What are the Ethical Approaches Used by Experts When Sharing Health Data? - An Interview Study

Jennifer Viberg Johansson (✉ jennifer.viberg@crb.uu.se)

Uppsala University

Heidi Beate Bentzen

University of Oslo

Deborah Mascalzoni

Uppsala University

Research Article

Keywords: Information technology, re-using and sharing health related data, experts' views, ethical approaches

Posted Date: July 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-691885/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Health data driven activities have become central in diverse areas (research, AI development, wearables, etc.), and new ethical challenges have arisen with regard to privacy, integrity, and appropriateness of use. To improve data subjects' privacy and security, we aim to identify ethically relevant issues experienced by experts in the data intensive exploitation area while collecting, using, or sharing peoples' health data.

Methods: Twelve experts, who were collecting, using, or sharing health data in different contexts in Sweden were interviewed. We used systematic expert interviews to access experts' specialist knowledge. Thereafter, thematic analysis was used to identify categories and subcategories. The codes targeted ethical issues and approaches reported by the interviewed experts.

Results: The main conceptual categories were 'Consideration of the consequences,' 'Respect for rights,' 'Procedural compliance,' and 'Professional conduct.' The respondents discussed and balanced different ethical approaches through several examples. They were morally sensitive to the problems involved in sharing health data.

Conclusions: These empirical findings suggest a need for practical procedures that make it easier for data collectors and sharers to follow the ethical principles and laws relating to data sharing. We suggest that the time is now ripe to move on from policy discussions to practical technological solutions of the principles.

Background

New data driven technologies are changing our societies, and the extensive and intensive use of health data can lead to positive outcomes in medicine and applications beneficial to our health. The rapid development of artificial intelligence and machine-learning techniques comes with the promises of finding patterns in big data (1) to facilitate new discoveries and save resources. Different actors ranging from technological and pharmaceutical companies, medical research institutions, healthcare and public health authorities are collecting, using —and increasingly sharing—people's individual-level health information (2, 3). The wide use of big data includes a broad range of data coming from applications in smartphones or wearables, medical data from health care, research data, but also social or demographic data that are used for different purposes, as individual level data for health, or merged, aggregated and linked in order to optimize products and services in our society.

However, data sharing and re-use of people's data entail risks and challenges for people's privacy, fair use of data, and justice (4–6). It becomes increasingly difficult to apply old safeguards and oversight in data-intensive contexts where the very notion of consent and anonymization change (7, 8).

In the developing field of medical informatics, how to ensure adequate protection of data subjects and communities while at the same time obtaining the full benefits that new data driven technology brings to

our society is a dilemma. There is a need to develop solutions and common practices for handling people's data. Despite the recent adoption of the EU General Data Protection Regulation (GDPR), and the proposal for an EU Artificial Intelligence Act (9), the procedures for how to govern data use and data-sharing responsibly are not all practically in place yet (10).

Hence, different domains being dependent on public data must maintain public trust for the collection and use of data. There is growing literature about the public's, research participants', and patients' preferences and attitudes for data sharing. Different factors contribute to people's willingness to share data. Factors like the level of identifiability, whether there will be an overview of the sharing practices and the extent to which the participants will be informed are central elements that play an important role for people's privacy concerns (11–15).

The expert processing people's data is the one that needs to reflect on and translate standards into practice. Since privacy concerns are prominently considered in the discourse of cyber governance, we found it pertinent to explore what ethical views and approaches have been experienced by experts in relation to re-use and sharing of health data. It will be valuable to identify the experts' views on the situation of sharing data with different organizations, for different purposes, in order to provide insights to policy implementations and give indications of areas to be improved to live up to various ethical and legal standards (e.g., GDPR).

The experts' ethical approaches are the ethical posture and decision-making framework they express. Identifying the ethical approaches and reasoning behind choices can provide insights for the implementation of new technical and legal solutions for re-use of health data. The aim of this study was to describe what ethical approaches are expressed by experts in relation to data use and sharing.

Method

This study is a secondary analysis of Swedish expert interviews, which ensued from a previous study that examined users of health data's views on governance mechanisms that are and should be in place when sharing health data for new purposes with other actors [ref added after review]. The focus of that analysis was procedural issues, policies and practical experience of sharing health data related to governance mechanisms.

