Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240431

Sustainability in Secondary Use of Health Data - A Scoping Review

Tim KILGUS[a,1], Alessia NOWAK[b], Martin GERSCH[a] and Daniel FÜRSTENAU[a,b]
[a] *Freie Universität Berlin, School of Business & Economics, Germany*
[b] *Institute of Medical Informatics, Charité - Universitätsmedizin, Germany*
ORCiD ID: Tim Kilgus 0009-0002-2908-8272, Alessia Nowak 0009-0009-5194-5585, Martin Gersch 0000-0002-3468-0097, Daniel Fürstenau 0000-0001-8490-7707

Abstract. This scoping review investigates sustainability in the reuse of health data on a technological, intra-organizational, inter-organizational, and regulatory level. Thereby, it focuses on the evolutionary, relational, and durational perspective of sustainability. The study highlights various challenges in achieving data sustainability, from regulatory norms such as FAIR principles towards data governance processes and responsibilities in organizations that facilitate data sharing. By highlighting the need for economic sustainability of health data sharing platforms and adapted principles for data sharing, this study aims to analyze current practices that aim for sustainability in the secondary use of health data.

Keywords. Data sustainability, Data Governance, Secondary Use, Data Reuse

1. Introduction

Within the last two decades, data have become a valuable asset for researchers to enhance healthcare services. In particular, the secondary use of health data has gained significant traction [1]. Secondary use of health data is defined as data being reused for a purpose that differs from the intention for which the data were collected [2]. Although it is seen as important, 97% of generated data in hospitals are not utilized for secondary use [3].

In March 2024, the Council of the European Union and the European Parliament reached a provisional agreement on the regulatory proposal for European Health Data Space (EHDS). Its objective is to foster research, innovation, policy-making and regulatory activities through secondary use of health data [4,5].

Besides regulatory initiatives, there are normative rules, such as the FAIR-principles that aim to increase the potential of initially collected scholarly data through reuse. The principles proclaim that data should be findable, accessible, interoperable, and reusable [6]. While those principles have been adopted by academia, industry, funding agencies, and publishers, those principles remain normative and many challenges persist. To make data FAIR whilst preserving them over time, the principles are sufficient, missing, e.g., sustainable governance, and organizational frameworks [7].

An analysis of 35 data sharing platforms, looking at factors such as sharing mechanisms, data network, and presence of project management features indicates that

---

[1] Corresponding Author: Tim Kilgus, Department of Information Systems, Freie Universität Berlin; E-Mail: t.kilgus@fu-berlin.de.

data is findable and accessible. However, according to Guillot and colleagues [8], those platforms do not provide appropriate standards to ensure interoperability. Although regulatory initiatives and FAIR principles contribute to the secondary use of health data, i.e., data management practices are required, too. To investigate the regulatory initiatives, normative rules and practices involved in a sustainable secondary use of health data, we adapt the concept of data sustainability for analysis.

Data sustainability is defined as 'the capacity of data to endure across technological and human generations. This capacity includes data's potentiality for transcending technological and social arrangements and facilitating knowledge advances beyond current questions.' [9, p. 2] This definition is focusing on three perspectives. The evolutionary perspective highlights challenges to data sustainability caused through technological change. The relational perspective underlines the domain-specificity of data and challenges regarding shifting context. Finally, the durational perspective elaborates the required work associated with achieving value from past data in the present and the future. However, it remains unclear what regulatory initiatives, normative rules, and practices contribute to data sustainability. Therefore, we seek to address the following research question: *How is data sustainability implemented through regulatory initiatives, normative rules, processes, and practices?*

2. Methods

This scoping review aimed to overview key concepts of sustainability in health data reuse. We followed Arksey & O'Malley's method [10] and PRISMA guidelines [11]. The five steps are: identifying the research question, relevant studies, selecting studies, charting data, and summarizing results. See online appendix for detailed methods and the study selection process[2].

