MENA OBSERVATORY
ON RESPONSIBLE AI
مرصد الشرق الأوسط وشمال إفريقيا للذكاء الاصطناعي المسؤول

Governing Responsible Artificial Intelligence and Data
in the Middle East and North Africa (MENA)

The American University in Cairo
Onsi Sawiris School of Business
Access to Knowledge for Development Center

IDRC · CRDI
International Development Research Centre
Centre de recherches pour le développement international
Canada

Governance of Data for Responsible AI in the Health Sector

THE CASE OF EGYPT

March 2025

GOVERNANCE OF DATA FOR RESPONSIBLE AI IN THE HEALTH SECTOR THE CASE OF EGYPT

Developed and Conceptualized by the Access to Knowledge for Development Center (A2K4D)

Nagla Rizk and Farah Elbehairy
Access to Knowledge for Development Center (A2K4D)
The American University in Cairo

Acknowledgements

This study/report/publication/policy brief was carried out in line with the conceptual framework developed by The Access to Knowledge for Development Center (A2K4D) at the American University in Cairo (AUC)'s Onsi Sawiris School of Business, as part of the project titled "Governing Responsible Artificial Intelligence and Data in the Middle East and North Africa." This project is held as a partnership between A2K4D and Birzeit University Palestine (BZU), with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada. The views expressed herein do not necessarily represent those of A2K4D, BZU, IDRC or its Board of Governors.

This research was conducted with financial support from Canada's International Development Research Center (IDRC) in partnership with Birzeit University.

We are grateful to our team at the Access to Knowledge for Development Center at the American University in Cairo's Onsi Sawiris School of Business, with special thanks to Nada Nassar for her contribution to this report. We also thank our fieldwork partners, Shamseya for Innovative Community Healthcare Solutions, for their support throughout this collaboration.

Last but not least, we extend our thanks to our interviewees from the Egyptian public health centers (PHCs) and the private health sector for their time and cooperation.

TABLE OF CONTENTS

I. INTRODUCTION

1.1. Background and Objectives

This fieldwork was done as part of the *Governing Responsible Data and Artificial Intelligence in the Middle East and North Africa (MENA)*[1] project, which has an overarching objective to strengthen the development, deployment and governance of responsible AI and data through knowledge creation, networking, policy and regulatory support, and interdisciplinary collaboration and capacity development. The research objective of this fieldwork is to assess the governance of health data in different healthcare settings in urban Egypt, with a focus on data on women's health. The goal is to identify gaps and provide policy recommendations for the governance of responsible data and AI in MENA.

The objective of these exploratory visits was to observe the function and processes at the public health centers (PHCs) and identify possible study participants to finalize the fieldwork interview questions and data catalog in preparation for the next phase of the fieldwork.

Fieldwork was divided into two phases, a preliminary fieldwork phase in which exploratory visits were conducted to identify potential study participants and finalize fieldwork interview questions and a second phase in which the interviews were conducted. In the second phase, we interviewed relevant stakeholders at public healthcare centers identified by our partner, Shamseya, in underprivileged areas in Cairo and Giza governorates.

Health-related data is a focal point in this report due to its role as a prerequisite for implementing responsible AI technologies. In order for AI to provide equitable and accurate outcomes within the healthcare system and its services, responsible data governance processes need to be strategically implemented. Datasets also need to be more inclusive and representative of all people within the region. AI has the potential to improve and enhance health services, but in order to do so, Egypt's health datasets need major improvement. Utilizing the current World Bank Health Data Governance Principles,[2] this research maps Egypt's health data governance mechanisms to identify the gaps that hinder the responsible governance of AI in the healthcare system.

The two-week preliminary fieldwork phase consisted of exploratory field visits to five PHCs in Cairo and Giza. Informed by this preliminary phase, the second fieldwork phase consisted of multiple semi-structured interviews conducted on a purposive sample from 12 entities in the healthcare sector: public healthcare centers, radiology centers, private clinics, pharmaceutical labs, and multiple startups in health innovation. Fieldwork was guided by the Data Catalog, which is a dashboard/checklist we developed to note the types of health-related data that are available per institution. Secondly, we referred to a set of open-ended questions for semi-structured interviews on data governance practices/frameworks.

The material collected and analyzed following the two fieldwork phases aims to map the existing health data practices and governance frameworks, and solicits stakeholder insights on what changes are needed to work toward more responsible governance of health data for development in Egypt.

1.2. Methodology

The current research follows a purposeful sampling method to select the participants for key stakeholder or expert interviews to provide empirical evidence on the realities of health data governance processes and mechanisms in Egypt. The sampling is purposeful, as the study aims to interview the most relevant participants to the research objective, based on availability and willingness to participate. Therefore, we selected certain stakeholders to capture narratives around health data governing practices, including representatives from the public and private healthcare sector. They were selected based on their engagement and/or expertise in healthcare and data management. The semi-structured expert interviews were conducted with representatives from 12 different health entities across the public and private sectors: five public health centers, hospitals, radiology centers, private clinics, pharmaceutical labs, and various startups in health innovation.

To interrogate the existing state of data governance in women's health in Egypt, an in-depth review of regional and international questions around data governance processes has been conducted to identify key questions that pertain to data collection, data management, data sharing, and data security. The identified questions guided the development of a semi-structured interview tool for qualitative fieldwork questions that assess the process, challenges, gaps, and opportunities regarding health data governance. Thus, our research strategy is broken down into a

1 A2K4D, in partnership with the Center for Continuing Education (CCE) at Birzeit University (BZU) in Palestine, is currently leading a project titled "Governing Responsible AI and Data in the MENA Region," with a focus on Egypt, Lebanon, Morocco, Tunisia, Jordan, and Palestine. The project is supported by Canada's International Development Research Center (IDRC).

2 Health Data Governance: Universalising the benefits of health digitization, World Bank, 2022, https://healthdataprinciples.org/images/English/%5BEN%5D%20Health%20Data%20Governance%20Principles.pdf

METADATA	DESCRIPTION
Data point	This is the actual data item or field that is collected; e.g., a patient's name, or an institution's address, and whether or not a case needs to be referred, etc.
Consent	Is patient (or another data owner) consent required for this data? Was it obtained?
Data type	This is used to distinguish between different classes of data as we define them. An initial classification can be "clinical" and "non-clinical" or "personal" and "non-personal", etc.
Data format	The data collection/storage format; e.g., numerical, date, text, image, table, etc.
Data source	The data sources ; e.g., patient, doctor, other organization, internal, etc.
Data storage type	Manual or digitized, local storage or cloud, etc.
Data access	Who has access to the data?
Data sharing	Is the data shared with any person or entity by default?
Data update frequency	How often is the data updated? How often does it need to be updated?
Data use	What is the data being/going to be used for?

Table 1: Data catalog showcasing the data collection.

series of open-ended and non-leading questions so that respondents can share their contextualized perspectives and experiences of their governance practices of health data without being forced to certain interpretations. The extensive list of questions[3] is then simplified to suit the participants of semi-structured interviews in each healthcare setting and be meaningful and engaging to them.

The second data collection tool is a data catalog,[4] where we lay out the types of data that are currently being collected and/or would be helpful to collect from the various healthcare providers. The data catalog includes the following metadata parameters: data point, consent, data type, data format, data source, data storage type and location, data access and sharing, data update frequency, data use, and data protection. (see Table 1)

II. THEORETICAL FRAMEWORK:

This report employs the World Bank's Health Data Governance Principles[5] as its main framework, with the research methodology, questions, and discussions reflecting and referencing them. The principles provide an integral perspective to the discourse on data and public health systems, one that centers around human rights, equity, and sustainability. Due to the emergence and growth of data-driven approaches in the health sector, also accelerated by the COVID-19 pan-

3 See Annex 1

4 The data catalog is a spreadsheet viewed as a living document and is updated regularly based on new thinking and learning.

5 Health Data Governance: Universalising the benefits of health digitization, World Bank, 2022, https://healthdataprinciples.org/images/English/%5BEN%5D%20Health%20Data%20Governance%20Principles.pdf

demic, numerous countries and regions have started to implement policies that pertain to health data governance and regulation. However, public health systems still lack a unified global set of principles that ensure that data governance is achieved equitably and fairly.

There are three main objectives of the principles, all intersectional and interconnected:

1) Protect People: acknowledges the importance of data protection on individual, group, and community levels
2) Promote Health Value: considers the collective benefits and needs of public health systems and aims to maximize them
3) Prioritize Equality: ensures that individuals and communities have an equitable stake in the health value created by data use

Each objective has several pillars that guide health data governance as presented in the figure below.

The following section summarizes and cites the objectives of the World Bank's Health Data Governance Principles and was utilized as a framework for this study.[6]

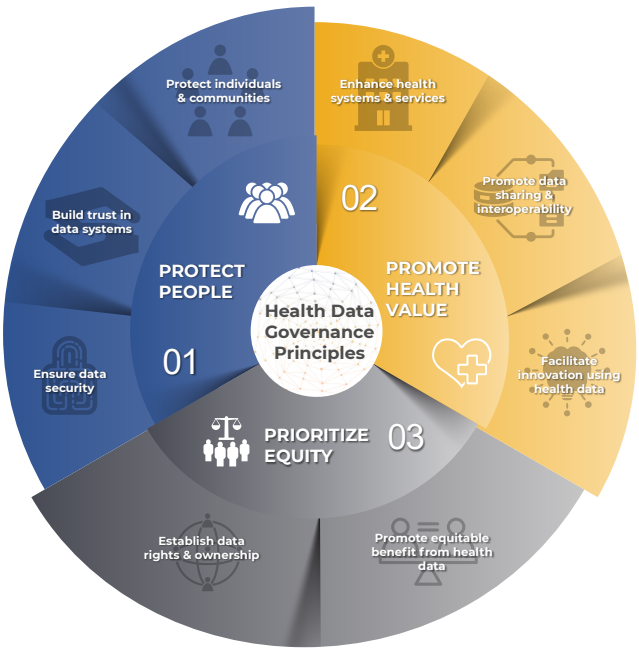


Figure 1: The Health Data Governance Principles framework.[6]

Objective 1: Protect People

The first objective of the Principles is composed of several pillars that convey the importance of individual, group, and community health data protection against the different types of harm that can be attributed to the sensitive nature of health data.

First Pillar: Protect Individuals and Communities

The core elements of this pillar span numerous aspects that concern the protection of people's health data. These elements emphasize the need for guidelines to define the limitations of data collection and highlight the importance of gathering informed consent from individuals and communities, which includes parties understanding their data rights and how their health data may be used.

Other core elements focus on the utilization of secure data collection and storage mechanisms, such as employing data collection tools with robust data protection features and ensuring secure data storage practices like encryption and use of cloud servers.

Second Pillar: Build Trust in Data Systems

This pillar highlights the significance of achieving public trust in health data governance, which can be fostered through a transparent and inclusive governance system. Data system trust-building encompasses many elements: protecting data, maintaining privacy, and establishing transparent and inclusive processes throughout the data lifecycle. There is a need to employ technical measures, such as encryption and two-factor authentication, to increase data security. Moreover, systems should acknowledge the data subject's right to comprehend the purpose of data collection. By taking all of this into consideration, maintaining transparency and inclusive stakeholder engagement, and implementing mechanisms that respond to feedback and concerns over data misuse, health data governance can foster trust and uphold standards of integrity and respect for individual and collective rights.

6 The World Bank's Health Data Governance Principles (https://healthdatagovernance.org/principles/), revisualised

Third Pillar: Ensure Data Security

The third and final pillar of this objective focuses on the importance of developing security practices with the emergence of new technologies. Its core elements mandate the adoption of robust security mechanisms throughout the data lifecycle, which includes the utilization of guidelines to define and enforce measures like password standards, data encryption, and two-factor authentication, as well as the mitigation of potential consequences of security breaches.

One core element also draws attention to the need for transparency around data breaches, meaning that in the event of a breach, the responsible entity should not only report it to relevant authorities but also promptly inform the people who are affected.

Finally, the elements suggest the implementation of federated data systems, which maintain data close to their origin while enabling consent-based viewing and analysis across the health system. All of this can reinforce data security, ensuring that protective measures are in place to mitigate risk while upholding the framework's objective to protect people.

Objective 2: Promote Health Value

The Principles' second objective focuses on the efficiency and resilience of health systems. It emphasizes the critical role of deriving the maximal value from health data use and analysis to enhance outcomes for both individuals and society, which entails responsible data sharing and aggregation.

First Pillar: Enhance Health Systems and Services

This pillar's core elements are concerned with the potential that health data holds in enhancing the quality, efficiency, effectiveness, and sustainability of healthcare systems, which will, in turn, maximize the value of health outcomes. The first, and perhaps main, element discusses the ways in which data can promote health and wellbeing for individuals and communities; enhanced health services can encompass improved access, robust surveillance, accurate diagnostics, etc. Health data governance should also acknowledge and facilitate the role and agency of frontline health workers; proper governance has the capability of empowering these workers to leverage data-based insights in their daily tasks, which will guarantee that their roles are fortified rather than undermined by data systems.

Second Pillar: Promote Data Sharing and Interoperability

This pillar underscores the significance of data sharing as a fundamental aspect of extracting maximum value from health data while maintaining ethical and equitable practices. It emphasizes that data sharing should be carried out in a manner that upholds human rights, both on local and global scales. It also characterizes data sharing as a catalyst for obtaining deeper insights into health needs, which is crucial for fostering a more effective healthcare ecosystem. In addition, the pillar's core elements address the importance of interoperability for efficient and secure data sharing, which involves designing data and digital health systems in a manner that simplifies the process and minimizes errors, enabling seamless exchange of information and enhancing data portability and access. Moreover, systems should focus on defining common data structures across health systems, such as standardized data fields and system architecture, which will facilitate interoperability and streamline data consolidation from several sources.

Another core element of this pillar addresses the varying levels of data access, advocating that health data governance should outline and clarify which stakeholders can access de-identified or anonymized data.

Third Pillar: Facilitate Innovation Using Health Data

This pillar emphasizes the need for forward-looking health data governance that is able to adapt to new technology such as AI and to support innovation in the health sector. Its elements are concerned with the imperative of emerging technologies' adherence to the governance framework, which entails integrating health data governance principles and policies during the preliminary design stages of such technologies.

Other elements address the use of non-health data in health contexts. Proper health data governance should consider the integration of different data types and sources

into health contexts, upholding and aligning with well-defined health data governance principles. They also note the need for developing and enhancing digital public infrastructure, as it is crucial for facilitating health service delivery.

Most importantly, this pillar recognizes that policy and legislation might lag behind rapid technological advancements, underscoring the urgency of crafting new policy frameworks to address the capabilities of data collection and processing.

Objective 3: Prioritize Equity

The Principles' final objective draws special attention to the importance of ensuring that the health value and benefits derived from health data are distributed equitably among all people.

First Pillar: Promote Equitable Benefits from Health Data

Health data governance must ensure that data collection is inclusive and representative of all demographic and social groups, promoting fair and comprehensive insights into health trends and needs. To achieve equitable representation in data, data collectors need to consider the unique needs of marginalized groups. To that end, intersectional and cross-cutting approaches should be implemented when collecting and analyzing data for marginalized communities, with the consideration of several categories, including gender, age, class, ability, citizenship, etc. Mitigating data bias is also a core element of prioritizing equity. Health data governance should identify and counteract bias at all stages of the data lifecycle to prevent discrimination reinforcement through misuse of health data.

Health data governance should be written in an understandable, gender-neutral language and be accessible to diverse populations. This inclusiveness should also be evident in the feedback mechanisms of data systems, meaning that data contributors and health workers should comprehend the purpose and outcomes of data use, thus ensuring informed agency over their data and fostering meaningful community engagement.

Second Pillar: Establish Data Rights and Ownership

This pillar emphasizes the necessity of strong and clear data-related rights as the foundation of health data governance. The pillar highlights the centrality of human rights in shaping data principles, laws, and policies, including those that pertain to collection, processing, and ownership. Through this human-rights framework, governance roles and responsibilities will be defined, which is crucial to upholding rights and ownership. Yet, in order to implement data rights and ownership effectively, institutions like health data trusts and cooperatives should be established, which will define rules for data-related activity while respecting individual and community rights. Most importantly, health data governance should integrate with formal public accountability mechanisms to ensure adherence to health-related policies, laws, and rights, thus enhancing transparency and trust.

III. FIELDWORK

3.1. Public Health Centers

Preliminary Fieldwork

Duration: November 17–December 01, 2022

Objective:

During a two-week preliminary fieldwork phase, exploratory field visits to five PHCs were conducted in Cairo and Giza. The objective of these exploratory visits was to observe the function and processes at the PHCs and identify possible study participants to finalize the fieldwork interview questions and data catalog in preparation for the next phase of the fieldwork from March to April 2023. The interview questions for the final fieldwork phase were tailored and finalized based on our observations at these preliminary visits. This phase aimed to collect insights on and map the governance of health data in PHCs, with a focus on data on women's health, identify gaps, and provide policy recommendations for the governance of responsible health data for Egypt.

Observations and Analysis

Mapping of Public Health Centers (PHCs)

The PHCs are composed of a registration counter, a registration office for birth and death certificates, a pharmacy, laboratories including a blood testing

FIELDWORK TIMELINE

PUBLIC HEALTH CENTERS (PHCS)

PHASE 01

TWO-WEEK PRELIMINARY FIELDWORK PHASE

HEALTH CENTER VISITS
Exploratory field visits to five public health centers in Cairo and Giza.

OBSERVATION
Observe the function and processes at the PHCs.

IDENTIFY PARTICIPANTS
Identify possible study participants.

FINALIZE INTERVIEWS
Finalize the fieldwork interview questions for the next phase.

PHASE 02

SECOND ROUND OF FIELDWORK

INTERVIEWS
Second round of fieldwork conducted to carry out qualitative semi-structured interviews with nurses and administrators.

PRIVATE SECTOR

SEMI-STRUCTURED EXPERT INTERVIEWS

INTERVIEWEE SELECTION
Interviewees included representatives of health startups, radiology centers, private clinics, as well as actors from the pharmaceutical industry.

EXPERT INTERVIEWS
Semi-structured expert interviews were conducted online to survey private health settings.


Shubra al Khayma
شبرا الخيمة
Ausim Public Health Center
مركز صحة أوسيم
Bashteel Public Health Center
رعاية أولية بشتيل
Aziz Ezzat Public Health Center
مركز صحة أرض عزيز عزت
Cairo
القاهرة
Giza
الجيزة
Abu Al Nomros Public Health Center
مركز صحة أبو النمرس
Hawamdeya Public Health Center
مركز صحة الحوامدية

Figure 2: Women waiting for their medical examinations in front of the Women's Health Clinic, Source: A2K4D's illustration from field visits.

unit, and a dental lab. Moreover, there is a care unit for pregnant women, a family planning unit, and a childcare clinic. As part of the Egyptian Women Health Initiative, there is a women's health clinic (see Figure 2) where sugar levels and blood pressure levels are measured. They also have a sanitization room, a vaccination room, an emergency room, as well as a filing room to store paper files. Each center has a women's health club that aims to raise women's awareness of family planning, and reproductive health. Some public health centers provide neonatal hearing examinations, premarital screening, and hepatitis detection services.

Patient Data Collection Mechanisms

Data collection is digitized when it is part of specific government health initiatives, while routine patient visits or cases where the patient is not registered in the PHC system are still managed on paper.