Due to the richness of the material collected in Sweden, and considering that the respondents discussed deeply the ethical dimension of sharing subjects' data, another analysis was performed to scrutinize interests distinct from the original analysis (16).

Design and theoretical framework

Empirical results can be helpful to identify moral and ethical issues as described by experts in the data sharing field, unpacking their beliefs and their attitudes that can directly impact approaches to data use and data sharing. How people's health data should be handled is not only a practical issue, but also a

normative issue. For the ethical discussion, we invoke empirical findings that can improve the context-sensitivity to the issue of how to handle health data (17). The article focuses on the ethical approaches expressed by the experts based on their experience of collecting, using, and sharing health information. We performed systematic expert interviews with the aim to access knowledge and specialized information acquired by an expert in a certain field (18).

Participants

We invited experts in the area of health data use to describe how data driven activities are conducted and managed, and to identify additional issues that might need to be considered in policymaking and practical solutions for secured data intense activities. Twelve experts participated in this study. Our sample comprised three female and nine male respondents. They had different professional and organizational backgrounds; see Table 1.

Table 1
Demographics of the experts in the study

Profession	Organization	Gender
<ul style="list-style-type: none"> • 2 data managers of a bioinformatics infrastructure • 1 medical scientist developing AI based tools • 1 geneticist • 1 neuroscientist • 2 epidemiologists • 1 nephrologist • 1 scientist in applied mathematics • 1 project coordinator • 1 legal adviser to research projects and a hospital • 1 data protection officer 	<ul style="list-style-type: none"> • 6 from universities • 2 from a national center for molecular biosciences with a focus on health and environmental research • 2 from hospitals • 1 from a county council • 1 from a cancer research consortium 	<ul style="list-style-type: none"> • 9 males • 3 females

Data collection

The interviews were performed in September 2019 to February 2020 by the first author (JVJ). First, we performed purposive sampling with criterion sampling, which means that the approached interview candidates meet a particular criterion (19). We consulted contacts and webpages for different areas that collect, store, and transfer health data in Sweden (research groups, research infrastructure, and health care). Thereafter, we used snowball sampling to reach new interview candidates (n = 2) from a specific

area. All of the candidates we approached were positive and eager to participate; they emphasized the pressing interview topic of health data handling. Two planned interviews had to be cancelled due to time constraints of the participants. The interviews lasted 29 to 72 minutes each and were conducted in Swedish, apart from one which was conducted in English. The interviews were conducted in a quiet room, either at the first author's work place or in the expert's office. One interview took place via Zoom due to the expert being abroad. Each interview began by asking about the interviewee's experience of using and sharing health data in their organization (e.g., what type of data do you collect and with whom is it shared, for which purpose, whether the data subject is able to give consent, which digital technique is used, and how do you forecast the future in your field). A semi-structured interview guide with open-ended questions was developed with the broader research team that this study was part of (20, 21).

One pilot interview was conducted with two colleagues to assess whether the questions stimulated reflection. Minor adjustments were made, such as having more of the questions as probing questions, and the order of the questions was changed.

Analysis

The reordered interviews were transcribed verbatim by a professional transcription company. The first author (JVJ) listened to the analysis in their entirety to verify the transcription. After reading the transcripts again, segments for further scrutiny were identified where ethical values and norms were explicitly or implicitly expressed. All data from the twelve interviews were given equal consideration. The material contained 86 such segments in total. Atlas.ti software (22) and Excel were used to assist in the data management. At the next stage, the first and last author (JVJ and DM) continued with a comparison of the segments, in terms of similarities and differences, with the perceptive of ethical approaches. Each segment was coded with an initial code. This was done separately, followed by a joint discussion of our interpretations. Codes that reflected similar concepts were grouped together and sub-categories and categories were formulated together; see Table 2 (23).

Table 2

Example of the analytical process in a study on ethical approaches used by experts collecting and sharing health data.