To identify relevant studies, our search strategy covered three electronic databases, EBSCOhost Academic Ultimate, PubMed/Medline and Web of Science. Following a preliminary search focusing on the search term 'data sustainability', we decided on the following search terms and relevant synonyms in Title/Abstract/Keywords: (1) sustainability AND, (2) data reuse OR data governance AND, (3) health. The complete search string and the PRISMA flow diagram is provided in the online appendix. Only studies in English published after 2014 were included. The search resulted in 497 studies, from which 163 duplicates were removed. The 334 resulting articles were considered for abstract-screening. For the initial abstract-screening, the inclusion criteria were defined twofold. The studies have to focus on the secondary use of health data and the healthcare domain. Two authors conducted the abstract-screening and subsequent full text reading. To ensure reliability of the application of the inclusion criteria, 10% (34/334) of the studies were initially discussed to ensure common understanding. In the second round, 66 articles were read full-text, focusing, applying the same inclusion criteria with an additional focus on the aspect of data sustainability. From those articles, 20 were included for analysis. In the following section, the results of the analysis are presented.

---

[2] https://osf.io/scnzm/?view_only=00aeeed36ff54b0fa9244c80006e816e

3. Results

The 20 articles included in the analysis, discussed data sustainability on different levels. While seven articles focused on the technological level, eleven discussed on intra-organizational level, two on inter-organizational, and two on regulatory level. The assignment of the articles to the levels is provided in the online appendix. Two of the articles discussed more than one level. In the following, the results are discussed in the light of the three sustainability perspectives provided by Jarvenpaa and Essén [9].

From a **technological level** perspective, the identified articles highlight the need for interoperability to achieve sustainability of data. Although not explicitly mentioning data sustainability, the articles highlight the need of interoperability standards to achieve evolutionary and relational sustainability. The four identified articles focus on different aspects, such as comparison of different interoperability standards [12], the development of organizational standards for EHR [13], data models including metadata for study and project specific information [14], and even standard-based architectures for Data Integration Centers [15]. The evolutionary perspective is included by extensively focusing on metadata to increase available information and machine-readability that can be potentially useful in changing technological regimes. This also enables the use of data in shifting contexts, which therefore includes the relational perspective.

On an **intra-organizational level**, there are several processes and practices involved that contribute to the sustainability of data. Several studies focus on data governance [16-18], data management and sharing plans [14-16, 21] data stewardship [19], data access rules [14-16, 20], and organizational structures required [21]. These mechanisms address the temporal relevance and future viability of data sustainability, which is embodied in a durational perspective. This could include long-term data storage and management plans that are independent of individuals and organizations but also consider special requirements of authorized end-users [16, 20]. This supplements the level with a domain-specific perspective shaped by specific contextual factors and covers the relational perspective of sustainability. Although those articles highlight the complex processes in place to contribute to data sustainability, Jarvenpaa & Essèn [9] highlighted that data governance is often implicitly considering sustainability by suggesting roles and procedures to ensure that data will sustain as an asset in an organization. This could be demanding, considering that several studies highlight challenges for sustainable funding and limited resources [16, 18, 19, 22]. Devriendt et al. [21] propose mixed model funding instead of project-based funding to resolve this challenge, although this solution might have downsides, too.

On an **inter-organizational level**, understood as the cooperation between diverse actors, several normative rules have emerged in recent years. In particular, the FAIR-principles [6] facilitated the **F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability of data that were highlighted in several identified articles [14, 21]. Besides that, Coelho et al. [20] highlight the benefits of adapting the TRUST principles to incorporate a durational perspective. Lin et al. [7] developed those principles to actively preserve data to changing technological and stakeholder-specific requirements. With **T**ransparency, **R**esponsibility, **U**ser focus, **S**ustainability, and **T**echnology, those principles provide actionable recommendations for organizations to ensure sustainability from all three perspectives. From the 12 included studies after 2020 considering FAIR-principles, only one study incorporated TRUST-principles [20]. Although this sample is not representative for the adoption of the TRUST-principles in the scientific community, it highlights the potential for further considerations regarding the durational perspective of sustainability.