Patients are expected to bring their medical card that they received from previous PHCs visits. If they do not have one, they must wait in line at the registration counter to get a ticket. A clerk then writes down the patient's information and requested service on the ticket. When the patient arrives at the clinic and completes their check-ups, all collected information is recorded on the ticket or in the patient's medical file. The data is recorded on the ticket in cases where the patient is not registered in the PHC's records, such as when they live in a different geographical area. A filing room is used for storing these paper records.

In contrast, digitized data collection occurs through government health initiatives like the 100 Million Healthy Lives and the Egyptian Women's Health initiatives, where tablets are used to directly enter women's data into a digital database. In these initiatives, women are registered using their national ID numbers, after which their height and weight are measured. They are then grouped by age: women under 25 receive awareness on natural versus cesarean births; those between 25 and 29 undergo breast examinations; and those between 30 and 40 have blood tests for creatinine and sugar levels.

7 https://www.100millionseha.eg/main/about .100 مليون صحة

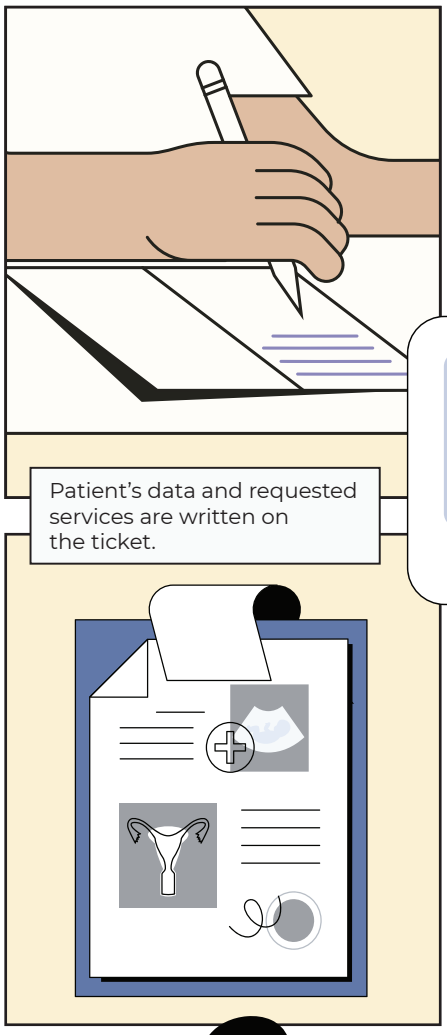
8 Hassanin, A., Kamel, S., Waked, I., & Fort, M. (2021). Egypt's ambitious strategy to eliminate hepatitis C virus: a case study. *Global Health: Science and Practice*, *9*(1), 187-200.

9 Ibid.

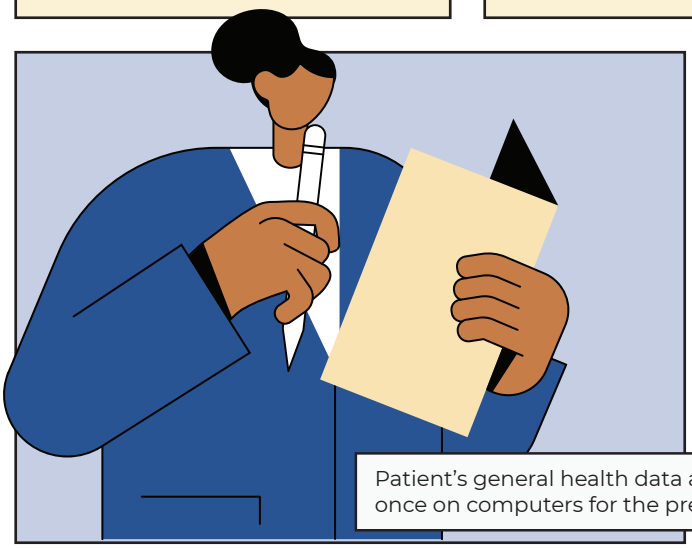
100 MILLION HEALTHY LIVES INITIATIVE

The presidential initiative was launched in October 2018 in order to eliminate Hepatitis C (HCV) and detect noncommunicable diseases.[7] It provides free HCV examination to all Egyptians above the age of 18, as well as free treatment for those who test positive.[8] In addition, part of the purpose of the campaign is to eliminate waiting lists, as well as raise health awareness and promote healthy lifestyle habits to prevent noncommunicable diseases.[9]


Patient goes
to registration
desk to get a ticket.


Patient's data and requested
services are written on
the ticket.


Patient goes to clinic for checkups.




Patient's general health data are recorded. Once on paper and
once on computers for the presidential digital health initiatives.

Potential Study Participants

There are few health professionals at the PHC: a pharmacist, a dentist, a pediatrician, and an internist. Generally, the main personnel are the administrators who are responsible for tickets and payments, in addition to nurses who usually conduct the medical checkups and examinations and then transfer the patient to a Women's Health Initiative hospital, if needed.

Evaluation/Insights and Limitations

Informed by the preliminary fieldwork phase and exploratory visits, semi-structured interview questions were developed. Questions were to be directed to two groups of potential interviewees:

Nurses would be asked a set of questions on data acquisition

Administrators would be asked a set of questions on data management and health data workflow

Second Round of Fieldwork

Based on the preliminary phase, the second round of fieldwork was conducted, between March and April 2023 to carry out qualitative semi-structured interviews with nurses and administrators at the PHCs.

The interview questions were conceptualized based on the context of the public healthcare centers in Egypt. They cover several areas, including: (1) patient data acquisition practices, including data collection methods, the types of data collected on paper versus the government's digitized initiatives, and their (2) data management practices, including data workflow, the sequence of data-related tasks performed by the different healthcare workers within the public health center, data access , storage, retrieval, update, quality, sharing, security as well as their data-related challenges and the best practices to overcome these challenges. Furthermore, questions related to the use/ potential use of digital technologies in the future were also included.

3.2 Private Sector Interviews

Semi-structured expert interviews were conducted to survey private health settings. Interviewees included representatives of health startups, radiology centers, private clinics, as well as actors from the pharmaceutical industry.

10 Egyptian Women's Health Initiative. Presidency of the Arab Republic of Egypt. https://www.presidency.eg/en/%D8%A7%D9%84%D8%B1%D8%A6%D8%A7%D8%B3%D8%A9/%D9%85%D8%A8%D8%A7%D8%AF%D8%B1%D8%A9-%D8%B1%D8%A6%D9%8A%D8%B3-%D8%A7%D9%84%D8%AC%D9%85%D9%87%D9%88%D8%B1%D9%8A%D8%A9-%D9%84%D8%AF%D8%B9%D9%85-%D8%B5%D8%AD%D8%A9-%D8%A7%D9%84%D9%85%D8%B1%D8%A3%D8%A9-%D8%A7%D9%84%D9%85%D8%B5%D8%B1%D9%8A%D8%A9/

11 Mahmoud Ali, E., Mohammed Attya, A., Ragab Abo Shabana, K., & Gouda Nasr, E. (2022). Women's Perception Regarding the Ministry of Health Initiatives Plan for Early Detection of Breast Cancer. *Egyptian Journal of Health Care*, *13*(4), 449-458.

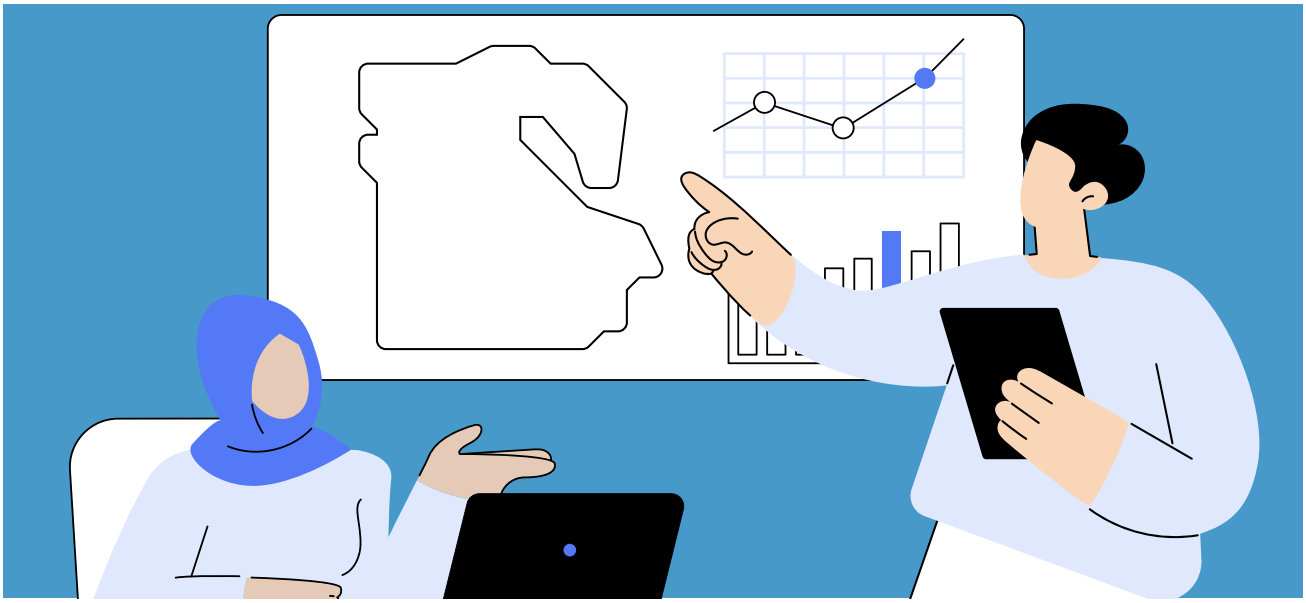
12 https://www.100millionseha.eg/women/about .صحة المرأة .100 مليون صحة

13 Ibid.

EGYPTIAN WOMEN'S HEALTH INITIATIVE

With statistics from the World Health Organization showing an alarmingly high incidence of breast cancer among Egyptian women, this presidential initiative was launched with a focus on reproductive health. Egyptian women account for 35% of all cancer cases among women.[10] The goal of the initiative is to detect non-communicable diseases and promote early detection of breast cancer, which has proven to increase the likelihood of successful treatment.[11]

This presidential initiative targets women over the age of 18 by providing clinical assessment, breast examination services and free treatment. The examination includes a 28-question questionnaire regarding the patient's social and health data, and is conducted in "an atmosphere of confidentiality and privacy."[12] The examination is coupled with an awareness session on the importance of early detection, encouraging women to periodically come in for examination and teaching them how to self-examine.[13]



Modified versions of the interviews were used as appropriate. Consent of the interviewees was secured ahead of the interview and the guiding questions were shared in advance. Interviews were held online and recorded with the interviewees' consent, lasting approximately 60 minutes each, with answers documented and transcribed. The key takeaways were shared with the interviewees for approval prior to their use for this report.

IV. FIELDWORK FINDINGS

Through the lens of the World Bank's Health Data Governance Principles,[14] the semi-structured interview questions are tied to and organized into their corresponding principles to explore and describe the qualitative insights on different aspects of the existing health data practices. See Annex 1 to view the interview questions vs their corresponding WHO Health Data Governance Principles.

The insights collected from the semi-structured interviews are written in textual form before analysis. These texts are produced by transcribing the interview data as well as creating notes based on observations from the field. In the following section, these fieldwork findings are organized based on the World Bank's health data governance structure, which is the theoretical framework of this research.

14 Health Data Governance: Universalising the benefits of health digitization, World Bank, 2022, https://healthdataprinciples.org/images/English/%5BEN%5D%20Health%20Data%20Governance%20Principles.pdf

Figure 3: A woman patient purchasing a ticket at the registration counter. Source: A2K4D's illustration from field visits.

4.1. Data Practices in PHCs: Data Collection

P1. Protect People

P.1.1 Protect Individuals & Communities

The first pillar focuses on protecting data subjects and communities through collecting data with defined purposes, collecting personal and sensitive data with informed consent, and deploying secured data collection methods.

In the context of PHCs, the patient goes to the registration counter to purchase a ticket. The clerk writes down the patient's name, date, ID number, and requested service on the ticket (see Figure 3). The patient then goes to the requested clinic and has the checkups done (see Figure 4).

For regular examinations in the different care units, general data such as name, age, date of birth, ID

Figure 4: At the care unit, a nurse writes down the data/measurements of the patient after the check-up. Source: A2K4D's illustration from field visits.

number, and address are noted, and health-related data such as height, weight, and blood pressure are measured during each visit to assess the overall health of the patient. Also, information on the patient's medical history (e.g., chronic illnesses), medications that the patient has previously taken, and symptoms are collected to keep up with the patient's history. For instance, in family planning units, nurses ask the woman patient for information such as the number of children, years of marriage, number of previous deliveries, number of c-sections versus natural births, the contraceptive methods used, previous pregnancy complications, as well as the date of the last menstrual period to calculate the expected delivery date. Furthermore, body mass index, obesity, and weight gain are also measured. As for the government's digitized initiatives, specific questions, based on the aim of each initiative, are asked to the patient for diagnostic purposes.

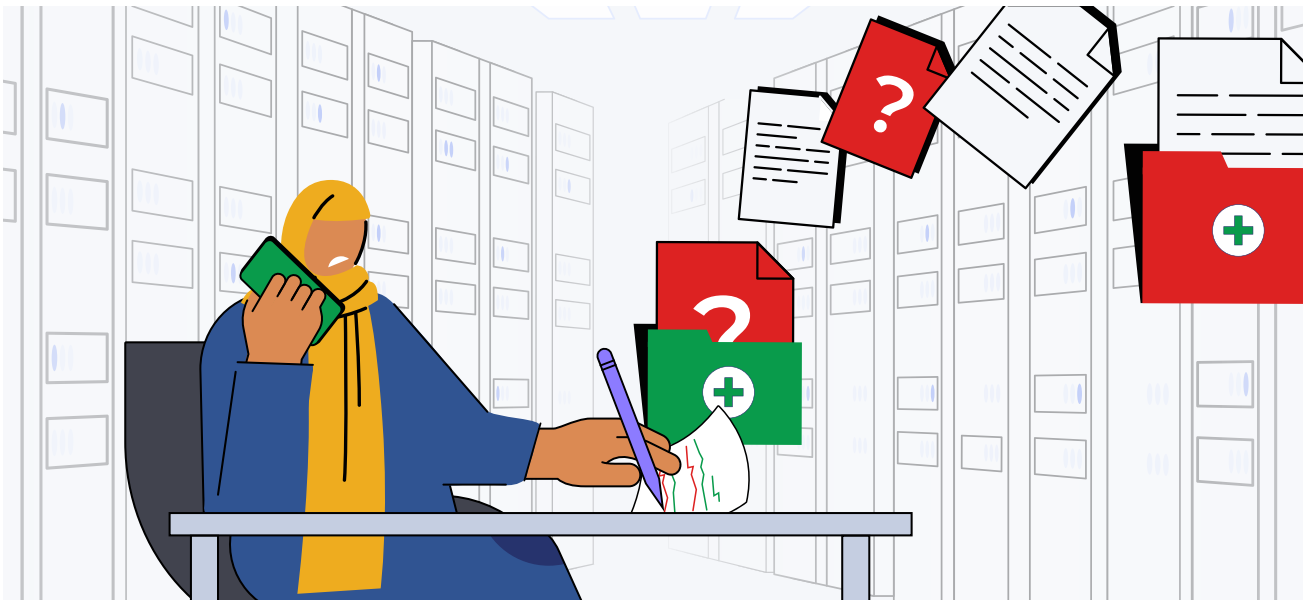
Reflecting on patient consent during data collection, nurses are careful to protect the patients' privacy. However, consent is only obtained in specific cases, including surgical interventions such as contraceptive-device placements at family planning units or dental procedures such as tooth extractions, before which patients must sign consent forms that outline potential side effects.

Generally, patients do not fully understand the implications of having their data collected and recorded. In this context, data is treated as confidential, meaning that the files are only accessible to authorized personnel, such as the PHC manager, the treating physician, the clinic nurse who collects the data, and the head nurse. Although the data is handled with confidentiality, it is not categorized as "sensitive," implying that it does not include information considered highly private or requiring additional protection under data protection laws.

P1.2 Build Trust in Data Systems

The second pillar focuses on inclusive data collection mechanisms, ensuring that consent is informed and securing high-quality and accurate data collection processes.

Reflecting on data quality, only authorized personnel are permitted to collect and/ or register patients' data. Random data checks are conducted periodically, where data are picked arbitrarily to cross-check their accuracy. In repeated visits, the nurses who input the data usually revise it manually during every recurring patient visit. Health-related data like weight, body temperature, and blood pressure are measured during every visit. The collected analog data is revised



regularly by the head nurse to detect data entry errors, including entering data in wrong fields, entering incorrect numbers, and missing/empty data fields. Moreover, periodic oversight is conducted by the administration, directorate, or the Ministry of Health. For paper-based records, the data errors are corrected manually.

As for the presidential initiatives, such as the Egyptian Women's Health Initiative and the 100 Million Healthy Lives initiative, minimal data errors occur in digital data entry, as there are predefined fields that control data entry formats to help the nurses enter the right type of data. The system detects illogical or incorrect information and requests its correction. In cases of digital data errors, data corrections can only be made through the Ministry of Health's Information Center. IT personnel are in charge of keeping track of the government's information systems, and periodic review is also conducted to identify data errors.

P2. Promote Health Value

P2.1 Enhance Health Systems Services

The first pillar focuses on enhancing health services by increasing community engagement during data collection and empowering frontline health workers with the needed skills to support their collection of health data.

Reflecting on the context of the public health centers, all the data collection is still done exclusively on paper in registers or notebooks dedicated to each clinic (see Figure 5). Nurses rely on patients' verbally communicated information in data collection. Registered patients have their own patient codes and a medical notebook that they are expected to bring with them on each visit (see Figure 6). The data of unregistered patients are recorded on their tickets and cannot be subsequently retrieved.

A limited number of tablets/computers are available for use across all the presidential initiatives, resulting in a time-consuming process to collect all the data and navigate between different initiative systems. In many cases, nurses resort to using their personal mobile phones. Although the initiatives have digital data entry systems, nurses still think paper is essential, as they are used to paper-based data collection methods. They initially record data on paper for reference, despite the digital infrastructure of the presidential initiatives.

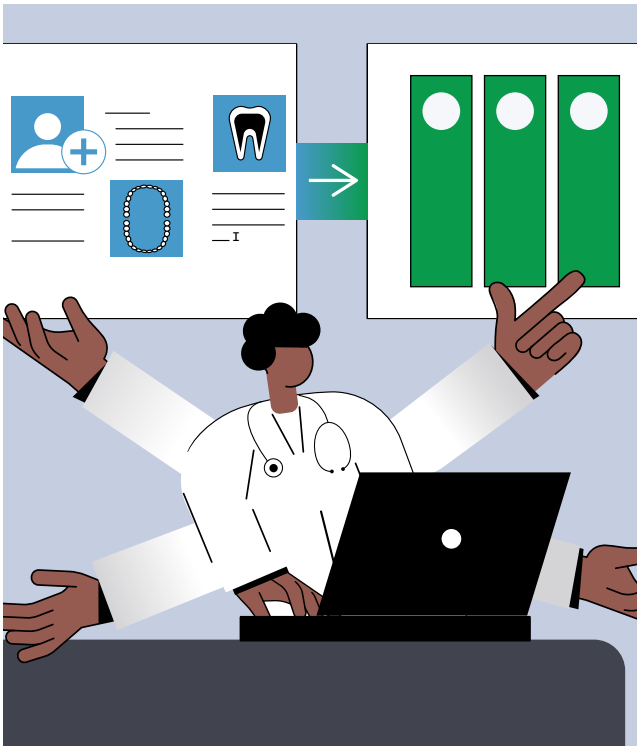


Figure 5: A nurse writing down the woman patient's data on paper. Source: A2K4D's illustration from field visits.

