Legal and technological aspects for the creation of a European Health Data Space

André Caravela Machado
Economics, Management, Industrial Engineering and Tourism Department of the University of Aveiro
Aveiro, Portugal
amachado@ua.pt

Daniel Ferreira Polónia
Economics, Management, Industrial Engineering and Tourism Department of the University of Aveiro
GovCOPP (Governance, Competitiveness and Public Policies) Research Group
Aveiro, Portugal
dpolonia@ua.pt

Abstract — **This paper presents review of the legal and technological topics surrounding the introduction of a European Health Data Space, currently under development. From the legal perspective, the limitations imposed by the GDPR are analysed, with a specific analysis on the privacy and anonymization issues, and question regarding the secondary use of data from existing health records are also approached. From the technological perspective, an analysis is made on how data should be governed and how data quality impacts the realization of the intended Data Space. As a conclusion, it is suggested a mechanism so that each person can define how and when its health data can be accessed and (re)used for secondary purposes.**

Keywords – European Strategy for Data, European Health Data Space, GDPR, health data, secondary use, privacy, data quality and data governance.

I. INTRODUCTION

In February 2020, the European Parliament [1] published the European Strategy for Data, thus communicating the elements of the data strategy to be adopted in the European Union for the next five years, challenging all technological agents, governments, researchers, and academics to reflect on the use of data, with the main objective of developing common data space in strategic and public interest sectors across Europe.

This reflection intends to obtain answers in the various domains, combining as many infrastructures as possible and useful governance mechanisms, to promote the interoperability of data and its better use to contribute to the well-being of European citizens.

Although these common data spaces cover several sectors, all of them relevant to European society, special attention is devoted to the creation of a European Health Data Space (EHDS). This data space will store, manage and operationalize the use of health data at the European level, contributing to the emergence of wealth creation mechanisms and the promotion of advances in health. Also, the creation of such space is based on the fact that only by significantly revamping health and delivery systems of care it will be possible to ensure that they remain suitable for their purpose [2].

The creation of structured mechanisms on a European scale to leverage data at the health level is also essential for the development of technologies and innovations that contribute to advances in the prevention and cure of diseases and to support the practice of evidence-based medicine. In this type of practice, the appropriate and interventional use of data will support preventive mechanisms and disease treatment, bringing benefits to human beings through a new scientific and technological dynamic.

The strategy that supports the creation of a European Health Data Space aims to strengthen the use and reuse of health data, providing support to the various authorities, within the scope of decisions that improve the accessibility, effectiveness, and sustainability of existing health systems. This strategy advocates that this approach will contribute to improving citizens' access to healthcare, improving the management of healthcare systems, developing new medicines, conducting clinical trials, increasing safety and the evaluation of medical products and, finally, for the increase of competition between economic actors and for the competitiveness of the European Union in the health sector worldwide.

In short, it is intended to enhance a set of benefits for all citizens, which should be weighted according to the costs of creating and maintaining a system capable of pursuing these ends, with the implementation of innovations and technologies that can reach an adequate level within the European space. Currently, each Member State defines its data use policy, so creating a harmonization of these policies among them, in addition to being challenging, can only be achieved through control mechanisms that are under the rules of protection of personal data, that guarantee its use with the proper quality and that are interconnected as a single space of health data.

Since health is one of the areas that will potentially benefit most from the creation of a common European data space, which may be reflected in greater access to health services, a reduction in costs in the provision of services and an increase in the quality of service provided, it will be up to the Member States to lay the foundations so that models can be developed

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

and created to ensure the sustainability of the European Health Data Space.

To contribute to this objective, two of the main aspects will be analyzed:

1. Legal Aspect – In this aspect, the implications in terms of the General Data Protection Regulation (GDPR) of the implementation of a system as proposed by the European Commission and the existing legal limitations regarding privacy and secondary use of data will be analyzed.

2. Technological Aspect – The technological issues underlying the sharing of health data from different institutions providing health care in different European countries will be analyzed and how it will be possible to integrate this information in a single health data space that guarantees the privacy, discussing aspects such as data quality and governance.

II. European Health Data Space

At the beginning of the 21st century, the development of data-centric strategies changed the paradigm of existing business models [3].