Interview transcript	Initial coding	Sub-category	Category
...and that the patient has approved and is informed that it is shared and that it is understandable patient information for example. That the patient understands that it is voluntary and that they can say no as well.	Give participants the opportunity to make an autonomous decision	Autonomy	Respect for rights (rights-based approach)
that is what people expect from us.	Act in accordance with what is said	Keep promises	Respect for rights (rights-based approach)
...the risk is that information will come out about people's diseases, which should not be in the public domain.	Protect against bad consequences	Do no harm	Consideration of the consequences (consequentialism)

Ethical considerations

According to the Swedish Ethical Review Act, the study did not meet the need for ethical review as it did not involve special categories of personal data. Nevertheless, all procedures performed in the study were in accordance with the ethical conduct as described by Swedish law (SFS 2003: 460). E-mail invitations were sent to the prospective candidates. The invitation included an overview of the project as well as the expected questions. After the expert approved participation, a time and a calm interview location were agreed upon. We requested oral consent from all participants to record the interviews, transcribe the interviews, and save the transcripts for 10 years. Along with the consent, the participants were informed that they could withdraw from the study at any time with no explanation as to why. Their name was replaced with a code. Only the first author had access to the data. All personal identifiers have been removed so the persons described are not identifiable, and care has also been taken so that the interview subjects cannot be identified through the details of the story.

Results

The results describe the ethical approaches used by experts in collecting, using, and sharing health data. Through the analysis, four main categories and fourteen sub-categories were used to classify the discussions. 1) Consideration of the consequences (consequentialism), 2) Respect for rights (rights-based approach), 3) Having a good procedure (procedural ethics), and 4) Professional conduct (virtue ethics). An overview of the categories and sub-categories is presented in Table 3. In the following, categories and sub-categories will be described and illustrated by quotes.

Table 3. Categories and sub-categories of ethical approaches used by experts related to sharing health data.

Category	Sub-category
1) Consideration of the consequences (consequentialism)	Benefit for society
	Benefit for science
	Do no harm to individuals/non-maleficence
2) Respect for rights (rights-based approach)	Right to a private sphere
	Autonomy
	Freedom
	Human dignity
	Keep promises
	Justice
3) Procedural compliance (procedural ethics)	Effective approach
	Transparent to data subject and society
	Do the right thing by default
4) Professional conduct (virtue ethics)	Be responsible
	Be respectful

Category 1: Consideration of the consequences (consequentialism)

The consequentialist approach was expressed in three different ways: benefit for society, benefit for science, and do no harm to individuals/non-maleficence.

Benefit for society

One perspective of sharing data is that it will benefit society in one way or another: data sharing may help explain the origin of diseases, have utility for the population due to new development of treatments and lives saved. Some participants expressed that the more collaboration there is with academic scientific research and other academic projects or even commercial companies developing new medical devices or drugs, the more benefit there will be to society. A concern was raised that having too strict legal rules does

not benefit society, as it may hinder beneficial research and technical development for society. The respondents pushed for open data sharing in the research community as a means to improve health care down the line.

Another consequentialist argument for the wide re-use of health data was the need to maximize benefits to the society by maximizing data re-use as much as possible. It was argued that people's tax money needs to be used in a manner that is beneficial for the population. And because hospitals, universities, and national consortia in Sweden are financed with tax money, the data collected or generated need to be widely used to maximize the benefits.

[...] data has been produced by tax money for example, then they [tax payers] want them to be used as much as possible. There should be no obstacles to that. So that one [taxpayers] gets as much bang for the buck of data that has been produced. It must then be shared with more researchers and so on. (Respondent 1, data manager, male)

Benefit for science

All participants stressed that sharing data is not only beneficial for research, but also a necessity to answer some research questions. Sharing and re-use of data increase the possibility of making new discoveries. If the legal rules are too strict, the view expressed was that the law will hinder important research.

I understand that there are reasons for laws and stuff like that. But it just seems obvious to me that Facebook should be allowed to do some basic experiments on us. It is not really harming us, and it is often a hindrance I think, being constrained to what you are allowed to do in this way. (Respondent 12, scientist in applied mathematics, male)

Since data collection is expensive, a view is that the research community needs to preserve existing data. Re-use is a way to utilize. Even under this sub-category, two participants expressed that they see no great risk for harm. Rather, their view was that the good consequences for better research outweigh the risk of people being harmed. They could imagine potential bad consequences, but they perceived the risk as being highly unlikely. Some of the participants expressed engagement in having access to data and freedom to perform research. They were triggered by and expressed a great motivation to perform data intensive research tasks.

