

Overcoming ethical and legal obstacles to data linkage in health research: stakeholder perspectives

Julie-Anne R. Smit^{1,*}, Rieke van der Graaf¹, Menno Mostert¹, Ilonca Vaartjes¹, Mira Zuidgeest¹, Diederik E. Grobbee¹, and Johannes J.M. van Delden¹

Submission History

Submitted:	03/04/2023
Accepted:	14/07/2023
Published:	25/09/2023

¹Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Abstract

Introduction

Data linkage for health research purposes enables the answering of countless new research questions, is said to be cost effective and less intrusive than other means of data collection. Nevertheless, health researchers are currently dealing with a complicated, fragmented, and inconsistent regulatory landscape with regard to the processing of data, and progress in health research is hindered.

Aim

We designed a qualitative study to assess what different stakeholders perceive as ethical and legal obstacles to data linkage for health research purposes, and how these obstacles could be overcome.

Methods

Two focus groups and eighteen semi-structured in-depth interviews were held to collect opinions and insights of various stakeholders. An inductive thematic analysis approach was used to identify overarching themes.

Results

This study showed that the ambiguity regarding the 'correct' interpretation of the law, the fragmentation of policies governing the processing of personal health data, and the demandingness of legal requirements are experienced as causes for the impediment of data linkage for research purposes by the participating stakeholders. To remove or reduce these obstacles authoritative interpretations of the laws and regulations governing data linkage should be issued. The participants furthermore encouraged the harmonisation of data linkage policies, as well as promoting trust and transparency and the enhancement of technical and organisational measures. Lastly, there is a demand for legislative and regulatory modifications amongst the participants.

Conclusions

To overcome the obstacles in data linkage for scientific research purposes, perhaps we should shift the focus from adapting the current laws and regulations governing data linkage, or even designing completely new laws, towards creating a more thorough understanding of the law and making better use of the flexibilities within the existing legislation. Important steps in achieving this shift could be clarification of the legal provisions governing data linkage by issuing authoritative interpretations, as well as the strengthening of ethical-legal oversight bodies.

Keywords

health research; scientific research; data linkage; data processing; GDPR; ethical, legal, stakeholder perspectives; obstacles; solutions

*Corresponding Author:

Email Address: J.A.R.Smit-14@umcutrecht.nl (Julie-Anne R. Smit)

Introduction

In recent years, using, re-using, linking, sharing, and analysing of health and genomic data has occurred on an unprecedented scale, and is expected to only keep expanding [1–3]. With this surge of data use, the processing of (health) data for scientific research purposes has been extensively debated in academic literature [4–6]. Among other things, optimizing the use of data poses challenges on an ethical and legal level, and raises questions and concerns regarding privacy and data protection [7–11].

Data linkage is a technique that allows for the establishing of links between data from different sources relating to, for instance, the same person, family, place or event, and bringing them together in a single file [2, 12]. In health research, data linkage can be used for e.g. the merging of routine care data with census data, administrative data and/or health insurance data. Maximizing the use of existing data collections can be done effectively by using data linkage [8]. Some of the benefits of data linkage are its cost effectiveness, its whole population reach, its avoidance of bias, its timeliness, and its possibility to make use of real-world data [13, 14]. It furthermore enables researchers to ask and answer new research questions that cannot be answered with a single dataset [15]. It has been said that data linkage could be instrumental in the development of policy and research design, as well as be of substantial significance in the medical, epidemiological, and economic field [16].

In Europe, linkage of personal data is governed by the provisions laid down in the General Data Protection Regulation (GDPR). Data linkage can be subsumed under the broader term data processing, as defined in Article 4(2) of the GDPR. In the Netherlands, the implementation of the GDPR has been further specified and elaborated in The Dutch GDPR Execution Act (UAVG). Other national provisions on (health) data processing can be found in the Dutch Medical Treatment Contracts Act.

The GDPR strives to '*harmonise the protection of fundamental rights and freedoms of natural persons in respect of processing activities and to ensure the free flow of personal data* (Recital 3 GDPR).' However, a recent assessment by the European Commission on the Member States' rules on health data in light of the GDPR, showed that the margin of appreciation granted to the Member States has caused fragmentation in data protection legislation, standards and guidelines throughout the EU [17]. Scholars have pointed out that researchers are currently maneuvering a complicated and inconsistent regulatory landscape with regard to the processing of data, which hinders progress in health research [11, 18]. It has been argued that the legal and ethical frameworks governing data processing are contradictory [19]. Furthermore, the lack of interoperability between policies and processes in various countries and institutions, which became especially apparent during the COVID-19 pandemic, has been pointed out [19–23].