The concept of Data-Driven Innovation (DDI) refers to using data and conducting analysis to improve and promote new products, processes, organizational methods, and markets in digital ecosystems [4], which are defined as being digital markets where the information is generated because of user actions is stored in the form of data, and this data can then be analyzed to find patterns and trends [5].

The search for these patterns and trends enables the structuring of strategies based on DDI and that has led to the emergence and development of new products, business models and opportunities in the digital ecosystem of various sectors of society and business [6].

In the field of health, the European Commission states that DDI will bring enormous benefits to European citizens in personalized medicine, allowing an improvement in the health care provided and the control of the costs of providing these services [7].

However, these improvements should consider the growing ageing of the European population, which has been putting pressure on health and care systems to carry out reforms and find innovative solutions to become more resilient, accessible and effective in providing quality care to European citizens [8].

In this context, the European Health Data Space emerges as a common European space for health data that aims to promote the exchange and access to different types of data (electronic health records, genomic data, data on patients' clinical records, etc.) not only to support the provision of care (primary use of data) but also for research and policy-making purposes in the health field (secondary use of data) [9].

This plan is based on four main pillars (prevention, early diagnosis, treatment, and follow-up care) aiming to reduce health inequalities and be the first step towards a true European Union in terms of Health [10].

As a result of these initiatives, it is expected that by 2030 there will be a common European Health Data Area across the Union, which will allow the development of symbiotic relationships between European authorities, regulators, providers, health professionals and citizens, enhancing the emergence of innovative solutions based on the data made available.

III. Legal Aspects of EHDS

A. Privacy of Personal Data

The General Data Protection Regulation (GDPR) [11], which aims to protect natural persons concerning the processing of personal data and their free movement, states in its preamble that "the principles of, and rules on the protection of natural persons with regard to the processing of their personal data should, whatever their nationality or residence, respect their fundamental rights and freedoms, in particular their right to the protection of personal data.", reinforcing that such regulation is "intended to contribute to the accomplishment of an area of freedom, security and justice and of an economic union, to economic and social progress, to the strengthening and the convergence of the economies within the internal market, and to the well-being of natural persons".

Following the publication of this regulation, several authors have been analyzing its application in terms of health data. Larrucea et al. [12] note that transfers of personal data to third countries are considered one of the main challenges to be addressed, with informed consent being a fundamental aspect, since the law emphasizes this role as central to the processing of data. Also, Doetsch et al. [13] argue that the GDPR operates as the main legal framework for the protection of personal data and data privacy between EU countries and its application should be complemented in health data by elements such as the Helsinki declaration of ethical principles for clinical research.

Regarding the use of data, in article 5, paragraph 1 of the GDPR, safeguards and provides derogations to the rights of data subjects are provided when data are processed for scientific and statistical research purposes, including sensitive data. The use of information to characterize a collective phenomenon in a given population and the processing of personal data for statistical, scientific or historical purposes is permitted, subject to adequate safeguards and the adoption of technical and organizational measures (eg pseudonymization, anonymization).

Hence, from the GDPR, and regarding data and health, it is possible to extract that:

a) the processing of Community statistics on public health and health and safety at work is allowed;

b) it is prohibited to make decisions or actions related to a specific individual;

c) the public health interest is defined as all the fundamentals that are linked to health [14]*;*

d) if a statistical analysis cannot be carried out with anonymized data, data collected for a particular purpose must be anonymized as soon as possible.

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

These conditions, when applied to data-based solutions in the healthcare field, imply that any application that can store, process, and analyze user data must comply with the essential requirements mentioned in the GDPR.

It is also clear from the analysis of the GDPR that the penalties to be applied to those who violate the rules provided for in the regulation are foreseen in the preamble of the regulation, where it is stated that "in order to strengthen the enforcement of the rules of this Regulation, penalties including administrative fines should be imposed for any infringement of this Regulation, in addition to, or instead of appropriate measures imposed by the supervisory authority pursuant to this Regulation", making the supervisory authority the competent entity for the application of criminal sanctions legally in force in terms of data protection.

