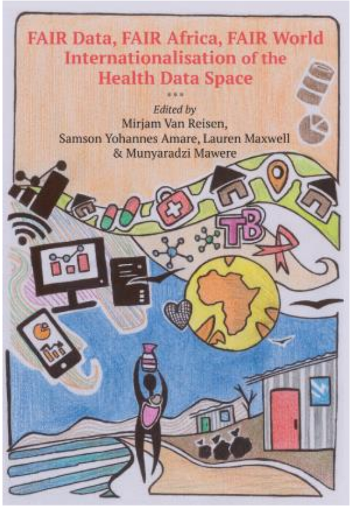
The Potential of Adoption of FAIR Guidelines in Digital Healthcare in Kazakhstan

Aliya Aktau

Chapter in:
Fair Data Fair Africa Fair World:
Internationalisation of the Health Data Space



Cite as: Aktau, A. (2025). The Potential of Adoption of FAIR Guidelines in Digital Healthcare in Kazakhstan. https://doi.org/10.5281/zenodo.15382913. In M. Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space. pp. 277–300). Bamenda: Langaa. URL: https://www.researchgate.net/publication/391750151_FAIR_Data_FAIR_Africa_FAIR_World_The_Internationalisation_of_the_Health_Data_Space

The About the Authors note can be found here: https://raee.eu/wp-content/uploads/2025/05/About-the-Authors-and-Editors.pdf
The list of figures and tables can be found here: https://raee.eu/wp-content/uploads/2025/05/List-of-Figures-and-Tables.pdf

Contents

Chapter 9

The Potential of Adoption of FAIR Guidelines in Digital Healthcare in Kazakhstan

Aliya Aktau

Abstract

Expanding the implementation of the FAIR Guidelines to more countries can diversify existing data and help researchers draw right insights from it. This exploratory study focuses on the use of the FAIR Guidelines for health data in Kazakhstan. For this, initiatives in the field of digital health were studied to understand how digital health is organised in Kazakhstan. Kazakhstan is a country with a big land surface, and comparatively small population and digital health can help solve problems of remoteness. The study investigated whether the FAIR Guidelines are aligned with the regulatory and policy framework for health data in Kazakhstan. It is concluded that implementing the FAIR Guidelines in the digital health system in Kazakhstan would solve problems at a technical and policy level, as well as assist integration of digital healthcare solutions in smarter ways. While the technical advantages are clear in terms of its relevance for public health and scientific research, and it can improve personalised medicine in Kazakhstan, the study also found that political and policy uptake of FAIR in Kazakhstan requires an improved understanding. This is mostly an issue of capacity building which requires training, not only of technical leadership, but especially also in administration and political leadership. The capacity building will also require adapting the FAIR architecture to meet the specific needs and requirements for data integration in the context of Kazakhstan.

Keywords: FAIR Guidelines, Kazakhstan, data management, scientific data, scientific research, public health, patient data

Introduction

Genome-wide association studies (GWAS) have increased our understanding of the genetic determinants of disease risks. However, large-scale studies have not captured the diversity of the global population, and these are mostly based on individuals of European ancestry (Popejoy & Fullerton, 2016), with limited representation of other groups, including from Africa, North and South America, Asia and Oceania (Gurdasani, Barroso, Zeggini & Sandhu, 2019). There were attempts to change this situation in the early 2010s, and the '1000 Genomes' project was launched to diversify genetic data and create a broad understanding of human genetic variation across multiple population groups (Clarke et al., 2012). However, as of 2018, most of the GWAS were still conducted among European and Asian population groups (among which East Asian population groups are dominant), with 78% conducted among those of European ethnicity (Sirugo, Williams & Tishkoff, 2019). This European bias has significant consequences for predicting the risks of diseases. In such a situation, the translation of genetic research into clinical practice may be inaccurate for many populations, due to the lack of ethnic diversity in human genome studies (Sirugo et al., 2019).

A critical moment has now been reached – going forward, the analysis and storage of annotated clinical and genomic information in disconnected bunkers will hinder the development and growth of research (Siu et al., 2016). Data in most electronic health record (EHR) systems is not checked for quality, nor is it structured so that it can be easily retrieved (Siu et al., 2016). These issues are compounded when data needs to be compared and used across institutions, forming a significant barrier to cross-border data exchange (Siu et al., 2016). As researchers are not usually able to generate all the data, they need to be able to reuse data from other projects (Grossman et al., 2016). Making health data available and useful to the scientific community is, therefore, a critical issue (Van Reisen et al., 2021).

The Findable, Accessible, Interoperable and Reusable (FAIR) Guidelines are designed to make data reusable. In its initial conceptualisation, this was a requirement identified mainly for researchers. For data to be 'FAIR', data should be 'Findable', ‘Accessible’ under well-defined conditions, ‘Interoperable’ without data munging[1] and ‘Reusable’ to better serve purposes (Mons et al., 2018).

In a subsequent development, the ‘FAIRification’ of data, that is the process of making data FAIR, has been introduced to the domain of health and health services. Digital health data has been identified as a way of overcoming the challenges posed by the lack of diversified medical health data in global research. FAIR is a strategy for overcoming fragmentation and lack of data integration across medical health data, both within and across countries (Van Reisen et al., 2021). This would not only serve better research, but it would also serve better health outcomes. The insights based on FAIR data could inform health workers of patient risks and help them by improved guidance in decision-making while attending to patients (Van Reisen et al. 2024).