We designed a qualitative study to assess what different stakeholders perceive as ethical and legal obstacles to data linkage for health research purposes, and how these obstacles could be overcome. Firstly, because enriching the normative claims from literature with detailed information on practical experiences can be of great value. Gathering a variety of

perspectives of different stakeholders allows us to examine whether the claims that are being made in the academic literature coincide with practical experiences of those involved in linking datasets on a daily basis. Secondly, because a lot has been written about the obstacles in data linkage, but relatively little assessment has been made regarding (practical) solutions. This qualitative study warrants exploring possible solutions for the flagged obstacles, since those with practical experience are well suited for pointing out where there is room for, or a need for improvement.

Methods

We used an inductive thematic analysis for our qualitative study. Opinions and insights of various stakeholders on this topic were collected through focus groups and in-depth interviews. The methods and reporting of this study follow the Consolidated Criteria for Reporting Qualitative Research (COREQ) [24].

Sample

A purposeful sample was selected comprising of relevant stakeholders in the field of health data linkage. To capture a wide range of perspectives, stakeholders were selected based on their variation in background and involvement in dealing with the linking of datasets for health research purposes. The following areas of expertise or backgrounds were represented among the stakeholders: healthcare providers, scientists, legal counsels, (healthcare) industrialists, data providers and (healthcare) policymakers (for specific participant characteristics see Appendices 1, 2). Prospective candidates were invited to take part in the study via e-mail. E-mail addresses were obtained through public websites and via research team or consortium members. Ultimately, two focus groups consisting of six and respectively seven participants were held, and eighteen semi-structured in-depth interviews were conducted. All participants were of Dutch nationality.

Data collection

The focus groups and interviews were conducted through digital video meetings using MS Teams. This mode of convening was chosen because of Covid-19 and the related governmental restrictions. The focus groups were ninety minutes long and took place in February 2021. The duration of the interviews was approximately one hour, and they took place between May and November 2021. All meetings were recorded and transcribed verbatim. The transcriptions were pseudonymized to protect the identity of the participants. After transcription all recordings were deleted. Informed consent has been obtained from all participants. Apart from the researchers and the participants no other persons were present.

This type of study does not fall under the scope of the Dutch Medical Research Involving Human Subjects Act (WMO) and therefore did not require approval from an accredited ethics committee in the Netherlands [25]. An independent quality check was carried out to ensure compliance with legislation and regulations (including

informed consent procedure, data management, privacy aspects and legal aspects).

The focus groups were moderated by RG and MM (both part of the research team and experienced researchers), the interviews were conducted by JS (also part of the research team and trained in qualitative research). A topic list was compiled for the focus groups. The topic list included scenarios that reflect common practical situations in which researchers encounter obstacles when linking different (medical) datasets (see Appendix 3). These scenarios were developed by the research team and were instrumental in identifying the perceived obstacles. The study's consortium and thinktank members (formation of a thinktank with experts in the field of data linkage was part of the study design), were consulted during the development of the scenarios.

During the focus groups, opinions on the recognisability of the scenarios were collected first. Subsequently, the participants were asked what obstacles they encounter in situations as reflected in the scenarios. Lastly, the participants were asked what they envision as possible solutions to these perceived obstacles. After analyzing the transcripts of the focus groups, the topic list was slightly adjusted for the interviews. Upon reaching saturation on the most important obstacles, the focus of the remaining interviews shifted towards exploring possible solutions for the identified obstacles.

Data analysis

The research team consisted of seven researchers (JS, G, MM, MZ, IV, DG and JD) that were involved in the process of analyzing the collected data. All transcripts were coded using NVivo12 qualitative data analysis software. An inductive thematic analysis approach was used to identify different overarching themes arising from the transcriptions. Quotes regarding perceived obstacles and possible solutions for the linkage of (medical) datasets were retrieved. Each quote was assigned one or more codes.

Five researchers were involved in the primary data analysis process (JS, RG, MM, MZ and IV). One member of the research team (JS) coded all transcriptions, four other researchers (RG, MM, IV and MZ) each coded one-fourth of the transcriptions to check the codes for consistency. During the process of analysis, through constant comparison and through discussion within the research team, the code tree was adjusted. Consecutively, to reach agreement on the interpretation of the data and findings, deliberations were held with JS, RG, MM, DG and JD. After reaching consensus on the coding, the data analysis resulted in the identification of the themes mentioned below.

