Open Data Management Plan
Middle East and North Africa
A Guide

Nagla Rizk, Youmna Hashem, Nancy Salem

Briefing Note

Published: 17 October 2018

TABLE OF CONTENTS

Authors

Dr. Nagla Rizk
Founding Director
Access to Knowledge for Development Center (A2K4D)
Professor of Economics, School of Business
The American University in Cairo (AUC)
Principal Investigator, Open Data for Development Node Middle East and North Africa (ODMENA)

Youmna Hashem
Researcher
Access to Knowledge for Development Center (A2K4D)
School of Business
The American University in Cairo (AUC)
Open Data for Development Node Middle East and North Africa (ODMENA)

Nancy Salem
Senior Researcher
Access to Knowledge for Development Center (A2K4D)
School of Business
The American University in Cairo (AUC)
Open Data for Development Node Middle East and North Africa (ODMENA)

Acknowledgments

This work was carried out under the auspices of the Open Data for Development (OD4D) Network, as the regional node for Middle East and North Africa.

We acknowledge the support provided for this research by the OD4D Network supported by the International Development Research Centre (IDRC). The views expressed in this work are those of the authors and do not necessarily represent those of the research funders.

The authors are grateful for the guidance, expertise and input of Sherif El Kassas, Professor of Computer Science at the American University in Cairo, and Ahmed Hussien, IT expert and consultant, in developing this Open Data Management plan. The authors further express gratitude to Sarah El Saeed and Hussein Mahfouz from the Access to Knowledge for Development Center for their contribution to the research.

INTRODUCTION

This guide consists of a set of three documents developed by the Access to Knowledge for Development team at the American University in Cairo. The first - a background paper - delves into open data as it relates to the fields of research and development and introduces open data management plans and their key characteristics. The background paper sets out to outline the most important concepts that relate to data management, security, storage, and long-term preservation, and discusses the Middle East and North Africa region's legal framework that researchers dealing with data should operate within. The second document - an open data management plan template - consists of a set of questions that, when answered, constitute the Open Data Management Plan (ODMP). This template was developed by Ahmed Hussien and the team at A2K4D, and prompts those filling it out to consider the key concepts introduced in the background document in the ways that they relate to their research. The questions were developed with local context in mind and are intended to help guide and encourage researchers to consider various contingency plans that may arise when dealing with open data in the regional context. The third and final document - the Solar Data Platform ODMP - is the ODMP filled out by the researchers at A2K4D for the center's own open data platform. The initiative - which mapped out the solar energy sector in Egypt and made that data open and available online - was used to showcase a real example of how the ODMP can be filled out.

BACKGROUND PAPER

LIST OF ABBREVIATIONS

Internet of Things	IoT
Middle East and North Africa	MENA
Open Data Management Plan	ODMP
Open Knowledge Foundation	OKF
Australian National Data Service	ANDS

I. A Primer on Open Data

Digital technologies are present in almost all aspects of human activity, including many economic, social and political processes. This digitization has resulted in the generation of an unprecedented amount of data - as of 2017, there were a recorded 4.9 billion mobile users around the world, contributing to an average of 8 exabytes (a unit of measurement equating to billions of gigabytes) of data traffic per month.[1] Recent developments in the field of technology have resulted in the evolvement of what is now commonly referred to as the Internet of Things (IoT). Used to describe the emergence of smart devices - physical electronic devices that connected to the Internet – the IoT bridges the gap between the physical world and the information world.[2] This increased connectivity has undoubtedly contributed to the exponential rise in the amount of data being generated daily.

Data - defined as a collection of information comprised mainly of facts and statistics[3] - is considered by many one of the world's most abundant and powerful resources.[4] In his book *The Data Revolution,* author and academic Rob Kitchin describes the power and utility of data as lying in their ability to provide valuable insights into a multitude of various facets of human activity, "which in turn are used to create innovations, products, policies and knowledge that shape how people live their lives." [5] Fields of research dedicated to the conceptualization of data have grown exponentially over the past couple of decades, and as a result the concept of *open data* emerged.

The open data movement looks to break down barriers to access to knowledge and information by advocating the free and open dissemination of data. According to the Open Knowledge Foundation (OKF), for data to be considered *open* it must be made easily accessible, usable, and distributable, free of restrictions "beyond a requirement for attribution."[6]The OKF emphasizes the importance of ensuring that the dataset or body of work is made available through a widely accessible medium (preferably the Internet), that it is available in its entirety, and that its format is open or can be opened using free open source software – going on to state that "any further restrictions make an item closed knowledge."[7]In her article on the uses of open data in research, Jennifer C. Molloy stresses the key role that open access plays in maximizing on the utility and impact of data. She explains that in breaking down barriers to

---

[1] Simon Kemp, "The global state of the internet in April 2017," *The Next Web,* April 11, 2017, https://tnw.to/2nZHhVb.
[2] Hermann Kopetz, "Internet of Things," in *Real-Time Systems,* (Boston, MA: Springer, 2011), 307-323, https://doi.org/10.1007/978-1-4419-8237-7_13.
[3] *Oxford Dictionaries,* "data," accessed March 31, 2018, https://en.oxforddictionaries.com/definition/data.
[4] Bob Picciano, "IBMVoice: Why Big Data Is The New Natural Resource," *Forbes,* June 30, 2014, https://www.forbes.com/sites/ibm/2014/06/30/why-big-data-is-the-new-natural-resource.
[5] Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (SAGE, 2014), 1.
[6] "Open Definition 2.0," Open Definition, accessed March 31, 2018, https://opendefinition.org/od/2.0/en/.
[7] Ibid.

access and usage of information, open data contributes to the development and growth of a well-informed, well-researched body of knowledge that builds on a wealth of data.[8]

With data proliferation taking place at such a large scale all around the globe, questions of the future of data as a commodity arise. In his article on privacy and property, Paul M. Schwartz delves into the implications that the Internet has on users' individual data and their privacies. Schwartz discusses the growing capacity the Internet has for gathering large quantities of data, and the subsequent emersion of a market place in which this data is a commodity. The commodification of data has proven to be increasingly profitable. When users opt into free services, they may authorize the service provider to collect and often monetize their data. [9] The harvesting and commodification of personal information has recently come under public scrutiny. A recent example is that of Cambridge Analytica, in which it was revealed that an estimated 50 million Facebook users' data was collected and utilized by a data analytics firm. In what is being called by some critics one of the largest breaches of Facebook users' trust, the software attempted to systematically predict and influence users' political choices through highly targeted advertisements.[10]

Whistleblowers and policy-makers are calling for stricter regulations concerning the collection and control of data, emphasizing the need for more proactive measures of data collection and management.[11] Bruce Schneier, security technologist and cryptographer, argues that surveillance is the current business model of the Internet. Citing the constant tracking and monetization of users and their data – oftentimes without their knowledge, and sometimes without their consent – Schneier calls into question the future of data collection and management on the Internet, and the role that large businesses play in exploiting personal information for commercial gain.[12] Although Schneier believes that surveillance is prevalent across the Internet, he also argues that data is starting to lose its commercial value. Anticipating a crash in the surveillance capitalism market, Schneier suggests that data may lose its value to abundance, arguing that the more data is generated and collected, the less each piece of data will start to be worth.[13] Given the proven record of Internet tracking and surveillance , a conscious awareness and understanding of the current business model of the Internet and its possible repercussions is integral. This holds especially true in the Middle East and North Africa (MENA) region

---

[8] Jennifer C. Molloy, "The Open Knowledge Foundation: Open Data Means Better Science," *PLoS Biology* 9, (December 2011): https://doi.org/10.1371/journal.pbio.1001195.
[9] Ben Kepes, "Google Users - You're The Product, Not The Customer," *Forbes,* December 4, 2013, https://www.forbes.com/sites/benkepes/2013/12/04/google-users-youre-the-product-not-the-customer.
[10] Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach," *The Guardian*, March 17, 2018, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.
[11] The Economist, "The world's most valuable resource is no longer oil, but data - Regulating the internet giants," *The Economist*, May 6, 2017, https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.
[12] Bruce Schneier, "'Stalker economy' here to stay," *CNN*, November 26, 2013, https://edition.cnn.com/2013/11/20/opinion/schneier-stalker-economy/index.html.
[13] Hal Berghel, "Bruce Schneier on Future Digital Threats," *Computer* 51, no. 2 (February 2018): 64-67, https://doi.org/10.1109/MC.2018.1451653.

and the rest of the developing world, where there are fewer regulations for the control and protection of individual's private data.