The FAIR Guidelines give direction of how data management practices should develop, including: (i) data serving the public interest and governed by public policy; (ii) data science enhancing collective knowledge, (iii) science serving as practical solutions and services, and (iv) the close attention and participation of citizens and the general public in the discovery of knowledge (Mons, 2018). The aspiration is that this will lead to the democratisation of science and its use. The FAIR Guidelines have been included as a tool for developing and promoting regional guidelines for data exchange in East Africa and also as a regional real-time data warehouse for collecting, storing, retrieving, analysing, and managing national and regional healthcare in East Africa (Van Reisen et al., 2021; Van Reisen, Stokmans, Mawere et al., 2020).

Despite these few examples, studies show that the FAIR Guidelines have mostly been implemented in European countries, and to a lesser degree in the United States (Van Reisen, Stokmans, Basajja et al., 2020). The concern is that the dominance of the use of FAIR data in the western hemisphere, suggests that data used in research are biased

---

[1] “Data munging is the initial process of refining raw data into content or formats better-suited for consumption by downstream systems and users” (https://www.talend.com/resources/what-is-data-munging/).

towards Western geographies or people of Caucasian descent, causing bias in the representativeness of data. The integration of more diverse populations for empirical and theoretical reasoning can address this issue (Van Reisen, Stokmans, Basajja et al., 2020). Expanding the implementation of FAIR Guidelines to non-European geographies is necessary to overcome biases in health and medical studies (Van Reisen et al., 2021). Failing to integrate data from non-European geographies may negatively affect the prediction and treatment of various diseases, because AI solutions, among other things, will be based on biased data if data is inherently biased.

Research gap

The FAIR Guidelines have been implemented in Europe, Australia, North America and Africa. The inclusion of countries not commonly available for research is important for the diversification of data and to provide a richer and more inclusive perspective on health data at the global level. Thus, the inclusion of Kazakhstan has been chosen for this research as a case study to test the feasibility of FAIR adoption in Kazakhstan. Up until the time of writing (2021), there have been no articles or publications in Kazakhstan that mention the FAIR Guidelines (Lin et al., 2022). A search for references to publications on FAIR in Kazakhstan did not yield any results. Kazakhstan is a non-English speaking country, hence, the keywords searched were in both official languages of the country: Kazakh and Russian. A few Russian language articles were found that mention FAIR in research communities, but these did not mention Kazakhstan (Lin et al., 2022).

From the lack of literature on FAIR in Kazakhstan, it can be deduced that FAIR-based research and medical data from Kazakhstan are not included in international research. Hence, the adoption of the FAIR Guidelines for research data would contribute to global data diversification and potential improvements to make solutions relevant to Kazakh populations. As the FAIRification of data would benefit the information position of Kazakhstan, we could expect that it would lead to better informed tools for decision-making in health in the country.

Study objectives and research questions

This research determines how FAIR could contribute to diversifying data and reducing bias in the life sciences towards those of European descent. It particularly looks at the issue of data fragmentation within digital health and how the FAIR Guidelines might address this issue.

The research objectives were as follows:

- To highlight the current situation of digital health in Kazakhstan and to understand how digital health is set up in Kazakhstan;
- To find out what kind of digital health data is available in Kazakhstan and assess the FAIRness of digital data;
- To check health data from the Ministry of Health of Kazakhstan for alignment with global data;
- To develop a FAIR data-based model that illustrates how health data in Kazakhstan can be linked to, and benefit from, global open science.

The focus for data in the third sub-question is on cancer data, since this is one of the areas where integration in global research is beneficial. For Kazakhstan, the implementation of the FAIR Guidelines would provide access to global scientific data, while simultaneously improving science and bringing local scientists to the global level. FAIR-compliant (meta)data can reduce the cost of scientific projects, which require much effort and time in the data munging process (Mons et al., 2018). Thus, Kazakhstan may benefit significantly from adopting FAIR Guidelines for its digital health initiatives. Furthermore, there is a need for global analytics related to health data for the analysis of certain diseases and epidemic or pandemic outbreaks. The availability of medical data from various resources can reveal important patterns and lead to more accurate decision-making. Hence, improving the availability of medical data for research will improve our understanding of disease and our ability to deal with disease outbreaks around the world, as well as in Kazakhstan. This study examined the feasibility of the adoption and implementation of FAIR Guidelines using Kingdon's theory of agenda-setting and Ajzen's theory of planned behaviour as a theoretical lens.

The following specific research questions were formulated:

- Q1. What are the challenges of digitisation in Kazakhstan? Can digital health data be accessible from outside the country and, if so, how? What is the level of FAIRness of digital health data for cancer in Kazakhstan?
- Q2. How is digital health data organised in Kazakhstan? What are the public policies on digital health in Kazakhstan and how are these implemented?
- Q3. How could FAIR be implemented in Kazakhstan? What is the adoption potential for the deployment of FAIR structures? What is the political momentum for FAIR-policies in Kazakhstan?

Theoretical framework

The FAIR Guidelines are a way of facilitating knowledge discovery from data. FAIR is not a standard, but a set of guidelines that can facilitate the process of reusing data (Mons, 2018). The FAIR Guidelines provide guidance for scientific data management and stewardship and are relevant to all stakeholders in the current digital ecosystem. They call on data producers and data publishers to promote the maximum use of research data.