Figure 6: A registered patient showing her medical notebook to the nurse. Source: A2K4D's illustration from field visits.

Figure 7: In a typical scene at any presidential health initiative, a nurse enters data on a tablet while two other nurses, working on different initiatives, await their turn to use the device. Source: A2K4D's illustration from field visits.

Figure 8: In a presidential initiative care unit, a doctor is using a computer to enter the patient's data on the initiative's digital system instead of a nurse/healthcare data entry employee. Source: A2K4D's illustration from field visits.

Furthermore, most nurses are not digitally literate and few are trained to use technology, meaning the unit is entirely dependent on the few that can use the tablets and computers (see Figure 7). In some cases, doctors conduct the data entry, as they may be more digitally literate than the nurses, which is a misuse of human resources (see Figure 8). As such, more efficient allocation of online digital data entry tasks is needed.

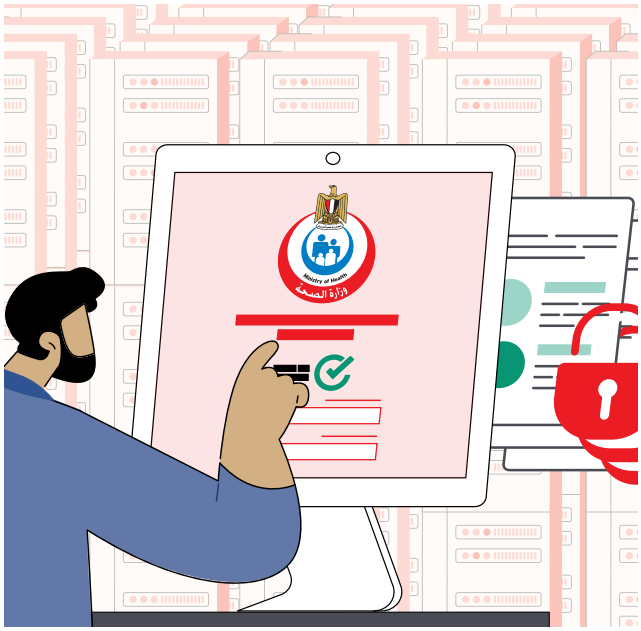
In addition, there are internet connectivity issues due to unreliable internet infrastructure in Egypt. Hence, nurses may encounter difficulties in accessing the internet, which is required to complete their online data entries on the initiatives' digital systems. In the event of an internet outage, nurses record data on paper for subsequent entry into the online system once the internet connection is restored, as no offline data entry option is afforded. Nurses recommend enabling offline digital data collection so that they can use the digital system during an internet outage. Enabling this would require the system to automatically synchronize and update the recorded data when internet connectivity is restored. Nurses also noted their need for reliable, wired internet connections.

P2.2. Promote Data Sharing & Interoperability

The second pillar focuses on promoting data sharing by using previously collected data to reduce the need for new data collection, designing interoperable data collection systems, and defining common data collection structures and fields.

In the context of the PHCs, the data of registered patients who have a medical file may be manually retrieved from previous records using their registration number. Cases that cannot be treated within the PHC are referred and their data is recorded in a referral log that usually contains incomplete data fields due to the lack of communication and absence of a feedback loop between the PHCs and other health entities. Furthermore, unregistered patients, such as those who do not live near the PHC, have no medical data history recorded or stored at the center.

For the presidential digital initiatives, digital systems are designed with specific data fields. However, nurses are not able to retrieve any of the previously collected data since they do not have access to the data collected on their computers and tablets. The collected data is only accessed



or retrieved centrally through the Ministry of Health's Information Center. Moreover, presidential initiatives are working in silos, wherein the initiatives have separate data collection systems that do not complement each other.

P2.3. Facilitate Innovation Using Health Data

The third pillar focuses on adopting policies that support innovative collection of health data to enhance innovation in the healthcare sector.

Reflecting on the context of the PHCs, there are several data collection-related problems that hinder the potential deployment of innovative data collection tools, as these health centers are still mainly reliant on paperwork. For instance, nurses have to comb through a substantial amount of paper files to retrieve patient records, which is inefficient. Furthermore, digitizing paper records requires substantial time and effort, as each PHC has approximately 6000 paper-based files.

Moreover, there are still some barriers related to the utilization of digital technologies in collecting health-related data due to the awareness and capacities of healthcare workers with regard to using computers and/or tablets in digital data collection processes. The limited number of digitally trained nurses and available tablets results in time-consuming and effort-intensive data collection processes. Furthermore, nurses do not have the ability to directly modify or address entry errors. They are instead required to go through a time-consuming process of writing a memorandum to the relevant Ministry of Health unit, explaining the data error and required modification to arrive at a suitable solution. As previously mentioned, nurses may also encounter difficulties in accessing the internet if there is a high demand for online digital services and internet usage in the different PHC units.

Hence, there is a need to expand the human resources at the PHC level, in addition to increasing the number of tablets and overcoming the internet access and connectivity issues through the installation of reliable internet connection.

P3. Prioritize Equity

P3.1 Promote Equitable Benefits from Health Data

The first pillar focuses on advancing equitable benefits from health data through inclusive data collection methodologies, in terms of which individuals/groups are asked to provide data, as well as which data categories are collected to avoid data bias in health data collection. This means cross-cutting data collection processes along categories like gender, age, socio-economic status, abilities, citizenship status, and ethnicity.

In PHCs, the primary focus is on women. However, there are challenges in collecting cross-cutting data due to a lack of awareness about its value and importance. Priority is given to registered women living in the geographical area served by the PHC. Their data is collected in a specified notebook designated for each individual. However, women who do not reside in the same area do not have their data history collected in the same comprehensive manner. Moreover, in both cases, the data collected does not yield any benefits or value because it is not analyzed. The data is primarily recorded in paper notebooks, making digital analysis impossible. Therefore, there are limitations in deriving benefits from the collected data due to the lack of analysis and digitalization.

P3.2 Establish Data Rights & Ownership

This pillar focuses on establishing data rights through developing data trusts and health data cooperatives to define the rules of data collection. Currently, there are no formalized data rights, data trusts, or defined rules for data collection at the PHC level. Patients lose their rights concerning their personal health information, as they are unaware of them. There are no clear guidelines or mechanisms in place that allow patients to know how their health data will be used, who will have access to it, or how it will be protected. This absence of data rights undermines patient trust in the healthcare system. Moreover, the rules and guidelines for data collection at PHCs are not well-defined, as the data is collected primarily for immediate healthcare needs. This lack of defined rules may cause inconsistent data collection practices, making it difficult to ensure data quality and security.

4.2. Data Practices in PHCs: Data Processing

P1. Protect People

P.1.3 Ensure Data Security

This pillar focuses on protecting patients and their data by establishing strong technical security measures for data processing. In PHCs, data is collected and stored in paper notebooks designated for each individual and securely locked in a dedicated filing room. While this method ensures some level of physical security by re-

stricting access to authorized personnel only, it does not support advanced data security measures that are typically employed during digital data processing, such as encryption or de-identification. Moreover, there is no data processing happening at the PHC level since the collected data is not digitized.

P.2. Promote Health Value

P.2.3 Facilitate Innovation Using Health Data

This pillar focuses on enhancing innovation in healthcare through the deployment of innovative health data processing mechanisms. Reflecting on the context of the PHCs, tablets and computers are used in the presidential digital initiatives to manage the center's data. Nurses believe that digital technologies may help them in data operations, namely updating, maintaining and retrieving data. They are looking forward to the full digitalisation of work and reliance on the digital system instead of paper-based processes. The lack of digitization means that data cannot be easily analyzed, shared or processed in any meaningful way. The inability to process data digitally means that valuable insights cannot be derived from the collected information. Data processing, which could inform better and innovative healthcare decisions and policies, is not feasible within the current PHC system.

4.3. Data Practices in PHCs: Data Sharing

P1. Protect People

P.1.2 Build Trust in Data Systems

This pillar focuses on establishing trust in data systems by allowing data subjects to accept or decline further sharing of their data for purposes beyond its initial intended use. At the PHC level, this concept is not being implemented for several reasons. First, the collected data at PHCs is not intended for further use beyond its primary purpose. The current data collection approach focuses solely on immediate healthcare needs without considering the potential for further data uses. As a result, there is no mechanism in place for patients to consent to or decline the further sharing of their data. This reflects a broader issue amid a perception that the collected health data may never be utilized, which hinders innovation.

Moreover, there is a significant lack of awareness about the importance of health data among both healthcare workers and the women patients at PHCs. Healthcare workers are not adequately informed about the potential value and benefits of utilizing health data for broader health insights, research, or policy-making. Hence, they do not communicate this importance to the patients.

Furthermore, the women patients themselves lack awareness and trust in data and AI technologies. This lack of trust and understanding means that even if there was a system in place to seek their consent for data sharing, it is most likely that they would be hesitant or refuse to agree to the sharing and re-sharing of their data. This is due to the limited understanding of how their data could be used, the benefits it could bring, and the measures that would be in place to protect their privacy.

P.1.3 Ensure Data Security

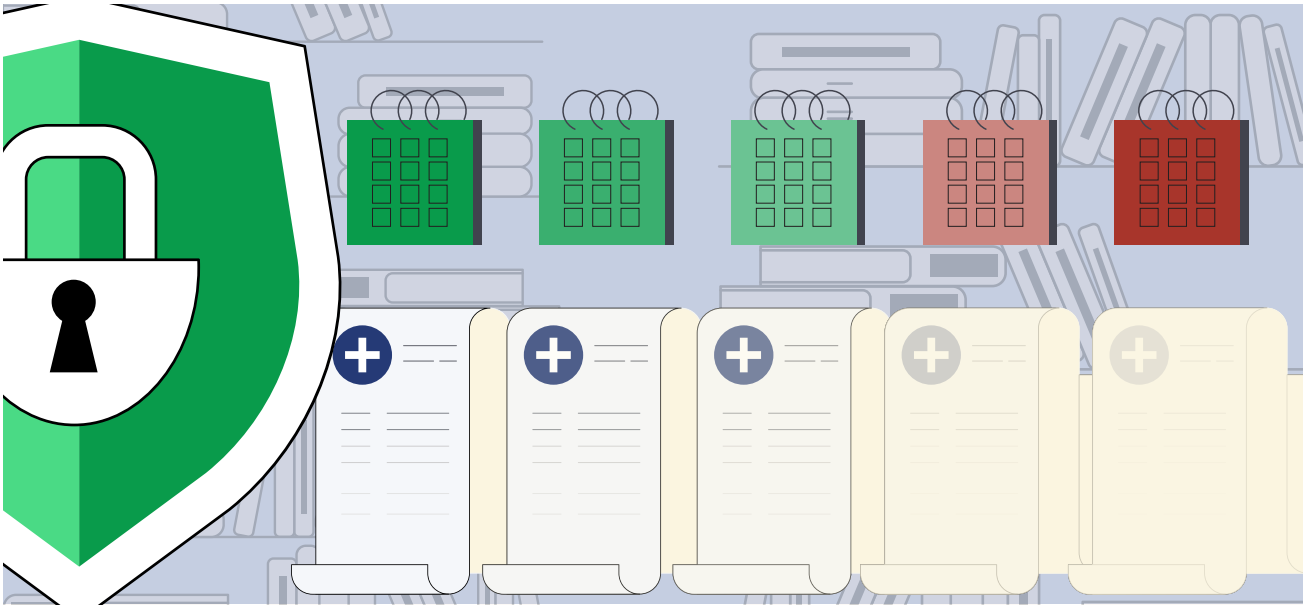
This pillar focuses on protecting data security by establishing federated data systems across the health system that allow secure cross-system sharing. In the context of PHCs, data is primarily collected and stored in paper notebooks, which means there is no digital data to share in the first place. Even when data is digitized in the presidential health initiatives, it is not stored in an interchangeable format that facilitates sharing. This absence of digitization and reluctance of data sharing hinders any possibility of collaboration between PHCs and other healthcare entities.

Moreover, digital safeguards such as data encryption and backups cannot be implemented in paper records. Therefore, the starting point for better data security is the digitization of the health records. Digital records will allow for better access control, ensuring only authorized personnel can view sensitive information. Digitization will also provide for better standardization of data formats, facilitating secure data exchange between the different entities and thus reducing the risk of data breaches.

P.2. Promote Health Value

P.2.1 Enhance Health Systems Services

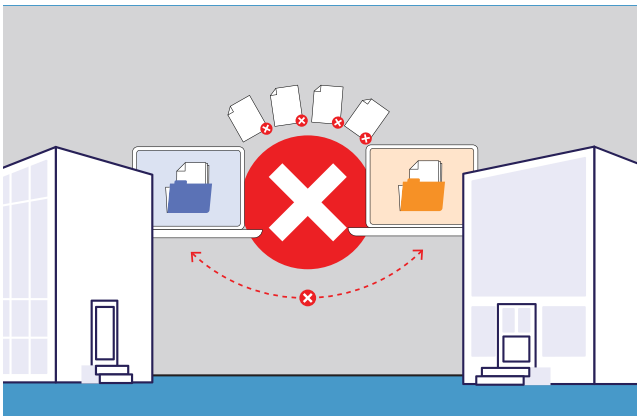
This pillar focuses on improving health services through enhancing data sharing between health facilities. As mentioned above, there is a lack of established data-sharing practices between PHCs and other healthcare entities such as labs and radiology centers or data from health professionals and physicians. Therefore, PHCs do not complement their data with any other external data sources. In cases of referral to government hospitals, an administrative referral letter is issued from the PHC to transfer the patient.



The hospital treats the referred patient as a new case since their data is not transferred, resulting in incomplete data fields in the PHCs' referral logs, as there are no proper communication and feedback loops between both entities.

On the other hand, the digital data collected at the PHCs as part of the presidential initiatives can only be accessed by the Ministry of Health as well as government hospitals associated with the initiative. It is retrieved from the digital system by entering the patient's national ID number.

Therefore, even when the data is digitized, there is a lack of data sharing between PHCs and other healthcare entities. They operate in silos with separate data collection systems that do not integrate or complement one another. Moreover, there is a lack of interoperability between different data systems, which means that even if the data is shared, it would not be easily usable across different entities. This fragmented approach and absence of data interoperability hinders any potential collaborative efforts, data sharing initiatives, and research toward better health systems and services.



4.4. Data Practices in PHCs: Data Storage and Security

P1. Protect People

P.1.1 Protect Individuals & Communities

This pillar focuses on protecting data subjects by using secure data storage mechanisms such as data encryption and cloud servers.

Reflecting on the context of the PHCs, the collected data is stored in paper files or separate notebooks for each clinic in dedicated filing rooms that are securely locked (see Figure 9). Only authorized personnel such as the treating physician, the clinic nurse that collects the data and/or the head nurse, the PHC manager, as well as government inspectors, have access to the filing room and are allowed to access the raw data in the registers (see Figure 10).

Paper data files are kept for a period of five years, which makes it difficult to maintain the quality of the paperwork due to issues like ink fading and paper decomposition. The nurses resort to stapling additional copies to the original papers in order to ensure that the data remains again. Furthermore, the stored registers take a lot of physical space.

For the presidential digital initiatives, the collected digitized data is stored on a server shared with and managed by the Ministry of Health to ensure its security. In case of any technical issues, the Ministry's Information Center is contacted for support.

Figure 9: Shelves that contain the paper records and notebooks with information concerning patient health. Source: A2K4D's illustration from field visits.

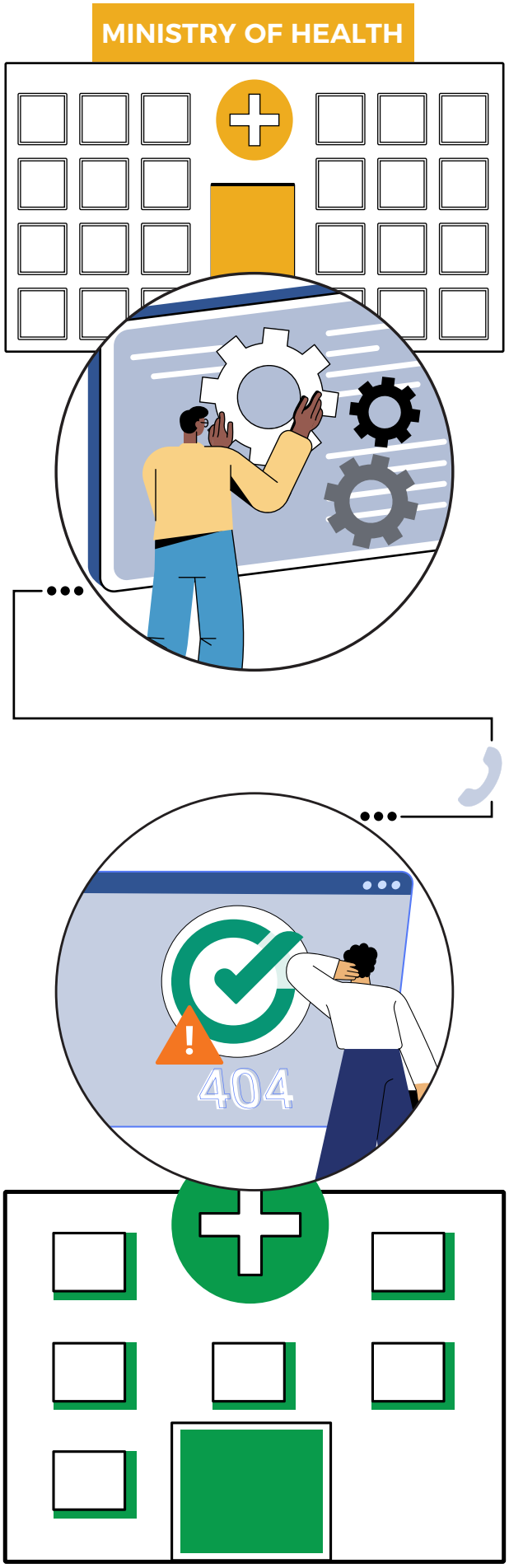
Figure 10: A doctor searches through patient records stored and kept in the filing room. Source: A2K4D's illustration from field visits.

P.1.2 Build Trust in Data Systems

This pillar focuses on deploying data protection and privacy practices for data storage such as two-factor authentication, encryption, and data de-identification, so that the stored information does not identify the patient with whom the data is associated.

In the context of PHCs, the collected data is not anonymized or encrypted. Hence, patients' personal details are not removed and can still be identified. However, as previously mentioned, only authorized personnel are permitted to view stored patient data.

As for the presidential digital initiatives, once the collected data is stored on the Ministry of Health system, healthcare workers at the centers do not have access to it through their computers and tablets. IT personnel are in charge of keeping track of the government's information systems, including data pertaining to the initiatives.