Data management is therefore an important aspect of promoting open data without infringing on data rights, whilst maintaining its fair use. In their article on the FAIR Data Principles - discussed in further detail in the coming section of this guide - Mark D. Wilkinson et al. argue for the necessity of an infrastructure to be put in place in order to support the discovery, utility and reproducibility of data that is made open. Such an infrastructure should take into account the key aspects of open data and should be seen not as "a goal in itself, but rather the key conduit leading to knowledge discovery and innovation."[14] This point is further emphasized by Molloy who discusses the need for infrastructure that supports the wider implementation of open data.[15] Not only is good data management seen as a means through which to ensure the structured integration of data and knowledge into the wider digital universe, it is also seen as highly beneficial in helping individuals, academics, researchers, organizations, governments etc. maximize on and ensure the efficiency of their knowledge databases.

II. What is an Open Data Management Plan?

In recent years, several open data management plans (ODMP) have been developed. Simply put, an ODMP is a living document that consists of a set of questions relating to the data that is being collected, stored, and disseminated. It is intended to be filled out by the responsible individual and acts primarily as a guide for facilitating the management of data on any given project. Through its Q&A format, the guide prompts the individual filling it out to consider a variety of issues pertaining to the data, ranging from questions of ownership and custodianship to considerations of required equipment and facilities. Although not by any means exhaustive, the plan acts as a basis through which researchers and individuals can begin thinking about the ways in which they wish to structure their data, the timeline of their projects, and the preservation of their data over the long run.

Further, the plan is intended to encourage researchers to think about the type of data they plan on availing to the public, the necessity of the act, and any implications doing so might have on individuals or the community at large. Whilst there is a due tendency to gravitate towards enabling access at a wide scale, this plan prompts individuals to think critically about the data they are handling and the different restrictions they may wish to place in order to protect individual privacy or maintain the integrity of their data. This ODMP has been developed by the Access to Knowledge for Development team at the American University in Cairo and sets out to localize existent ODMPs and contextualize them to the MENA region.

III. Why is Important to Develop a Plan?

[14] Wilkinson, M. D. et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific. Data* 3, (2016), https://doi.org/10.1038/sdata.2016.18.
[15] Molloy, "Open Knowledge Foundation."

There are a diverse range of benefits that individuals and organizations stand to gain when developing an ODMP. According to the Australian National Data Service (ANDS), the creation and implementation of an ODMP helps to eliminate errors throughout any data project and minimizes the amount of time spent on containing damage should anything go wrong with the data. In other words, it puts into place a robust and structured system of accountability, efficiency, and management in order to ensure the smooth execution of data projects.[16] The ODMP can be considered a ledger through which all information relating to the data is stored and can be referred back to at any point in time. Aside from maximizing efficiency in the way that the data is stored and handled, an ODMP is structured in a way that allows for greater collaboration between organizations and individuals. It ensures that there is some form of "continuity [should] project staff leave, or new researchers join," and aids in ensuring that the decision-making process throughout the project takes into account "the wider context and consequences of different options."[17] If the plan is executed correctly, it helps to map out the greater structure of the data and how it relates to the project in a way that ensures greater visibility and greater impact.[18] Ultimately – the plan is beneficial to the overall efficiency and impact of the project.

Given the versatility of projects across different disciplines, each plan will undoubtedly be unique to its project. That being said, there are a set of concepts that are considered a cornerstone for any open data project, and these concepts should be integrated into open data management plans. These concepts emerge from FORCE11's FAIR Guiding Principles. FORCE11 is a community comprised mainly of academics, scholars, and researchers, created with the aim of improving access to knowledge and knowledge sharing.[19] FAIR - an acronym standing for ***F****indable,* ***A****ccessible,* ***I****nteroperable and* ***R****eusable* - is a framework that sets out to enable researchers and organizations to maximize the utility and reuse of their data. Core to these principles is the importance placed on ensuring the data being stored and managed is machine-readable, allowing for maximized utility, visibility, and the effective analysis of multiple datasets.[20]

[16] ANDS Guide, "Data management plans," Australian National Data Service, October, 2017, http://www.ands.org.au/guides/data-management-plans.
[17] DCC, "Checklist for a Data Management Plan, v4.0," Edinburgh: Digital Curation Center, 2013, http://www.dcc.ac.uk/resources/data-management-plans.
[18] Ibid.
[19] "About FORCE11," FORCE11, accessed March 31, 2018, https://www.force11.org/about.
[20] "The FAIR data principles: Working with data," Australian National Data Service, accessed March 31, 2018, https://www.ands.org.au/working-with-data/fairdata.

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Fig. 1: The FAIR Guiding Principles. Table from The FAIR Guiding Principles for scientific data management and stewardship publication.[21]

IV. Data Storage and Security

Although rapid digitization occurs around the world, gaps in access and use remain. According to research carried out by the World Wide Web Foundation, the UN Sustainable Development Goal of ensuring equal access to technology by 2030 is - based on current socioeconomic trends - unattainable before the year 2042, highlighting the severity of the digital divide between countries across the world.[22] In the MENA region, many challenges remain to achieving equitable and sustainable access to technologies. When developing programs and projects, it is important to keep these limitations in mind and to have an understanding of the available local capacities. The following section will highlight key points relating to data storage and security, focusing on the aspects that are unique to the region.

a. Data Formats

There are various different ways to categorize and classify data. Generally, data is classified based on whether it was collected or generated, qualitative or quantitative, and structured or non-structured. Depending on this classification, the data's file format is likely to be determined, affecting the sections of the ODMP that relate to structure, storage and data preservation. While data comes in various

[21] Wilkinson et al., "FAIR Guiding Principles."
[22] Web Foundation, "Closing the Digital Divide: A Briefing Note," *World Wide Web Foundation,* April 14, 2016, https://webfoundation.org/2016/04/closing-the-digital-divide-a-briefing-note/.

different forms, the most common ones tend to be text, databases, images, videos and audios. Amongst the considerations during the preliminary stages of data collection is the data's initial file format, and whether that format will change or remain the same when the data is made open. In keeping with the open data standards as well as the FAIR principles, the best methods of practice when choosing file formats would be choosing non-proprietary formats (such as .cmv for database files), open source software, or software commonly used by the surrounding community.[23]

b. Data Storage

Depending on the type and volume of data an individual or company is working with, different storage options might need to be taken into consideration. When dealing with data that is large in volume, such as videos and photos, storage requirements might go up to hundreds of terabytes. Although terabyte drives are more commercially available today than they were a decade ago, they can be costly. Barring the cost barrier - whilst having data on external drives might prove to be beneficial for some - the management of several external drives can prove to be quite challenging. Ultimately, having several drives will entail a greater degree of management and logging than having all the data in a centralized location. For researchers and their entities who handle large volumes of data, but lack the storage capacity, there exist storage data center facilities that provide storage. Such data centers exist in Egypt, allowing individuals or entities the opportunity to rent servers or storage spaces at a monthly fee. These centers - amongst which are Link, TEData, Egypt Cyber Center, to name a few - house the necessary infrastructure and power capacity to provide a large amount of storage and server space for individuals and companies.