FAIR stands for the following (Wilkinson, 2016):

- 'Findability': (Meta)data must be discoverable by humans and machines. For the automatic discovery of datasets and services, machine readable metadata is necessary.
- 'Accessibility': After the required (meta)data is found, it is necessary to provide access to them using open universal protocols, including authorisation and authentication.
- 'Interoperability': It must be possible to integrate data from various sources for analysis, storage and processing.
- 'Reusability': Data must be adequately and precisely described to be reused.

The terms used in this study are defined as follows:

- FAIRness: This refers to the extent to which the data adheres to the FAIR Guidelines. To determine the FAIRness of the data, it is evaluated in terms of its compliance with the FAIR Guidelines.

- FAIRification: the process of making data FAIR.
- FAIR Maturity Evaluation Service: The FAIR Maturity Evaluation Service is an application or system to determine the FAIRness of digital resources. There are several services available.
- FAIR Data Point (FDP): This is a way of presenting (meta)data on the web in accordance with the FAIR Guidelines, in which (meta)data must be 'Findable', 'Accessible' (under well-defined conditions), 'Interoperable' and 'Reusable' (https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/).
- Persistent unique identifier (PID): A PID allows for the identification of digital objects and can provide information about the object, regardless of where it is located. PIDs must be global, standardised and widely used in the digital environment (e.g., Open Researcher and Contributor ID (ORCID), Digital Object Identifier (DOI) (Bazzanella, Bortoli & Bouquet, 2013).

To investigate the public policy agenda and how it changes, the multiple streams model (Kingdon, 1984) was used as a theoretical framework. This model has been at the forefront of theories of public agenda-setting for more than 30 years. The multiple streams model is composed of three streams: the problem stream, policy stream and political stream. He purports that policy change can happen when all three streams are aligned and a 'policy window' opens, allowing new policies to be put on the agenda (Kingdon, 1984).

The first stream is the problem stream, in which issues are identified as needing the attention of society (Kingdon, 1984). The second stream is the policy stream, in which policies (solutions) are formed and proposed by people in or around government (including civil society and lobby groups) (Kingdon, 1984). In this stream, experts – including people from the government, researchers, and academics – gather ideas and refine and propose solutions to the problems identified in the problem stream (Kingdon, 1984). In this stage, various solutions and alternatives may be offered and ideas tried out, generating completely new solutions (Kingdon, 1984). The third stream is the political stream, which is composed of the environment, including the national mood, internal and external changes in the country, changes in administration (which can affect the national mood) and political willingness to develop an agenda. When all streams come together, a 'policy window' is opened (Quirk, 1986), which is the point in time when new ideas can enter the policy agenda (Kingdon, 1984).

According to the multiple streams model, policy entrepreneurs play a crucial role in 'softening up' the system and connecting the problem, policy and politics streams. Policy change is unlikely to happen without policy entrepreneurs (Kingdon, 1984; Gagnon & Labonté, 2013). This group of people push their proposals or direct attention to problems. Proposals are made when the timing is right, and an open policy window presents a special opportunity; thus, proposals are often kept ready by policy entrepreneurs to launch when the political climate is right (Kingdon, 1984; Wilson, 1993).

Ajzen's theory of planned behaviour predicts and explains intention and human behaviour. The theory addresses three independent variables: attitude toward the behaviour, subjective norms, and perceived behavioural control, all of which affect the intention and behaviour to perform certain actions (Ajzen, 1991). Attitude refers to the positive or negative feelings that a person may have about performing a given behaviour (Fishbein & Ajzen, 1975). As a variable of attitude, the attitude of users regarding the adoption of FAIR is addressed in this research. A subjective norm is "a person's perception that most people who are important to him think that he should or should not perform this behaviour" (Fishbein & Ajzen, 1975). As a subjective norm, policies and regulatory frameworks on digital health are analysed in terms of the feasibility of applying FAIR Guidelines in Kazakhstan. Perceived behavioural control refers to people's perception of their ability to perform the target behaviour (Fishbein & Ajzen, 1975). As a variable of perceived behavioural control, the technical infrastructure of the healthcare sector, specifically health-related data, is analysed. Understanding these three independent variables will help to determine the feasibility of adopting the FAIR Guidelines in the healthcare sector in Kazakhstan.

Methodology

An exploratory design was used in this case study to analyse the current state of digital health in Kazakhstan and ascertain the feasibility of adopting the FAIR Guidelines, as well as explore the difficulties, challenges and opportunities involved in deploying FAIR in Kazakhstan. The research sought to understand the entire picture of digital healthcare in Kazakhstan through in-depth research into state programmes and through the investigation of the accessibility of digital health data and platforms. Based on this, the study sought to determine the conditions for the adoption and implementation of FAIR. Lastly, any obstacles and limitations on the adoption and implementation of FAIR were identified to provide direction for future research. To facilitate the research, an internship was obtained with the Ministry of Health of Kazakhstan to investigate how digital healthcare is set up in Kazakhstan and the challenges Kazakhstan is facing in relation to this.