Another aspect of sharing data responsibly is that science needs a good reputation to maintain trust. It is important to maintain people's confidence in the research community for financial purposes since most research funding comes from taxpayers.

Do no harm to individuals/non-maleficence

The majority of the respondents acknowledged that there are threats to participants' private sphere in the form of bad consequences if people's personal health information falls in the wrong hands. They recognize that data can be lost and end up in the wrong hands; people can be identified, and data can be misused. Someone can sell the data, earn money by blackmailing individuals by threatening to disseminate information about being in a risk group or having a certain disease. Insurance companies are mentioned as interested in this kind of information. There is therefore a risk that data ending up in the wrong hands can give people an economic disadvantage. Some of the participants think this is very hypothetical, but it could happen, and therefore protective measures must be taken to protect participants.

However, two respondents perceived the risk of individuals getting harmed as being so small that it almost does not count. One respondent could not see how anyone could be interested in the participants' health data.

I do not really think that there are very many who are super interested in this data in that way.
(Respondent 7, epidemiologist, male)

People [researchers] are generally very unnecessarily anxious. People sit behind desks and behind paper and are very anxious about... 'what if data comes out and what if I do wrong and I do not know exactly what is right or wrong.' There seems to be a bit of chaos and GDPR has not exactly made things clearer and eased nervousness either. My very personal attitude is that people are a little too anxious for their own health about this. (Respondent 4, project coordinator, male)

Category 2: Respect for rights (rights-based approach)

In this category, respondents described certain rights that need to be respected regardless of the consequences. The respondents described that it may be unpleasant and create a bad feeling among the data subjects that the information is in the hands of others. Under this category, we have also included respondents' views on the researchers' right to perform research in the name of freedom of research.

Right to a private sphere

Participants strongly emphasized people's right to be protected, the right not to be identified without consent. It was viewed as a matter of respecting other people and their integrity. Respondents expressed that if people's health data is spread and comes in the wrong hands, personal integrity is violated. Another opinion strengthening this view was that personal data is not open to anyone: people have the right to control information that relates to them. One respondent thought that personal information known of any random person is not sensitive, but if it is a neighbor or colleague that knows the same information, it makes a difference. Hence, as it is difficult to know what is a person's level of acceptance

to share information and his or her level of vulnerability beforehand, the respondents expressed that actions and level of security need to be such that it meets those of the most vulnerable.

So, in general, I think the risks are quite low, but it is more to protect personal integrity, ... some care very much and do not want to give out their social security number or do not want to have such information everywhere, and you have to respect that. Then, there are others who do not care. (Respondent 6, epidemiologist, female)

Autonomy

The respondents described that people are entitled to decide according to their own wishes; autonomy is the core. It is the right of participants to decide for themselves whether or not health data should be shared, and with whom. Careful attention should be paid to the information stage before collecting data, so that potential participants are aware of the purpose, scientific method, re-use of the data, and expected result before consenting. It is essential for people to know whether the data will be used by others or for other purposes to be able to make an autonomous decision. Some of the respondents viewed that the responsibility for re-use of data lay on the potential participants' agreement. If participants consented to the potential risks and benefits, then it is for a good reason. People's different preferences were also brought up as a reason to leave the responsibility to decide to the individual. Moreover, to give them the opportunity to decide was seen as a way of respecting them as a person, the life they have, and their experiences.

who is responsible for?... I would say the data subject. So, who is responsible for the data that is shared, yes, it is the individual [...] it is the individual who decides for him or herself. Do I want to accept this, that is, or do I want to sign this consent? I want to buy a smart watch. I want a Google Home like this at home. Somewhere consciously or unconsciously, the individual makes a lot of choices in our society. So, it must always start from the individual. [...] Yes and some people glide around in some form of ignorance as well. But I do not see that the government should be allowed to take responsibility for the fact that there happens to be ignorant individuals... (Respondent 4, project coordinator, male)

...and that the patient has approved and is informed that it is shared and that it is understandable patient information for example. That the patient understands that it is voluntary and that they can say no as well. (Respondent 11, nephrologist, male)

Freedom

Moreover, the respondents voiced the importance of people being free to choose whether they want to be part of data collections. Not solely in order to exercise their right to autonomy, but also for the value of living a free life. It was argued that the state should not regulate everything and protect all possible incidents. That was perceived as an undesirable situation for our society.