Results

Perceived obstacles

Analysis of the focus group and interview transcripts resulted in the identification of the main themes of perceived ethical and legal obstacles to data linkage for health research purposes:

- I. The prevailing ambiguity regarding the interpretation of the law;
- II. The fragmentation of policies governing the processing of health data;
- III. Demandingness of legal compliance.

The prevailing ambiguity regarding the interpretation of the law

The obstacle most referred to by the participants is the prevailing lack of clarity about what is, and what is not, allowed within the regulatory framework governing the processing of health data. Multiple participants recognized that the lack of clarity on how to interpret certain legislative provisions and open norms results in diverging interpretations of the law. According to them, the way of interpreting the legal provisions laid down in the GDPR or national legislation may range from very strict to very broad. The interpretation can be influenced by different factors, such as the sensitivity of the data or the interests of the person that has to interpret the legal provisions or principles.

Participants gave multiple examples of legal provisions or principles they perceive as unclear. For instance, they spoke of ambiguity about the principle of purpose limitation. This principle can be found in Article 5 of the GDPR, which states that personal data shall be '*collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.*' Participants stated that discussions are being held among researchers about the purpose of clinical data collection. That they debate whether the purpose of certain types of data collection should be labelled as healthcare improvement or scientific research, or perhaps for a dual purpose. Other examples were uncertainty about how to obtain valid consent, and confusion about the interpretation of the GDPR's research exemption in Article 9(2)(j) GDPR. The Dutch GDPR Execution Act (UAVG) states that when invoking this research exemption, data processors should be able to demonstrate that obtaining informed consent is '*impossible or no longer realistically attainable.*' Participants indicated that in practice it is unclear how to interpret this requirement correctly.

This ambiguity regarding the interpretation of certain legal requirements is perceived by different parties. Multiple respondents, legal councils as well as other stakeholders, felt that since the introduction of the GDPR legal councils appear to be more afraid of making mistakes, and that their adherence to the law tends to be more cautious and stricter than before the introduction. Risk aversion was flagged as one of the causes that hinder data linkage for scientific research purposes. Some referred to the fear of reputational damage, others felt that the fear of being fined by the Dutch Data Protection Authority contributes to restrictive interpretations of the law.

Also, scientists indicated that they lack (necessary) legal knowledge. It was stated that because there is so much uncertainty about the 'right' way of interpreting the applicable legal provisions, scientists have developed workarounds that might not be in line with the law but are currently being adopted as standing practice.

The fragmentation of policies governing the processing of personal health data

The second theme relates to the policies that institutions have adopted for the processing of data and with it, the linkage of datasets. The main concern appears to be that different institutions and facilities have adopted different systems and procedures, resulting in a lack of coherence.

It was indicated by multiple participants that this fragmentation of policies seems to stem from the ambiguity regarding the interpretation of the legislation and its open norms. Differences in interpretation lead to differences in implementation. For example, because of the lack of clarity of how to obtain valid consent, some institutions or studies require specific consent, some have implemented opt-out procedures and others ask for broad consent.

Some participants see the importance that healthcare providers attribute to their autonomy as the cause for the fragmentation. Accordingly, healthcare providers want to be able to design their own policies. It was furthermore suggested that the diversity in data use and re-use policies is caused by a lack of mutual trust between institutions and therefore, each institution feels that they are best suited to develop their own policies.

Multiple respondents felt that an explanation for the fragmentation can be found in the decentralisation of the healthcare system in the Netherlands, which has been taking place since 2015. Because of this decentralisation, some of the healthcare tasks previously attributed to the central government have been taken over by local governments and private organisations. The feeling that there is a lack of centralised control and guidance regarding policies governing the processing of personal health data for the purpose of scientific research was broadly shared amongst the participants. One of the respondents, working as a Data Protection Officer in a healthcare institution, described the current situation as "different institutions all sitting on their own closed data silos, with barriers obstructing exchange."