Digitisation of health data in Kazakhstan

Kazakhstan is an emerging economy which seeks to become a member of the Organisation for Economic Co-operation and Development (OECD) by 2050. In 2012, Kazakhstan began a reform and investment programme aimed at bringing the country onto the list of most developed countries in the world. This strategic goal, set out in the 'Kazakhstan-2050 Strategy', established a new political course for the social and economic development of the country (Nazarbayev, 2012). The Strategy set new goals in the field of healthcare and served as a basis for later health-related programmes. Currently, Kazakhstan is working on improving primary healthcare, embedding obligatory social health insurance, improving the accessibility and quality of health services, and harmonising health data infrastructure. However, in comparison to the OECD and post-Soviet states, key health indicators in Kazakhstan are still low (Obermann et al., 2016). Thus, reforming healthcare and addressing several health-related issues still require work, including the development of a data infrastructure for digital health and scientific research.

Digital health and digital health data is seen as critical areas for development in Kazakhstan. This can be seen from several recent reforms and frameworks within the digital health field, including the E-health Development Concept for 2013–2020 (Ministry of Health, 2013) Densaulyk 2016–2019 (Republic of Kazakhstan, 2016) and the State Health Development Program 2020–2025 years (Ministry of Justice, 2019). All these reforms are aimed at creating an integrated information environment for data analytics. Despite these reforms, data at the level of healthcare organisations remain fragmented, do not interact with each other (Ministry of Justice, 2019), and are not accessible to researchers.

Most digital health policies in Kazakhstan are concerned with achieving data interoperability, while the ability to find, access and reuse data is still a challenge for the scientific community, healthcare organisations and other stakeholders. In total, Kazakhstan has 22 health information systems that have been developed to collect health-related data from healthcare organisations for statistical and analytical purposes. They store health data on patients, diseases, drugs and the allocation of health facilities; however, these systems are not fully integrated and do not communicate with health medical information systems in health facilities (Republic of Kazakhstan, 2020). Although the accessibility of data is regulated by the Ministry of Health, there is no platform to which interested parties can send their requests and where they can access data.

Opening state depositories to the public is required in Kazakhstan by the Law on Access to Information (Ministry of Justice, 2015). Hence, the government has established the Open Data Portal (data.egov.kz), which is a component of the Open Government project, to provide citizens with access to state data. Although it provides about 4,000 government datasets, and data is findable, accessible, interoperable (machine-readable) and reusable on this portal, there is no research or clinical data, and the portal provides only general statistics, mainly administrative health data. Furthermore, although citizens can request data from the portal, what data can be requested is not specified.

An essential aspect of implementing FAIR Guidelines is to measure the level of FAIRness using specific metrics that quantify the FAIRness of data (Wilkinson, 2018). For this study, data from the

Open Data Portal were used to assess the availability of existing health data in Kazakhstan and measure the level of FAIRness. In addition, data was obtained from the Ministry of Health of Kazakhstan to investigate the feasibility of implementing FAIR and designing a mock FDP.

Digital health data in Kazakhstan is available to medical personnel and the public, who can access their own data from their Electronic Health Passport on the government portal. Existing health information systems are only described on the website of the Ministry of Health but are not accessible in terms of specific attributes and how they might be accessed. The datasets presented on the Open Data Portal mainly provide general health statistics, but do not contain research or clinical data.

Health data is difficult to find, access, interact with, and reuse in Kazakhstan, by both humans and machines. Thus, the informational value of data is not sufficiently used. A report published by the European Union concluded that the cost of not having FAIR (Findable, Accessible, Interoperable, and Reusable) research data could run into billions (European Commission, 2018). Applying the FAIR Guidelines in Kazakhstan's healthcare infrastructure means that the data services included in it will eventually match certain infrastructural and quality standards, enabling data to be used more efficiently, both domestically and internationally, and saving significant amounts of money.

Comparison of cancer data from Kazakhstan with international cancer data

To determine the level of FAIRness of health data in Kazakhstan, an anonymised aggregated dataset obtained from the Ministry of Health in Kazakhstan was compared to a large-scale collaborative cancer project—the Cancer Genome Atlas (TCGA). The TCGA is registered on FAIRsharing, which lists over a thousand FAIR-related data and metadata standards (Sansone et al., 2019). The dataset used for this study is part of the cancer data held in Kazakhstan's cancer information system. A more detailed description and the existing parameters of the cancer information system are not publicly available. Therefore, we only requested patient parameters such as age, gender, and ethnicity, as well as the medical parameters of the cancer. For this study, an explicit comparison of cancer data with global cancer datasets cannot be made, as the parameters that could be accessed were limited.

The data obtained from the Ministry of Health has the same categories as the global cancer data, including personal data and medical data. However, the personal data could be improved by adding more of the patient's personal history, including history of smoking, alcohol consumption, and genetic history. In addition, the medical data could be improved by adding more specific details on tumours and genomic information. Thus, the inclusion of genomic data and the personal history of patients in the FDP would provide a more accurate picture of the causes of different diseases, resulting in better treatment for cancer patients. Moreover, most international data repositories are unlocking the genomic data of cancer patients, which helps to better understand the genetic causes of various diseases. This study did not look at the genomic data of patients, due to the scope of the project and the limited time. Despite this, including the genomic data of patients in future research could significantly affect the treatment of different diseases, both globally and within Kazakhstan. This might influence the entire treatment of cancer in terms of precision oncology, as it uses precise knowledge of the structure and activity of a patient's tumour genome to suggest therapies, thereby providing meaningful therapeutic responses (Jensen, Ferretti, Grossman & Staudt, 2017). As Kazakhstan intends to provide patients with personalised treatment (Dyussembekova, 2017), applying FAIR data guidelines can make a significant contribution. A combination of digitising clinical data, as well as genomics data, using a FAIR data workflow can provide a good infrastructure, which can lead to more accurate individualised treatment.