...it is so dynamic [the technical development] so I feel very skeptical about appointing the state as responsible; I think it is the wrong way to go. One must be able to protect oneself in the first place. I am a little reluctant to leave it to a government to decide what kind of data I may share or what I allow someone else to do with it. (Respondent 4, project coordinator, male)

Two respondents made the opposite declaration that all participants are not autonomous and therefore not free to decide on their own, due to the fact that they are patients and dependent on care (a power imbalance). Therefore, they need the protection of rules, processes, and guidelines that make the collection and use of people's health data safe and rigorous. Thus, they are free because of the rules.

Freedom to perform research was mentioned as an important norm for the research community. That is the basis for scientific research and needs to be protected; too strict and complicated rules hinder free research. Additionally, rules were viewed as potentially limiting the knowledge we can achieve, which goes against the nature of doing research.

Human dignity

Human dignity reflects the inherent worthiness of being a human. Respondents are of the view that respect for human dignity may be less of a concern with the extensive data sets that researchers are working with, since an individual becomes just one in a crowd. One reason can be the social distance between the participant and the user of the data. The importance of remembering that there is a real person behind the numbers was emphasized, and therefore it is important to be careful when collecting, storing, and using data.

I think it is important that the researcher who uses data or those who share data should be aware that there are actually people attached to the information, that you should be very careful about how to use it and how to store it, and there should not exist any names and so on. (Respondent 6, epidemiologist, female)

Keeping promises

Keeping promises and following what has been agreed upon were mentioned in two contexts. In the first context, it was described as if there is an agreement in place; nothing hinders the use of people's health data. If people agree to something, there is no need to question whether that is harmful or wrong. In the second context, participants mentioned the importance of acting according to what has been promised. What has been said should be followed:

[...] that is what people expect from us. (Respondent 8, geneticist, female)

Justice

Another reason for having a high standard of control over participants' health data is that some people will be in a more vulnerable situation. Therefore, justice based on needs should be applied.

... there should be very high demands [on how we treat data]. I think there are some people that are weak and extra vulnerable. (Respondent 7, epidemiologist, male)

In contrast, the idea of justice was also deemed to motivate openness of data in research and care. The argument follows: by having open access to data, it will be possible to achieve a representativeness of all people in the scientific research. In addition, there was a concern that that inequalities, which exist when it comes to people's health, will be maintained due to a lack of representativeness of data. The concern is that in the long run, there will be an unequal distribution of care. A more open sharing of health data will benefit us all, and mostly the underrepresented groups.

And that is a driving force for why it is so damn important that open sharing and equal sharing of data are as broad as possible [...]it is because this group has not been represented in the data on which this algorithm was practiced... (Respondent 10, medical scientist developing AI based tools, male)

On the other hand, having a totally open access to data was viewed as problematic because the people performing the job should be recognized for their accomplishment. A researcher needs to get credit for the work performed before sharing.

You have to get credit for what you have done, but still you can share it with others. (Respondent 3, neuroscientist, female)

Category 3: Procedural compliance (procedural ethics)

Many respondents voiced the need to have a good procedure for good professional behavior. They expressed a will to do the right thing, but they wanted less obstacles that hold back beneficial development and make the administrative work inefficient. They asked for a routine that embedded the right action; it is simple to follow rules and not think all the time about what is right or wrong. Rules that are implemented in a good procedure were perceived as giving freedom to do the research and innovate.

[...]but it has happened that I get emails with social security numbers and addresses to people, it should not happen. I do not want that in my mailbox. (Respondent 6, epidemiologist, female)

As a new researcher, I think you become overwhelmed, and then I think that many start to cheat. Many do not apply [for ethical approval], but they run their race. (Respondent 8, geneticist, female)

However, one respondent emphasized that the rules need to be adaptive to comply with a changing reality (e.g., technological development and new research questions). This can be difficult to foresee; therefore, the need to be transparent was viewed as important. A transparent process is a good process for a

changing world. To be transparent to the community and to the participants is also viewed as a good path for an ethically sustainable research environment.