On the **regulatory level,** only two included articles focused on data sustainability. The studies elaborated the 21st Century Cures Act, that supports, as part of a more general promotion of developments in the US healthcare sector, the creation of a robust ecosystem of initiatives that facilitate data sharing and the challenges of establishing and sustaining so called information commons [22].

4. Discussion and Conclusions

The provided overview on data sustainability in the secondary use of health data highlights the need to deepen understanding on each of the presented levels.

On the technological level, the durational perspective is not incorporated, considering that this requires organizational processes. This is particularly relevant for data access and the role of data access committees. On the organizational level, further research should be conducted on data governance independent from focal organizations towards e.g., community-based approaches [16]. On the intra-organizational level, financial sustainability for data sharing platforms should be ensured. While Devriendt et al. [21] argue that mixed-models, other models suggest the funding from philanthropic organizations, private-public partnerships, or sufficient public funding [18]. The inter-organizational level is characterized through normative rules, with the dominant paradigm of FAIR principles. Although Lin et al. [7] highlight the need to extend those principles to incorporate the durational perspective, the adoption of TRUST principles is insufficient and further investigations regarding incentives to adopt and the operationalization of those principles should be facilitated. Furthermore, the literature is yet to acknowledge the consequences of the newly established regulatory frameworks, such as the European Health Data Space [23] or the Data Management and Sharing Policy of the National Institute of Health (NIH) in the US [24]. Questions arise, how e.g. the policy of the NIH could contribute to economic sustainability of data sharing platforms. While the EHDS defines usage areas, prohibited areas, legitimate purposes, data categories, and intellectual property, aspects of data sustainability remain limited to FAIR-principles [23].

The results show data sustainability is addressed in healthcare research, but challenges remain. Governance should focus on organizational sustainability, while economic sustainability needs inter-organizational and regulatory support. Data sharing platforms often lack long-term funding, highlighting the need for mixed funding models, public-private partnerships, and public funding.

Complementing the FAIR principles with the TRUST principles can provide a more comprehensive approach to data sustainability, ensuring governance frameworks are both technically and ethically robust. While this review focuses on technological, organizational, and regulatory aspects, it is important to note that patient needs and perspectives are often considered in data sharing practices. Ethical and patient advisory boards help address concerns about privacy, data security, and consent. Principles like the TRUST principles emphasize transparency and user focus, aligning with patient expectations. Future research should explicitly include patient perspectives to align data governance with regulatory requirements and patient values, enhancing sustainability and acceptance of health data reuse. Addressing these challenges can improve the effectiveness and ethical soundness of data reuse.

References

[1] Riley M, Robinson K, Kilkenny MF, Leggat SG. The suitability of government health information assets for secondary use in research: A fit-for-purpose analysis. Health Information Management Journal. 2023;52(3):157-166. doi:10.1177/18333583221078377

[2] Näher AF, Vorisek CN, Klopfenstein SAI, et al. Secondary data for global health digitalisation. The Lancet Digital Health. 2023;5(2):e93-e101. doi:10.1016/S2589-7500(22)00195-9

[3] World Economic Forum. 4 Ways Data Is Improving Healthcare. December 2019. Accessed April 4, 2024. https://www.weforum.org/agenda/2019/12/four-ways-data-is-improving-healthcare/.

[4] Council of the EU. European Health Data Space: Council and Parliament strike deal. https://www.consilium.europa.eu/en/press/press-releases/2024/03/15/european-health-data-space-council-and-parliament-strike-provisional-deal/. Accessed April 4, 2024.