Applying Kingdon's theory of agenda-setting

Problem stream

According to the World Health Organization (WHO) in Kazakhstan (WHO, 2018), one of Kazakhstan's top three challenges is increasing the country's overall data handling capabilities. The amount of digital data in health is rapidly increasing and, to deal with this volume of data, emphasis needs to be placed on data quality. In addition, data should be made easily accessible for interested parties and medical personnel, who should be able to interact with one another to provide better medical services and conduct medical research. Accordingly, since 2013, Kazakhstan has been focusing on the development of an integrated information environment that can serve as the foundation for tailored and preventive medicine (Ministry of Health, 2013). To strengthen people-centred health systems, the country has started developing an integrated data infrastructure. Efforts have been made to introduce the Interoperable Platform (OECD, 2018), which was initially planned to be launched in 2020, and has been promoted as a solution to the problem of data fragmentation (Abishev & Spatayev, 2019). The platform contains instruments and services that will ensure interoperability throughout the digital healthcare sector by using a common set of registers and classifiers. This platform contains an electronic health record repository, analytic data storage using business intelligence tools, and a patient portal. However, at the time of writing, the platform has not yet been launched, thus, data remains fragmented on the level of healthcare organisations and do not interact with each other. As data is still not findable, accessible, or interoperable, it cannot be reused by stakeholders. Hence, the informational value of health data is not fully utilised in Kazakhstan.

The cost of resolving this issue in a FAIR manner would be offset by the saving of billions of euros (European Commission, 2018). In addition, making data FAIR will address data management issues in healthcare and research and allow meaningful patterns to be discovered, which will contribute to better decision-making. The key today is not the problem of collecting, publishing, and storing information but ensuring the findability, accessibility, interoperability, and reusability of data. Thus, the effective management of data by making it FAIR can help to address issues within the healthcare sector and medical research in Kazakhstan.

Policy stream

The Ministry of Health in Kazakhstan has recognised most of the problems related to digital health, as set out above, and is pursuing changes and policies in accordance with national strategies

(Obermann et al., 2016). Although several healthcare reforms have been implemented since 2013 to strengthen health data infrastructure in Kazakhstan, a solution to overcome the heterogeneity of databases in healthcare is needed. The Ministry of Health announced a new State Health Development Program for 2020–2025 aimed at improving health infrastructure to combat data fragmentation (Ministry of Justice, 2019). Hence, there is a need for an FDP infrastructure for digital healthcare in Kazakhstan, to contribute to research and healthcare, both nationally and globally.

Politics stream

Achieving sustainability requires the creation of a favourable environment, including a good institutional framework, funding for standards, and high-quality data and use cases, rather than merely investing in a particular technology (Abishev & Spatayev, 2019). In this regard, Kazakhstan's government is eager to develop an advanced health data infrastructure, which could be critical for high-quality medical care and medical research. A strong political intent to enhance health outcomes can be seen in recent reforms, such as the E-health Development Concept for 2013–2020 (Ministry of Health, 2013), Densaulyk 2016–2019 (Republic of Kazakhstan, 2016), and the State Health Development Programme for 2020–2025 (Ministry of Justice, 2019). These reforms are all in favour of data interoperability and efficient use for future outcomes.

The inclusion of Kazakhstan's health data can make a significant contribution to global open science. Although it benefits open science around the world, health infrastructure development has the potential to give much more to Kazakhstan, including improving the wellbeing of its citizens. Although there is an Open Data Portal that provides government datasets in Kazakhstan, it only contains general health statistics and administrative health-related data and does not provide any data that can be used for research. Although the portal is called 'open', the data on the portal is not entirely open and does not conform to the FAIR Guidelines.

Since there has not been any information on Guidelines in Kazakhstan, there has not been any promotion of it. Kingdon contends that the uptake of a new agenda in the political stream requires policy entrepreneurs to engage with the issue to encourage political interest (Kingdon, 1984). In the FAIR ecosystem, it is recognised that communities of practice on FAIR are needed to support uptake. The uptake of FAIR is promoted by FAIR Implementation Networks (INs). In Kazakhstan, this work was carried out under the FAIR Implementation Network 'Ambassadors', which introduces FAIR in countries and geographies that are new to the concept.

In this way, political willingness to accept FAIR can come from the bottom up, by gradually exposing people from the government, researchers, and scientists to the idea. These stakeholders will then become the 'policy entrepreneurs' for the promotion of FAIR, towards placing it on the policy agenda. The introduction of FAIR must be carried out by promoting its value, including how it can improve the scientific environment and contribute to society so that stakeholders understand the value that FAIR Guidelines can bring to Kazakhstan. In this way, it is conceivable that a 'political window' could be opened, and circumstances could be created that are conducive to moving FAIR Guidelines onto the policy agenda in Kazakhstan.

Feasibility of deploying FAIR in Kazakhstan through Ajzen's theoretical lens

An FDP is a data registry that provides data and metadata using FAIR Guidelines (GitHub, n.d.). It helps data owners to expose their data in a FAIR manner and, if licence conditions permit, also allows data users to find metadata and access them. Although an FDP could be applied to many knowledge domains, in this research we will be focusing on the data of patients with breast cancer and pancreatic cancer. The FDP prototype will use these diseases to test the feasibility of implementing FAIR in Kazakhstan.