Since a European Health Data Space requires a harmonized health data privacy management framework, GDPR makes the sharing and flow of healthcare data across borders possible by laying the foundations of a framework of trust for patients, consumers, and other stakeholders. The nature of the data involved (personal data vs. non-personal data) is covered by the GDPR and its processing is subject to numerous legal data protection restrictions, which do not apply to non-personal data [15].

B. Secondary Use of Data

The primary use of data consists of its use for a purpose, determined by the organization that collected it [14]. On the other hand, the secondary use of data, namely in health, is defined by Geissbuhler et al. [16] as personal health information that is used outside the direct provision of health care, which includes activities such as research and development of new products or processes, measurement of the quality and safety of the service provided, public health and payments. In addition, the secondary use of health data can improve individuals' health experiences, expand knowledge about diseases and appropriate treatments, enrich understanding of the effectiveness and efficiency of health systems, and support public health and health goals. safety.

In the specific case of research and development, the secondary use of data makes it possible to transform the forms of care delivery, by being able to capture, integrate, explore, and connect multiple stakeholders in improving the diagnosis, treatment and prevention of diseases in individuals, allowing innovative research techniques reveal relevant biomarkers [17].

Typically, hospitals, pharmacies, laboratories, and other healthcare organizations generate clinical data as a by-product of the service provided. The widespread adoption of electronic health records (EHRs) is also accelerating the collection of sensitive clinical data, raising concerns about their use, particularly in terms of privacy.

However, the improvement of health services involves sharing data so that patient care can be carried out and better clinical research results can be obtained. These benefits increase with the scale of data sharing and the variety of healthcare environments. However, due to the diversity of cultures, languages, policies, regulations and operational arrangements for data access and collection in the European Union, the re-use of health data is limited, not only across borders but also within each country [16].

Another area that is highly dependent on the use of health data is clinical trials. Dupont et al. [18] carried out a survey, in the form of interviews, having identified gaps and areas for improvement in this regard. From the results, it was found that improvements can go through optimizing the process of identifying patients for recruitment into clinical trials (70% of respondents), reducing the time needed to carry out clinical trials (59%), controlling high costs and workload in clinical research (54%) and to validate the feasibility of the evaluation protocol (50%).

In terms of managing the relationship with patients, McCormack et al. [19] emphasize that patients value the request for consent whenever it is intended to make additional use of health data beyond its primary objective of providing health services, with patients affirming that consent is a social agreement and decisions taken on research are not automatically assigned to the research team or ethics panel. In this regard, Emam et al. [20] state that current practice in many countries and institutions has often been based on the premise that data reuse is less ethically questionable when data are previously anonymized.

There are also other ways in which individuals provide medical information voluntarily and often unconsciously, such as online use. This case occurs when the individual searches for information about diseases, discussing their medical experiences in emails, blogs, social networks, search engines, including those dedicated to specific diseases or even when they register to obtain coupons on pharmaceutical advertising websites, providing specific information to third parties. About three-quarters of consumers who use the Internet look for health information, and about three-quarters of health websites contain one-third of the elements necessary for proper screening [21].

In terms of electronic health records (Electronic Health Records - EHR), Hyppönen et al. [22] argue that the integrity and interoperability of health data are essential requirements when patient information is intended to be used for secondary purposes. With the increasing availability of large databases of electronic health records (EHR), clinical investigators are increasingly interested in secondary use of data, although future data collection will be expensive and time-consuming. The use of an EHR can allow a medical institution to develop a clinical data repository with extensive records for many patients, thus allowing for more efficient retrospective research, and these data are a promising resource for comparative efficacy research, outcome research, epidemiology, drug surveillance, and public health research [23].

Danciu et al. [24] highlight that the transition from paper to electronic support has created new opportunities for the secondary use of clinical data in biomedical research to assess the quality of care. The use of data collected during patient care for clinical research, based on secondary use of data, is becoming increasingly important because they operationalize the possibility of collecting and analyzing clinical data collected over decades, and that would be impossible to

2022 17th Iberian Conference on Information Systems and Technologies (CISTI)
22 – 25 June 2022, Madrid, Spain
ISBN: 978-989-33-3436-2

analyze in terms of time and cost with the refusal of traditional methods. As an example of this application, the first large study linked to early use of erythromycin for infantile hypertrophic pyloric stenosis was based on a retrospective review of over 14,000 infant records based on a clinical EHR [25].