According to Ahmed Hussien, IT expert and consultant, taking this route has a number of advantages and disadvantages that need to be carefully considered. In terms of storage space, accessibility, and access to the internet, utilizing a data center's facilities is the best option. These data centers are equipped with the necessary infrastructure required for the upkeep of large servers, some of which are virtual and require a strong connection to the Internet. By renting out a space - be it virtual or physical - at one of these centers, individuals or companies are able to centralize large volumes of data in one place, easing the management process.

This option - whilst convenient and reliable - is not the most affordable for an individual with a limited budget for their project. According to Hussien, the cost is manageable for enterprises and large companies, but might prove to be a barrier for those with limited funding options. Another consideration is the issue of confidentiality and privacy. Researchers and entities who opt to utilize these facilities need to be aware of the risk of a data breach at the center. Although not common, this may be a potential consideration for users. Further, those interested in pursuing this option need to acquaint themselves with the policies and regulations of the respective center with regards to

[23] "Data Types & File Formats," University of Virginia Library, accessed March 31, 2018, https://data.library.virginia.edu/data-management/plan/format-types/.

ownership of the data and third-party access. These issues are further discussed in the Data Security section of this paper.

c. Metadata

Although a relatively simple concept, metadata plays a fundamental role in ensuring the discovery, interoperability and reusability of data. Most commonly explained as *data about data*[24], metadata refers to the documentation of data. It is used as an identifier or tag, containing descriptive information about any given dataset including its formatting and its content, as well as any other relevant information. Metadata is especially important when dealing with unstructured data – data that is generated and easily understood by humans, but not machines. An example could be an audio recording of a lecture given at a university – a human might be able to listen to the audio and easily identify the speaker, the topic, and the tone being set. A machine, on the other hand, would face great difficulty in breaking down the various characteristics.[25] Metadata describes the contextual details of the piece of data. According to reports, 80-90% of data in organizations is unstructured,[26] amassing to a sizable amount of information that cannot be processed by machines. In order to maximize efficiency, it is important to take the time to *tag* the different data elements as they are being collected. In doing so at the onset of the project, there is a higher likelihood that the data will be understood and discovered by others once it is made open.

d. Data Security

Data holds immense value and oftentimes carries sensitive information that needs to be anonymized or kept secure until the time of its publishing. Regardless of whether an individual or organization plans on making their data open, data security is a fundamental aspect that needs to be taken into consideration at the onset of any project. During the earliest stages of data collection, some form of risk assessment should be carried out in order to determine the relevant security measures that should be put in place, bearing in mind limitations such as access to the relevant technology as well as potential cost. A thorough assessment of the datasets in possession must be carried out - taking into consideration data privacy and ownership - and the appropriate form of security measures applied. Ahmed Hussien recommends that individuals weigh out the possible risks that might be associated with a security breach of the data, assess the extent of these risks or threats, and choose the relevant measure accordingly. Taking into account the high costs associated with implementing stringent security protocols, Hussien recommends carrying out a cost analysis prior to choosing the security and storage method. In order to justify the cost, the cost of the data damage or data infringement should be higher than the value of the data on its own.

---

[24] Further reading available through this link.
[25] It is worth noting, however, that this field is constantly evolving, with various progressions and breakthroughs taking place regularly. Although this is currently the case, it might not be the case a few years from now.
[26] Vangie Beal, "Unstructured Data," *Webopedia,* accessed March 31, 2018, https://www.webopedia.com/TERM/U/unstructured_data.html.

In her open data release toolkit for the release of sensitive or protected datasets, Erica Finkle identifies three data classifications that aim to guide individuals dealing with open data. Generally, data can be classified as either public, meaning there are no concerns over the sensitivity of the data and that it may be published in its entirety; sensitive, meaning that the data carries information that may, in its raw form, pose a security concern; or protected, meaning that the dataset in its entirety carries sensitive information that only be shared or accessed "internally and per organizational procedures".[27]Depending on the different classifications, different forms of securing the data can be pursued.

In order to ensure the security of the data before, during, and after its release, several measures can be taken. Hussien recommends ensuring that the terminals used to input the data are all in a secure location, and that there is a clear system of accountability in place. During the initial phases of the project, a clear hierarchy should be determined as to who is granted access to data, and who will be held accountable for any damage that might occur to it. Having a clear logging system in place is a further recommended step which increases transparency and minimizes the risk of a fraudulent attack or a leak of the data. Depending on the sensitivity of the data being managed, Hussien further recommends adding two parameter authentications in order to grant access to the data. This added layer of security consists of a biometric authentication (such as a finger print or eye scan) and a credential authentication (such as a username and password). Further, individuals or organizations might consider encrypting their data during the initial phases of the project in order to ensure its security before being made open.

Sherif El-Kassas, professor of computer science at the American University in Cairo, warns against rushing into a single "fit all" solution. Whilst there are different ways of classifying data, and while those classification methods help determine whether a dataset should be made public or not, El-Kassas recommends looking at cases on an individual basis. Some datasets can easily be classified as *public* in their entirety, whilst others may contain a mix of attributes – some that can be made public, and others that should be kept private. Giving the example of data relating to student grades, El-Kassas explains that a professor may wish to share specific trends related to grades but might need to anonymize the identities of the grade holders. In this case, two different classification methods can be applied. Those dealing with data are recommended to think critically about the datasets they wish to make public and decide on the classification method accordingly.

V. Ethical and Legal Provisions

When dealing with any form of data, exercising caution with regards to privacy and security and complying with the surrounding legal framework is rudimentary. While laws pertaining to data and data protection are scarce in the region, there exist ethical guidelines that can generally be followed in order to maintain the integrity of the data whilst minimizing potential harm or breaching confidentiality. In

[27] Erica Finkle, "Open Data Release Toolkit," Version 1.2, San Francisco: DataSF.org, https://datasf.org/resources/open-data-release-toolkit/.

their analytical report on Open Data and Privacy, written for the European Data Portal, authors Elena SImperl et al., delve into some of ethical considerations that should guide researchers in the field of open data. When operating in an environment void of specific data protection laws and regulations, researchers and individuals dealing with data must employ rigorous and critical thought. The authors discuss the importance of striking a balance between the greater good for society, and individual rights to privacy, explaining that "…[the] releases of data in the open must be scrutinized carefully to ensure that the risk of a privacy breach is as low as possible, unless there was some other reason why privacy was not a priority."[28]

Although region specific data laws are sparse, there exist some legal codes and regulations that pertain to the uses and publication of data and personal information that will affect how an individual who deals with data might choose to disseminate their work. Individuals partaking in open data activities are encouraged to take the articles in the following section into consideration, and to further research the implications these articles might have on their work. There are a number of legal provisions that exist around the region, however, the following section focuses on Egypt, Morocco, Jordan, and Tunisia as four primary examples. The section is not exhaustive but can nonetheless prove to be useful introduction.