5.1 Data Practices in the Private Sector: Data Collection

P1. Protect People

P.1.1 Protect Individuals & Communities

In the context of health startups, the type of data collected from the patient depends on the service provided. At one of the surveyed start-ups , an optional patient feedback system collects the patients' contact information in the form of textual data, including phone number, email, and name. This startup allows patients to submit their feedback responses regarding their experience at the health entity. Patients are also able to submit audio data in the form of voice recordings in the case of a complaint.

In addition to collecting patient contact information, many healthcare startups also collect demographic data such as age and gender, as well as specific medical data related to the patient's health. This data could include blood tests, laboratory and radiology results, symptoms, medications, medical history, and previous diagnoses. Data is also collected from the healthcare providers, including the doctor's name, phone number, email, accreditations and certificates, and the health services they provide, to be showcased to patients on the startups' website.

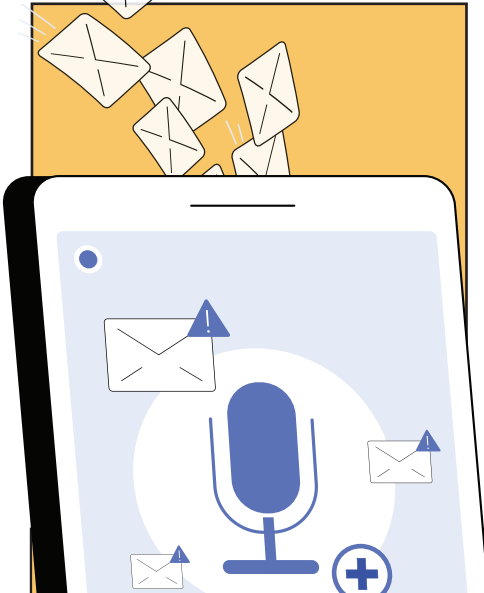
Radiology centers begin by collecting personal identifying data from the patients such as their full name, age, address, email, phone number, occupation, and national ID. One of the surveyed radiology clinics, which specializes in mammography, provides several forms that the patient must fill out to give the doctor a better understanding of their condition. The data collected includes weight, family medical history, smoking and drinking habits, and women's health-related data regarding menstrual period, number of children, cases of miscarriage, etc. When mammography scans are carried out, medical data is also collected in the form of radiological imagery to obtain information about mass size and location. After the patient is examined, the doctor records their Breast Imaging Reporting & Data System (BI-RADS) score, which ranges from 1 (not cancer) to 6 (high chance of cancer), as well as their breast density rating to produce the mammogram reports. Some radiology centers also noted the collection of non-health data such as customer satisfaction data.

The entities surveyed in the pharmaceutical sector collect the patient's name, age, phone number, email, and prescriptions. Other data included allergies, medical history, family medical history, other medications the patient may be using, blood type, and any diseases they may have. In addition to health-related data, one of the pharmaceutical start-ups surveyed also collected user behavior data on their website, including topics read, services used, favorite products, and the user location.

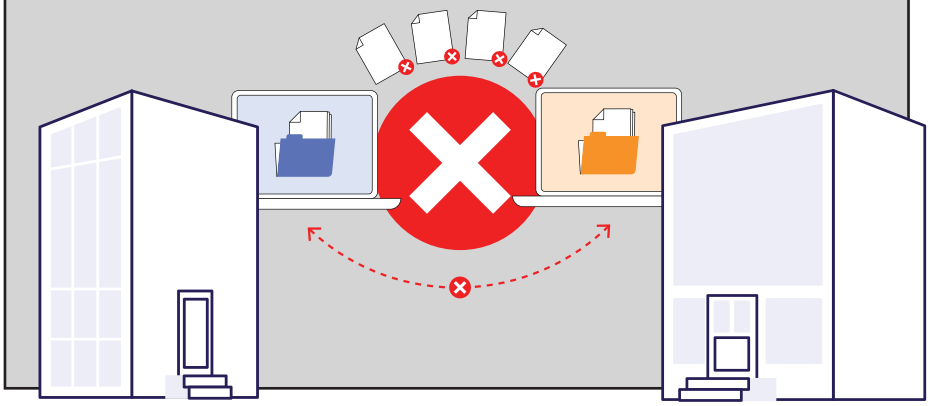
In terms of data collection mechanisms, some entities opt for analogue forms, while others use digital technologies via the entity's website, on-site tablets, computers, radiology machines, or email. Most of the entities give patients the option to enter their data in English or Arabic, while some only collect data in English due to ease of processing and global standards in medical reporting. Almost all the data collected

PATIENTS ENTER THEIR DATA ON TABLETS THROUGH SURVEY LINKS IN BOTH ENGLISH AND ARABIC.

COLLECTION OF PATIENTS' DATA THROUGH EMAILS.



THERE IS NO DATA SHARING, SENDING OR RECEIVING BETWEEN PRIVATE ENTITITES.



AUDIO FORMAT COMPLAINTS AREN'T PROCESSED AS CUSTOMER SERVICE TEAMS HAVE TO MANUALLY LISTEN AND TRANSCRIBE EVERY RECORDING.

from the patient directly, except when the patient provides a previous prescription or medical report. The majority of the entities surveyed do not check the accuracy of the data collected from patients. They are only able to check the data that they collect directly, such as weight and radiology images, for accuracy.

Most of the entities surveyed do not define or notify the patient of a specific purpose for their data collection. They argue that the purpose of sharing their data with the health entity is obvious to the patient due to the nature of the health service they are seeking. One lab collects additional data from patients to create customized medications. Another entity provides dosage reminders for medication and therefore needs the name of the medication, the required dosage, and the phone number of the patient's caregiver.

As for data collection with informed consent, the interviewed health entities do not usually receive written consent from their data subjects. However, they acquire patients' consent in the form of disclaimers noting that their data will not be used for any other purposes, disclaimers to doctors that their data will be shared on the entity's website for marketing purposes, permission from patients to transfer their data to another doctor within the same entity, written consent for data to be used in research, and consent to save credit card information. On the other hand, some entities do not ask for consent since they don't use or share any of the collected patient data. Some entities also anticipate that they would ask for data subjects' consent in the future when introducing new services that require further data sharing.

Some of the recommendations suggested by the surveyed entities on innovative ways to leverage technology in their data collection practices include transcription of voice recordings, AI to detect fake data entries, chatbots, AI referral system, automated transcription and compilation of medical reports, AI that can track trends and relations among variables, programs to identify masses in radiology images, and computer-aided diagnosis. The entities agree that the use of technology in data collection is a much more efficient option, as it saves time and resources. However, one entity, a radiology clinic, highlights that the use of technology-driven data collection is not always applicable in the medical field. They emphasize that technology and AI are meant to act as complementary services that can facilitate objective tasks such as locating a tumor, measuring its size, and establishing a preliminary diagnosis. They believe that such technologies should not replace radiologists or doctors, with common agreement among radiologists against the use of AI for subjective decisions such as whether a patient should undergo a certain treatment procedure.

P.1.2 Build Trust in Data Systems

Data bias is mitigated in different ways, including the anonymization of therapy sessions and deployment of an analytics team to detect and archive any aberrant data entries. Several entities, however, suggest that there is no bias or discrimination when dealing with a patient and their data, due to the medical nature of their work, stating that they "deal with patients as cases." They believe that they cannot be biased while collecting data since medicine is scientific and objective.

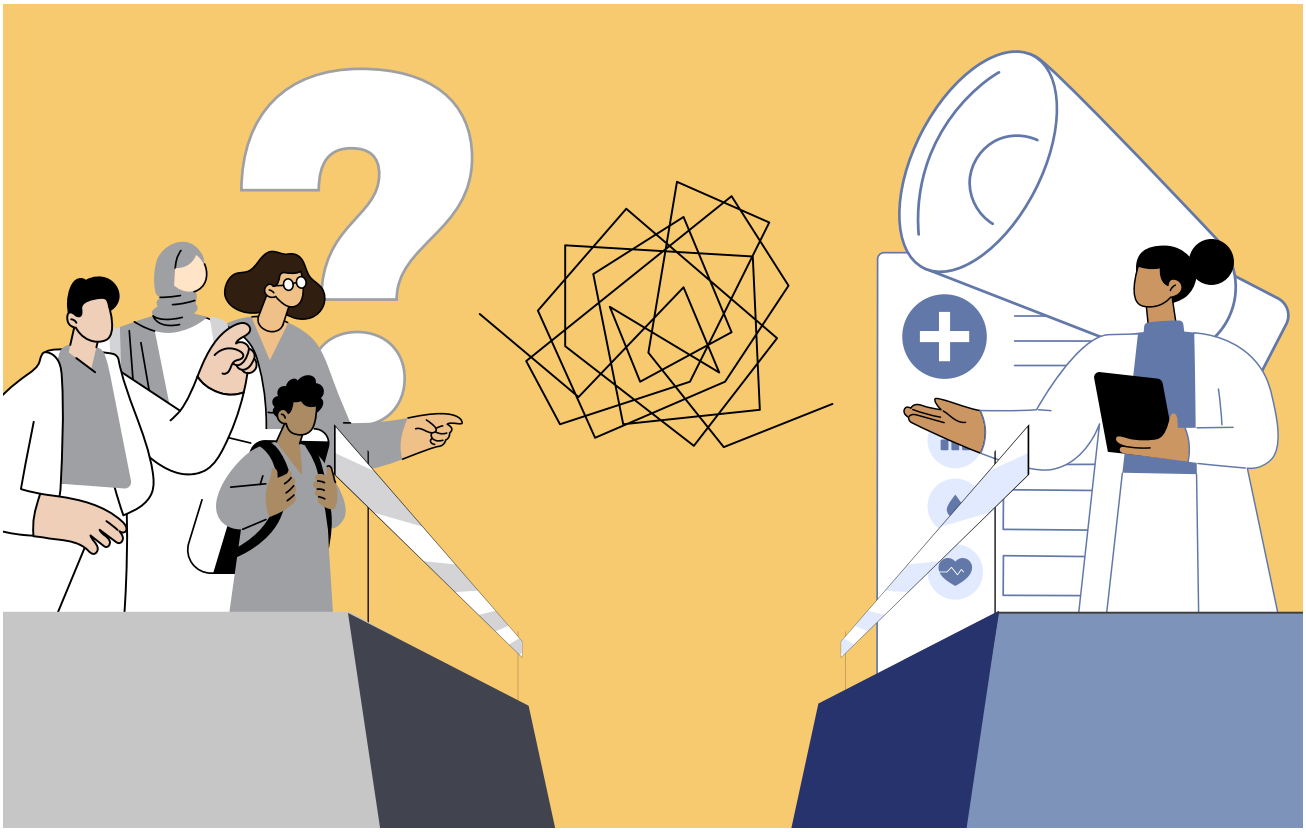
Data bias is mitigated in different ways, including the anonymization of therapy sessions and deployment of an analytics team to detect and archive any aberrant data entries. Several entities, however, suggest that there is no bias or discrimination when dealing with a patient and their data, stating that they "deal with patients as cases." They believe that they cannot be biased while collecting data since medicine is scientific and objective.

Some entities state that they have experienced little to no errors in data entry because they receive their data in the form of multiple-choice answers or email attachments. One health startup offering anonymous therapy anticipates that some patients provide inaccurate data due to cultural reasons that discourage the disclosure of personal information. For example, patients might enter false symptoms or diagnoses in a pre-survey to avoid revealing sensitive details. However, these inaccuracies are typically identified and corrected by the doctor during the subsequent patient consultation.

A radiology clinic states that they sometimes mix up the size and location of a mass on their spreadsheet dataset. They hired personnel to "clean up the data" six years prior to the interview, and they suggest that these errors are unavoidable since "anything done manually will contain errors."

Other pharmaceutical entities highlight the issue of receiving incomplete prescriptions and encountering medication spelling errors. They mitigate these problems by contacting the physician or patient and deploying a spell-check and autocomplete function that unifies the database.

To ensure high-quality data is obtained, the entities use various methods, including the deployment of analytics teams to detect and archive any aberrant, incorrect or false data entries;



manual data checks by doctors; calls to patients by technicians to ensure accurate, up-to-date data and to file missing entries; regular communication with patients; implementation of mandatory data-entry fields; and a spell-check and autocomplete function that unifies the database.

P.1.3 Ensure Data Security

To ensure security and patient privacy during data collection, one organization collects patient names solely for user experience purposes, and all names in their database are encrypted to protect privacy. The other organizations interviewed, however, did not implement practices to safeguard data security during the data collection process

P.2. Promote Health Value

P.2.2 Promote Data Sharing & Interoperability

Some entities do not incorporate data from external sources with their own datasets, while others do receive data from labs and the patients themselves. In some cases, patients bring in previous hard-copy medical reports and blood work, which are then stored at the facility. Some have internal data transfers, such as when a patient decides to see another doctor, and their medical records are therefore consensually shared with the new physician to avoid starting anew.

The organizations that collect data in different formats do not integrate such data. Audio, for example, is not integrated with textual data, as the recordings are not transcribed. The audio and textual data may or may not be mutually exclusive, but the recordings tend to give more clarity to the text data collected by the entity. Those who collect image data also do not integrate it with their textual data and store it on separate databases connected to the same patient ID.

In terms of data retrieval, each entity has their own methods. These include retrieval via their database dashboard, spreadsheets, emails, and their administrative systems. Some entities give their patients the ability to retrieve their data through their own profile on a digital platform or through a portal link. In most cases, the data is retrieved using the patient's unique ID.

P.2.3 Facilitate Innovation using Health Data

The main data-collection-related problem identified by the different organizations is the result of cultural barriers and the lack of awareness in Egypt about the importance of data. A reiterated statement by the different surveyed organizations is, "there is not enough data in Egypt." Medical forms, reports, and prescriptions in

Egypt are often filled incorrectly or contain data points that are ignored. Patients are often hesitant to share their data due to a lack of understanding about the benefits of the process. Educating patients about how sharing data can contribute to improved personalized care, research advancement, and enhanced treatment outcomes may help alleviate their concerns and encourage more openness to data sharing.

Another issue brought up by one of the pharmaceutical labs we interviewed is the lack of communication among different healthcare providers, namely pharmacists and doctors. When there are issues with the prescriptions that pharmacists receive from patients, such as missing data, pharmacists tend to be hesitant to call doctors for clarification. There appears to be hierarchical barriers that make accurate data collection more difficult.

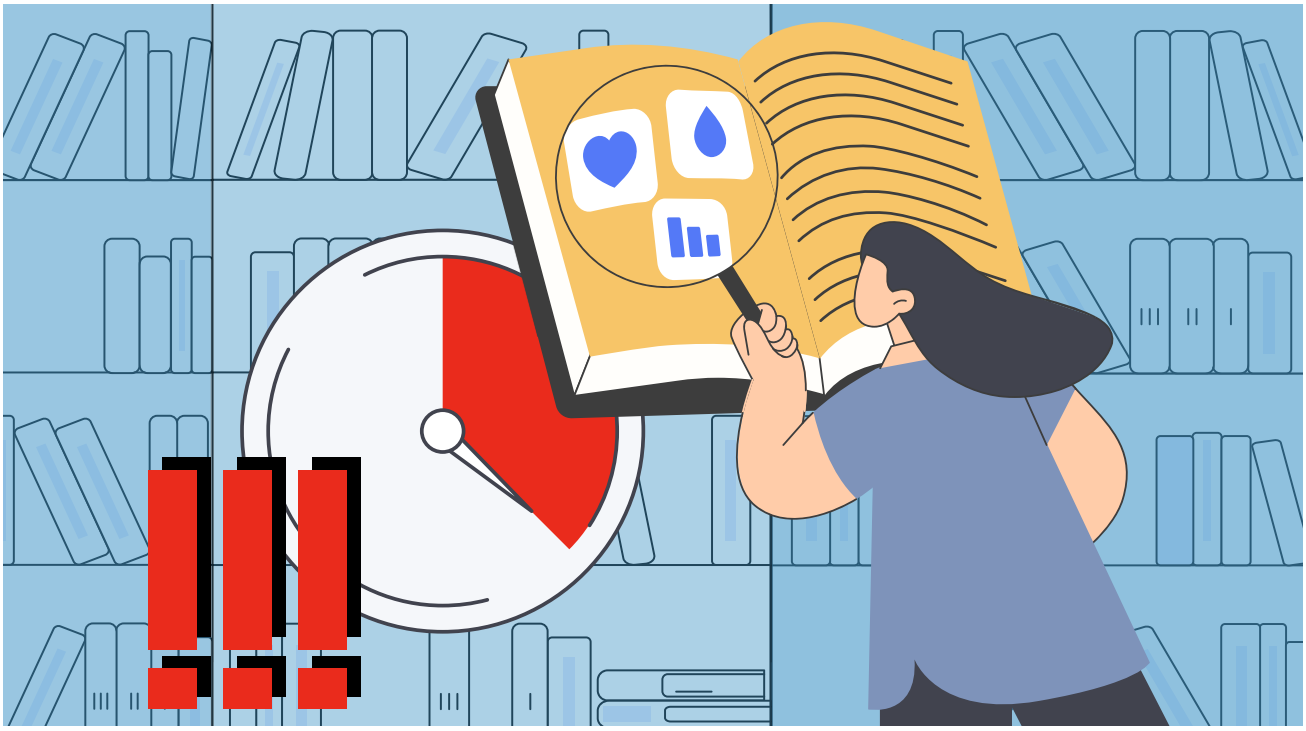
Other issues include the lack of digitized medical forms, reports, and prescriptions, such that large amounts of paper must be used in the process of data collection. This analog data then needs to be manually digitized into the organization's system and database in an extremely time-consuming process. Sometimes the data is written in unclear handwriting, or the paper is in bad condition, making it more difficult to digitize and/or store.

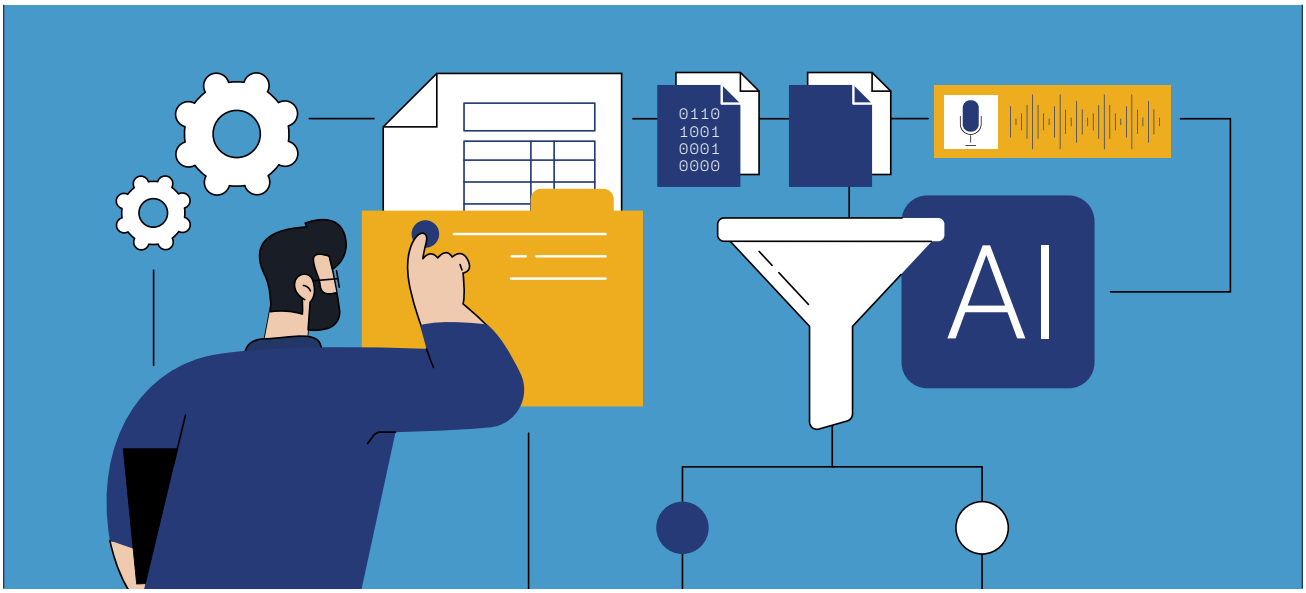
The costliness of certain technologies is also highlighted by several organizations. As much as they would like to digitize their data-collection methods, it would be extremely difficult to do so due to budgetary constraints. One of the interviewed radiology labs states that they have considered replacing data-collection paper forms with tablets connected to their database but have not had enough funds or the technical expertise to do so.