So, this is a balancing act; it [the regulations and processes] must be alive ... the ethical regulations must be alive so that it can adapt, because if there is new technology that makes it possible to do something we cannot today, we must be able to say, you cannot do it this way, you have to do it this way or protect the individual in some way. (Respondent 8, geneticist, female)

Category 4: Professional conduct (virtue ethics)

Finally, the analysis revealed expressions of professional conduct. It was mostly related to the character that a data collector and user has or should have. One view was that a researcher (data collector) is not interested in individuals' health status, and therefore the risk of something going wrong is negligible.

A researcher needs to be respectful and responsible. That is part of the profession of being a researcher. Another view is that data users can be affected by the interest of discovery and forget or stretch the rules because of being curious.

[...] and then these people start to fumble a bit and hand data over to some company they collaborated with, which they thought was very exciting, and so it starts to slide. (Respondent 4, project coordinator, male)

Discussion

The aim of the study was to describe the ethical approaches used by experts in relation to data use and further sharing of peoples' health data with a new user. The experts in the present study primarily referred to and described norms and values in their practical work, when handling peoples' health data. The category of 'Considering the consequences' and its sub-categories (benefit for society, benefit for science, do no harm to individuals/non-maleficence) and the category of 'Respect for rights' with its sub-categories (right to a private sphere and autonomy) were highlighted by all experts, making up the major share of the interview material. The sub-categories of 'Respect for rights,' including freedom, human dignity, keeping promises, and justice were not as prominent for all respondents. Tagging behind those categories were different kinds of approaches attached to the work of handling data. Firstly, many respondents discussed the need for good procedures regarding good professionally ethical behavior. The need for good procedures was often described with strong emotional expressions (e.g., irritated or frustrated), both from the point of view of the safety of the data subjects, but also to ensure a good working environment. Secondly, the last category of 'Professional conduct' was not mentioned by all respondents, but it was very prominent for a few.

The EU General Data Protection Regulation 2016/679 (GDPR) strengthens individuals' rights regarding peoples personal data (24). The GDPR advocates informational self-determination by increasing

transparency requirements for data collection practices. The respondents in this study were focusing greatly on consent, one of the lawful basis for personal data processing in the GDPR. In accordance with the data subjects' expressed values, they express that participants should be able to decide with whom to share, what type of data, for which purposes, and consequently what type of risk they would be willing to take (15, 25). However, some of the respondents in this study seem to place too much responsibility on the individual data subject, disregarding the power imbalance between the data subject and the data controller. Relying solely on what a data subject consents to provides insufficient protection; consent is necessary but not sufficient, and additional safeguards need to be in place (26). Thus, the GDPR points to further safeguards, for example, encryption and pseudonymization. Further, some respondents express a societal obligation to use the data to the furthest extent possible for the benefit of science and society, particularly where the data generation is publicly funded. We can see a risk that data users place the benefit for science and society above that of the individual's fundamental rights, wanting to maximize taxpayers' financial contribution to benefit the health field. The risk to the individual is by some considered so small that it is negligible and outweighed by the benefits that result from the data processing. This seemingly disregards that in today's data driven society, the risk to the individual is no longer just the potential for physical harm, but also for informational harm.

There were mixed views regarding whether subjects' data can be of interest for other people than the collector and user of the data. Some respondents felt that there is a real threat that data can fall in the wrong hands and thereby harm people. Some of the respondents in this study could not fathom others being interested in the data. They perceived the risk as being so small that it could be ignored. In the academic risk literature, the concept is referred to as the *de minimis* (27). It is also a legal term meaning that insignificant matters do not need to be considered – “the law does not concern itself with trifles” (28). From an ethical point of view, one should not ignore even small risks or treat them very differently from other risks. Lundgren and Stefánsson (29) argue that there “is no probability threshold below which risks can rationally be treated categorically differently from other risks.” It is furthermore well established that subjects' health data can be misused or processed for unintended purposes (30, 31). Moreover, the consequences on an individual level can be large and overrule the small risk. Health data have been compared to the *new oil* of its lucrative potential (32) and therefore might attract fraudsters. Another comparison in healthcare is that health data is the *new blood*. The proposed comparison is that health data are “digital specimens and should be treated with the same rigour, care, and caution afforded to physical medical specimens” (6). Harm is not only physical but also psychological and emotional (33). One ongoing debate is whether one's digital twin should be viewed as an extension of the patient's body (34, 35). Even if the digital twin is only a reduction of the person and cannot be considered as an autonomous person, its integrity should be respected with regard to how data is used, shared, and secured. The reason for this, as mentioned earlier, is that data that are not protected adequately have the possibility to attract fraudsters.

