?

- ○ Yes
- ○ No
- ○ Don't know

The next group of questions refers to your research and research data.

Q9 Does your primary funding agency require you to provide a data management plan?

- ○ Yes
- ○ No
- ○ Don't Know

Q10 What metadata standards do you currently use to describe your data, if any? (Choose all that apply.)

- DC (Dublin Core)
- DwC (Darwin Core)
- DIF (Directory Interchange Format)
- EML (Ecological Metadata Language)
- FGDC (Federal Geographic Data Committee) CSDGM (content standard)
- ISO 19115.xx (Geographic Information – Metadata)
- Other ISO metadata standard
- OGIS (Open GIS)
- ANZLIC metadata profile
- Net CDF
- Metadata standardized within my institution
- Metadata standardized within my lab
- None
- Other

Q10_Other If you selected other, please specify

_____

Q11 Which of the following best describes the primary funding agency for your research?

- Federal/national government
- State/regional government
- Local government
- Corporation
- Private foundation
- Internal/my institution
- Other

Q11_Other If you selected other, please specify

_____

Q12 How much of your data do you **currently store or deposit** in the following locations? (For each location, choose only the one best answer.)

	None of my data	Some of my data	Most of my Data	All of my data	Not sure
On my institution's server	○	○	○	○	○
On the principal investigator's server	○	○	○	○	○
On a departmental server	○	○	○	○	○
On my personal computer	○	○	○	○	○
On paper in my office	○	○	○	○	○
Thumb/external drive	○	○	○	○	○
In a discipline-based repository, (e.g. NEON or LTER)	○	○	○	○	○
In a publisher or publisher-related repository (e.g., specific publisher or Dryad)	○	○	○	○	○
Other data repository or archive (e.g., national data center)	○	○	○	○	○
In my institution's repository	○	○	○	○	○
Cloud storage	○	○	○	○	○
Other	○	○	○	○	○

Q12_Other If you selected other, please specify

_____

Q13 If all or part of your data are not available to others, why not? (Choose all that apply.)

- Lack of funding
- Lack of standards
- People don’t need them
- There is insufficient time to make them available
- There is no place to put them
- Sponsor doesn't require it
- Don't have the rights to make the data public
- I would lose control of the data
- I need to publish first
- I have insufficient skills to make my data available
- Other

Q13_Other If you selected other, please specify

_____

Q14 The following statements relate to how you store and manage your data. Tell us how much you agree with the following ways to complete this sentence:
I am satisfied with the…

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not Sure
...process for storing my data during the life of the project (short-term).	○	○	○	○	○	○
…process for storing my data beyond the life of the project (long-term).	○	○	○	○	○	○
…tools for preparing metadata	○	○	○	○	○	○
... ability to track & verify provenance information	○	○	○	○	○	○
...ease of locating a suitable repository for the deposit of data	○	○	○	○	○	○

Q15 The following statements relate to **conditions for use of your data**. Indicate whether you agree or disagree with each condition.
For others to use my data, I would expect the following in exchange:

	Yes	No	Not Sure
Co-authorship on publications resulting from use of the data.	○	○	○
Acknowledgement of the data providers in all disseminated work making use of the data.	○	○	○
Citation of the data providers in all disseminated work making use of the data.	○	○	○
The opportunity to collaborate on a project using the data.	○	○	○
Results based (at least in part) on the data could not be disseminated in any format without the data provider's approval.	○	○	○
Results based (at least in part) on the data could not be disseminated without the data provider having the opportunity to review the results and make suggestions or comments, but approval not required.	○	○	○
Reprints of articles that make use of the data must be provided to the data provider.	○	○	○
The data provider is given a complete list of all products that make use of the data, including articles, presentations, educational materials, etc.	○	○	○
Legal permission for data use is obtained.	○	○	○
Mutual agreement on reciprocal sharing of data.	○	○	○
The data provider is given and agrees to a statement of uses to which the data will be put.	○	○	○

We are also interested in how your organization is involved with research data.

Q16 The following statements relate to how your organization is involved with **managing and storing data**.