a. Egypt

The Egyptian legal code addresses the issues of data protection and privacy under two main legal codes: the Egyptian Constitution of 2014 and the Telecommunication Regulation Law of Egypt No. 10 of 2003. Egypt is also a signatory of the Arab Convention on Combatting Information Technology Offences.

i. The Egyptian Constitution of 2014

Various freedoms related to data protection and privacy are mentioned in the 2014 Egyptian Constitution, under two main articles. Article 57 of the constitution guarantees the right to "private life" as well the right of citizens to use all forms of public means of communication without them being "arbitrarily disrupted". The article also ensures that various forms of communication, including electronic communication are "inviolable" and confidential. Although broadly phrased, Article 57 of the Constitution is central when discussing the rights to privacy, telecommunication and the regulation of evidence collection in the Egyptian legal context. Article 68 of the Constitution is the key and fundamental article in the Egyptian legal system that addresses access to information. Broadly outlining guarantees made by the law with regards to access to information and official documents, it expresses that state information, data, statistics and official documents are property of the people, and that all citizens have a right to access this information. [29]

[28] Elena Simperl, Kieron O'Hara and Richard Gomer, "Analytical Report 3: Open Data and Privacy," European Data Portal, June 2016, https://tinyurl.com/y92fv8jf.
[29] The Arab Republic of Egypt, 2014, *Constitution of the Arab Republic of Egypt,* retrieved from http://www.sis.gov.eg/Newvr/Dustor-en001.pdf.

ii. Telecommunication Regulation Law of Egypt No. 10 of 2003

Legal provisions most directly related to data, data protection and privacy can be found in the Telecommunication Regulation Law of Egypt No. 10 of 2003. This Law aims to create a legal framework for the regulation of communication networks and services (this includes both telecommunication and the Internet), guarantee the provision of communication services to all regions of the country, to safeguard the confidentiality of telecommunications, and to set up a regulatory authority for the sector of Internet and telecommunication. The Law creates the National Communications Regulatory Authority (NTRA) and gives it regulation and monitoring responsibilities. The NTRA is "subordinated to the Minister Concerned", i.e. the Minister of Communications and Information Technology. The State authorities and the armed forces are granted special powers to handle national security issues and in periods of general mobilization.

Article 64.2, of the Law provides that the operators and providers of networks and services should allow armed forces and national security entities to have complete access to their systems in order to allow them "to exercise their powers within the law." Article 64 also prohibits the use of encryption equipment in a very general way, unless authorization is obtained from the NTRA, the Armed Forces and national security entities. Finally, Article 64 of the Law broadly authorizes providers and operators to collect information about users from "individuals and various entities within the state", although the Article does not expand on the type of information that this includes, or the manner or limits of data collection that the Article allows for. Article 81 sets the criminal tariff for a violation of Article 64. [30]

iii. Arab Convention on Combatting Information Technology Offences

As of December 2010, Egypt is a signatory of the League of Arab States' Arab Convention on Combatting Information Technology Offences[31], the purpose of which is to enhance and strengthen cooperation between Arab States in the area of combatting information technology crimes to protect the security and interests of the Arab States and the safety of their communities and individuals. By 2014, a total of 18 Arab states had signed the Convention. The Articles of the Convention seem to provide a broad and sweeping overview of general provisions on privacy and data protection, rather than providing explicit stipulations on legal protection and regulation of data and privacy.[32]

b. Morocco

[30] The National Telecommunication Regulatory Authority of Egypt, *Egypt Telecommunication Regulation Law, Law No 10 of 2003*, retrieved from https://tinyurl.com/y7nrglod.
[31] Various other MENA countries are signatories including Bahrain, Jordan, Saudi Arabia and Kuwait, to name a few. For a full list of the signatories, the convention is available here.
[32] League of Arab States, General Secretariat, *Arab Convention on Combating Information Technology Offenses,* December 21, 2010, retrieved from https://dig.watch/actors/arab-league.

The legal framework in Morocco as it concerns data privacy and protection is very similar to the Egyptian legal context as it concerns privacy and protection of data.

i. Constitution of Morocco of 2011

Article 27 of the Constitution of Morocco of 2011 is the main, and general, legal provision that addresses citizens' right to access information. Article 27 of the Constitution guarantees citizens the right to access to information "held by the public administration, the elected institutions and the organs invested with missions of public service". The Article restricts citizens' right to access information in order to ensure the "protection of all which concerns national defense, the internal and external security of the State". It also restricts the access to information to ensure "the private life of persons, of preventing infringement to the fundamental freedoms and rights enounced in this Constitution and of protecting the sources and the domains determined with specificity by the law".[33]

ii. Access to Information Act of 2 January 2018

The draft bill on Morocco's controversial Access to Information Act was ratified on January 2 2018 by the Justice, Legislation and Human Rights Committee of the Second Chamber of the Moroccan Parliament. The Access to Information bill underwent a multitude of revisions over the last several years, and the delay in passing it can be mainly attributed to the controversial phrasing of various articles in the Act. The passing of the bill was met with strong opposition within civil society, notably from the Moroccan National Council for Human Rights and the United Nations Human Rights Council, whose main point of contention was that public institutions stated as obliged to provide information must be defined precisely in accordance with the objectives of the draft law in order to ensure access to information. The Human Rights Council also stressed the need to specify the status of institutions and private companies entrusted with public service missions.[34]

iii. Law No. 09-08 on the Protection of Physical Persons Regarding the Processing of Personal Data

On 21 May 2009, Decree No. 2-09-165 for the application of the Law No. 09-08 on the protection of physical persons regarding the processing of personal data was adopted. Law No. 09-08 constitutes the bulk of the law on privacy and data protection in the Moroccan legal context and is more similar to data protection laws of the European Union than legal provisions relating to data that exist in the rest of the MENA region.

---

[33] Kingdom of Morocco, *Morocco's Constitution of 2011*, retrieved from https://www.constituteproject.org/constitution/Morocco_2011.pdf.

[34] Latifa Elarasani, "البرلمان المغربي يصادق على قانون حق الحصول على المعلومات," *Ashraq Al-Awsat*, February 8, 2018, https://tinyurl.com/ydd96zrd.

The Law introduced a set of legal provisions that address online data privacy and protection. The main objective of Law 09-08, which governs personal data protection, is to facilitate the growth of the digital economy while protecting privacy. The Law defines the rights of data subjects (such as the right to access the data and to object to the processing of their data, the right of information, etc.), the obligations of the data controller (including prior authorization or declaration, security and limitations on data storage periods), and the rules governing the transfer of data abroad. The Law also creates the National Data Protection Authority and assigns it the responsibility of applying and ensuring compliance with the provisions of Law 09-08.

Article 1 of Law 09-08 defines personal data as any information relating to an identified or identifiable natural person (a "data subject"), regardless of the form or medium of said information. The definition also includes sound and images. Personal data processing is defined as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, block, erasure or destruction". Article 1 of the Law also aims at strengthening the protection of "sensitive data", defined as data relating to an individual's racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, health or genetic data. The same article states that "automatic processing includes the following operations if carried out in whole or in part by automated means: storage of data, carrying out of logical and/or arithmetical operations on those data, their alteration, erasure, retrieval or dissemination".

Article 3 of the Law specifies that the data controller has a "duty of quality" regarding the purposes of the data processing and the accuracy of the data collected. As such, data must be collected for specific, explicit and legitimate purposes. Moreover, the amount of data must be relevant and not disproportionate in relation to the purposes for which it is collected. Finally, the data must be accurate, up to date, and must be kept in a form which allows for the identification of data subjects for no longer than is necessary for the purposes for which the data is collected.