The objective of designing the FDP is to illustrate how oncological data on breast cancer and pancreatic cancer can be assigned machine-readable metadata to enable them to be discoverable by humans and machines. The basis of this project is to ensure that the Kazakhstan Cancer Center is ready to receive knowledge from the international level and share information and experience with others. The development of this FDP can contribute to cancer research worldwide, as well as in Kazakhstan, thereby improving the health of millions of people around the world. An FDP for healthcare in Kazakhstan could be deployed either: (i) as a stand-alone web application, through which data owners share their own data and data consumers gain access to the data, thus, contributing to Kazakhstani research and medicine, or (ii) as part of a larger interoperable system, providing accessibility functionality globally.

The FAIRification process for non-processed data involves several steps (GO FAIR, n.d.). Requirements for findability and accessibility can be achieved at the metadata level, whereas interoperability and reusability require more work at the data level. The FAIRification process for data includes: the analysis of the database structure together with field names, defining a semantic model that can determine the dataset accurately and in a computer-actionable way, capturing terms, Uniform Resource Identifiers (URIs) and descriptions, making data linkable (which can be done using the Semantic Web and Linked Data technologies), configuring access licences and usage rights, and rich metadata for the dataset. In addition, as a final step, the FAIRified data needs to be deployed or published together with the metadata and a licence. Thus, metadata can be found by search engines and the data can be accessed under well-defined licences.

For the FAIRification of cancer data, the OpenRefine tool was used. OpenRefine is a software tool that enables the pre-processing of data, i.e.: cleaning it, transforming it into a machine-readable format, and extending it with web services and external data. Thus, cancer data was converted to a machine-actionable format via OpenRefine, and made interoperable and reusable for further needs. The data in OpenRefine was then linked to the FDP and described using extended metadata.

By installing the mock FDP, it was concluded that the deployment of the FDP is feasible in terms of technical readiness compounded by the range of healthcare reforms undertaken to improve data infrastructure and health outcomes in Kazakhstan. However, the acceptance of FAIR requires a change in attitude and social norms among stakeholders, therefore, it needs to be introduced carefully, focusing on the value it can bring to healthcare and society in Kazakhstan and globally. An understanding of the relevance of the FAIR architecture for public health and scientific research in Kazakhstan is necessary. There must be deeper knowledge about how the FAIR architecture can improve science and healthcare, and an understanding of what the future benefits of the adoption of FAIR are.

Regarding subjective norms, these are critical for the national mood and the political relevance of any problem at a particular point in time. If a problem is perceived as urgent, researchers and entrepreneurs may reach out to decision-makers in the government. Regarding perceived behavioural control, the feasibility of the adoption of FAIR at the technical level is possible, however, it also requires a deeper understanding of the implementation of FAIR. This is critical for the formulation of policies in that it relates to the feasibility of the implementation of an agenda. Changing the perception of the relevance of FAIR data stewardship can be done by building the capacity of researchers and stakeholders interested in the development of scientific research and public health in Kazakhstan and around the world.

Conclusion

The inclusion of Kazakhstan in large-scale studies can diversify research data, which can contribute to better-quality data. This, in turn, can have implications for prevention, prediction of disease risks and treatment of diseases, in Kazakhstan and globally. The integration of more diverse populations is essential for empirical and theoretical reasoning on a global scale. While several digital health reforms have been launched in Kazakhstan in recent years, the findability, accessibility, interoperability, and reusability of health data remains a challenge. The opportunities presented by the FAIR Guidelines for healthcare in Kazakhstan are numerous and align with national interests in digital health.

This research looked at the current stewardship of digital data on cancer patients in Kazakhstan. Global research on cancer patients is advancing precision treatment which potentially would benefit health provision in the country. The evaluation of the FAIRness of cancer data on the available platforms in Kazakhstan showed that the current deployment does not comply with the FAIR Guidelines.

The lack of use of FAIR workflows may be because the FAIR Guidelines have not been promoted in the country. For the FAIR Guidelines to be implemented in relation to health and medical data in Kazakhstan it is necessary to combine the efforts of various parties, such as the government, researchers, and entrepreneurs, to make this possible and to formulate public policy based on FAIR. The acceptance of the FAIR Guidelines will depend on how stakeholders perceive these opportunities in terms of how effective the FAIR Guidelines are in overcoming data management issues, what impact they will have on their work, and whether they see the benefits that FAIR can bring.

The deployment of an FDP is necessary to address the issues in data management in Kazakhstan. This study found that the deployment of an FDP is feasible from the point of view of technical readiness. Deployment could contribute to the improvement of the data infrastructure and health data analytics in Kazakhstan. Adopting the FAIR Guidelines will help solve the problem of data fragmentation in Kazakhstan, which will lead to a better use of digital health data. This should ultimately improve treatment for cancer patients in Kazakhstan. Nevertheless, using Kingdon's theory of agenda setting (the problem stream, policy stream, and political stream), it was found that FAIR, while technically feasible (policy stream), and responding to existing problems in data management (problem stream), was lacking political momentum (political stream). Without political momentum, the solution (implementation of FAIR Guidelines) will not reach the policy agenda in Kazakhstan.

Investigating the lack of political interest in Kazakhstan, the study found that the understanding and acceptance of FAIR as a feasible solution requires a change in attitudes and social norms on data stewardship in the country. To effect such a change, capacity building and training are critical to clarify the issues surrounding digital health data and to build support for FAIR Guidelines among the various stakeholders within the health sector in Kazakhstan.