Despite the possibilities opened by this transition, the complexity associated with the massive use of clinical data for secondary purposes, as well as its collection, pose complex contractual issues and problems, as it becomes difficult to track ownership and responsibility rules for their use [26].

In European terms, the European Commission's data policy plan [27] includes the implementation of the data strategy that supports the reuse of public data, called secondary use of data.

Concerning actions already underway for this use, for the eHAction project [28], the secondary use of data is a pillar of the digital transformation in health and, to ensure its safe use through governance actions, collaborations are necessary. multi-sector with respect for data protection values. Since the creation of a Health Data Space in the EU aims to strengthen international cooperation in the field of data governance, based on GDPR, Member States, together with representatives of the European Commission and experts from various related entities, addressed models of governance and national strategies for secondary use of data, as well as policy, technical and legal issues for collecting and sharing data for secondary use.

In summary, Gallgher et al. [29] note that there are ethical issues underlying the secondary use of health data, and the possibility of opting out of data sharing with potential secondary use should be always made available to the patient.

IV. Technological Aspects of EHDS

In this section, complementing the legal issues previously discussed, technological issues will be addressed, especially regarding the governance and quality of the data to be shared.

A. Data Governance

Data governance involves a set of processes to improve consistency and accuracy, reduce the cost of management and increase the security of available data [30], and a good data governance structure ensures that organizational data provides efficient access to accurate information and tools for further analysis.

It is important to note that data governance is different from data management. Data governance complements data management, but never replaces it. Management concerns the decisions that the organization makes involving the implementation of these decisions, whereas governance concerns the decisions that need to be taken to ensure effective management while providing a framework for the fulfilment of these tasks [31].

In other words, governance encompasses not only decision domains but also the responsibility for decision making. An example of this is data quality: data governance provides a framework for identifying who has the right to make decisions to determine data quality standards, what necessary aspects of data quality need to be included and how this can be ensured. On the other hand, data management involves determining the actual metrics that will be used to assess predefined quality standards [32].

For good data governance, all key stakeholders need to be represented to ensure transparency, improve understanding of decisions made, as well as share strategic and operational processes. Regarding deficient and maladjusted data governance processes, they generally lead to systems with dysfunctional cultures and ineffectiveness [33].

An important component of implementing policies for data governance has to do with knowing how to coordinate participants and organizations that work together to achieve common goals [34]. Therefore, governance is critical to the overall development of health systems and understanding it is important to improve the performance of health systems [35].

B. Data Quality

Data quality is an important area of concern for practitioners and researchers, applying the terms data quality and information quality interchangeably to obtain relevant content [36].

For Kahn et al. [37] data quality terms, methods, and practices can establish a common understanding of the strengths and limitations of Electronic Health Record (EHR) data for operational analysis, quality improvement, and research. Data quality may be adequate when used for one task but may not be adequate for use in another. For each task, a set of data quality measures must be developed to determine whether the data are suitable for performing the task [38].

In the health data quality literature, various terminologies and definitions attempt to organize measures of data quality, but there is no consensus on what these measures should be [39].

For Khatri et al. [31], in general, data quality refers to the suitability to serve its purpose in each context. In the storage sense, data quality often involves multiple dimensions such as data accuracy and integrity. In most cases, these dimensions need to be defined in the context of data usage.

Brown et al. [40] state that as large-scale clinical databases based on EHR become established repositories of electronic health data, consistent methods for describing, evaluating and reporting based on data quality results, end up being a tool to assess the impact of data quality on its reuse and interpretation of results.

EHRs are collected by individuals with a wide range of experience and at different levels. As a result, EHRs are rarely subject to rigorous data quality assessment. If data quality issues with cross-site EHRs can be identified and resolved using a standard approach, this rich source of information can increase the efficiency and reduce the costs of observational studies and pragmatic clinical trials [41].

As reported by Cai et al. [42]), due to the influence of the high volume of information, it is difficult to assess the quality of the data in a reasonable period. So, collecting, cleaning,

ISBN: 978-989-33-3436-2

integrating and finally getting the high-quality data needed within a reasonable time frame becomes an arduous task. This is, therefore, a major challenge for existing data processing quality techniques.