Moreover, the risks increase as the interest in big data increases. Indeed, the interest in accessibility to data has increased extensively over the last few years (36–38). The reason for this increased interest in large amounts of health data is the need, for example, to have training sets for the development of AI

technologies (39, 40). The data economy is blossoming with start-up companies looking to earn money from data based innovations (41). The biggest threat is perhaps not primarily hackers and other criminals, but companies that have commercial profit as their top priority.

However, overprotection can also lead to harm. It is important to protect privacy, but not to the extent that it creates obstacles for big data, machine learning, and learning health systems (6). Advances in information technology are challenging the traditional view of consent and anonymization as the primary safeguards, showing that privacy enhancing technologies can contribute considerably to protecting people's privacy. It is a fine balancing that requires interdisciplinary collaboration. In this context, protecting people's privacy while providing flexible and fast technical solutions is invoked as a challenge. Focusing on FAIR principles of findability, accessibility, interoperability, and reusability point the way forward for facilitating data sharing more systematically (42). However, some methodological and organizational challenges remain. The FAIR principles only call for explication of access conditions, without specifying how data sharing should be facilitated or organized. The principles can provide guidance on how to think about all the "design choices embodied in data; developing new modes for respecting participants' rights; and coming up with robust measures for valuing data sharing which do not reproduce the problems related to current scientific reward systems" (43).

Privacy is traditionally defined as the right to be left alone (44). However, different contexts give rise to different expectations and preferences with regard to privacy (45). Contextual rules about how information flows depend on the actors involved, the accessibility of the data, and the purpose of that access. Nissenbaum describes that privacy is violated when the contextual rules are contravened (45). The response regarding why such violations are wrong can traditionally be divided into two categories – consequentialist and deontological concerns (6). Consequentialist concerns pertain to all possible bad things that can happen, the individual's uncomfortable feeling of being observed or the feeling of losing control, regardless of whether there is an actual threat. Deontological concerns do not depend on experiencing negative consequences. A privacy has been violated even if no one uses a person's information against him or her, or if the person never even becomes aware that a breach has occurred.

In addition to consequentialist and deontological concerns, the results of this study add to procedural ethics (46) as an important component of the potential harm in sharing people's data. Having a robust and harmonized procedure for sharing data does not only protect individuals, but also assures the experts that they are acting ethically and are legally compliant when it comes to the data subjects. Procedural ethics encompasses processes such as developing a protocol and having a framework for ethical behavior. The day-to-day practice needs to be built up in a way, so the data collector and the user follow the rules and the ethical concerns addressed by people and the law. This can be compared to the GDPR rules regarding data protection by design and by default, which aim to build the legal rules into the technology and make the legally compliant and privacy preserving technical option the default (47–49). The notion of ethics by design has sprung out of this, and similarly seeks to embed ethical principles into the technological design and development (50). Everyday procedures set up in a similar fashion, to

ensure legal and ethical compliance by default, ease the administrative burden of the data users and benefit the data subjects.

Ethical and legal issues in medical informatics grow in tandem with advances in digital health and computing technologies. This study combines empirical investigation with ethical reflection. The empirical investigation of ethical approaches used by experts can identify relevant moral issues and describe the experts' beliefs and attitudes that can be relevant for the ethical issue of sharing health data digitally. Empirical ethics can improve the context-sensitivity of the ethical deliberation (17). The results of this study stress that the rules and the ethical concerns need to be more easily incorporated into practice. Lack of a harmonized system and complex processes hinder beneficial technological development and the data subjects' security. Some of the ethical obstacles need to move from ethics to 'law-conceptions.' There is a risk that ethics becomes a sort of replica of law or a softer version of law (51). Therefore, we suggest that the time is now ripe to move forward from policy discussions to practical solutions of the principles. Our participants expressed a desire to do the right thing, but then again, the practical reality sometimes was a hindrance.