Article 4 begins defining the obligations of the data controller, which include prior notification to, and authorization from, the CNDP before performing any automated processing operation. Article 12 states that the authorization procedure applies to processing sensitive data and is aimed at strengthening the protection of this type of data in light of its nature. Article 4 outlines the data controller's obligations with respect to the data subject, including requesting consent from the data subject before collecting and processing any personal data. However, sub-articles 4(a) to 4(e) several exceptions to this general rule, including situations where the process is necessary for the performance of a "contractual obligation" or to "pursue the legitimate interests of the data controller provided that the fundamental rights and liberties of the data subject are respected". It is difficult to determine the exact limitations of these exceptions.

According to Article 5 of Law 09-08, prior to any data processing the data controller or his representative must provide the data subject from whom data is collected with at least the following information:

"a) the identity of the data controller and of its representative, if any;
b) the purposes of the processing for which the data are intended;
c) any further information such as the recipients or categories of recipients of the data, whether replies to the questions are obligatory or voluntary, and the possible consequences of a failure to reply;
d) the details of the notice of receipt of provided by the CNDP in the event of a prior notification or the details of the authorization issued by the CNDP".

Article 23 of the Law places an obligation on the data controller to comply with a "duty of confidentiality and security relating to the data processed". This means that the data processor is required to implement all technical and organizational measures to protect personal data in order to prevent it from being damaged, altered or used by a third party who is not authorized to gain access to it, as well as protect data against any form of illicit processing. In addition, if the data controller uses a sub-contractor for data processing, the obligation rests on the controller to ensure that it selects a service provider that can provide guaranteed technical security measures. Article 10 also prohibits any form of direct marketing that uses that contact details or private information of any person without their prior consent.

Law 09-08 also creates a system to protect transfers of personal data to foreign countries, including cases when the foreign countries concerned do not provide a sufficient level of protection of privacy, freedom and fundamental rights of individuals concerning personal data processing. Under Law 09-08, personal data must be subject to prior authorization from the CNDP before any transfer to a foreign state. Furthermore, the person in charge of the processing operation can transfer personal data to a foreign state only if the said state ensures under its applicable legal framework an adequate level of protection for the privacy and fundamental rights and freedoms of individuals regarding the processing to which these data is or might be subject. However, the data processor can transfer personal data to any foreign state which does not satisfy the conditions mentioned above (i.e. ensure an adequate level of protection of privacy and fundamental rights and freedoms of individuals), if the person to whom the data relates has expressly consented to the transfer. According to Article 3 of the Law, the CNDP has the power to establish a list of authorized countries that meet the criteria of providing sufficient levels of protection, but this list has yet to be drafted.

According to Article 44, authorization from the CNDP is not required when a transfer is considered "necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract" or if the transfer is carried out in connection with "the conclusion or performance of a contract concluded or to be concluded in the interest of the data subject between the data controller and a third party".

Articles 50 to 64 of the law outline the punishments for any violation of the provisions of Law 09-08. Finally, the Law does not include specific provisions on the collection of location and traffic data by public electronic communications service provides, or the use of cookies or similar technologies. [35]

iv. Law No. 03-03 Regarding Anti-Terrorism

Law No. 03-03 Regarding Anti-Terrorism of May 28 2003 provides authorities with legal powers to filter and delete content that is deemed to "disrupt public order by intimidation, force, violence, fear or terror". The law assigns legal liability to the author of online content, including site owners and Internet Service Providers (ISPs).[36]

c. Tunisia

i. Organic Act n°2004-63 of July 27th 2004 on the Protection of Personal Data

The majority of data privacy and protection legal provisions are set out in the 2004 Organic Act on the Protection of Personal Data. On one hand, the 2004 Organic Act seems to set a high standard of data protection for Tunisian citizens. The Tunisian data protection regime gives a range of rights to individuals whose data is processed and sets out certain obligations for organizations and individuals in charge of the data processing. According to the act, personal data processing must be declared by data processors or previously authorized by the National Authority for Protection of Personal Data (INPDP), which the act creates as the central body responsible for the control and enforcement of the legal framework on data protection. However, there are numerous exemptions provided by the Tunisian data protection regime for certain data processors that create loopholes in citizens' data protection. This includes the exemption of organizations with a "public personality" (this includes police stations, tribunals and universities) from the scope of the Act, meaning they are not bound by the obligations that that would normally apply to personal data processors in the Tunisian legal context. As such, "public organizations" do not have to declare data processing. In turn, citizens are unable to exercise their rights of access, opposition and informed consent in relation to personal data utilized by these public organizations.

Article 1 of the Act states that "Everyone has the right to the protection of personal data related to his privacy as one of the fundamental rights guaranteed by the Constitution. The processing of personal data shall respect transparency, fairness and the respect of human dignity, in accordance with the clauses of this act." Article 4 defines personal data as "any information whatever its origin or its means relating to an individual who can be identified, directly or indirectly, with the exception of any information related to public life or considered public life by law." Article 9 begins outlining the obligations of data controllers, in stating that "The processing of personal data shall be done as part of

[35] Kingdom of Morocco, *Law No. 09-08 on the protection of individuals with regard to the processing of personal data and its implementing Decree No. 1-09-15,* retrieved from https://tinyurl.com/ya72e36p.
[36] Kingdom of Morocco, *Law No. 03-03 on the Fight Against Terrorism, 2003, Art. 595-7*, retrieved from https://tinyurl.com/yc9kadv4.

the respect of human dignity, privacy and public liberties. The processing of personal data, whatever its origin or its methods shall not harm the human rights protected by the laws and the rules in force. In every case, it is forbidden to use personal data with the aim of infringing people's rights or damaging their reputation."

Article 11 stresses that personal data can only be processed "within the limits of the collecting purpose", and puts the responsibility on data controllers to ensure the accuracy, precision and update of the data. Article 13 outlaws the processing of personal data relating to criminal offenses, convictions, prosecutions or penalties. Article 14 prevents the processing of personal data that reveals, whether directly or indirectly, the racial and genetic origins, religious and political beliefs and trade union affiliation or health status, except in the case of explicit consent or if processing is necessary "for historical or scientific purposes" or "necessary for the protection of the data subject's vital interests". According to Article 18, data controllers are also responsible for ensuring that any sub-contractors entrusted with data processing "shall take all the required steps to ensure the safety of the data processing and prevent any third party from changing, modifying or consulting it without prior authorization of the data subject."

Article 27 begins by outlining the rights of the data subject. In terms of the consent of data subjects, the Article states that "With the exception of the cases regulated by the hereby Act and the laws in force, the processing of personal data cannot be carried out without the express and written consent of the data subject. This consent shall be governed by the general rules of law if the data subject is incompetent or unauthorized or incompetent to sign. The data subject is allowed to withdraw his consent, at anytime during the processing." Article 29 outlines exceptions for the requirement of consent provided by Article 27, and states that the processing of personal data can be carried out without the consent of the data subject in cases where "it has been proved without doubt that the processing is carried out in the data subject's own interest", "when it is impossible to contact the data subject", "when obtaining consent implicates disproportionate endeavor", or "when the processing of personal data is allowed by law or contract to which the data subject is a party." [37]

d. Jordan

Jordan has no data protection or privacy law, and it often remains difficult for Jordanians to exercise their information privacy rights. Jordan also does not currently have specific access to information laws. However, a few laws provide some level of privacy protections.

i. Constitution of the Hashemite Kingdom of Jordan 1952

The right to privacy is mainly protected by the Constitution. The Jordanian Constitution of 1952 (amended several times since, the last time being in 2011) recognizes a limited right to privacy. Article 7