According to some interviewees, what attributes to the fragmentation is that the policies on the processing of personal health data for research purposes are being developed by people who do not have any practical experience in the field of scientific research; *"(..) things are set in motion from an administrative perspective, from a managerial or an IT initiative, not from the functional need (..)."*

Demandingness of legal requirements

Many of the study participants perceive certain specific legal provisions governing data linkage for research purposes as an obstacle. The current processing of health data for research purposes is predominantly governed by the GDPR, the UAVG and the provisions of the Medical Treatment Contracts Act. With regard to the latter, one of the participants with a legal background argued that the provisions of the Medical Treatment Contracts Act are outdated: *"We have this abundance of data that we can do wonderful things with, and how stupid is it that we're being held back by a law dating from 1995, a time when people weren't even thinking about the possibility of using data for scientific research."* Regarding the GDPR views differed; according to

some of the interviewees the GDPR limits the possibilities of the processing of personal health data for research purposes. According to other interviewees, the GDPR provides ample room for scientific research, but they feel that the GDPR's research exemption has been implemented too strictly into our national law.

The requirement of obtaining informed consent was mentioned by the majority of the participants as an obstacle, not just by the participants requesting data but also by data providers. Multiple participants referred to the situation in which a new research question arises which can be answered by the linking of different datasets, but they are not able to use the data because consent for this type of data processing has not been obtained at the time of the data collection. Several participants said that obtaining consent in a way that is in line with the GDPR is incredibly difficult and requires a lot of effort to implement. It was also stated that there is too much emphasis put on consent by the provisions in the GDPR. It was suggested that we should attribute less importance to obtaining consent, and that there are less burdensome safeguards that will be more effective in achieving the aims of the GDPR.

The legal requirements of data minimization, purpose limitation, lawfulness, transparency, and confidentiality were also flagged as obstacles, mainly with regard to the re-use of data from past studies or data collections that started before the GDPR came into effect. One of the participants stated that some of these legal requirements were not properly taken into account in the past, which nowadays results in many uncertainties on handling earlier collected and stored data. Additionally, several participants raised the issue of the high costs that are associated with legal compliance. Moreover, it was stated by some that healthcare institutions have other priorities and do not want the minor profits that they make to be spent on scientific research.

Suggested solutions

The analysis of the focus group and interview transcripts resulted in the identification of five categories of suggested solutions for the perceived obstacles in data linkage:

- I. Issuing authoritative interpretations of the law;
- II. Harmonisation, collaboration, and communication;
- III. Promoting trust and transparency;
- IV. Enhancing technical and organisational measures;
- V. Legislative and regulatory modifications.

Issuing authoritative interpretations of the law

Many participants felt that issuing authoritative interpretations of the law would contribute to the removal of some of the obstacles that currently hinder data linkage for research purposes. It was suggested that providing clear-cut interpretations and guidance of the legal framework could assist in solving the ambiguity regarding the interpretation of the law and the fragmentation of policies on the processing of personal health data. Multiple respondents felt that such interpretations should be endorsed by authoritative bodies

such as, for instance, the Dutch Data Protection Authority. Others felt that producing authoritative legal interpretations is a task for the medical research field itself.

Codes of conduct, practical guidelines, professional summaries, and scientific papers were suggested as the appropriate means of conveying authoritative interpretations of the law. Multiple respondents indicated the need for accessible interpretations of the law in order to support scientists in their undertakings; *“(..) not complicated legal texts, but simply a practical guide indicating what is allowed and what is not allowed. That way at least, as a scientist, you know what to do without having to read or understand a legal text.”* It was also indicated that it would be helpful if the authoritative legal interpretations would be illustrated by accompanying practical examples.

Harmonisation, collaboration, and communication

The harmonising and standardising of data linkage policies was mentioned as a solution by several respondents. Multiple participants suggested the development of a ‘default method’ for the processing of data so that individuals and institutions can apply it similarly. Applying the same principles in data processing, for example by using the FAIR (Findable, Accessible, Interoperable and Reusable) data principles, and applying the same interpretation of the regulatory requirements could diminish the perceived obstacles.

Interrelated, optimising or enhancing communication was mentioned as a part of this solution. Participants did not solely point at the communication between different institutions, they also referred to institutions’ internal modes of communication and the correspondence between different departments and/or different specialties. It was suggested that, for instance, the CMIO (Chief Medical Information Officer) could assist in increasing communication inside institutions as well as between institutions.

Furthermore, several participants opted for the establishment of a (national) body for oversight and compliance with the policies governing data sharing and linkage. A suggestion was to install a centralized Research Ethics Committee (REC) or Data Access Committee (DAC), performing the task of reviewing data access requests. These types of bodies can independently assess whether the ethical and/or legal requirements for data linkage have been met.