The organizations suggest several ways to overcome these challenges, including improving data collection by encouraging doctors in Egypt to spend more time collecting and recording quality data, encouraging communication among different healthcare providers, improving education and awareness about data collection to remove cultural barriers and stigmas, and digitizing medical forms, reports, and prescriptions.

All organizations unanimously highlight the importance of digital technologies in overcoming these challenges. Suggestions include data collection using digital devices such as tablets; transcription of voice recordings and handwritten documents; implementation of AI tools to detect false data entries; chatbots; AI-powered referral systems; automated transcription and compilation of medical reports; AI tools that can track trends and relations among variables; programs to identify masses in radiology images; and computer-aided diagnosis. There is agreement that digital data collection would be more accurate and efficient, in addition to reducing the need for paper.

Although there is an understanding that technology and AI may help organizations in practice, there is also an awareness that such possibilities require resources. AI applications also require massive datasets, which Egypt is lacking.





5.2. Data Practices in Private Sector: Data Processing

P.1. Protect People

P.1.3 Ensure Data Security

To assess the degree to which the private sector has developed effective and reliable security measures, some of the interview questions focus on data auditing and detection of data breaches, both of which are crucial components of secure data processing. All interviewed entities adopt data auditing practices, with the exception of one startup working on developing healthcare feedback management systems.

There are, of course, variations in the auditing procedures implemented among startups. Some firms' data auditing requires manual processes, whereby technicians call patients to ensure that data is updated regularly. Other auditing process are undertaken by healthcare workers and data specialists, who analyze the data themselves.

As for data breaches, most startups don't have breach detection means. Two entities have elaborated on their data-security systems: one radiology center is equipped with software monitoring tools that send alerts to the IT team upon threat detection; another startup is notified when an account is accessed from multiple tabs, with the system merely registering the devices that are viewing the data.

Furthermore, the interviews touch on data-processing methods, which vary greatly among entities, though most do not have a unified process for all health data and therefore process data according to type. Some process their data into codes, with one firm coding binary responses as 1 or 0, while patient audio submissions are kept in their original format to be reviewed manually by employees. Another organization also processes data according to type; its healthcare provider and doctor data undergo frequent revision and updates to ensure accuracy, with the organization actively engaging with doctors to provide recent updates. On the other hand, patient data is not stored or updated in a standardized manner, as it is processed on a case-by-case basis.

When asked about the updating and maintenance of patient data, most entities state that their data is not updated regularly. Only one startup have their technicians contact patients for data updates, but this non-automated form of maintenance indicates that the data systems of all interviewed entities are not designed to ensure that the collected data is updated regularly.

This matches the findings of another element of focus in the interview: the prevalence, or lack thereof, of data officers and analysts in health entities. This aspect plays a pivotal role in explaining the features and patterns of data processing that were described earlier. Most firms do not have an employee who is responsible for data entry and processing. Instead, they rely on healthcare providers and other employees to process and store the data that they acquire. Only three firms have employees whose primary responsibilities concern health data, with each firm having a different title for said employment: "Senior Data Analyst," "Chief Data Officer," and "Data Scientist." The lack of consistency in health-data job titles, which can imply different responsibilities, is another indicator of the absence of a unified guideline that ensures that startups are adhering to proper data-security measures.

This notion is also evident in the final interview question of the data-security section, in which startups are asked about their compliance with external factors, such as regulations. All of them say that they do not follow any rules for data processing, with some even expressing a lack of awareness of the existence of national data laws.

P.2. Promote Health Value

P.2.3 Facilitate Innovation Using Health Data

Several organizations have their own systems and software dedicated to data processing. One radiology clinic uses a closed-source system linked to an Excel sheet. Another pharmaceutical lab uses compounding software that processes data inputs and produces customized medication. Other organizations either do not process their data or are in the process of building their own data-processing infrastructure. One of the interviewed radiology centers is developing a digital dashboard to visualize their data and insights more easily and efficiently.

While several of the organizations are attempting to implement AI and machine learning technologies, some have already deployed AI solutions. One startup has plans to automate their data collection and processing by compiling their data in one place and providing each patient with their own profile on their platform. Another has started collaborating with AI companies on creating a tool that automatically transcribes Arabic-language audio into text, including all dialects across the region. One organization states that they already use big data and machine learning to provide analytics and insights into patient behavior for other medical companies. Several organizations that have not yet adopted AI technology are aware of its possibilities and suggest that it could help them derive insights from their data while potentially resulting in operational and security improvements. While most organizations are open to the use of AI in certain aspects of the medical field, several are strictly opposed to the use of AI in the diagnostic process, advocating that AI should only be used as a complementary service for doctors.

Some of the data-processing-related problems that the organizations highlight include: the challenge of transcribing audio recordings of spoken Egyptian Arabic; a lack of data to produce insights; incomplete and incorrect data; and a lack of data-processing technologies and innovation, such as tools that can processes handwritten documents. Arabic data presents processing challenges to most of the organizations interviewed, several of which choose to use English, as it is easier to process. Moreover, several organizations note that due to global reporting systems standards, medical reports are usually written in English. One pharmaceutical startup processes its data in Arabic because it is the main language of communication with patients and on its website. Another startup says that they have started collaborating with AI companies on creating a technology that automatically transcribes the Arabic language, including all dialects across the region.

The organizations offer suggestions as to how AI technologies could facilitate the data-processing procedures. Suggestions include converting audio recordings into data points; detecting tone of voice from audio recordings; tracking trends and relations among variables; identifying masses in radiology images and assisting in diagnosis; assisting in Cognitive Behavioral Therapy (CBT); replacing human employees with AI in the data-processing stage, which may result in increased security; and detecting purchasing, behavioral, demographic, medical and geographic trends among customers and patients.

A pharmaceutical startup suggests implementing an AI tool that can recommend prescriptions and the re-ordering of medication, and a dosage reminder tool to alert patients and/or their caregivers to take their prescribed medication. Another startup proposes using AI to market their services more effectively and reach the right audience, though they also express concern about the ethical implications of using the technology in medical marketing. They note that while AI can optimize marketing practices, it must be ensured that it does not exploit patient trust and privacy. One startup states that some AI technologies are not necessarily needed at this stage, as the needed digital infrastructure is not yet adequate enough to leverage them, hence these new technologies do not guarantee better results or more efficiency.

Although most of the organizations are aware of the opportunities that AI technologies could provide, they are conscious of the challenges that are specific to Egypt in this domain. Most of the primary issues and challenges preventing the implementation of AI solutions link back to the absence of adequate data, in addition to inefficient data collection, archiving, and storage. For AI to be implemented effectively, a vast number of large datasets is required, and Egypt has not yet met that threshold. Medical entities struggle to receive data points from patients due to cultural stigmas and lack of education. Furthermore, doctors in public hospitals are often overwhelmed with patients and consequent-

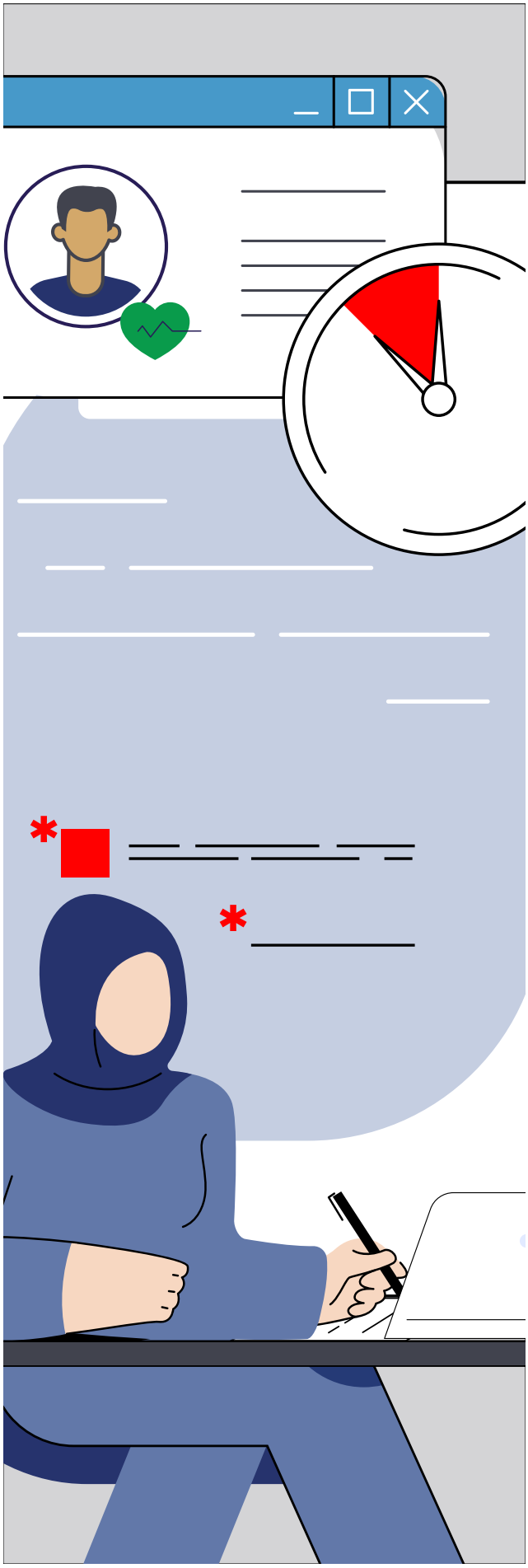
ly document data with little care, if they do so at all. The value of data is not emphasized in Egypt, and little time is put into authoring well-structured, high-quality medical reports.

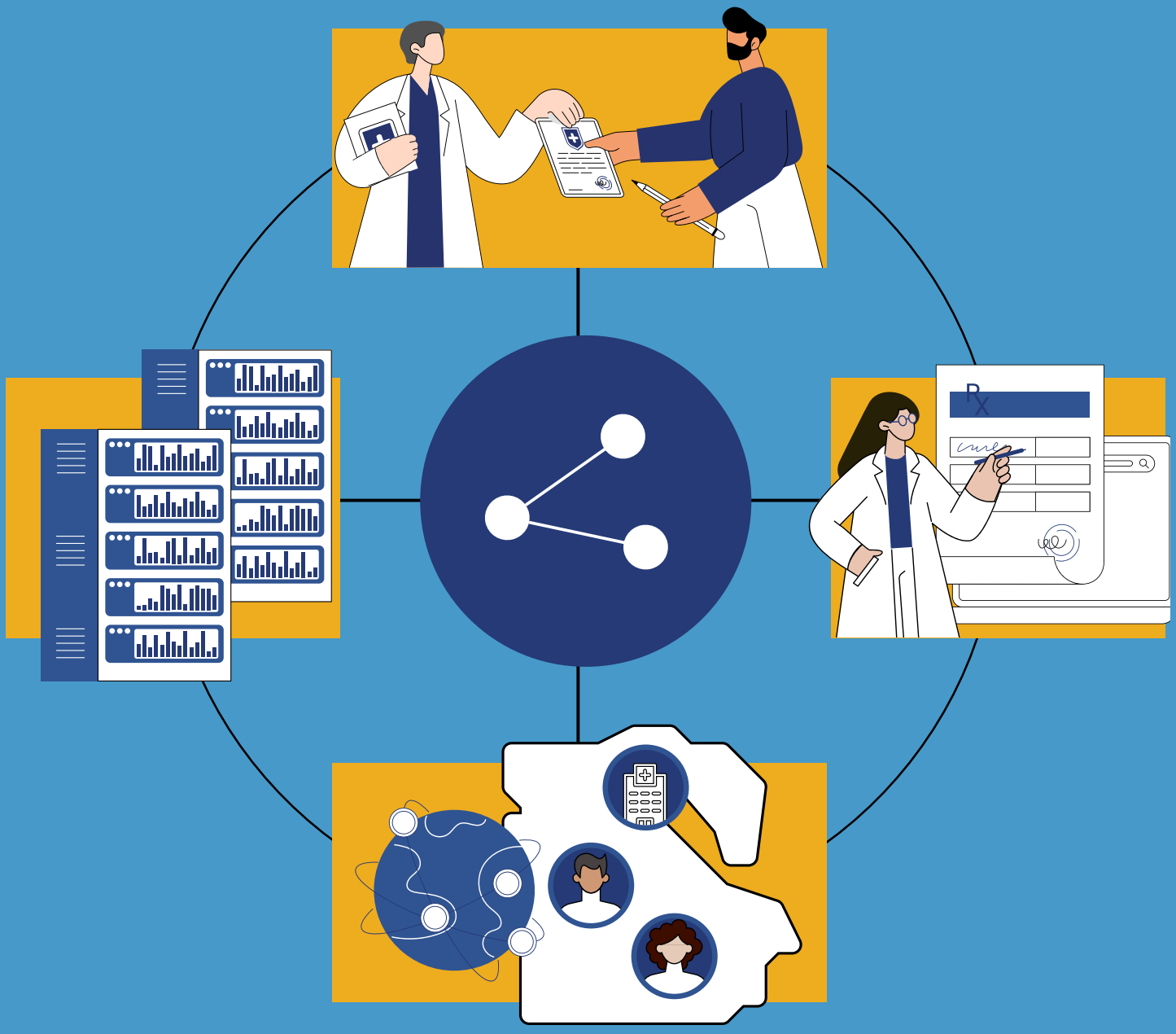
Another primary challenge lies in the vast amounts of paper-bound data that must be digitized if it is to be processed by AI. Data-collection standards are not unified across healthcare providers, making transfer and communication even more difficult. A further challenge is the large financial burden of developing and implementing AI technologies. One pharmaceutical lab expressed concerns about scientists in Big Pharma companies using AI in their labs, fearing that it could lead to the unauthorized leakage of their formulas.

The organizations offer several suggestions toward AI adoption in Egypt: improving and encouraging quality data collection, archiving and storage; creating a unified standard and format across all hospitals, clinics and PHCs; and digitizing the data-collection process. Healthcare professionals and patients should be encouraged to spend more time collecting and recording data accurately and comprehensively.

The cultural stigma in Egypt surrounding patient data needs to be considered when planning for AI adoption. Education on data and its importance to humanity must be provided to patients to enhance their cooperation when asked to provide their health data. Absent this cooperation, the lack of data will continue to pose a barrier against effective AI implementation and outcomes.

Most of the organizations surveyed are not aware of any health-data protection laws in Egypt and cite this lack as a barrier to implementing AI. Those who are aware of Egypt's data-privacy policies state that they are too vague and primarily focused on the protection of privacy and personal data such as name and address but not health data such as blood sugar levels. The absence of an e-pharmacy law is also brought up during an interview with a pharmaceutical start-up. The organizations stress that there needs to be a solid data-specific health framework that considers both the legal and technical aspects of healthcare and health data. This needs to be a priority of the Egyptian government, requiring solid legislation and enforcement.




Rx

5.3 Data Practices in Private Sector: Data Sharing

P1. Protect People

P.1.3 Ensure Data Security

Within each organization, the individuals who have access to raw data include doctors, managers, administrators, data scientists, chief medical officers and the operations teams. Most of the organizations interviewed have multiple levels of data access or limit the number of users with access. In most cases, employees only have access to the data necessary and relevant to their work and department. A marketing team, for example, only has access to data and insights from Google Analytics and other marketing tools. One organization has an internal data-transfer policy which gives employees access to certain data depending on their role. Another organization grants full access to all employees, given that their team is small and consists of six people or less at any given time.

Most of the organizations do not require employees to have permission or written consent to access data. However, one organization requires that authorization be granted by the facility manager. Employees at another organization are obliged to sign a confidentiality agreement stating that they will not share data from any software, reports, medical protocols, and templates externally, via email or texting applications. Some organizations provide software access rights via their IT department after approval from upper management. This provision of access depends on the employee's role and what data/software is applicable to it.

P.2. Promote Health Value

P.2.1 Enhance Health Systems Services

The organizations demonstrate different approaches to data sharing, depending on their needs and practices. One startup that acts as an intermediary third party between doctors and patients only shares a patient's data with the doctor responsible for their treatment. A radiology clinic specializing in mammography only shares a patient's data with another doctor if they require a second opinion and after the patient provides consent. Other organizations state that they share data with external entities only for the purpose of fulfilling orders and payments, thus only sharing data related to the customer's address, phone number, and credit card information. Some organizations share data with third parties for advertising purposes.

One organization only shares data with third parties that are Health Insurance Portability and Accountability Act (HIPAA) compliant to ensure high data privacy. HIPAA is a guideline for telehealth entities that ensures the protection and privacy of patient health data on numerous platforms, including digital medical records, websites, medical devices, and radiology imaging.[15] In order for an entity to qualify for HIPAA compliance, they must meet a checklist of detailed guidelines and requirements.

Some organizations do not share any data at all with external entities and are opposed to doing so, especially with regard to data that is identifiable. One pharmaceutical lab states that when they work with an external delivery company, they conceal the medication and patient's name on the receipt for the sake of privacy. Another pharmaceutical startup that provides analytical services to other companies states that they only share analytics and not raw identifying data.

Some entities receive patient data from external sources to complement their existing databases, including medical reports, blood work reports and radiology scans, either printed or on CDs. This data is usually brought in by patients and is sometimes stored on-site. Among the data-sharing-related issues highlighted by the organizations is the format of such externally received data, which in Egypt, usually takes the form of analog reports. Such data is difficult to store and integrate with the receiving party's own datasets. Similarly, when a new patient arrives at a medical facility, they spend considerable time filling in forms and may sometimes do so incompletely or incorrectly. Digitizing records and allowing for data-sharing between entities would be much more efficient, as doctors would have automatic access to the patient's complete medical history.