[37] Republic of Tunisia, *Organic Act n°2004-63 of July 27th 2004 on the protection of personal data,* retrieved from https://tinyurl.com/y8o76eau.

of the Constitution provides that every infringement on rights, public freedoms and the 'inviolability of the private life of Jordanians' is a crime punishable by law. Article 18 of the Constitution states that "all postal, telegraphic and telephonic communications shall be treated as secret and as such shall not be subject to censorship or suspension except in circumstances prescribed by law."[38]

ii. Draft Data Protection Law

In November 2016, the Jordanian Ministry of Information and Communications Technology launched a public consultation on a draft data protection law. The Draft Law states that an entity will be liable in respect of the personal data in its control and contains a range of conditions relating to the disclosure of personal data, data security, data transfer and data subject access. Under Article 9 of the Draft Law, there is a requirement to appoint a data protection officer who shall be responsible for creating and controlling data processing procedures for an entity as well as receiving and responding to complaints from data subjects. The Draft Law also requires that certain conditions must be met in respect of the transfer of data, including the consent of the data subject and adequate knowledge of the purpose of the transfer. It also stipulates that personal data must not be transferred outside of Jordan to entities operating in countries which are not deemed to offer sufficient protection levels. The law has yet to be promulgated, as it has not passed the sufficient requirement of passing by both the House of Representatives and the Senate, as well as confirmation by the King.[39]

VI. Conclusion

This paper has set out to provide an overview of the current technological landscape, as well as open data and the space it occupies within the larger data ecosystem. The paper provides insights into current trends surrounding the use of data in various sectors and for various purposes, and the debates surrounding the best methods of practice to ensure that no undue harm is caused to individuals or communities as a result of misuse of data. The background paper aims to provide a contextualized body of knowledge relating to open data and its management in the region.

Preliminary desk research carried out by the researchers prior to the development of the plan pointed to a gap in regional literature discussing issues pertaining to data collection, management and preservation. The creation of a localized plan was found to be an important step in supporting the emerging open data community in the MENA region. Whilst existent data management plans developed by governments and funding bodies in different contexts, can certainly be considered building blocks, they fail to accurately capture the experiences that are unique to the regional data community. The

[38] The Hashemite Kingdom of Jordan, *Constitution of the Hashemite Kingdom of Jordan 1952,* retrieved from http://www.wipo.int/wipolex/en/text.jsp?file_id=227814.
[39] "State of Privacy Jordan," Data Protection, Privacy International, January 2018, https://privacyinternational.org/state-privacy/1004/state-privacy-jordan#dataprotection.

paper illustrates the role a management plan can play in helping build a framework of accountability and efficiency of data projects, as well as the potential it has for ensuring the longevity of said project.

Intended to be read as part of a broader set of documents, this background paper acts as a supplementary piece of literature that will help researchers better understand and complete the Open Data Management Plan template developed by the team at Access to Knowledge for Development. Unlike other regions around the world, the MENA region lacks concrete laws and regulations that govern the creation, collection, and dissemination of data. Aware of this fact, the paper highlights some of the existent regulations in Egypt and Morocco that most closely relate to data and its dissemination as a means of laying the foundations for researchers to gain a sense of the legal environment they are operating within. In light of the scarcity of laws, the paper encourages researchers to think critically about the data they are aiming to avail and the implications this action may have on both an individual and community level.

Through the development of this series of documents, researchers are given a solid foundation upon which they can build in order to ensure the creation of a sustainable and efficient data project. The MENA region houses an immense amount of research potential and can contribute a significant body of data and knowledge to the greater data community - the creation of a framework through which researchers can operate in order to ensure the standard and quality of their work is integral.

VII. Bibliography

Australian National Data Service. "Data management plans." ANDS Guide, October 2017. http://www.ands.org.au/guides/data-management-plans.

Australian National Data Service. "The FAIR data principles: Working with data." Accessed March 31, 2018. https://www.ands.org.au/working-with-data/fairdata.

Beal, Vangie. "Unstructured Data." *Webopedia.* Accessed March 31, 2018. https://www.webopedia.com/TERM/U/unstructured_data.html.

Berghel, Hal. "Bruce Schneier on Future Digital Threats." *Computer* 51, no. 2 (February 2018): 64-67. https://doi.org/10.1109/MC.2018.1451653.

Cadwalladr, Carole and Emma Graham-Harrison. "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach." *The Guardian,* March 17, 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election.

DCC. "Checklist for a Data Management Plan, v4.0." Edinburgh: Digital Curation Center, 2013. http://www.dcc.ac.uk/resources/data-management-plans.

Elarasani, Latifa. "البرلمان المغربي يصادق على قانون حق الحصول على المعلومات." Ashraq Al-Awsat, February 8, 2018. https://tinyurl.com/ydd96zrd.

Finkle, Erica. "Open Data Release Toolkit." Version 1.2. San Fransisco: DataSF.org. https://datasf.org/resources/open-data-release-toolkit/.

FORCE11. "About FORCE11." Accessed March 31, 2018. https://www.force11.org/about.

Kemp, Simon. "The global state of the internet in April 2017." *The Next Web,* April 11, 2017. https://tnw.to/2nZHhVb.

Kepes, Ben. "Google Users - You're The Product, Not The Customer." *Forbes,* December 4, 2013. https://www.forbes.com/sites/benkepes/2013/12/04/google-users-youre-the-product-not-the-customer.

Kingdom of Morocco. *Law No. 03-03 on the Fight Against Terrorism, 2003, Art. 595-7*. Retrieved from https://tinyurl.com/yc9kadv4.

Kingdom of Morocco. *Law No. 09-08 on the protection of individuals with regard to the processing of personal data and its implementing Decree No. 1-09-15.* Retrieved from https://tinyurl.com/ya72e36p.

Kingdom of Morocco. *Morocco's Constitution of 2011.* Retrieved from https://www.constituteproject.org/constitution/Morocco_2011.pdf.

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.* SAGE, 2014.

Kopetz, Hermann. "Internet of Things." In *Real-Time Systems,* 307-323. Boston, MA: Springer, 2011. https://doi.org/10.1007/978-1-4419-8237-7_13.

League of Arab States, General Secretariat. *Arab Convention on Combating Information Technology Offenses.* December 21, 2010. Retrieved from https://dig.watch/actors/arab-league.

M.D., Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Sci. Data 3,* (2016). https://doi.org/10.1038/sdata.2016.18.

Molloy. C, Jennifer. "The Open Knowledge Foundation: Open Data Means Better Science." *PLoS Biology* 9, (December 2011): https://doi.org/10.1371/journal.pbio.1001195/.

Open Definition. "Open Definition 2.0." Accessed March 31, 2018. https://opendefinition.org/od/2.0/en/.

Picciano, Bob. "IBMVoice: Why Big Data Is The New Natural Resource." *Forbes,* June 30, 2014. https://www.forbes.com/sites/ibm/2014/06/30/why-big-data-is-the-new-natural-resource.

Privacy International. "State of Privacy Jordan." Data Protection. January 2018. https://privacyinternational.org/state-privacy/1004/state-privacy-jordan#dataprotection.

Republic of Tunisia. *Organic Act n°2004-63 of July 27th 2004 on the protection of personal data.* Retrieved from https://tinyurl.com/y8o76eau.

Schneier, Bruce. "'Stalker economy' here to stay." *CNN,* November 26, 2013. https://edition.cnn.com/2013/11/20/opinion/schneier-stalker-economy/index.html.