Typically, data quality standards are developed from the perspective of those who produce them. In the recent past, data consumers were either direct or indirect data producers, which guaranteed the quality of information. However, in the age of big data, with the diversity of information sources, users are not necessarily producers. Therefore, it is very difficult to measure quality, and there may be a hierarchical pattern of quality from the perspective of users.

Thus, while effective data quality assessment is necessary to accurately assess the impact of health interventions [43], it is not sufficient to consistently and comprehensively assess their use in contexts such as proposed for the EHDS.

V. Conclusions

The legal challenges for the implementation of the EHDS revolve around the disparate interpretations of the GDPR among EU member states and how personal private data such as healthcare data can be subject to an use other than its primary purpose. In order to achieve this, several authors [44], [45] point to the use of visual control mechanisms where each person can define the how and when data can be reused and the contexts where they can be used.

The technological challenge for the use of health data is the quality of the data, that is, how it can be sorted, selected and presented in an automated way with the quality and rigour that is imposed in the health area. Aligned with the legal challenges, the technological solution requires the development of data quality measures that guarantee that data can be shared and governance measures that enable the trust of the patients to share and re-use their personal data.

Acknowledgements

This work was supported by the INNHOSPITAL project. INNHOSPITAL: towards a new hospital model based on innovation and knowledge transfer; project code 0610_INNHOSPITAL_6_E; and financed by INTERREG V A Spain Portugal (POCTEP) program.

References

[1] European Commission, "EC Communication COM(2020) 66 final on 'A European strategy for data,'" European Commission, 2020. Accessed: Aug. 23, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN

[2] European Commission, "State of Health in the EU: report 2019," 2019. Accessed: Aug. 23, 2021. [Online]. Available: https://ec.europa.eu/health/state/companion_report_en

[3] J. Lies, "Marketing Intelligence and Big Data: Digital Marketing Techniques on their Way to Becoming Social Engineering Techniques in Marketing," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, p. 134, 2019, doi: 10.9781/ijimai.2019.05.002.

[4] Organisation for Economic Co-operation and Development, "Data-driven innovation for growth and well-being - OECD," 2017.

[5] P. de Camargo Fiorini, B. M. Roman Pais Seles, C. J. Chiappetta Jabbour, E. Barberio Mariano, and A. B. L. de Sousa Jabbour, "Management theory and big data literature: From a review to a research agenda," *International Journal of Information Management*, vol. 43. Elsevier Ltd, pp. 112–129, Dec. 2018. doi: 10.1016/j.ijinfomgt.2018.07.005.

[6] S. Akter, R. Bandara, U. Hani, S. Fosso Wamba, C. Foropon, and T. Papadopoulos, "Analytics-based decision-making for service systems: A qualitative study and agenda for future research," *International Journal of Information Management*, vol. 48, pp. 85–95, Oct. 2019, doi: 10.1016/j.ijinfomgt.2019.01.020.

[7] I. Graef, "Technology Regulation Paving the Way Forward for Data Governance : a Story of," pp. 24–28, 2020.

[8] European Commission, "COMMUNICATION FROM THE COMMISSION: On effective, accessible and resilient health systems," 2014.

[9] European Commission, "European Health Data Space," 2020.

[10] R. Pinto and V. Renda, "Recommendations on health data-processing," *DigitalEurope*, Brussels, Apr. 2020.

[11] European Parliament, "REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da," Brussels, Apr. 2016.

[12] X. Larrucea, M. Moffie, S. Asaf, and I. Santamaria, "Towards a GDPR compliant way to secure European cross border Healthcare Industry 4.0," *Computer Standards and Interfaces*, vol. 69, p. 103408, Mar. 2020, doi: 10.1016/j.csi.2019.103408.

[13] J. N. Doetsch *et al.*, "Record linkage of population-based cohort data from minors with national register data: a scoping review and comparative legal analysis of four European countries," *Open Research Europe*, vol. 1, p. 58, May 2021, doi: 10.12688/openreseurope.13689.1.