Promoting trust and transparency

Many respondents stated that promoting trust is key for successful and optimal data linkage and that currently there is a lack of trust between institutions engaged in data linkage. Participants referred to the need for the enhancement of trust between institutions, but also amongst the professionals processing (health) data.

It was furthermore indicated that patients and other data subjects highly value transparency and that transparency, in turn, can promote trust in the institutions and professionals processing (health) data. According to several participants, institutions should be transparent about their research policies; what type of research is being performed, what type of data is being used, what is being done to protect the data and what the consequences of data linkage processes are for patients.

This can be done, for example, by the publication of scientific research outcomes on the institutions website, distributing informative flyers, and providing information about hospital research policies to patients in person.

Enhancing technical and organisational measures

The enhancement of technical measures was referred to by several participants as a solution for the removal of some of the perceived obstacles: *“We’re going to have to go a little further with the technology. Maybe we need to de-identify and encrypt more, to make sure that you maintain those safeguards.”* Participants differed in their opinions of the most secure way to link data. Multiple participants referred to working with Trusted Third Parties (TTP’s), for instance with the Dutch Central Bureau of Statistics (CBS). One of the participants spoke about establishing an independent data-linkage authority, with a super secured cloud in which linkages could be performed. A different participant expressed the possibility of the creation of a central data warehouse.

Suggested organisational measures are the formulation of institutional protocols and/or standards regarding data linkage. Furthermore, multiple respondents stressed the importance of well trained and qualified personnel: *“If you are going to exchange data as an organisation you must also appoint a person who has a good understanding of the process, someone who thinks along, and who has the time to do a follow-up. It’s not just about throwing data over the fence, there is also aftercare involved.”* It was suggested that well-trained Privacy Officers with specialised knowledge of the use and re-use of (health) data for research purposes could be an asset for institutions engaged in data linkage. One of the respondents suggested that it could be worthwhile for institutions to offer schooling to their researchers and to give them the possibility to enhance their knowledge about data processing and the associated legislation.

Legislative and regulatory modifications

The last category regards the adaptation of the laws and regulations currently governing the processing of personal health data for research purposes. Multiple participants called on the government to take the lead in this matter, since leaving it up to the field is only attributing to the fragmentation. *“There are places in the world, Denmark, Sweden, and England for instance, where they’ve done some very good things [regarding scientific research with health data ed.]. And really, there’s just one explanation for it: explicit government interference.”*

Those participants that felt that the current legal and regulatory frameworks are a hindrance to data linkage and are obstructing scientific research, opted for modifications of these frameworks. One of the participants stated that it would be helpful if the applicable norms in the Medical Treatment Contracts Act would be revised and clarified, in a way that it clearly indicates the permitted legal bases for the processing of health data for scientific research purposes. A different respondent called for modification of the Health Insurance Act; broadening of the lawful bases for the processing of insurance data would make data linkage easier and could assist in fulfilling health insurers legal duty of improving the quality of care.

Discussion

This qualitative study showed that the ambiguity regarding the 'correct' interpretation of the law, the fragmentation of policies governing the processing of personal health data, and the demandingness of legal requirements are experienced as causes for the impediment of data linkage for research purposes by the participating stakeholders. We also found that in order to remove or reduce these obstacles, according to the participants, authoritative interpretations of the laws and regulations governing data linkage should be issued. They furthermore encouraged the harmonisation of data linkage policies, as well as promoting trust and transparency and the enhancement of technical and organisational measures. Lastly, the study showed that there is a demand for legislative and regulatory modifications amongst the participants.

Many of the participating stakeholders designated the open norms that are incorporated in the GDPR as the cause for the struggle with the current regulatory data linkage landscape. The European legislator decided to leave some of the norms in the GDPR 'open', providing Member States with discretionary powers for the implementation of certain specific provisions. These open norms mainly concern provisions that the Member States were unable to reach agreement on during the drafting stage of the GDPR, such as the research exemption. However, by solely focusing on the fragmentation and confusion that these open norms bring, the advantages are being overlooked. The fast-evolving field of data intensive health research requires a certain degree of flexibility for the law to be congruent with the continuous changes and advancements in the field. Without open norms, we would be left with rigid and static laws, unable to adapt to the developments in the field [26]. So perhaps the true problem lies not such much in the open norms, as in the failure to understand and make proper use of flexibilities within the law that allow data linkage [27].