Tell us how much you agree with the following ways to complete this sentence:

My organization has a formal process for:

	Yes *	No	Don't know
...managing data **during** the life of the project (short-term).	○	○	○
...storing data **beyond** the life of the project (long-term).	○	○	○

***If yes is selected, skip to Q17**

Q17 You have expressed agreement that your organization or project has a formal process for managing or storing data during or beyond the life of the project (short-term or long-term). Which of the following are involved with this process? (Choose all that apply.)

- ○ Research support unit(s) (e.g. Office of Research, Office of Sponsored Programs and Contracts)
- ○ The library
- ○ Information technology support unit(s) (e.g. Office of Information Technology, IT Support Center)
- ○ Administrative office(s) (e.g. Department Heads, Deans, Provosts, Program Offices, Research Offices, Divisions, Directorates / Directors, Managers)
- ○ Designated data manager(s)
- ○ Colleagues in my own unit / department
- ○ Other

Q17_Other If you selected other, please specify

_____

Q18 My organization or project provides:

	Disagree Strongly	Disagree Somewhat	Neither agree nor disagree	Agree Somewhat	Agree Strongly	Not sure
...training on best practices for data management.	○	○	○	○	○	○
...assistance on creating data management plans.	○	○	○	○	○	○
...assistance on creating metadata to describe my data or datasets.	○	○	○	○	○	○
...training on how to cite datasets.	○	○	○	○	○	○

Next we would like to ask you about your views on data use and reuse in science.

Q19 The following statements relate to **sharing scientific data**. Tell us how much you agree with each statement.

	Disagree Strongly	Disagree Somewhat	Neither agree nor disagree	Agree somewhat	Agree Strongly	Not sure
I would use other researchers' datasets if their datasets were easily accessible.	○	○	○	○	○	○
I would be willing to place at least some of my data into a central data repository with no restrictions.	○	○	○	○	○	○
I would be willing to place all of my data into a central data repository with no restrictions.	○	○	○	○	○	○
I would be more likely to make my data available if I could place conditions on access.	○	○	○	○	○	○
I am satisfied with my ability to integrate data from disparate sources to address research questions.	○	○	○	○	○	○
I would be willing to share data across a broad group of researchers.	○	○	○	○	○	○
It is important that my data are cited when used by other researchers.	○	○	○	○	○	○
It is appropriate to create new datasets from shared data.	○	○	○	○	○	○

Q20 The following statements relate to your views on the **use of scientific research data**. Tell us how much you agree with each statement.

	Disagree Strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure
Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.	○	○	○	○	○	○
Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions.	○	○	○	○	○	○
Data may be misinterpreted due to complexity of the data.	○	○	○	○	○	○
Data may be misinterpreted due to poor quality of the data.	○	○	○	○	○	○
Data may be used in other ways than intended.	○	○	○	○	○	○

Q21 The following statements relate to your views on the **reuse of scientific research data**. Tell us how much you agree with the following ways to complete this sentence: I would have increased confidence in re-using data collected by others if…

	Disagree Strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure
The metadata standard(s) utilized were explicitly stated with the data	○	○	○	○	○	○
The data were accompanied by written details about collection and quality assurance methods	○	○	○	○	○	○
A recorded workflow from a standard workflow system (Kepler, VisTrails, Taverna, etc) were also available with the data	○	○	○	○	○	○
Detailed information about the provenance (data lineage, chain of custody) were available with data	○	○	○	○	○	○
Other	○	○	○	○	○	○

Q21_Other If you selected other, please specify

____

Q22
Tell us how much you agree with the following ways to complete this sentence:

When I need data to answer a research question, I...

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure
...make a plan to generate or collect the data I need myself.	○	○	○	○	○	○
...make a plan to generate or collect the data I need within my research team.	○	○	○	○	○	○
...ask colleagues if they have data I can use for analysis.	○	○	○	○	○	○
...search for data to use for analysis.	○	○	○	○	○	○
...ask colleagues if they know of data I can use for analysis.	○	○	○	○	○	○
...talk to a librarian about my data needs.	○	○	○	○	○	○
...consult my data manager.	○	○	○	○	○	○

Q23 How often do you conduct research in which **some or all of the data analyzed was collected by someone besides yourself or members of your immediate research team**? (Choose the one best answer.)