[5] Ferretti F. A Single European Data Space and Data Act for the Digital Single Market: On Datafication and the Viability of a PSD2-Like Access Regime for the Platform Economy. Eur J Legal Stud. 2022;14(1):173–218. doi: 10.2924/EJLS.2022.015

[6] Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3(1):160018. doi:10.1038/sdata.2016.18

[7] Lin D, Crabtree J, Dillo I, et al. The TRUST Principles for digital repositories. Scientific Data. 2020;7(1). doi:10.1038/s41597-020-0486-7

[8] Guillot P, Bøgsted M, Vesteghem C. FAIR sharing of health data: a systematic review of applicable solutions. Health Technol. 2023;13(6):869-882. doi:10.1007/s12553-023-00789-5.

[9] Jarvenpaa SL, Essén A. Data sustainability: Data governance in data infrastructures across technological and human generations. Information and Organization. 2023;33(1). doi:10.1016/j.infoandorg.2023.100449

[10] Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology: Theory and Practice. 2005;8(1):19-32. doi:10.1080/1364557032000119616

[11] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ. 2021;372. doi:10.1136/bmj.n71

[12] Klopfenstein SAI, Sass J, Vorisek CN, et al. Bringing Communities Together: Mapping the Investigation-Study-Assay-Model (ISA) to Fast Healthcare Interoperability Resources (FHIR). Studies in health technology and informatics. 2024;310:18-22. doi:10.3233/SHTI230919

[13] Collins SA, Gesner E, Morgan S, et al. A Practical Approach to Governance and Optimization of Structured Data Elements. Studies in health technology and informatics. 2015;216:7-11.

[14] Welter D, Rocca-Serra P, Groues V, et al. The Translational Data Catalog-discoverable biomedical datasets. SCIENTIFIC DATA. 2023;10(1). doi:10.1038/s41597-023-02258-0

[15] Winter A, Stäubert S, Ammon D, et al. Smart Medical Information Technology for Healthcare (SMITH). Methods of information in medicine. 2018;57(S 01):e92-e105. doi:10.3414/ME18-02-0004

[16] Tiffin N, George A, LeFevre AE. How to use relevant data for maximal benefit with minimal risk: digital health data governance to protect vulnerable populations in low-income and middle-income countries. BMJ GLOBAL HEALTH. 2019;4(2). doi:10.1136/bmjgh-2019-001395

[17] Plana A, Furner B, Palese M, et al. Pediatric Cancer Data Commons: Federating and Democratizing Data for Childhood Cancer Research. JCO CLINICAL CANCER INFORMATICS. 2021;5:1034-1043. doi:10.1200/CCI.21.00075

[18] Labkoff SE, Quintana Y, Rozenblit L. Identifying the capabilities for creating next-generation registries: a guide for data leaders and a case for "registry science." JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION. Published online 2024. doi:10.1093/jamia/ocae024

[19] Robinson JO, Daoud A, Geary J, et al. Policy options to facilitate cancer genomic variant data sharing: outcomes of a modified policy Delphi. JOURNAL OF LAW AND THE BIOSCIENCES. 2023;10(2). doi:10.1093/jlb/lsad022

[20] Coelho F, Câmara DCP, Araújo EC, et al. A Platform for Data-Centric, Continuous Epidemiological Analyses (EpiGraphHub): Descriptive Analysis. JOURNAL OF MEDICAL INTERNET RESEARCH. 2023;25. doi:10.2196/40554

[21] Devriendt T, Ammann C, W Asselbergs F, et al. An agenda-setting paper on data sharing platforms: euCanSHare workshop. Open research Europe. 2021;1:80. doi:10.12688/openreseurope.13860.2

[22] Majumder MA, Guerrini CJ, Bollinger JM, Cook-Deegan R, McGuire AL. Sharing data under the 21st Century Cures Act. GENETICS IN MEDICINE. 2017;19(12):1289-1294. doi:10.1038/gim.2017.59

[23] Council of the European Union Proposal for a Regulation on the European Health Data Space - Analysis of the final compromise text with a view to agreement. Brussels; March 18, 2024.

[24] Ross JS, Waldstreicher J, Krumholz HM. Data Sharing — A New Era for Research Funded by the U.S. Government. New England Journal of Medicine. 2023;389(26):2408-2410. doi:10.1056/nejmp2308792