With regards to the obstacle 'demandingness of legal requirements' a similar premise arises. The demandingness of legal requirements and the amount of leeway provided by law for data linkage was a cause for debate and the participants of our study did not agree on this topic. Multiple participants identified the current legal provisions as a demanding obstacle. Notable is that most of these participants have a background in science, not in law. At the same time, a large part of the legally educated participants felt that there is ample room for data linkage in the laws governing this topic. This leads us to wonder if the law is the actual obstacle, or whether it is a lack of knowledge and understanding of the law by the people engaged in data linkage practices. If we can agree that the law already allows for a lot in terms of data linkage and sharing [27], our focus can shift from enforcing new laws into making proper use of the existing legislation.

A solution that was often mentioned and which could assist in making proper use of the existing legislation is the issuing of authoritative interpretations to clarify the provisions governing data linkage for scientific research purposes. It was suggested that these interpretations could come in the form of codes of conduct, guidelines, statements and/or reports. Ideally, the issuing of authoritative interpretations would be done by the European Data Protection Board [28]. But solely issuing authoritative interpretations will probably not suffice,

although they could indeed help reduce legal uncertainty, it will remain necessary to deal with legal complexity on the level of the initiative [26]. The open norms in the GDPR, including the research exemption, and the context-dependent influences of different cases imply that judgments about the legal and ethical permissibility of linking datasets will always have to be made on a case-by-case basis.

Therefore, we would like to encourage the creation, or enforcement, of bodies that can accommodate and interpret that flexibility. Strengthening the position of ethical-legal oversight bodies, including Data Access Committees (DACs) is a solution mentioned by several participants that we strongly support. If well-educated and equipped, these bodies could play an important role in case-by-case decisions on data linkage requests. Several European countries have already incorporated review by a REC or DAC as a legal requirement into their national legislation [17], indicating that multiple national legislators consider review by those types of bodies as an appropriate measure to ensure legally and ethically sound data processing. Moreover, due to fast evolving technical developments, some of the more traditional safeguards such as obtaining consent and anonymizing data are becoming more and more unattainable, and therefore, other types of safeguards will have to be adapted [29]. Oversight bodies such as DACs could address the GDPR's requirement of adopting organisational measures and safeguards when processing personal data for research purposes [30].

Conclusion

Perhaps the problem lies not so much in the law itself, but in how the research community makes use of it. In order to overcome the obstacles in data linkage for scientific research purposes, maybe we should shift the focus from adapting the current laws and regulations governing data linkage, or even designing completely new laws, towards creating a more thorough understanding of the law and making better use of the flexibilities within the existing legislation. Important steps in achieving this shift could be clarification of the legal provisions governing data linkage by issuing authoritative interpretations, as well as the strengthening of ethical-legal oversight bodies.

Acknowledgments

The funding for this article was provided by Health-Holland Top Sector Life Sciences and Health (LSH).

Conflicts of interest

The authors have no conflicts of interest to declare.

Ethical approval

This type of study does not fall under the scope of the Dutch Medical Research Involving Human Subjects Act (WMO) and therefore did not require approval from an accredited ethics committee in the Netherlands. An independent quality check

has been carried out to ensure compliance with legislation and regulations (including informed consent procedure, data management, privacy aspects and legal aspects).