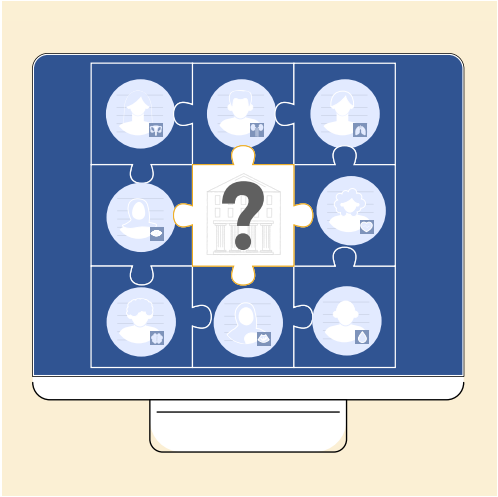
The organizations stress the importance of data sharing in the medical field and see potential in AI-enhanced data sharing in Egypt. To understand Egypt's national health indicators (cases of cancer, diabetes, etc.), data-sharing is key. Healthcare providers in Egypt should be sharing their data among themselves to help attain a more comprehensive grasp of the population's health needs and to make development possible. Egypt does not have a populations genetics studies in relation to public health, and the importance of having an overview of the population's health needs seems to be ignored. One organization notes that a health crisis may emerge in the next 15 to 20 years due to poor nutrition attributable to the current economic crisis, un-

15 Mbonihankuye, S., Nkunzimana, A., & Ndagijimana, A. (2019). *Healthcare data security technology: HIPAA compliance*. Wireless communications and mobile computing, 2019(1), 1927495.

derscoring the potential implications of not keeping track of these needs.

Firstly, a unified and digitized structure for data collection across all health facilities in Egypt is suggested, allowing for easier synchronization among different health entities. This system would act as a digital, national medical record and would compile a patient's entire medical history while making it accessible to and updatable by any healthcare provider in Egypt. Such a data-sharing process would be far more efficient than dealing with damage-prone physical copies of medical reports that are difficult to store and manage.

Secondly, several organizations emphasize the importance of education and the need for patients to understand where their data is going when sharing it with a healthcare provider. Patients are often not medically literate or tech-savvy and therefore need to understand what the implications are when they enter and share their data.



Thirdly, entities that have not shared their collected data with any external parties do show interest in doing so in the future for the benefit of Egypt's healthcare needs. They are open to sharing aggregate data in the form of case studies, analytics on diseases and medication (number of heart transplants, surgeries, units of medication sold per year, etc.), and heatmaps illustrating the distribution of diseases. However, sharing identifiable patient data is out of the question.

Finally, when asked what could be improved in the law to provide better support for data sharing, some were unsure of how drafting new laws could impact the domain. One organization suggests that the issue of data sharing and compilation has more to do with culture than legislation, stating: "Laws do not prevent Egyptians from creating datasets; it's a matter of culture and what we prioritize."[16] Nevertheless, a data-specific health framework that promotes data sharing between entities as well as a national medical database are put forward as recommendations by several organizations.

P.2.2 Promote Data Sharing & Interoperability

Almost all the organizations do not store their data in an interchangeable format and have not yet considered implementing one, mainly due to the lack of data-sharing with other health entities. However, one radiology center uses Reporting and Data Systems (RADS) and Breast Imaging Reporting and Data Systems (BI-RADS), both of which are unified communication and assessment systems that are used globally.

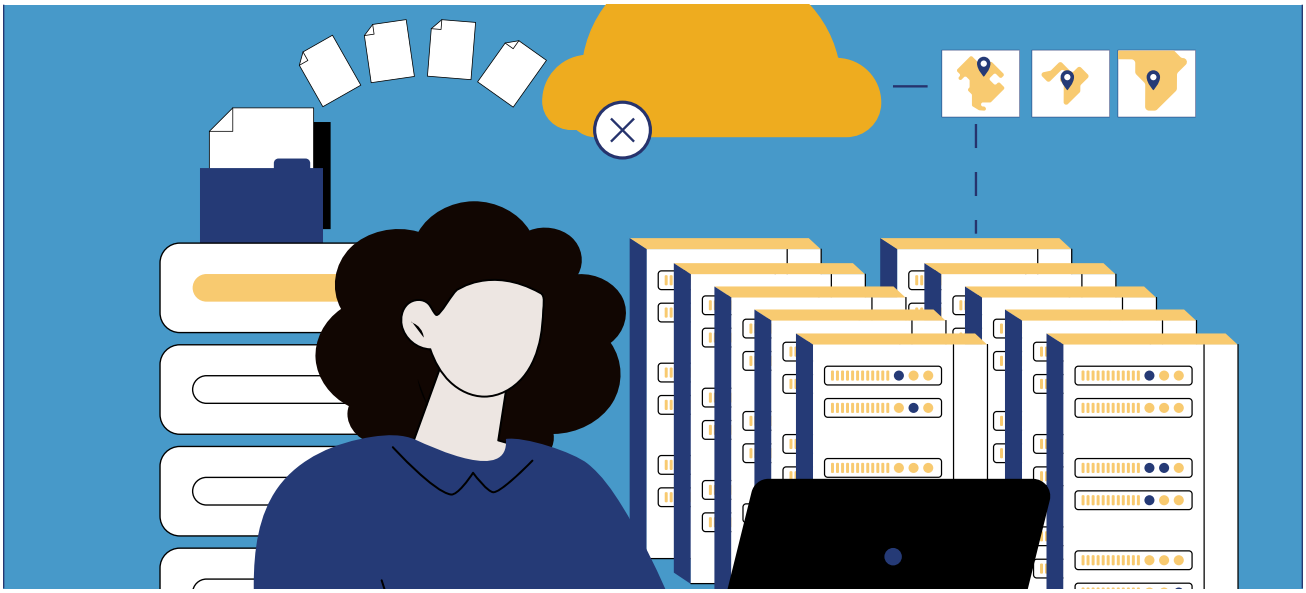
Almost all the organizations interviewed state that they do not have a person or entity responsible for defining the data-sharing rules, except two organizations that have data officers who make sure data is recorded appropriately, determine access privileges to data and software, and define data-sharing rules within their respective companies.

When asked if their data-sharing practices are influenced by external factors such as national regulations, most organizations say they are unaware of any health-data-protection laws. Some organizations cite Egypt's Law on the Protection of Personal Data and Pharmacy Profession Practice Law as external regulations that influence their practices, though neither are health-data specific. One organization that is not aware of any regulations states that they do not depend on laws to secure a patient's data privacy, affirming that "it goes without saying," and "doctors must respect patient privacy."[17]

On the other hand, one startup that operates in several countries outside of Egypt states that they are obliged to follow each country's health-data laws. When the company decided to start operations in South Africa, they learned that the government there requires that the parent company be registered in the country to ensure it is subject to the local health-data laws. Moreover, health data in South Africa must be recorded and stored on local data servers. The authorities there also require that all third parties to any company operating within their borders must be South African or have offices in the country as a further guarantee of compliance with these laws. Another organization states

16 Private sector interview. (2023, July 13).

17 Private Sector Interview. (2023, June 5).



that they are HIPAA compliant and therefore do not share identifying information. They also follow the data-related terms and conditions set by Amazon Web Services (AWS), Apple and Google.

5.4 Data Practices in Private Sector: Data Storage

P1. Protect People

P.1.1 Protect Individuals & Communities

The interviewed organizations use different methods of data storage. One organization uses their local data center, which consists of several servers and is connected to all its branches via underground fiberoptic links. A radiology center uses digital systems such as Radiology Information Systems (RIS) and Picture Archiving and Communication Systems (PACS) to store their data. This entity has archived 10 years' worth of data belonging to approximately 20,000 patients.

Some organizations store their data on the cloud, though several others suggest that cloud storage is not ideal due to a few factors, including Egypt's unreliable connectivity and infrastructure, difficult access and retrieval due to heavy datasets, high costs, and lack of security. Other forms of data storage include the use of local computers, hard drives, third-party software (Google Analytics, Business Suite, etc.), and custom-made servers and software. One organization does not store their data at all and deletes all patient files once they are done providing them with services.

One organization notes that Egypt does not have adequate data centers, requiring them to store their data on servers abroad, in countries such as Saudi Arabia, the USA, and South Africa. This raises questions around Egypt's data infrastructure. The interviewee clarifies that most companies in Egypt store their data in countries with a high volume of data centers, such as Ireland, the USA, and Bahrain, which prompts further inquiry as to how Egypt could enforce data-governance laws when the data isn't stored within its borders.

Stored data is retrieved in several ways, including through patient ID, via a data dashboard, offline management system, old emails, or the patient's profile on the organization's website. A surveyed radiology clinic has launched a portal link for patients, allowing them to access all their reports digitally instead of having to print them.

P.1.2 Build Trust in Data Systems

Almost all the organizations interviewed state that they do not employ a person or entity responsible for defining the rules of data storage, except one that employs a data officer who is responsible for the setup and maintenance of laptops and their 2-factor authentication procedure.

P.1.3 Ensure Data Security

The surveyed organizations use several data-protection and privacy measures to ensure the security of their data, including anonymizing and/or encrypting their data and monitoring access and logins to their systems. Some encrypt identifying information to ensure that the data points are not linked; accordingly, in the event of a leak, the data could not be linked to any specific individual. In some cases, the patient's name is replaced with a code alongside their medical details to ensure anonymity

and privacy. One organization created their own encryption keys through another company to ensure maximum security.

The cloud software used by some organizations have their own standard data security systems, which must be updated regularly. One organization is HIPAA compliant and therefore must follow its data storage guidelines, as well as the rules set by third-party services that they utilize such as Amazon Web Services (AWS), Apple and Google. Another organization chooses to store their data in countries they perceive to have strong data protection laws, such as the USA, which they cite as having a low risk of data misuse. On the other hand, several organizations have no data-protection measures at all, which points to wide gaps among the organizations in terms of data governance.

5.5 Data Practices in Private Sector: Data Security

P1. Protect People

P.1.1 Protect Individuals & Communities

This pillar focuses on using secure data-storage mechanisms such as encryption and cloud servers to protect the data subjects' privacy. In terms of private-sector practices, the majority of the surveyed entities does not consider the data they collect to be sensitive, but rather personal or confidential. Some entities, however, do regard medical data and data that may identify its subject as sensitive. Most of the entities state that personal and identifying data such as a patient's name is anonymized or encrypted to preserve the patient's privacy and prevent its viewing, recognition or linking to any individual. In some instances, even the treating physician does not know the name of their patient and refers to them by a designated code.

All the interviewed entities state that they have safe data-storage mechanisms, as they utilize cloud solutions such as Cloudflare and Google Cloud to mitigate security risks and protect their data from being hacked. One of the entities states that they have a round-the-clock monitoring system that sends automated notifications to the IT department when any security breaches or threats are detected. This monitoring software is regularly updated to ensure the highest dataset security. Another entity notes that they aim to always obtain the necessary licensing to secure their data and ensure that their platform's performance is under constant monitoring. A third entity states that they conduct regular checks of log-in history to observe any unauthorized access to the systems.

Reflecting on the security standards that they follow, one interviewee explains that there are no clear guidelines to follow in their data-storage and retrieval practices in the Egyptian context. On the other hand, he notes that, when working with South African partners, they are required to comply with the country's healthcare and governance laws. For example, South Africans' health data must be recorded and stored on data servers within the country's borders. Moreover, all healthcare entities and any third parties they engage with must have offices based in the country to ensure that they are subject to South African health-data laws. As previously noted, another entity states that they are compliant with the Health Insurance Portability & Accountability Act (HIPAA), which prohibits the sharing

of identifiable information. Additionally, they abide by the data policies set by Apple, Google, and Amazon Web Services (AWS).

P.1.2 Build Trust in Data Systems

This pillar focuses on following the best data-protection and privacy practices to build trust in data systems. Reflecting on the private sector practices, most of the interviewed entities do not have a chief data officer position, and the teams responsible for cloud servers are often also responsible for data security. Personal data such as patient name, gender, diagnosis, and medical history are treated as highly confidential. The entities add that they regularly back up their databases to ensure that their data is secured against destructive cyberattacks. One entity notes that they regularly back up their data onto hard drives, storing it for five years as a standard, though they still possess data that is over 10 years old. Another says that their data is automatically backed up every three hours, while a third reports that they have a biannual backup system. A surveyed radiology center reported working with an imaging company that uses PACS software and its backup system. They also mentioned having backups for their internet connection, prioritizing fiber optic lines for stability. In case of issues with the fiber connection, they resort to using 4G lines as a backup due to frequent internet connection problems in Egypt.

P.1.3 Ensure Data Security

This pillar requires strong technical measures to enhance data security and mitigate related threats. The interviewed private-sector entities indicate that their monitoring systems allow them to detect data breaches. One interviewee explains that all their devices are monitored, and notifications are sent if an account is being accessed on multiple browser tabs at once.

V. TAKEAWAYS AND POLICY-ORIENTED RECOMMENDATIONS

Through the fieldwork interviews at PHCs and with private-sector stakeholders, several gaps are highlighted regarding data governance in healthcare and have been used to guide the development of policy-oriented recommendations.

For PHCs, the health-data collection system is a critical area that must be addressed for proper governance. While PHCs use identical forms to collect data, the process is often scattered, incomplete, unstructured, and almost entirely done on paper, with the exception of the presidential health initiatives. Entities lack digitized data-collection mechanisms, and the mitigation of data bias across all health processes is often ignored. The establishment of data rights and ownership is also lacking, with informed consent rarely obtained from patients and almost no data-sharing rules put in place.

The healthcare system in Egypt also suffers from significant fragmentation, with little to no data sharing and interoperability. Moreover, very few mechanisms have been put in place within PHCs to tackle the issue of data security in line with the best data-protection and privacy practices, with analog data simply being stored in a locked room. Finally, PHC personnel do not possess the required digital capacity or access, nor are they provided with mandatory and well-defined data rules due to a lack of health-sector-specific governing laws and regulations pertaining to data governance.

The private sector echoes similar concerns, with some entities engaging in scattered and unstructured data collection while foregoing informed patient consent. They operate amid a lack of health-sector-specific governing laws and a fragmented health sector, with little to no data sharing and a disregard of data bias. While the private sector shows promising progress in terms of digitized data collection and storage, more needs to be done in terms of securing such datasets.

VI. POLICY-ORIENTED RECOMMENDATIONS

Establish proper digital health data collection mechanisms

Paper-based data collection, a lack of digitized records, unstructured data-collection systems, and the difficulty of collecting data due to cultural barriers and a lack of awareness are factors that have created an ecosystem devoid of proper digital health-data-collection mechanisms. To address these issues, there is a need for a unified, digital data-collection standard across all healthcare providers to facilitate data transfer and synchronization among discrete health entities. Previously collected analogue data, including health records and medical documentation, will also need to be digitized to transition into a digital health system with data that is ready to be utilized by AI. The digitization of forms, reports, and prescriptions will allow for efficient data sharing and the creation of comprehensive, accessible medical history. For robust and accurate

outcomes, innovative data-collection tools must be created, and the pool of available data that is to be fed into AI models must be maximized.

Several countries have begun implementing similar recommendations to modernize their health-data systems. In the United Kingdom, the General Practice Data for Planning and Research (GPDPR) program proposed in 2021 by NHS Digital, exemplifies a unified approach to data collection. This initiative aims to replace the existing, outdated data-collection system by consolidating 300 separate data collections into a single process.[18] The new system seeks to gather primary-care health data from general practitioners (GPs)[19] and sufficient public-health data to monitor and protect public health, produce comprehensive health reports, support research and innovation, and improve services.[20] To ensure the system is effective, NHS Digital collaborates closely with GPs, patients, data experts, and public-sector partners to develop a solution that is efficient, secure, and reduces the administrative burden on GPs, allowing them to focus more on patient care.[21]

Additionally, there is a need to raise awareness around the importance of data, among both patients and healthcare personnel. Most importantly, since digitalization is very expensive, incentivizing the practice of accurate digital data collection within health entities to save their resources and increase their revenues is crucial, especially due to financial constraints.

Implement and standardize consent practices for data collection

Since no written consent is obtained from patients, and only oral consent is sometimes obtained due to a perception that the collected health data may never be utilized, the importance of consent must be highlighted to both patients and healthcare personnel. To promote ethical data practices, awareness must be raised among stakeholders about the imperative of obtaining patient consent.

Improve the quality of health data

Health entities in Egypt collect poor-quality data that contains incomplete fields, entry errors, and inaccuracies while relying on manual updates and patient-provided information, with limited verification and auditing. There is a need for prioritizing quality health-data recording. The value of data should be maximized, with more effort put into writing quality medical reports by prioritizing the digitization of patient data, unifying datasets, promoting complete data entry, and analyzing data for its eventual utilization to improve health services and outcomes. In addition to digitalization's potential to improve the quality of data, AI should also be leveraged in the healthcare system on the policy level and, more importantly, on the practical level. AI presents opportunities for the improvement of data quality through more efficient collection processes, including detection of false entries, transcription of records, trend tracking, and diagnosis assistance through radiology image analysis.

Introduce efficient data-storage mechanisms

Data-storage mechanisms in public health entities are often inefficient, with analog data stored in paper files for several years on end, posing challenges in maintaining data quality and visibility, as well as storage space. Digital data is not stored in an interchangeable format, amid limited data-sharing among healthcare entities. On the other hand, cloud storage is inaccessible, resulting in entities choosing to store their data abroad. To redress this, interoperable standards for digital data storage should be established, ensuring the possibility of exchange and access across different healthcare entities. There is also a need to facilitate the migration of analog data into digital formats while ensuring data integrity and quality. Training and capacity-building initiatives should be implemented to ensure that healthcare professionals and IT staff can effectively manage and utilize digital data storage systems.

Introduce efficient data retrieval mechanisms

Data retrieval processes are time-consuming due to the difficulty of manually locating archived paper files. On the digital side, healthcare providers in PHCs lack access to previously collected data on their devices, with correction and retrieval features centralized through the Ministry of Health. There is a need to streamline and digitize data-retrieval processes and establish decentralized mechanisms for data correction and retrieval at the local level within PHCs. In addition, healthcare providers should be empowered and given the authority to update and retrieve data, reducing dependency on centralized processes.

18 NHS England Digital. (2024, June 6). *About the GPDPR programme*. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research/about-the-gpdpr-programme

19 Ibid

20 NHS England Digital. (2023, December 15). *General practice data for planning and research (GPDPR)*. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/general-practice-data-for-planning-and-research

21 Ibid.

Promote data-sharing practices

Since public and private health entities both operate in silos with separate data collection systems that do not integrate or complement one another, resulting in fragmented efforts and reluctance around data exchange, there is a need to promote data-sharing practices among healthcare providers. This fragmentation of the healthcare system due to the lack of data-sharing and interoperability hinders collaboration and research among institutions. Entities should allow for data sharing between departments, external entities and across medical providers. As a best-practice example, the creation of a national digitized medical record that compiles patients' complete medical histories and is accessible to any healthcare provider could transform the health data-sharing process. While data sharing is important, it is crucial to strike a balance between ease of access and data sovereignty. To create a unified system across all healthcare providers, there is a need to counteract the prevailing perception of health data as a personal project rather than a tool for public-health policy.

Implement an efficient legal framework

Personnel working at PHCs and some private entities lack awareness of data protection laws. Moreover, data privacy policies are vague and tend to focus on personal rather than health data. In the private sector, some entities that operate internationally follow each respective country's health data laws, and there is no unified Egyptian guideline for them to ensure proper data measures. Mainstream dialogue should be initiated between healthcare institutions and legal stakeholders to create laws for governing health data. Health-sector-specific governing laws and regulations should define data-governance rules and promote the innovation of new technologies and AI tools, without restricting innovation, development, and advancement. Such laws should aim to push for the fulfillment of the components of the World Bank's Health Data Governance Principles framework and implement policies that protect data on both an institutional and national level. To ensure the implementation and enforcement of these laws, the formation of a regulatory body is also needed.