Simperl, Elena, Kireon O'Hara and Richard Gomer. "Analytical Report 3: Open Data and Privacy." European Data Portal, June 2016. https://tinyurl.com/y92fv8jf.

The Arab Republic of Egypt. 2014. *Constitution of the Arab Republic of Egypt.* Retrieved from http://www.sis.gov.eg/Newvr/Dustor-en001.pdf.

The Economist. “The world’s most valuable resource is no longer oil, but data - Regulating the internet giants.” *The Economist,* May 6, 2017. https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource.

The Hashemite Kingdom of Jordan. *Constitution of the Hashemite Kingdom of Jordan 1952.* Retrieved from http://www.wipo.int/wipolex/en/text.jsp?file_id=227814.

The National Telecommunication Regulatory Authority of Egypt. *Egypt Telecommunications Regulation Law, Law No. 10 of 2003.* Retrieved from https://tinyurl.com/y7nrglod.

University of Virginia Library. “Data Types & File Formats.” Accessed March 31, 2018. https://data.library.virginia.edu/data-management/plan/format-types/.

Web Foundation. “Closing the Digital Divide: A Briefing Note.” *World Wide Web Foundation,* April 14, 2016. https://webfoundation.org/2016/04/closing-the-digital-divide-a-briefing-note/.

OPEN DATA MANAGEMENT PLAN TEMPLATE

Title of the project

Date of this plan

Description

- What data is required for the research?
  Provide a clear description of the data that needs to be collected.
- What is the project timeline?

Principal researchers

- Who are the main researchers involved in this research?
  Provide contact details for those researchers.
- Who is the main contact person for this project?
- Do the researchers have specific roles?

Data collection

- How will the data be obtained?
  i. Will it be produced in a controlled environment (by conducting laboratory experiments, clinical trials, etc.), or will it be collected (using sensors for measuring natural phenomena, interviewing people, taking pictures, collecting data from medical records, etc.)?
- If the data is collected, not produced, which sources will it be collected from?
- Will the data be collected by human beings, or by automated systems?
  i. If the data is collected by automated systems, can these systems produce an open format that can be imported into other systems and platforms?
- If a number of individuals are involved in the actual data collection activities, what is their role and accountability for the validity and quality of the data they collect?
- If data is collected using online surveys/questionnaires, what measures are taken to ensure the authenticity and integrity of the collected data?

Data responsibility and privacy

- Who is ultimately responsible for the validity and quality of the data set?
  i. What is the limit of responsibility for the data?
  ii. Who is the owner or the custodian of the data set?
- Will the collected data contain aspects that might be considered private (for example, confidential data about people and their living environments, medical records, etc.)?
  i. Is collecting this private data mandatory, or can alternative data be used to avoid any potential privacy risk?
  ii. Does the consent of the individuals concerned need to be obtained before collecting this data?

iii. Does consent for publishing private data cover the original data, an anonymized derivate of the original data, or only summary data?

Metadata and data formats

- How will the data be organized and/or tagged in order to be able to uniquely identify each individual piece (establishing file naming rules, creating a temporary storage hierarchy, setting rules for record identification etc.)?
    i. How will data collected from different locations or by different people be aggregated?
- What metadata do you need to be able to describe or establish the meaning of the collected data?
- In what formats will the data be collected?
    i. Are these formats open (meaning they can be processed by industry standard open source software) or do they need specialized proprietary software?
- Will the data collected be processed to produce new data elements?

Legal considerations

- Are there any legal implications to the collected data?
- Are there any laws, rules, or regulations concerning data collection in your country?
- Do you need to obtain prior approvals or authorizations before initiating any data collection activities?
- Do the data collection activities have the potential of provoking responses from the local community?
    i. If so, can this pose a safety or security risk to the data collectors?

Data transmission

- Will the data be obtained and processed at the same location **OR,**
    i. Will it be collected at various locations and processed centrally?
    ii. Will it be collected and processed at multiple different locations?
- If the data is being collected at a location different to its processing location, how will the data be delivered to the location in which it will be processed?
    i. Will it be delivered physically using some form of storage medium (USB flash drives, SD cards, hard disks, paper documents, etc.)? Will it be transmitted over the Internet?
        - Based on the volume of data collected, is it practical to deliver the data over the Internet?
- If the data is to be transmitted? over the Internet, the following are some considerations:
    i. How much bandwidth and time is required to transmit the data over the Internet?
    ii. How will the authenticity and integrity of the data collected be guaranteed?
    iii. If the transmitted data has privacy requirements, how will the confidentiality of the data in transit be guaranteed?

iv. What measures are taken in order to ensure the continuity of the data collection activities in the case of network unavailability?

- If the data is to be transmitted? using a physical medium, the following are some considerations:
    i. How many physical storage devices are required to store the data in transit? What is the required storage capacity for each device?
    ii. What measures are in place to ensure that all of the collected data is delivered to the storage and processing locations? What measures in place to ensure that data lost in transit can be recovered?
    iii. What measures are in place to ensure the authenticity and integrity of the delivered data?

Data storage and management

- How much data will be collected daily?
    i. If the data is being collected at a location different to its processing location, is there sufficient local storage at the data collection point to store the data until it is delivered to the processing location?
- What types of data will the master data set include (video, audio, pictures, text documents, structured data, etc.)?
- What is the average size of each data element?
    i. How many data elements are anticipated to be collected?
    ii. Is the data volume expected to grow with time?
    iii. How much storage is required to store a single copy of the master data?
    iv. Can this storage be availed using normal, personal computers? Or does it require enterprise grade data storage facilities?
        - If storage facilities provided by a storage service provider be used, what implications might this have on ensuring that the privacy and confidentiality concerns are met?
    v. What measures are in place to intend the continuous availability of the master data?
    vi. How many backup copies will be kept?
        - What is the frequency of backing up the data?
        - How many historical copies of the backup will be kept?
        - Are backups full, or incremental?
    vii. Will backup copies be dispersed geographically?
        - Do these copies need to be dispersed in different countries?
        - How will backup copies be delivered to their storage locations?
- How will the authenticity and integrity of each copy of the master data be validated?
- What types of mediums will be used to store the master data set and any derivative data sets?
    i. What is the expected lifetime of the selected storage medium?
    ii. What process and methods will be employed to ensure the continuous availability of the data?

Data publishing

- Is the data going to be published in its original form, or in a derivate form (for example, anonymized or aggregate results and statistics based on the data)? The following can be considered:
    i. Can the original data set be published, or are there privacy concerns?
        - Does the derivative data set address all privacy concerns?
    ii. Is the size of the master data set suitable for access over the Internet from the format and size points of view?
        - Does the derivative data set sufficiently address the format and size aspects?
    iii. In what formats is the master data set stored?
        - Are all the formats open and suitable for public access?
    iv. Is the data set to be published intended for consumption by humans, machines, or both?
        - What format requirements and publishing standards are needed to ensure it achieves its intended target (csv., XML., JSON., etc)?
        - Does the master data set meet these requirements?
    v. Under what license will the data be published?
        - Are there any aspects of the original data set that would limit the license used to publish the data?
- Is the data set published intended for public access, or is it intended for a specific audience?
    i. If it is intended for a specific audience, what kind of authentication and authorization is required to support the required access rights?
    ii. How will the authentication credentials be created, disseminated and managed?
- Is the data set intended for publishing accompanied with sufficient meta-data to clearly describe it?
    i. Which data naming conventions does the meta-data use?
    ii. Does this naming convention comply with industry standards and best practices?
- Will the data set be made searchable from the interface provided by the open data publishing platform, or will it be accessible as a single unit?
    i. If the data set is to be searchable, is there sufficient meta-data and tags to provide the required search capabilities?
- Who owns and controls the platform that the data set will be published on (public open data platform, university open data platform, private open data platform?
    i. What plans are in place to ensure the continuous availability of the published data in the case that the selected platform is no longer available?
- What plans are in place to identify and handle incidents that compromise the integrity, availability, and/or accessibility of the data?