[14] L. A. Sprague, G. P. Oelsner, and D. M. Argue, "Challenges with secondary use of multi-source water-quality data in the United States," *Water Research*, vol. 110, pp. 252–261, Mar. 2017, doi: 10.1016/j.watres.2016.12.024.

[15] DigitalEurope, "Common European Health Data Space Including recommendations on the Improving lives and managing diseases through a data-driven EU healthcare system," 2020.

[16] A. Geissbuhler *et al.*, "Trustworthy reuse of health data: A transnational perspective," *International Journal of Medical Informatics*, vol. 82, no. 1, pp. 1–9, Jan. 2013, doi: 10.1016/j.ijmedinf.2012.11.003.

[17] H. Kondylakis *et al.*, "Donor's support tool: Enabling informed secondary use of patient's biomaterial and personal data," *International Journal of Medical Informatics*, vol. 97, pp. 282–292, Jan. 2017, doi: 10.1016/j.ijmedinf.2016.10.019.

[18] D. Dupont *et al.*, "Business analysis for a sustainable, multi-stakeholder ecosystem for leveraging the Electronic Health Records for Clinical Research (EHR4CR) platform in Europe," *International Journal of Medical Informatics*, vol.

ISBN: 978-989-33-3436-2

97, pp. 341–352, Jan. 2017, doi: 10.1016/j.ijmedinf.2016.11.003.
[19] P. McCormack *et al.*, "'You should at least ask'. the expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research," *European Journal of Human Genetics*, vol. 24, no. 10, pp. 1403–1408, Oct. 2016, doi: 10.1038/ejhg.2016.30.
[20] K. el Emam, S. Rodgers, and B. Malin, "Anonymising and sharing individual patient data," *BMJ (Online)*, vol. 350, Mar. 2015, doi: 10.1136/bmj.h1139.
[21] T. Glenn and S. Monteith, "Privacy in the Digital World: Medical and Health Data Outside of HIPAA Protections," *PSYCHIATRY IN THE DIGITAL AGE*, 2014, doi: 10.1007/s11920-014-0494-4.
[22] H. Hyppönen *et al.*, "Impacts of structuring the electronic health record: A systematic review protocol and results of previous reviews," *International Journal of Medical Informatics*, vol. 83, no. 3. Elsevier, pp. 159–169, Mar. 2014. doi: 10.1016/j.ijmedinf.2013.11.006.
[23] N. G. Weiskopf, G. Hripcsak, S. Swaminathan, and C. Weng, "Defining and measuring completeness of electronic health records for secondary use," *Journal of Biomedical Informatics*, vol. 46, no. 5, pp. 830–836, Oct. 2013, doi: 10.1016/j.jbi.2013.06.010.
[24] I. Danciu *et al.*, "Secondary use of clinical data: The Vanderbilt approach," *Journal of Biomedical Informatics*, vol. 52, pp. 28–35, Dec. 2014, doi: 10.1016/j.jbi.2014.02.003.
[25] S. Abhyankar, D. Demner-Fushman, and C. J. McDonald, "Standardizing clinical laboratory data for secondary use," *Journal of Biomedical Informatics*, vol. 45, no. 4, pp. 642–650, Aug. 2012, doi: 10.1016/j.jbi.2012.04.012.
[26] G. Schneider, "Health Data Pools under European Policy and Data Protection Law: Research as a New Efficiency Defence?," *JIPITEC*, vol. 11, no. 1, pp. 49–67, Apr. 2020, Accessed: Sep. 15, 2021. [Online]. Available: https://www.jipitec.eu/issues/jipitec-11-1-2020/5082
[27] European Commission, "COMMITTEE AND THE COMMITTEE OF THE REGIONS Commission Work Programme 2021 A Union of vitality in a world of fragility," Brussels, Oct. 2020.
[28] Ehaction, "Towards the European Health Data Space workshop - eHealth Action," 2020.
[29] T. Gallgher, K. Dube, and S. McLachlan, "Ethical Issues in Secondary Use of Personal Health Information – IEEE Future Directions," *IEEE Future Directions*, May 2018.