Address issues of limited technological infrastructure and resources

Among the challenges presented by inadequate healthcare funding and the limited technological infrastructure are device shortages, connectivity problems, inefficient digital data-entry systems, and system malfunctions. These shortcomings must be mitigated by enabling offline data collection within digital systems, among other measures. There is a need for more efficient digital and data infrastructure that enables the utilization of AI in the healthcare system, which requires more support and funds dedicated to innovation in public health. Moreover, equal access to digital resources and the internet in remote areas should be promoted to ensure that potential AI technologies will not be used to deny patient care based on analysis that is biased toward the currently available datasets.

Advocate for funding digital healthcare technologies and AI implementation

Healthcare financing should be improved, with public funds allocated toward digitizing the Egyptian healthcare system and maintaining proper data-collection, storage, and protection mechanisms, in addition to promoting medical research, innovation, and the development of new digital tools to enhance healthcare-service quality and efficiency. This can be done through providing sufficient digital devices to medical providers, training medical personnel involved in data collection and maintenance, investing in innovation and research, and financing new medical technologies. Due to Egypt's economic challenges, it is imperative to facilitate and incentivize the innovation of home-grown, sustainable, and scalable healthcare technologies.

Provide capacity building for frontline healthcare workers

It is evident from the field work that few frontline healthcare workers are digitally literate and/or trained to use technology. In some cases, doctors, who are likely to be more digitally literate, are the ones entering and filling in data forms, which is a major misuse of human resources. Capacity building and continuous training for healthcare professionals on the concept of health-data governance is needed, in addition to technical training on data processing, protection, security, and utilization to ensure that medical professionals are equipped to utilize emerging technologies. Further, there is a need to raise awareness around the importance of data, as well as more efficient allocation of online digital data-entry tasks.

Several countries have prioritized digital literacy and capacity-building efforts to enhance their healthcare systems. For example, training and capacity building of human resources is listed as one of the seven main priorities of Brazil's Digital Health Strategy 2020–28 to improve data collection mechanisms and enhance data quality and

reliability.[22] Morocco has also taken steps to improve digital literacy among its healthcare workforce. Medical students are taught basic computer skills and biostatistics, and medical informatics courses are offered to physicians interested in this specialty.[23] The Moroccan Society of Medical Informatics and Health (SMIMS) has actively promoted IT education among both clinical and non-clinical specialists to improve their digital capacities.[24] Moreover, Morocco collaborates with developed countries, such as France, to provide further opportunities to train Moroccan informatics specialists.[25]

The following table synthesizes and summarizes the findings and recommendations alongside the corresponding World Bank Health Data Governance Principle.[26]

WORLD BANK PRINCIPLE	KEY FINDING	POLICY RECOMMENDATION
Objective 1: Protect People	Lack of proper health-data collection mechanisms, leading to low quality data and fragmented collection standards while hindering digitization efforts.	Establish proper digital health-data collection mechanisms.
	Lack of standardized consent practices for data collection, partly due to a lack of trust in data systems, resulting in inadequate protection of patients and data subjects.	Implement and standardize consent practices for data collection.
	Inefficient data storage mechanisms, with analogue data stored in filing rooms and in non-interchangeable format due to limited data sharing between healthcare entities. This limits the ability to secure data efficiently while restricting data-sharing, innovation and enhancement of healthcare systems.	Introduce efficient data storage mechanisms.
	Inefficient data retrieval mechanisms due to time-consuming manual processes and lack of access to centralized datasets, resulting in restricted data sharing and inefficient health services.	Introduce efficient data-retrieval mechanisms.

22 OECD (2021). *OECD reviews of health systems: Brazil 2021*. OECD Publishing, Paris. https://doi.org/10.1787/146d0dea-en.

23 Bouhaddou, O., Othmani, M. B., & Diouny, S. (2013). *Medical informatics in morocco*. Yearbook of Medical Informatics, 22(01), 190-196.

24 Ibid.

25 Ibid.

26 Transform Health. (2022). *Health data governance principles: Universalising the benefits of health digitalisation*. World Bank. https://healthdataprinciples.org/images/English/%5BEN%5D%20Health%20Data%20Governance%20Principles.pdf

	Absence of a unified guideline that ensures proper health-data measures and establishes data rights, as well as a lack of awareness of data-protection laws among healthcare personnel.	Implement an efficient legal framework.
Objective 2: Promote Health Value	Poor health-data quality due to incomplete and inaccurate entries, limited data auditing, and manual updates. This limits innovation and delays the enhancement of Egypt's healthcare system and services.	Improve the quality of health data and promote regular auditing.
	Lack of data-sharing practices, with entities operating in silos, resulting in fragmented efforts and an absence of data sharing and interoperability	Promote data-sharing practices
	Limited healthcare funding allocated to providing digital devices, training medical personnel, digitizing and maintaining proper data, as well as promoting medical research and innovation.	Advocate for funding digital healthcare technologies and AI implementation.
	Lack of the needed capacities and skills among frontline healthcare workers, resulting in a misuse of human resources and inefficient healthcare systems.	Promote capacity building for frontline healthcare workers.
Objective 3: Prioritize Equity	Limited technological infrastructure and needed resources, including a shortage of devices, connectivity issues due to unreliable internet infrastructure, and inefficient and malfunctioning digital data-entry systems.	Address issues of limited technological infrastructure and resources.

ANNEX 1: THEORETICAL FRAMEWORK–WHO HEALTH DATA GOVERNANCE PRINCIPLES AND CORRESPONDING INTERVIEW QUESTIONS

HEALTH DATA GOVERNANCE PRINCIPLES BY WORLD BANK 2022			DATA COLLECTION	
Health Data Governance Principles		Aim and Objectives	Public Health Center Questions	Private Sector Questions
P.1. Protect People	P.1.1 Protect Individuals & Communities	Collect data with defined purposes	What types of physiological measurements do you obtain from patients (temperature, blood pressure, sugar levels, etc.)?	What types of data do you collect from your patients/customers/users/etc.?
			What types of health data do you obtain from the patient (medical history, symptoms, etc.)?	
			What types of data do you collect as part of the government's digitized initiatives (e.g., 100 Million Healthy Lives)?	
		Collect personal/sensitive data with informed consent	Is there data that you consider sensitive?	Do you collect, store or process sensitive personal data?
			How do you obtain patients' consent when collecting personal health data (written forms, verbally, etc.)?	How do you obtain the consent of data subjects when collecting their health data? Do you get written consent from data subjects?
		Secured data collection mechanism: Data collection methods with robust data-collection functionality		
	P.1.2 Build Trust in Data Systems	Inclusive data collection mechanisms		

		Ensure consent is informed	How do you obtain patients' consent when collecting personal health data (written forms, verbally, etc.)?	How do you obtain the consent of data subjects when collecting personal health data? Do you get written consent from data subjects?
		High-quality & accurate data collection processes	How do you ensure data is of high quality?	How do you ensure data is of high quality?
			Do you encounter a lot of errors in data entry (field misentry, entry error, etc.)? How do you remedy this?	Do you encounter a lot of errors in data entry (field misentry, entry error, etc.)? How do you remedy this?
	P.1.3 Ensure Data Security	Ensure data security in data collection practices		
P.2. Promote Health Value	P.2.1 Enhance Health Systems Services	Increase community engagement during data collection		
		Empower frontline health workers: Continual skill-building to support their collection of health data		
	P.2.2 Promote Data Sharing & Interoperability	Use previously collected data to reduce the need for new data collection	For repeat visits, do you retrieve the patient's data ask them the same questions on every occasion?	Do you incorporate patient data collected by external sources? If so, what types of data?
			How is the data retrieved? Can data be retrieved from previous records?	How is the data retrieved? Can data be retrieved from previous records?
		Data-collection systems to be designed with interoperability in mind		

		Define common data structures (e.g., specific fields for data collection)	How do you integrate digital datasets with offline/hard copy? Do they contain the same data? Are they mutually exclusive or overlapping?	
	P.2.3 Facilitate Innovation using Health Data	New types of policies are needed for innovative collection of health data	What data-related problems have you managed to identify at your organization?	What data-related problems have you managed to identify at your organization?
			What do you suggest is the best way to overcome these challenges?	What do you suggest is the best way to overcome these challenges?
			Do you think using digital technologies can help overcome these challenges?	Do you think using digital technologies can help overcome these challenges?
P.3. Prioritise Equity	P.3.1 Promote Equitable Benefits from Health Data	Inclusive data-collection methodologies that consider which individuals/groups are asked to provide data, which data categories are collected, and what the use of the collected data is		
		Cross-cutting data collection along categories like gender, sex, age, socio-economic status, abilities, citizenship status, class, race, and ethnicity	What types of data do you collect from the patient?	
		Mitigate data bias in health-data collection		

		Individuals involved in health-data collection, such as frontline health workers, should understand the purpose of data use		
	P.3.2 Establish Data Rights & Ownerships	Develop health data trusts and health data cooperatives to define rules of data collection		
HEALTH DATA GOVERNANCE PRINCIPLES BY WORLD BANK 2022			**DATA PROCESSING**	
Health Data Governance Principles		Aim and Objectives	Public Health Center Questions	Private Sector Questions
P.1. Protect People	P.1.3 Ensure Data Security	Strong technical security measures for data processing (e.g., data security audits)	Is the data audited? How, by whom, and how frequently?	How do you detect data breaches?
				Is the data audited? How, by whom, and how frequently?
		Consider federated data processing across the health system		
P.2. Promote Health Value	P.2.3 Facilitate Innovation using Health Data	New types of policies are needed for innovative processing of health data	What technologies do you use to manage the PHC's data?	What technologies do you use to manage the organization's data?
			How can these technologies help you? In what areas can the use of artificial intelligence (AI) and machine learning (ML) technologies help? (interviewer to explain with an example)	How can these technologies help you? In what areas can the use of artificial intelligence (AI) and machine learning (ML) technologies specifically help? (interviewer to explain with an example)

			What is the biggest obstacle facing the use of AI, based on the status of the available datasets?	What is the biggest obstacle facing the use of AI, based on the status of the available datasets?
			Do you anticipate using digital technologies like AI in the future?	Do you anticipate using digital technologies like AI in the future?
P.3. Prioritize Equity	P.3.1 Promote Equitable Benefits from Health Data	Mitigate data bias in health data processing		Is the data anonymized and/or encrypted? Can patients' details be viewed/identified?
	P.3.2 Establish Data Rights & Ownerships	Develop health-data trusts and health-data cooperatives to define rules of data processing		
HEALTH DATA GOVERNANCE PRINCIPLES BY WORLD BANK 2022			**DATA SHARING**	
Health Data Governance Principles		Aim and Objectives	Public Health Center Questions	Private Sector Questions
P.1. Protect People	P.1.2 Build Trust in Data Systems	Data subjects may accept or decline further sharing of their data for purposes other than the initial intended use		
	P.1.3 Ensure Data Security	Consider federated data systems across the health sector for cross-system sharing and learning	Is the data shared with any external individuals and/or entities? Which?	Is the data shared with any external individuals and/or entities? Which?
			Do you incorporate patient data collected by external sources? If so, what types of data?	Do you incorporate patient data collected by external sources? If so, what types of data?
P.2. Promote Health Value	P.2.1 Enhance Health Systems Services	Enhance data sharing between health facilities and health providers for improved services	Do you obtain patient health data from other entities (labs, radiology, data from former providers, etc.)?	

			Do you incorporate patient data collected by external sources? If so, what types of data?	Do you incorporate patient data collected by external sources? If so, what types of data?
			Is the data shared with any external individuals and/or entities? Which?	Is the data shared with any external individuals and/or entities? Which?
			Is the patient data transferred to or shared with the government hospitals to which they are referred? Do the hospitals have access to the digital system?	
	P.2.2 Promote Data Sharing & Interoperability	Establish data-sharing rules and guidelines		As the collected health data may include sensitive information, do you implement data-sharing principles to avoid discrimination against individuals and certain groups?
		Validate informed consent before sharing data		
		Promote interoperability of data systems for simpler and more secure data-sharing (especially during manual data transfers) by implementing standards for basic data fields, open system design, etc.		
		Define common data structures (underlying architecture of data systems) across healthcare systems to support data sharing		

		Define multiple levels of data access to minimize the risk of exposure while maximizing sharing	Who has access to the raw data on the individual and entity levels?	
		Use common definitions and global standards for greater standardization and comparability of health data and greater systems interoperability		
		Support multi-sector partnerships		
P.3. Prioritise Equity	P.3.2 Establish Data Rights & Ownerships	Develop health-data trusts and health-data cooperatives to define rules of data sharing		
HEALTH DATA GOVERNANCE PRINCIPLES BY WORLD BANK 2022			**DATA STORAGE**	
Health Data Governance Principles		Aim and Objectives	Public Health Center Questions	Private Sector Questions
P.1. Protect People	P.1.1 Protect Individuals & Communities	Use secure data-storage mechanisms (encryption, cloud servers, etc.)	How do you store your data? Describe your data-storage architecture.	How do you store your data? Describe your data-storage architecture.
			Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?
	P.1.2 Build Trust in Data Systems	Align with best data-protection and privacy practices for storage (two- factor authentication, encryption, de-identification, etc.)	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?

	P.1.3 Ensure Data Security	Mitigate risks related to security threats through safe storage guidelines for confidential data		
		Consider federated data systems (federated storage) across the health system		
P.2. Promote Health Value	P.2.2 Promote Data Sharing & Interoperability	Support multi-sector partnerships		Is the data stored in any interchangeable or standard format?
P.3. Prioritise Equity	P.3.2 Establish Data Rights & Ownerships	Develop health-data trusts and health-data cooperatives to define rules of data storage		

HEALTH DATA GOVERNANCE PRINCIPLES BY WORLD BANK 2022			**DATA SECURITY**	
Health Data Governance Principles		Aim and Objectives	Public Health Center Questions	Private Sector Questions
P.1. Protect People	P.1.1 Protect Individuals & Communities	Use secure data-storage mechanisms (encryption, cloud servers, etc.)	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?
			What security standards are followed in data storage and retrieval for digital and analog data?	What security standards are followed in data storage and retrieval for digital and analog data?
	P.1.2 Build Trust in Data Systems	Align with best data-protection and privacy practices for storage of data (two- factor authentication, encryption, de-identification, etc.)	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?	Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?
	P.1.3 Ensure Data Security	Require strong technical security measures for data processing (data security audits)	Is the data audited? How, by whom, and how frequently?	Is the data audited? How, by whom, and how frequently?

		Mitigate risks related to security threats		How do you detect data breaches?
		Enhance data security through federated data systems		
		Apply a human rights lens to health-data governance: traditional rights such as security and health, and new, data-related rights such as privacy		

ANNEX 2: INTERVIEW QUESTIONS FOR PUBLIC HEALTHCARE CENTERS (PHCS)[27]

Please give a general overview of the patient's experience at the PHC:

يرجى شرح لمحة عامة عن رحلة المريض في مركز الرعاية الصحية الأولية

Group 1 – Nurses: Questions on Data Acquisition

1) What types of data do you collect from the patient?
ما هي أنواع البيانات التي تجمعها من المريض ؟

2) How do you collect patient data (on paper, etc.)?
كيف تجمع بيانات المرضى (على الورق، وما إلى ذلك) ؟

3) What types of data do you use a tablet/computer to collect?
ما هي أنواع البيانات التي تستخدم الجهاز اللوحي أو الكمبيوتر لجمعها ؟ ما الفرق بين البيانات اللى يتم جمعها على الجهاز اللوحي أو الكمبيوتر والبيانات اللتى لا يتم جمعها على أى منهما ؟

4) Are there any other types of data that are captured on paper/offline?
هل هناك أي أنواع أخرى من البيانات التي يتم التقاطها على الورق/ دون اتصال بالإنترنت؟
فى حالة جمع البيانات على الجهاز اللوحي/ الكمبيوتر، هل يكون الجهاز متصلاً بالإنترنت؟

5) In the case of offline data collection on a tablet/computer, when does the data synchronization between your offline database and the online server occur?
في حالة جمع البيانات غير المتصلة بالإنترنت على جهاز لوحي/كمبيوتر، متى يحدث تزامن البيانات بين قاعدة بياناتك في الوضع الإلكتروني وغير المتصل بالإنترنت ؟

6) Can the offline data be edited before synchronization?
هل يمكن لشخص ما تعديل البيانات غير المتصلة بالإنترنت قبل التزامن عبر الإنترنت ؟

7) What types of data do you collect as part of the government's digitized initiatives (e.g., 100 Million Healthy Lives)?
ما هى أنواع البيانات التى يتم جمعها فى المبادرات الحكومية المميكنة (مثل مبادرة ا.. مليون صحة)؟

8) If a woman has other health problems that are not covered by the government's digitized initiatives, do you still register her data onto the digital system?
في الحالة التي تعاني فيها المرأة من مشاكل صحية أخرى لا تغطيها المبادرات الحكومية المميكنة، هل تلتقط بياناتها أيضًا في النظام الرقمي ؟

9) How do you integrate the digital and offline/hard copy datasets? Do they contain the same data? Are they mutually exclusive or overlapping?

27 Both sets of questions on data acquisition and data management/workflow should be repeated to the PHC managers for cross-checking after interviewing nurses and administrators.

10) What types of physiological measurements do you obtain from patients (temperature, blood pressure readings, sugar levels, etc.)?

ما هي أنواع قياسات الجسم التي تتناولها للمرضى (على سبيل المثال درجة الحرارة وقراءات ضغط الدم ومستويات السكر) ؟

11) What types of health data do you obtain from the patient (e.g. medical history, symptoms, etc.)? Do you have a method for verifying this data?

ما هي أنواع البيانات الصحية التي تعرفها من المريض (على سبيل المثال التاريخ الطبي والأعراض وما إلى ذلك)؟

هل لديك طريقة للتحقق من صحة هذه البيانات أم أنك تعتمد على كلام المريض (على سبيل المثال هل تطلب من المريض أن يقوم بتأكيد بياناته عن طريق إحضار ورقة عليها آخر قياس ضغط أو شهادة من معالج سابق أم أنك تقوم بالقياسات بنفسك للتأكد)؟

12) Do you obtain health data from other entities (labs, radiology, former provider, etc.)?

هل تحصل على بيانات صحية أخرى من كيانات صحية أخرى (على سبيل المثال نتائج المعامل و الأشعة، وبيانات من المعالج السابق ، إلخ) ؟

Or

Do you incorporate data collected by external sources? If so, what types of data?

هل تكمل بيانات المرضى الموجودة من مصادر أخرى ؟ إذا كان الأمر كذلك، أيهما ؟

13) For repeat visits[28], do you retrieve the patient's data or do you ask them the same questions on every occasion?

في الزيارات المتكررة، هل تسترجع بيانات المريض أم تسأل المريض نفس الأسئلة مرة أخرى؟

14) How do you obtain patients' consent when collecting personal health data (written forms, verbally, etc.)?

كيف تأخذ موافقة المرضى عند جمع البيانات الصحية الشخصية (على سبيل المثال في نماذج مكتوبة، شفهية، إلخ) ؟

15) What data-related problems have you managed to identify at your organization?[29]

ما هي مشاكل مؤسستك المتعلقة بالبيانات التي تمكنت من تحديدها ؟

16) What do you suggest is the best way to overcome these challenges?