It is concluded that the introduction of FAIR Guidelines should be carried out from the bottom up, with support of people in high offices, by promoting FAIR workflows and communicating what it can bring to healthcare in Kazakhstan. The acceptance and adoption of innovation can only be achieved if national actors can relate to the issues and solutions at hand. Public policy formulation may then ensure the inclusion and even mainstreaming of the FAIR Guidelines in digital healthcare. The idea of making data FAIR, should be seen as a practical proposal to help create a favourable environment in which high-quality data is the foundation of better medical treatment and research.

Acknowledgements

I would like to express my gratitude and appreciation to my supervisors, Professor Mirjam van Reisen and Dr. Katy Wolstencroft, for their excellent leadership and inspiring mentoring during this study, which enriched my knowledge about FAIR data science and helped me to complete this work. I would also like to thank Susan Sellars for copyediting and proofreading. Finally, I would like to acknowledge VODAN-Africa, the Philips Foundation, the Dutch Development Bank FMO, CORDAID, and the GO FAIR Foundation for supporting this research.

Authors' Contribution

The author is responsible for this study and preparation of this article.

Ethical Considerations

Tilburg University, Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences REDC#2020/013, June 1, 2020-May 31, 2024, on Social Dynamics of Digital Innovation in remote non-western communities. Uganda National Council for Science and Technology, Reference IS18ES, July 23, 2019-July 23, 2023.

Conflict of Interest

The author declares that they have no competing interests.

References

Abishev, O., & Spatayev, Y. (2019). The future development of digital health in Kazakhstan. *Eurohealth, 25*(2), 24–26. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/332524/Eurohealth-25-2-24-26-eng.pdf

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Bazzanella, B., Bortoli, S., & Bouquet, P. (2013). Can persistent identifiers be cool? *International Journal of Digital Curation, 8*(1), 14–28. https://doi.org/10.2218/ijdc.v8i1.246

Clarke, L., Zheng-Bradley, X., Smith, R., Kulesha, E., Xiao, C., Toneva, I., et al. (2012). The 1000 Genomes project: Data management and community access. *Nature Methods, 9*(5), 459–462. https://doi.org/10.1038/nmeth.1974

Dyussembekova, Z. (2017, November 17). AI to be used in Kazakh healthcare. *The Astana Times*. Retrieved from https://astanatimes.com/2017/11/ai-to-be-used-in-kazakh-healthcare/

European Commission, Directorate-General for Research and Innovation. (2018). *Cost-benefit analysis for FAIR research data: Cost of not having FAIR research data*. Publications Office of the European Union. Retrieved from http://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1

FAIRsharing. (n.d.). *FAIR evaluator frontend.* FAIRsharing. Retrieved from https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#!/

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: An introduction to theory and research*. Addison-Wesley.

Gagnon, M. L., & Labonté, R. (2013). Understanding how and why health is integrated into foreign policy – A case study of Health is Global, a UK government strategy 2008–2013. *Globalization and Health, 9*(1), 24. https://doi.org/10.1186/1744-8603-9-24

GitHub. (n.d.). *FAIR Data Point design specification* [Online]. Retrieved from https://github.com/FAIRDataTeam/FAIRDataPoint-Spec

GO FAIR. (n.d.). *FAIRification process* [Online]. Retrieved from https://www.go-fair.org/fair-principles/fairification-process/

Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., & Staudt, L. M. (2016). Toward a shared vision for cancer genomic data. *New England Journal of Medicine, 375*(12), 1109–1112. https://doi.org/10.1056/NEJMp1607591

Gurdasani, D., Barroso, I., Zeggini, E., & Sandhu, M. S. (2019). Genomics of disease risk in globally diverse populations. *Nature Reviews Genetics, 20*(9), 520–535. https://doi.org/10.1038/s41576-019-0144-0

Jensen, M. A., Ferretti, V., Grossman, R. L., & Staudt, L. M. (2017). The NCI genomic data commons as an engine for precision medicine.

Blood, 130(4), 453–459. https://doi.org/10.1182/blood-2017-03-735654

Kingdon, J. W. (1984). *Agendas, alternatives, and public policies*. Little, Brown.

Lin, Y., Purnama Jati, P. H., Aktau, A., Ghardallou, M. A., Nodehi, S., & Van Reisen, M. (2022). Implementation of FAIR guidelines in selected non-Western geographies. *Data Intelligence, 4*.

Ministry of Health. (2013). *E-health development concept for 2013–2020*. Republican Center for Health Development, Ministry of Health, Republic of Kazakhstan. Retrieved April 11, 2020, from http://www.rcrz.kz/index.php/ru/kontseptsiya-razvitiya-elektronnogo-zdravookhraneniya

Ministry of Justice. (2015). *On access to information: Law of the Republic of Kazakhstan dated November 16, 2015, No. 401-V*. Ministry of Justice of the Republic of Kazakhstan. Retrieved April 5, 2020, from http://adilet.zan.kz/eng/docs/Z1500000401

Ministry of Justice. (2019). *State health development program for 2020–2025*. Ministry of Justice of the Republic of Kazakhstan, Institute of Legislation and Legal Information. Retrieved April 11, 2020, from http://adilet.zan.kz/rus/docs/P1900000982#z17