- Never
- Seldom
- Occasionally
- Frequently
- Always

Q24 Tell us how much you agree with the following ways to complete this sentence:

Conducting research in which some or all of the data analyzed was collected by others besides myself or members of my immediate research team...

	Disagree strongly	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Agree strongly	Not sure
...saves time	○	○	○	○	○	○
...is efficient	○	○	○	○	○	○
...is easier than having to collect all my own data for analysis	○	○	○	○	○	○
...is hard to explain in methods section	○	○	○	○	○	○
...requires too much trust in others' methods	○	○	○	○	○	○
...improves my results	○	○	○	○	○	○
...helps me answer my research questions	○	○	○	○	○	○
...is harder than conducting research using only my own data	○	○	○	○	○	○
...takes longer than conducting research with only my own data	○	○	○	○	○	○

Appendix B – Interview Instrument

INTERVIEW QUESTIONS

Interview#: _____ (alpha/numeric code)

Introduction:
This interview is to help understand how and why research data are currently being shared by scientists in the Persian Gulf states of the GCC.

Demographic Info

1. Country (where you currently live/work)
    - Is this your country of origin?
2. Job Description
    - What is your primary/secondary subject discipline?
    - Which work sector best describes this job? Academic, Government, Commercial, Non-Profit, Other?
    - Is your time primarily dedicated to research? Other?
    - How long have you held this position?
3. Which terminal Degree (s) do you hold?
    - Year(s) completed
    - Country where Degree(s) were completed
4. Languages you speak in addition to English?
   (have at least basic speaking/reading/writing proficiency)
    - Primary language you use at work?
    - Primary language you use for research collaboration with others?
5. Gender
6. Age

Data Sharing – How

7. Can you describe the types of data sets you work with/produce in your research?
    - Types of data sets (Abiotic, biotics, data models, experimental, observational, social science data (surveys/interviews, etc.), Other)?
    - File formats (structured/unstructured, free text/spreadsheet, specialized software data format)?
    - Do you capture Metadata? Use any standards?
8. What size/complexity are your data?
    - Instrumentation used to collect data?
9. Discuss your satisfaction with your current process of collecting/analyzing your data?
    - Proper equipment/tools?
10. Where do you store your data?

- (networked computer system, laptop/desktop, organizationally owned/personal storage service, repositories, other)
- How much is stored in each of the above locations?
- Do you backup your data? Where? How often?

11. Do you have a data management plan in place for your data?
    - Requires Long-term preservation/sharing/re-use?
    - Required by your funding agency?

Data Sharing – Why

12. What are your thoughts on sharing your research data with others?
    - No one, your research team, members of your organization, broadly in your discipline, funding organizations, scientific community at large, other
13. If you do not share, why not?
14. What would be an important incentive(s) for you to share?
15. What do you think are the benefits of sharing data?
16. What are some of the barriers for you to share / re-use data?
17. Do you think sharing data is important for the progress of science?
    - Important for your country/region?
    - Important to your institution?
    - Important to your discipline/research?
18. How do you see your research data in relation to "big data" ?
19. Does your sponsor/funding agency require you to share your data?

20. Do you have the sole authority to share your data? Requires institutional approval? Other approvals?
21. Additional comments?

VITA

Jim Malone was born August 13, 1966 in rural middle Tennessee, on the Upper Cumberland Plateau in Fentress County, USA. His parents (Elmer and Geneva Malone) did not come from an academic background, being from poor farming communities, and never had the opportunity to complete elementary school, but they always encouraged and wished their 11 children to have a better life through the pursuit of education. Jim completed his bachelor's degree in General Studies (liberal arts, with a technology studies emphasis) in 2003 at East Tennessee State University, and immediately followed this up with his master's in Information Sciences at the University of Tennessee, Knoxville. He then worked in higher education libraries in Kabul, Afghanistan and Doha, Qatar, before returning to Tennessee to pursue his Doctoral studies, while helping care for an aging parent. He began his doctoral course work in 2010, and finally completing his dissertation research and defense in 2018, with a formal graduation ceremony in May 2019.