OPEN DATA MANAGEMENT PLAN TEMPLATE

SOLAR DATA PLATFORM

Title of the project

Harnessing the Economic Power of Data in the Middle East and North Africa

---

Date of this plan

March 31st, 2018

Description

- What data is required for your research?
  Provide a clear description of the data that needs to be collected.
  This project looks to map out the value chain for the solar energy sector in Egypt. As such, the data required relates to actors along the value chain, and their activities within the sector. Data collected comes from suppliers of solar panels, distributors, installers, and support providers. The data collected includes each actors' operational capacities, active projects, number of employees, clients, investment and financing, and capacity building activities, to name a few. Further data collected also includes information such as names, addresses, dates of birth, email addresses, and salary ranges.
- What is the project timeline?
  The platform was launched in October 2017 and has been ongoing ever since.

Principal researchers

- Who are the main researchers involved in this research?
  Nagla Rizk, Nancy Salem, Youmna Hashem, Stefanie Felsberger

---

Data collection

- How will the data be obtained?
  i. Will it be produced in a controlled environment (by conducting laboratory experiments, clinical trials, etc.), or will it be collected (using sensors for measuring natural phenomena, interviewing people, taking pictures, collecting data from medical records, etc.)?
  It will be collected.
- If the data is collected, not produced, which sources will it be collected from?
  Through online forms filled out by web users on the platform, and through face-to-face interviews.
- Will the data be collected by human beings, or by automated systems?
  i. If the data is collected by automated systems, can these systems produce an open format that can be imported into other systems and platforms?
  Yes, the online platform through which this data is collected is equipped with the technology to process the data into 5 different open formats.
- If data is collected using online surveys/questionnaires, what measures are taken to ensure the authenticity and integrity of the collected data?

Before the data is published onto the platform, platform administrators go through the provided information to ensure its accuracy before they approve its publishing.

Data responsibility and privacy

- Who is ultimately responsible for the validity and quality of the data set?
  The Access to Knowledge for Development Center at the American University in Cairo.
  i. Who is the owner or the custodian of the data set?
     The Access to Knowledge for Development Center at the American University in Cairo.
- Will the collected data contain aspects that might be considered private (for example, confidential data about people and their living environments, medical records, etc.)?
  No.
  i. Does the consent of the individuals concerned need to be obtained before collecting this data?
     Individuals opt to input their information based on their own free will and can at any point access their accounts and opt to remove parts of the information provided.

Metadata and data formats

- How will the data be organized and/or tagged in order to be able to uniquely identify each individual piece (establishing file naming rules, creating a temporary storage hierarchy, setting rules for record identification etc.)?
  The platform is equipped with technology that stores the data and provides URIs for each piece of data. This is done automatically.
- What metadata do you need to be able to describe or establish the meaning of the collected data?
  The platform is equipped with a technology that is able to access Dbpedia – a repository that pools together structured content and metadata from Wikipedia – and utilize the relevant metadata tags.
- In what formats will the data be collected?
  The platform automatically converts all data input into 5 different open formats.
  i. Are these formats open (meaning they can be processed by industry standard open source software) or do they need specialized proprietary software?
     They are open.
- Is the data set intended for publishing accompanied with sufficient meta-data to clearly describe it?
  Yes.
  i. Which data naming conventions does the meta-data use?
     The naming conventions set out by data.world
  ii. Does this naming convention comply with industry standards and best practices?
     Yes.

- Will the data set be made searchable from the interface provided by the open data publishing platform, or will it be accessible as a single unit?
  It will be searchable from the interface provided.
  i. If the data set is to be searchable, is there sufficient meta-data and tags to provide the required search capabilities?
     Yes.

Legal considerations

- Are there any laws, rules, or regulations, (formal or otherwise) concerning data collection in your country?
  Yes, the data collection was carried out in a way that complied with these regulations.
- Do you need to obtain prior approvals or authorizations before initiating any data collection activities?
  Yes, Institutional Review Board (IRB) authorization was obtained.

Data transmission

- Will the data be obtained and processed at the same location?
  Yes, the data is obtained and processed online through the platform. The only exception applies to the preliminary stages of the data collection, where data was obtained through face-to-face interviews where forms were physically filled out. These were then uploaded to the platform by researchers at the Access to Knowledge for Development Center.
- If the data is being collected at a location different to its processing location, how will the data be delivered to the location in which it will be processed?
  i. Will it be delivered physically using some form of storage medium (USB flash drives, SD cards, hard disks, paper documents, etc.)? Will it be transmitted over the Internet?
     Paper documents.

Data storage and management

- What types of data will the master data set include (video, audio, pictures, text documents, structured data, etc.)?
  Videos, pictures, documents including PDFs, word documents, PowerPoint presentations.
- What is the average size of each data element?
  Documents average at about 20kbs, pictures average at about 100-200kbs, and videos average at about 5mbs.
  i. Is the data volume expected to grow with time?
     Yes, but at a very small rate.
  ii. Can the storage be availed using normal, personal computers? Or does it require enterprise grade data storage facilities?
     It can be accessed using normal, personal computers. InsideOut Egypt, the web development company that developed the platform, are the cloud storage providers.

iii. What measures are in place to intend the continuous availability of the master data?
The platform creates and stores daily backups of all the data and information.

- What is the frequency of backing up the data?
Daily.

iv. Are backups full, or incremental?
Full.

- What process and methods will be employed to ensure the continuous availability of the data?
Everything is uploaded to the cloud, and the platform has several cloud mirror sites to ensure the continuous availability. Should one server go down, another mirror server will be up and running with the data.

Data publishing

- Is the data going to be published in its original form, or in a derivate form (for example, anonymized or aggregate results and statistics based on the data)? The following can be considered:
  i. Can the original data set be published, or are there privacy concerns?
  Certain data pieces are privatized and only made available to the admins of the platform. These data pieces are the personal data collected pertaining to individual names, addresses, salary ranges etc. The data is not aggregated.
  ii. Is the size of the master data set suitable for access over the Internet from the format and size points of view?
  Yes.
  iii. Is the data set to be published intended for consumption by humans, machines, or both?
  Both.
  iv. Under what license will the data be published?
  Creative Commons License.
- Is the data set published intended for public access, or is it intended for a specific audience?
The vast majority is intended for public access, with certain categories intended for a specific audience.
  i. If it is intended for a specific audience, what kind of authentication and authorization is required to support the required access rights?
  The data that is intended for specific audience requires users to register an account with the platform, after which they will be granted access to the data.
  ii. How will the authentication credentials be created, disseminated and managed?
  Anyone is free to register for an account. Account information will be managed by admins, who must approve any new registration.
- Who owns and controls the platform that the data set will be published on (public open data platform, university open data platform, private open data platform?
The Access to Knowledge for Development Center owns and controls the platform. The center

receives technical support from InsideOut Today, the web development company that developed the platform.

- What plans are in place to identify and handle incidents that compromise the integrity, availability, and/or accessibility of the data?
  The technical team at InsideOut are proactive in ensuring that the platform is always updated with the most recent software patches, bug fixes, and security fixes. In the worst-case scenario, should a breach occur, the company is able to access the last healthy copy of the data and can terminate the website if need be.