Mons, B. (2018). *Data stewardship for open science: Implementing FAIR principles*. CRC Press. Retrieved April 3, 2020, from https://www.crcpress.com/Data-Stewardship-for-Open-Science-Implementing-FAIR-Principles/Mons/p/book/9780815348184

Nazarbayev, N. (2012, December 14). *Address by the President of Kazakhstan: Leader of the Nation, N. Nazarbayev "Strategy Kazakhstan-2050": New political course of the established state"*. KazInform. Retrieved April 5, 2020, from https://www.inform.kz/en/article/2346141

Obermann, K., Chanturidze, T., Richardson, E., Tanirbergenov, S., Shoranov, M., & Nurgozhaev, A. (2016). Data for development in health: A case study and monitoring framework from Kazakhstan. *BMJ Global Health, 1*(1), e000003. https://doi.org/10.1136/bmjgh-2015-000003

OECD. (2018). *OECD reviews of health systems: Kazakhstan 2018*. OECD Publishing. Retrieved November 8, 2019, from https://read.oecd-ilibrary.org/social-issues-migration-health/oecd-reviews-of-health-systems-kazakhstan-2018_9789264289062-en

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature News, 538*(7624), 161. https://doi.org/10.1038/538161a

Quirk, P. (1986). Agendas, alternatives, and public policies. *Journal of Policy Analysis and Management, 5*(3), 607–613. Retrieved from http://search.proquest.com/docview/1761689531/

Republic of Kazakhstan. (2016). *Densaulyk State Program for 2016–2019 years*. Official Information Source of the Prime Minister of the Republic of Kazakhstan. Retrieved April 5, 2020, from https://primeminister.kz/ru/documents/gosprograms/gosudarstven naya-programma-razvitiya-zdravoohraneniya-respubliki-kazahstan-densaulyk-na-2016-2019-gody

Republic of Kazakhstan. (2020, February 14). *Implementation results of the Densaulyq State Program, modern medical technologies, social insurance, or how the healthcare system of Kazakhstan is improving*. Official Information Source of the Prime Minister of the Republic of Kazakhstan. Retrieved April 13, 2020, from https://primeminister.kz/en/news/reviews/implementation-results-of-densaulyq-state-program-modern-medical-technologies-social-insurance-or-how-healthcare-system-of-kazakhstan-is-improving

Sansone, S.-A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A. L., & Thurston, M. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology, 37*(4), 358–367. https://doi.org/10.1038/s41587-019-0080-8

Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell, 177*(1), 26–31. https://doi.org/10.1016/j.cell.2019.02.048

Siu, L. L., Lawler, M., Haussler, D., Knoppers, B. M., Lewin, J., Vis, D. J., et al. (2016). Facilitating a culture of responsible and effective sharing of cancer genome data. *Nature Medicine, 22*(5), 464–471. https://doi.org/10.1038/nm.4089

Van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihigo, M., Folorunso, S., Schultes, E., Basajja, M., Aktau, A., Amare, S. Y., Taye, G. T., Jati, P. H. P., Chindoza, K., Wirtz, M., Ghardallou, M., Van Stam, G., Ayele, W., Nalugala, R., Abdullahi, I., Osigwe, O., Graybeal, J., Medhanyie, A. A., Kawu, A. A., Liu, F., Wolstencroft, K., Flikkenschild, E., Lin, Y., Stocker, J., & Musen, M. A. (2021). Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Advanced Genetics, 2*(2). https://doi.org/10.1002/ggn2.10050

Van Reisen, M., Stokmans, M., Basajja, M., Ong'ayo, A., Kirkpatrick, K., Mons, B., (2020). Towards the Tipping Point of FAIR Implementation. In: *Data Intelligence 2(2020), 264–275*. doi: https://doi.org/10.1162/dint_a_00049 (eds. Mons, B., Jacobsen, A. & Schultes, E.). MIT Press, http://www.mitpressjournals.org/dint

Van Reisen, M., Stokmans, M., Mawere, M., Basajja, M., Ong'ayo, A. O., Nakazibwe, P., Kirkpatrick, C., & Chindoza, K. (2020). FAIR practices in Africa. *Data Intelligence, 2*(1–2), 246–256. https://doi.org/10.1162/dint_a_00047

Van Reisen, M., Yohannes, S., Plug, R., Tadele, G., Gebremeskel, T., Kawu, A. A., Smits, K., Woldu, L. M., Stocker, J., Heddema, F., Folorunso, S., Kievit, R., & Medhanyie, A. A. (2024). Curation of federated patient data: A proposed landscape for the Africa health data space. In A. L. Imoize, M. S. Obaidat, & H. Song (Eds.), *Federated learning for digital healthcare systems* (Intelligent Data-Centric Systems series). Elsevier. https://doi.org/10.1016/B978-0-443-13897-3.00013-8

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Wilkinson, M. D., Sansone, S.-A., Schultes, E., Doorn, P., da Silva Santos, L. O. B., & Dumontier, M. (2018). A design framework and exemplar metrics for FAIRness. *Scientific Data, 5*, 1–4. https://doi.org/10.1038/sdata.2018.118

Wilson, W. (1993). *Sociology and the public agenda*. Sage Publications. https://doi.org/10.4135/9781483325484

World Health Organization. (2018). *World Health Organization in Kazakhstan (Brochure)*. Retrieved April 11, 2020, from http://www.euro.who.int/en/countries/kazakhstan/publications/world-health-organization-in-kazakhstan-2018