References

- Meystre SM, Lovis C, Burkle T, Tognola G, Budrionis A, Lehmann CU. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. *Yearb Med Inform.* 2017;26(1):38-52. <https://doi.org/10.15265/IY-2017-007>
- Young A, Flack F. Recent trends in the use of linked data in Australia. *Aust Health Rev.* 2018;42(5):584-90. <https://doi.org/10.1071/AH18014>
- Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The Evolving Use of Electronic Health Records (EHR) for Research. *Semin Radiat Oncol.* 2019;29(4):354-61. <https://doi.org/10.1016/j.semradonc.2019.05.010>
- Holm S, Ploug T. Big Data and Health Research- The Governance Challenges in a Mixed Data Economy. *J Bioeth Inq.* 2017;14(4):515-25. <https://doi.org/10.1007/s11673-017-9810-0>
- Laurie G, Ainsworth J, Cunningham J, Dobbs C, Jones KH, Kalra D, et al. On moving targets and magic bullets: Can the UK lead the way with responsible data linkage for health research? *Int J Med Inform.* 2015;84(11):933-40. <https://doi.org/10.1016/j.ijmedinf.2015.08.011>
- Richter G, Borzikowsky C, Lieb W, Schreiber S, Krawczak M, Buyx A. Patient views on research use of clinical data without consent: Legal, but also acceptable? *Eur J Hum Genet.* 2019;27(6):841-7. <https://doi.org/10.1038/s41431-019-0340-6>
- Malin B, Goodman K, Section Editors for the IYSS. Between Access and Privacy: Challenges in Sharing Health Data. *Yearb Med Inform.* 2018;27(1):55-9. <https://doi.org/10.1055/s-0038-1641216>
- Andrew NE, Sundararajan V, Thrift AG, Kilkenny MF, Katzenellenbogen J, Flack F, et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. *Aust N Z J Public Health.* 2016;40(5):436-42. <https://doi.org/10.1111/1753-6405.12576>
- Bonomi L, Huang Y, Ohno-Machado L. Privacy challenges and research opportunities for genomic data sharing. *Nat Genet.* 2020;52(7):646-54. <https://doi.org/10.1038/s41588-020-0651-0>
- Wang S, Jiang X, Singh S, Marmor R, Bonomi L, Fox D, et al. Genome privacy: challenges, technical approaches to mitigate risk, and ethical considerations in the United States. *Ann N Y Acad Sci.* 2017;1387(1):73-83. <https://doi.org/10.1111/nyas.13259>
- Ballantyne A, Moore A, Bartholomew K, Aagaard N. Points of contention: Qualitative research identifying where researchers and research ethics committees disagree about consent waivers for secondary research with tissue and data. *PLoS One.* 2020;15(8):e0235618. <https://doi.org/10.1371/journal.pone.0235618>
- Hobbs MS, McCall MG. Health statistics and record linkage in Australia. *J Chronic Dis.* 1970;23(5):375-81. [https://doi.org/10.1016/0021-9681\(70\)90020-2](https://doi.org/10.1016/0021-9681(70)90020-2)
- Adams, C., Allen, J., & Flack, F. *Sharing Linked Data for Health Research: Toward Better Decision Making (Cambridge Bioethics and Law)*. Cambridge: Cambridge University Press (2022).
- Jorm L. Routinely collected data as a strategic resource for research: priorities for methods and workforce. *Public Health Res Pract.* 2015;25(4):e2541540. <https://doi.org/10.17061/phrp2541540>
- Boyd JH, Randall SM, Ferrante AM, Bauer JK, McInnery K, Brown AP, et al. Accuracy and completeness of patient pathways—the benefits of national data linkage in Australia. *BMC Health Serv Res.* 2015;15:312. <https://doi.org/10.1186/s12913-015-0981-2>
- Xafis V. The acceptability of conducting data linkage research without obtaining consent: lay people's views and justifications. *BMC Med Ethics.* 2015;16(1):79. <https://doi.org/10.1186/s12910-015-0070-4>
- Hansen, J., Wilson, P., Verhoeven, E., Kroneman, M., Kirwan, M., Verheij, R., Veen, E.B. van. *Assessment of the EU Member States' rules on health data in the light of GDPR*. Brussels: Publications Office of the European Union (2021).
- Vayena E. Value from health data: European opportunity to catalyse progress in digital health. *Lancet.* 2021;397(10275):652-3. [https://doi.org/10.1016/S0140-6736\(21\)00203-8](https://doi.org/10.1016/S0140-6736(21)00203-8)
- Mascalzoni D, Dove ES, Rubinstein Y, Dawkins HJ, Kole A, McCormack P, et al. International Charter of principles for sharing bio-specimens and data. *Eur J Hum Genet.* 2015;23(6):721-8. <https://doi.org/10.1038/ejhg.2014.197>
- Whitley EA, Kanellopoulou N, Kaye J. Consent and research governance in biobanks: evidence from focus groups with medical researchers. *Public Health Genomics.* 2012;15(5):232-42. <https://doi.org/10.1159/000336544>
- Bentzen HB, Castro R, Fears R, Griffin G, Ter Meulen V, Ursin G. Remove obstacles to sharing health data with researchers outside of the European Union. *Nat Med.* 2021;27(8):1329-33. <https://doi.org/10.1038/s41591-021-01460-0>
- Dron L, Kalatharan V, Gupta A, Haggstrom J, Zariffa N, Morris AD, et al. Data capture and sharing in the COVID-19 pandemic: a cause for