ما هي أفضل طريقة للتغلب على هذه التحديات ؟

17) Do you think using digital technologies can help overcome these challenges?

هل تعتقد أن استخدام التقنيات الرقمية يمكن أن يساعد في التغلب على هذه التحديات ؟

28 In case of repeated visits, the former therapist will also be from the same PHC.

29 Thomas, Gwen. (2006). *The DGI data governance. DGI data governance Framework*. Orlando, Florida, The Data Governance Institute. datagovernance.com/wp-content/uploads/2020/07/dgi_data_governance_framework.pdf.

Group 2 – Administrators: Questions on Data Management

Please explain the different types of administrator roles in the PHC:[30]

يرجى شرح مختلف أنواع أدوار المديرين في مركز الرعاية الصحية الأولية

A) Data-Related Questions:

1) What is the workflow in relation to the collected patient data? Explain the sequence of tasks performed by the different people within and across your organization.

ما هو سير عمل بيانات المريض التي تم جمعها ؟ اشرح تسلسل المهام التي يؤديها الأشخاص المختلفون داخل مؤسستك وعبرها

2) Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?

هل البيانات مجهولة الهوية و/أو مشفرة للحفاظ على خصوصية المرضى؟ هل يمكن الإطلاع على تفاصيل المرضى والتعرف عليها ؟

3) Is there data that you consider sensitive?

هل هناك نوع من البيانات تعتبره بيانات حساسة؟

4) How do you store your data?[31] Please describe your data storage architecture.

كيف تخزن بيانتك ؟ يرجى وصف بنية تخزين البيانات الخاصة بك.

5) Is the data stored in any interchangeable or standard format?

هل البيانات مخزنة في أي شكل قابل للتبديل أو في شكل موحد ؟ أي شكل ؟

6) Who has access to raw data on the individual and entity levels?

من لديه إمكانية الوصول إلى البيانات الأولية، من حيث الأفراد والكيانات ؟

7) Is the data shared with any external individuals and/or entities? Which?

هل تتم مشاركة البيانات مع أي أفراد و/أو كيانات خارجية ؟ ما هو ؟

8) Is the patient data transferred to or shared with the government hospitals to which they are referred? Do the hospitals have access to the digital system?

هل يتم مشاركة بيانات المريض مع المستشفيات الحكومية التي تمت إحالته إليها أو مشاركتها معها ؟ هل لديهم إمكانية الوصول إلى النظام الرقمي ؟

9) How is the data retrieved? Can the data be retrieved from previous records?

كيف يتم استرداد البيانات ؟ أو هل يمكن استرجاع البيانات من السجلات السابقة ؟

30 To assess whose role is most relevant to data.

31 Andanda, P. (2020). *Ethical and legal governance of health-related research that uses digital data from user-generated online health content. Information.* Communication & Society, 23(8), 1154-1169.

10) How do you ensure the data is up to date? Is there a way to update patients' records?

كيف تتأكد من تحديث البيانات ؟ هل هناك طريقة تقوموا من خلالها بتحديث بيانات المرضي؟

11) Is the data audited? How, by whom, and how frequently?

هل يتم مراجعة البيانات ؟ كيف، من قبل من، وكم مرة ؟

12) How do you ensure that data is of high quality?

كيف تضمن جودة بيانات عالية من خلال تدقيق البيانات (على سبيل المثال التأكد من أن قياس الوزن مسجَّل فى الخانة المخصصة له و ليس فى قياس الطول و العكس، أو التأكد من البيانات تخص نفس المريض)؟

13) Do you encounter a lot of errors in data entry (field misentry, entry error, etc.)? How do you remedy that?

هل تصادف الكثير من الأخطاء في إدخال البيانات (على سبيل المثال إدخال البيانات إلى مجالات خاطئة، أو إدخال أرقام خاطئة) ؟ كيف تعالج ذلك ؟

14) What security standards are followed in data storage and retrieval for digital and analog data?

ما هي معايير الأمن المتبعة في تخزين البيانات واسترجاعها (على سبيل المثال، في حالة البيانات الورقية هل غرفة الملفات تغلق بمفتاح؟ من الشخص المسئول عنها؟ ما المدة الزمنية المخصصة للإحتفاظ بالبيانات)؟

15) What data-related problems have you managed to identify at your organization?[32]

ما هي مشاكل مؤسستك المتعلقة بالبيانات التي تمكنت من تحديدها (على سبيل المثال تسريبات للبيانات أو سوء استخدام للبيانات)؟

16) What do you suggest is the best way to overcome these challenges?

ما هي أفضل طريقة للتغلب على هذه التحديات ؟

B) AI-Related Questions: (subject to expansion)

17) What technologies do you use to manage the PHC's data?

ما هي التقنيات التي تستخدمها لإدارة بيانات مركز مركز الرعاية الصحية الأولية ؟

18) How can these technologies help you?[33] In what areas can the use of artificial intelligence (AI) and machine learning (ML) technologies help? (interviewer to explain with an example)

كيف يمكن لهذه التقنيات أن تساعدك ؟ في أي مجالات يمكن أن يساعد استخدام تقنيات الذكاء الاصطناعي والتعلم الآلي على وجه التحديد ؟

32 Thomas, Gwen. (2006). *The DGI data governance. DGI data governance framework*. Orlando, Florida. The Data Governance Institute. datagovernance.com/wp-content/uploads/2020/07/dgi_data_governance_framework.pdf.

33 The interviewer needs to explain what is AI & ML to the public health unit interviewees.

19) What is the biggest obstacle facing the use of AI, based on the status of the available datasets? (interviewer to elaborate on q.18)

ما هي التحديات/التحدي الأكبر الذي يواجه استخدام الذكاء الاصطناعي في الصحة بناءً على حالة مجموعات البيانات المتاحة ؟

20) Do you anticipate using digital technologies like AI in the future?

هل تتطلعون لإستخدام التقنيات الرقمية مثل الذكاء الاصطناعي في المستقبل؟

ANNEX 2: PRIVATE SECTOR INTERVIEW QUESTIONS

Data Collection:

1) What types of data do you collect from the patient/customer?

ما هي أنواع البيانات التي تجمعها ؟

» ***Follow-up question(s):***

a) Which data is obtained from the patient directly (e.g. medical history, symptoms, etc.)? Do you have a way to veriify this data?

ما هي أنواع البيانات الصحية التي تعرفها من المريض (على سبيل المثال التاريخ الطبي والأعراض وما إلى ذلك)؟

b) Is the data collected with a defined purpose? Is the customer/patient notified of this purpose?

هل تجمع تلك بيانات لأغراض محددة؟ ما هو استخدام البيانات التي تم جمعها؟

هل يدرك المرضى/العملاء الاستخدام النهائي لبياناتهم أثناء جمعها ؟

2) How is patient/customer data collected?

كيف يتم جمع بيانات المريض/العملاء ؟

a) In what language is the data collected (Arabic and/or English)?

ما هي اللغة التي جمعت بها البيانات (مثل العربية و/أو الإنجليزية) ؟

b) Do you use digital technologies to collect data? If Yes, what are these digital technologies?

c) How can these technologies help you (e.g., for innovative collection of health data)?

3) How do you integrate two datasets: e.g., audio versus text datasets and/or digital versus offline/hard copy?

كيف يمكنك دمج مجموعتي البيانات: مثل مجموعات البيانات الصوتية مقابل النصوص و/أو الرقمية مقابل النسخ الورقية/غير المتصلة بالإنترنت ؟

a) Do both sets contain the same data? Are they mutually exclusive or overlapping?

هل قاعدتي البيانات لديهم نفس البيانات؟ هل قاعدتى البيانات حصرية بشكل متبادل أم متداخلة ؟

4) How do you obtain the consent of data subjects when collecting personal health data? (in written forms, verbally, etc.)?

كيف تأخذ موافقة الأشخاص المعنيين بالبيانات عند جمع البيانات الصحية الشخصية (على سبيل المثال في نماذج مكتوبة، شفهية، إلخ) ؟

5) How do you ensure that data and its collection processes are of high quality?

كيف تضمن جمع بيانات عالية الجودة من خلال تدقيق البيانات (على سبيل المثال التأكد من أن قياس الوزن مسجَّل فى الخانة المخصصة له و ليس فى قياس الطول و العكس، أو التأكد من البيانات تخص نفس المريض)؟

6) Do you use previously collected data to reduce the need for new data collection?

هل تستخدم البيانات التي تم جمعها مسبقًا لتقليل الحاجة إلى جمع بيانات جديدة؟ (على سبيل المثال في الزيارات المتكررة، هل تسترجع بيانات المريض أم تسأل المريض نفس الأسئلة مرة أخرى؟)

» ***Follow-up questions:***

a) What data-collection-related problems have you managed to identify at your organization?

ما هي مشاكل مؤسستك المتعلقة بجمع البيانات التي تمكنت من تحديدها ؟

b) Do you encounter a lot of errors in data entry (field misentry, entry errors)? How do you remedy that?

هل تصادف الكثير من الأخطاء في إدخال البيانات (على سبيل المثال إدخال البيانات إلى مجالات خاطئة، أو إدخال أرقام خاطئة) ؟ كيف تعالج ذلك ؟

c) How do you mitigate data bias in your health-data collection?

كيف تخفف من تحيز البيانات في جمع البيانات الصحية ؟

7) How do you suggest overcoming these challenges? What role can technologies like AI play in this regard?

Data Processing:

1) Can you describe your data-processing procedure (do you process audio and text datasets differently)?

2) Is the data audited? How, by whom, and how frequently?

هل يتم مراجعة البيانات ؟ كيف، من قبل من، وكم مرة ؟

» ***Follow-up questions:***

a) Do you have a chief data officer or similar role?

هل لديك كبير مسؤولي البيانات أو دور مماثل ؟

b) How are the records/data maintained and updated? Is there a way to update patients' records?

كيف تتأكد من تحديث البيانات ؟ هل هناك طريقة تقوموا من خلالها بتحديث بيانات المرضي؟

3) What data-processing-related problems have you managed to identify at your organization?

ما هي مشاكل مؤسستك المتعلقة بمعالجة البيانات التي تمكنت من تحديدها ؟

» ***Follow-up questions:***

a) How do you mitigate data bias in your health-data processing?

كيف تخفف من تحيز البيانات في جمع البيانات الصحية ؟

b) Do you think there is a risk that anonymized health-related and personal data can become identifiable during data processing? Which groups may be impacted (e.g., women and girls)?[34]

هل تعتقد أن هناك خطرًا يتمثل في إمكانية تحويل البيانات الشخصية الصحية المجهولة إلى بيانات شخصية يمكن التعرف عليها مرة أخرى أثناء معالجة البيانات ؟ ما هي المجموعات التي قد تتأثر (مثل النساء والفتيات) ؟

c) Does Arabic data present processing challenges?

هل تشكل البيانات العربية تحديات أمام معالجتها ؟

d) Does audio/voice data present processing challenges (Arabic versus English)?

هل تمثل البيانات السمعية/الصوتية تحديات أمام معالجتها (التسجيلات العربية مقابل الانكليزية) ؟

New Technologies, Data, and the Regulatory Environment

4) Do you use digital technologies to process data? If yes, what are these digital technologies?

هل تستخدم مؤسستكم التقنيات الرقمية لمعالجة البيانات ؟ ما هي هذ التقنيات الرقمية ؟

5) How can these technologies help you?[35]

كيف يمكن لهذه التقنيات أن تساعدك ؟

6) Are you working to adopt artificial intelligence (AI) and machine learning (ML) technologies? How can these technologies help you (e.g., innovative data processing)?

هل تحاول تبني تقنيات الذكاء الاصطناعي (AI) والتعلم الآلي (ML) ؟ كيف يمكن لهذه التقنيات أن تساعدك على وجه التحديد (مثل المعالجة المبتكرة للبيانات) ؟

34 Data rights check tool, https://digitalrights-check.bmz-digital.global/

35 The interviewer needs to explain what is AI & ML to the public-health unit interviewees.

7) What is the biggest obstacle facing the use of AI in health, based on the status of the available datasets?

ما هي التحديات/التحدي الأكبر الذي يواجه استخدام الذكاء الاصطناعي في الصحة بناءً على حالة مجموعات البيانات المتاحة ؟

8) In your opinion, what is needed in Egypt's health-data ecosystem for the adoption of AI?

في رأيك، ما المطلوب في نظام البيانات الصحية لتبني الذكاء الاصطناعي في مصر ؟

9) Are your data management/processing practices influenced by any external factors (e.g., regulations)?

هل تتأثر ممارسات إدارة/معالجة البيانات الخاصة بك بأي عوامل خارجية (مثل قوانين/سياسات البيانات) ؟

» ***Follow-up questions:***

a) What gaps do you see in the law/data-governance policies that may hinder your data-processing practices?

Or

What can potentially be improved in the law to provide better health-data processing/ support your data-processing practices?

10) What do you suggest is the best way to overcome these challenges?

ما هي أفضل طريقة للتغلب على هذه التحديات ؟

11) Do you think using digital technologies like AI may help you overcome these challenges?

هل تعتقد أن استخدام التقنيات الرقمية مثل الذكاء الاصطناعي يمكن أن يساعد في التغلب على هذه التحديات ؟

Data Sharing:

1) Is there a person/entity in your organization responsible for defining the rules of data sharing?

هل هناك شخص/ جهة مسئول(ة) في مؤسستك لتحديد قواعد مشاركة البيانات الصحية؟

2) Who has access to raw data?

من لديه إمكانية الوصول إلى البيانات الأولية، من حيث الأفراد والكيانات ؟

3) Do they need permission or written consent?

هل يحتاج لرخصة أو موافقة مكتوبة أو بدون رخصة؟

» ***Follow-up questions:***

a) What types of data require written consent/authorization for access?

ما هي البيانات التي يتطلب الوصول اليها موافقة مكتوبة؟

b) Do you have multiple levels of data access to minimize the risk of exposure? Explain.

هل لديك مستويات متعددة من الوصول إلى البيانات لتقليل مخاطر التعرض ؟ اشرح

4) Is the data shared with any external entity/individual? Which/Who? Do you receive any health data from other entities (labs, radiology, former provider, etc.)?

هل تتم مشاركة البيانات مع أي أفراد و/أو كيانات خارجية ؟ ما هو ؟ هل تحصل على بيانات صحية أخرى من
كيانات صحية أخرى (على سبيل المثال نتائج المعامل و الأشعة، وبيانات من المعالج الساب ، إلخ) ؟

Or

Do you incorporate existing patient data collected by other sources? If so, what types?

هل تكمل بيانات المرضى الموجودة من مصادر أخرى ؟ إذا كان الأمر كذلك، أيهما ؟

5) Do you think it's important to enhance data-sharing between health entities for improved health services?

في رأيك، هل تعتقد أنه من المهم تعزيز مشاركة البيانات بين المرافق الصحية لتحسين الخدمات الصحية ؟

6) Do you think using digital technologies like AI may help in enhancing data-sharing? How?

هل تعتقد أن استخدام التقنيات الرقمية مثل الذكاء الاصطناعي قد يساعد في تعزيز مشاركة البيانات ؟ كيف ؟

7) Are your data-sharing practices influenced by any external factors (e.g., regulations)?

هل تتأثر ممارسات مشاركة البيانات الخاصة بك بأي عوامل خارجية (مثل قوانين/سياسات البيانات) ؟

» ***Follow-up question:***

a) What can potentially be improved in the law to provide better health-data processing/ support for your data-sharing practices?

8) What data-sharing-related problems have you managed to identify at your organization?

ما هي مشاكل مؤسستك المتعلقة بجمع البيانات التي تمكنت من تحديدها ؟

9) How do you suggest overcoming these challenges? What role can technologies like AI play in this regard?

ما هي أفضل طريقة للتغلب على هذه التحديات ؟ هل تعتقد أن استخدام التقنيات الرقمية مثل الذكاء الاصطناعي يمكن أن يساعد في التغلب على هذه التحديات ؟

Data Storage:

1) How do you store your data?[36] Please describe your data-storage architecture.

كيف تخزن بيانتك ؟ يرجى وصف بنية تخزين البيانات الخاصة بك

» ***Follow-up questions:***

a) Do you use local or cloud storage?

هل تستخدم التخزين المحلي أو السحابي ؟

b) Is the data stored in any interchangeable or standard format? Which?

هل البيانات مخزنة أو في شكل موحد ؟ أي شكل ؟

2) What data protection and privacy practices do you use for the storage of data (e.g., using secure data-storage mechanisms like encryption)?

ما هي معايير الأمن المتبعة في تخزين البيانات واسترجاعها لتخفيف مخاطر التهديدات الأمنية (على سبيل المثال، استخدام آليات آمنة لتخزين البيانات مثل التشفير، أو في حالة البيانات الورقية هل غرفة الملفات تغلق بمفتاح؟ من الشخص المسئول عنها؟ ما المدة الزمنية المخصصة للإحتفاظ بالبيانات)؟

» ***Follow-up question:***

a) How is the data retrieved? Can the data be retrieved from previous records?

ما هي الوسائل التي تعتمدها لاسترجاع بيانات المرضى ؟ هل يمكن استرجاع البيانات التي تم جمعها مسبقًا من السجلات السابقة ؟

3) Is there a person/entity in your organization responsible for defining the rules of data storage?

هل هناك شخص/ جهة مسئول(ة) في مؤسستك لتحديد قواعد تخزين البيانات الصحية؟

Data Security:

1) Is there data that you consider sensitive?

هل هناك نوع من البيانات تعتبره بيانات حساسة؟

2) Do you have a chief security officer or similar role?

هل لديك كبير ضباط الأمن أو دور مماثل ؟

3) How do you secure your patient/customer data during collection, processing, and sharing? How do you mitigate security risks (e.g., through safe data-storage guidelines)?

كيف تؤمن مؤسستكم بيانات المرضى في ممارسات جمع، معالجة، و مشاركة البيانات؟

كيف تقلل من المخاطر المتعلقة بالتهديدات الأمنية؟

36 Andanda, P. (2020). Ethical and legal governance of health-related research that uses digital data from user-generated online health content. Information, Communication & Society, 23(8), 1154-1169.

4) How do you detect data security breaches?

كيف تكتشف اختراقات أمن البيانات ؟

5) Is the data anonymized and/or encrypted to preserve patients' privacy? Can patients' details be viewed/identified?

هل البيانات مجهولة الهوية و/أو مشفرة للحفاظ على خصوصية المرضى؟ هل يمكن الإطلاع على تفاصيل المرضى والتعرف عليها ؟

6) What security standards are followed in data storage and retrieval?

ما هي المعايير الأمنية المتبعة في تخزين البيانات واسترجاعها ؟

7) How often do you back up your database to ensure that your data is protected against cyberattacks?

كم مرة تقوم بعمل نسخ احتياطية من قاعدة البيانات الخاصة بك للتأكد من أن المعلومات القيمة لمؤسستك آمنة ضد الهجمات الإلكترونية ؟

MENA OBSERVATORY
ON RESPONSIBLE AI
مرصد الشرق الأوسط وشمال إفريقيا للذكاء الاصطناعي المسؤول

Governing Responsible Artificial Intelligence and Data in the Middle East and North Africa (MENA)

The American University in Cairo
Onsi Sawiris School of Business
Access to Knowledge for Development Center

IDRC · CRDI
International Development Research Centre
Centre de recherches pour le développement international

Canada