[30] M. N. Kooper, R. Maes, and E. E. O. R. Lindgreen, "On the governance of information: Introducing a new concept of governance to support the management of information," *International Journal of Information Management*, vol. 31, no. 3, pp. 195–200, Jun. 2011, doi: 10.1016/j.ijinfomgt.2010.05.009.
[31] V. Khatri and C. v. Brown, "Designing data governance," *Communications of the ACM*, vol. 53, no. 1, pp. 148–152, Jan. 2010, doi: 10.1145/1629175.1629210.
[32] X. Fu, A. Wojak, D. Neagu, M. Ridley, and T. Kim, "Data governance in predictive toxicology: A review," *Journal of Cheminformatics*, vol. 3, no. 7. BioMed Central, p. 24, Jul. 2011. doi: 10.1186/1758-2946-3-24.
[33] L. Delaney, "The challenges of an integrated governance process in healthcare," *Clinical Governance*, vol. 20, no. 2. Emerald Group Publishing Ltd., pp. 74–81, Jan. 2015. doi: 10.1108/CGIJ-02-2015-0005.
[34] P. A. Campos and M. R. Reich, "Political Analysis for Health Policy Implementation," *Health Systems & Reform*, vol. 5, no. 3, pp. 224–235, Jul. 2019, doi: 10.1080/23288604.2019.1625251.
[35] N. Upadhaya *et al.*, "Current situations and future directions for mental health system governance in Nepal: Findings from a qualitative study," *International Journal of Mental Health Systems*, vol. 11, no. 1, p. 37, Jun. 2017, doi: 10.1186/s13033-017-0145-3.
[36] S. Watts, G. Shankaranarayanan, and A. Even, "Data quality assessment in context: A cognitive perspective," *Decision Support Systems*, vol. 48, no. 1, pp. 202–211, Jan. 2009, doi: 10.1016/j.dss.2009.07.012.
[37] M. G. Kahn *et al.*, "A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data," *eGEMs*, vol. 4, no. 1, p. 18, Sep. 2016, doi: 10.13063/2327-9214.1244.
[38] S. G. Johnson, S. Speedie, G. Simon, V. Kumar, and B. L. Westra, "A Data Quality Ontology for the Secondary Use of EHR Data," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2015, pp. 1937–1946, 2015.
[39] S. T. Liaw *et al.*, "Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature," *International Journal of Medical Informatics*, vol. 82, no. 1. Int J Med Inform, pp. 10–24, Jan. 2013. doi: 10.1016/j.ijmedinf.2012.10.001.
[40] J. S. Brown, M. Kahn, and D. Toh, "Data quality assessment for comparative effectiveness research in distributed data networks," *Medical Care*, vol. 51, no. 8 SUPPL.3, Aug. 2013, doi: 10.1097/MLR.0b013e31829b1e2c.
[41] M. G. Kahn, M. A. Raebel, J. M. Glanz, K. Riedlinger, and J. F. Steiner, "A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research," *Medical Care*, vol. 50, no. SUPPL. 1, Jul. 2012, doi: 10.1097/MLR.0b013e318257dd67.
[42] L. Cai and Y. Zhu, "The Challenges of Data Quality and Data Quality Assessment in the Big Data Era," *Data Science Journal*, vol. 14, no. 0, p. 2, May 2015, doi: 10.5334/dsj-2015-002.
[43] H. Chen, D. Hailey, N. Wang, and P. Yu, "A Review of Data Quality Assessment Methods for Public Health Information Systems," *International Journal of Environmental Research and Public Health*, vol. 11, no. 5, pp. 5170–5207, May 2014, doi: 10.3390/ijerph110505170.
[44] G. Schneider and G. Schneider, "Health Data Pools under European Policy and Data Protection Law: Research as a New Efficiency Defence?," *JIPITEC*, vol. 11, no. 1, Apr. 2020, Accessed: Sep. 15, 2021. [Online]. Available: https://www.jipitec.eu/issues/jipitec-11-1-2020/5082
[45] H. Pettenpohl, D. Tebernum, and B. Otto, "WFDU-net: A workflow notation for sovereign data exchange," *Proceedings of the 10th International Conference on Data Science, Technology and Applications, DATA 2021*, pp. 231–240, 2021, doi: 10.5220/0010550402310240.

ISBN: 978-989-33-3436-2