- concern. *Lancet Digit Health*. 2022;4(10):e748-e56. [https://doi.org/10.1016/S2589-7500\(22\)00147-9](https://doi.org/10.1016/S2589-7500(22)00147-9)
23. McLennan S, Rachut S, Lange J, Fiske A, Heckmann D, Buyx A. Practices and Attitudes of Bavarian Stakeholders Regarding the Secondary Use of Health Data for Research Purposes During the COVID-19 Pandemic: Qualitative Interview Study. *J Med Internet Res*. 2022;24(6):e38754. <https://doi.org/10.2196/38754>
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349-57. <https://doi.org/10.1093/intqhc/mzm042>
25. Non-WMO Statement by the MREC Utrecht reference number MvdL/mb/20/500564.
26. Mostert M, Koomen BM, van Delden JJM, Bredenoord AL. Privacy in Big Data psychiatric and behavioural research: A multiple-case study. *Int J Law Psychiatry*. 2018;60:40-4. <https://doi.org/10.1016/j.ijlp.2018.07.002>
27. Sethi N, Laurie GT. Delivering proportionate governance in the era of eHealth: Making linkage and privacy work together. *Med Law Int*. 2013;13(2-3):168-204. <https://doi.org/10.1177/0968533213508974>
28. Donnelly M, McDonagh M. Health Research, Consent and the GDPR Exemption. *Eur J Health Law*. 2019;26(2):97-119. <https://doi.org/10.1163/15718093-12262427>
29. Smit JR, Mostert M, van Delden JJM. Protecting Privacy While Optimizing the Use of (Health)Data: The Importance of Measures and Safeguards. *Am J Bioeth*. 2022;22(7):79-81. <https://doi.org/10.1080/15265161.2022.2075973>
30. Shabani M, Borry P. Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation. *Eur J Hum Genet*. 2018;26(2):149-56. <https://doi.org/10.1038/s41431-017-0045-7>



Appendix 1: Stakeholder categories focus group participants

Focusgroup 1

Participant no.	Stakeholder category
1.	Legal council
2.	Policymaker
3.	Scientist
4.	Data provider
5.	Scientist
6.	Scientist

Focusgroup 2

Participant no.	Stakeholder category
1.	Healthcare provider/Scientist
2.	Data provider
3.	Scientist
4.	Healthcare provider/Scientist
5.	Scientist
6.	Scientist
7.	Legal council

Appendix 2: Stakeholder categories in-depth interview participants

Participant no.	Stakeholder category
1.	Scientist
2.	Policy maker
3.	Scientist
4.	Healthcare provider
5.	Scientist
6.	Legal counsel
7.	Legal counsel
8.	Healthcare provider
9.	Legal counsel
10.	Data provider
11.	Healthcare industrialist
12.	Scientist
13.	Healthcare provider/scientist
14.	Legal counsel
15.	Data provider
16.	Scientist
17.	Scientist
18.	Healthcare provider

Appendix 3: Illustrative scenarios

Scenario 1 – Setting up new data linkage studies in a particular research area without consent for secondary processing and linkage

A physician-researcher in a UMC decides to set up a new database with patient-data for scientific studies in the field of cardiovascular disease. Subsequently, the researcher plans to link the collected patient data with data from health insurance companies for both current and future studies in the field of cardiovascular diseases. Consent for the processing of data for specific research questions of current and future studies, as well as consent for linking with external datasets by the patients involved, has not (yet) been obtained.

Scenario 2 – Enriching an existing cohort

As part of a long-term cohort study, a researcher at a UMC has been collecting a variety of patient data for years. With these data she has created a database for research on lung diseases. For the collection and processing of these data, the researcher asked permission from the patients involved at the start of the study. The results generated within the cohort provide new insights and give rise to new research questions. To answer one of these new research questions it is necessary to add health insurance data from Vektis, data from the Julius GP Network and the Dutch Central Statistics Office to the database of the cohort. However, for this new research question and for the linkage with the external databases, no informed consent has been given by the patients when they originally consented to participate in the cohort.

Scenario 3 – New research with existing databases for which consent for reuse was not sought at the time of primary data collection

A researcher from a UMC is setting up a new scientific study. For this study, the researcher wants to use patient data stored in his own patients' medical records. Subsequently, the researcher wants to link these data with data from multiple already existing databases containing health data, health insurance data, and/or demographic data. The researcher wants to be able to establish a relationship between personal characteristics from the health record and the data from the databases. Therefore, working with anonymous data is not an option. Consent for reuse of the patient data for scientific research has not been obtained during the time of collection.