Conclusion

These empirical findings suggest a need to have practical procedures that make it easier for data collectors and sharers to follow the ethical principles and laws regarding data sharing. We suggest that the time is now ripe to move forward from policy discussions to practical technological solutions of the principles. Data collectors need better technical and practical conditions to follow the ethical and legal demands. Moreover, to focus on good ethical practice will help the experts to do what they want and are good at – innovate.

Abbreviations

GDPR: General Data Protection Regulation

Declarations

Ethics approval and consent to participate

According to the Swedish Ethical Review Act, the study did not meet the need for ethical review as it did not involve special categories of personal data. Nevertheless, all procedures performed in the study were in accordance with the ethical conduct as described by Swedish law (SFS 2003: 460). E-mail invitations were sent to the prospective candidates. The invitation included an overview of the project as well as the expected questions. After the expert approved participation, a time and a calm interview location were agreed upon. We requested oral consent from all participants to record the interviews, transcribe the interviews, and save the transcripts for 10 years. Along with the consent, the participants were informed that they could withdraw from the study at any time with no explanation as to why. Their name was

replaced with a code. Only the first author had access to the data. All personal identifiers have been removed so the persons described are not identifiable, and care has also been taken so that the interview subjects cannot be identified through the details of the story.

Consent for publication

Not applicable

Availability of data and materials

The datasets in form of interviews, generated and analysed during the current study are not publicly available due that individuals can be identified and consent for to make that available was not obtained. However, it will be partly available from corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by NordForsk [grant number 81105] and the Economic and Social Research Council (part of UK Research and Innovation). The funders had no influence on the design or writing of the article.

Authors' contributions

JVJ, DM and HBB made the conceptualisation of the study. JVJ analysed the interviews. JVJ and DM interpreted the data. JVJ was the major contributor in writing the manuscript. DM and HBB have substantively revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This paper reports on research conducted as part of the 'Governance of Health Data in Cyberspace' project. The project is a joint undertaking between the University of Oxford, the University of Iceland, Uppsala University, and the University of Oslo. In addition to the named authors, the following individuals have contributed to the chapter through group discussions, peer support, and/or document review: Lee Andrew Bygrave, Sarah Coy, Eik Haraldsdóttir, Guðbjörg Andrea Jónsdóttir, Jane Kaye, Salvör Nordal, Nisha Shah, and Jorien Veldwijk. Our greatest gratitude goes to all the experts participating in the interviews. Thank you for your time and for sharing your experiences and ideas so generously.

References

1. Costa FF. Big data in biomedicine. Drug Discovery Today. 2014;19(4):433-40.

2. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, et al. Making big data useful for health care: a summary of the inaugural mit critical data conference. *J JMIR medical informatics*. 2014;2(2).
3. Gadde SS, Kalli VDR. Descriptive Analysis of Machine Learning and Its Application in Healthcare. *International Journal of Computer Science Trends and Technology (IJCST)* 2020;8(2).
4. Geneviève LD, Martani A, Wangmo T, Paolotti D, Koppeschaar C, Kjelsø C, et al. Participatory disease surveillance systems: ethical framework. *J Med Internet Res*. 2019;21(5):e12273.
5. Petersen C, Subbian V. Special Section on Ethics in Health Informatics. *J Yearbook of Medical Informatics*. 2020;29(1):77.
6. Price WN, Cohen IG. Privacy in the age of medical big data. *J Nature medicine*. 2019;25(1):37-43.
7. Mostert M, Bredenoord AL, Biesaat MC, Van Delden JJ. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *J European Journal of Human Genetics*. 2016;24(7):956-60.
8. Mascalzoni D, Bentzen HB, Budin-Ljøsne I, Bygrave LA, Bell J, Dove ES, et al. Are requirements to deposit data in research repositories compatible with the European Union's general data protection regulation? *J Annals of internal medicine*. 2019;170(5):332-4.
9. European Commission. Proposal for a Regulation laying down harmonised rules on artificial intelligence: POLICY AND LEGISLATION; 2021 [Available from: <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence>].
10. Kalkman S, Mostert M, Gerlinger C, Van Delden JJ, Van Thiel GJ. Responsible data sharing in international health research: a systematic review of principles and norms. *J BMC Medical Ethics*. 2019;20(1):1-13.
11. Clayton E, Halverson C, Sathe N, Malin B. A systematic literature review of individuals' perspectives on privacy and genetic information in the United States. *PLoS One*. 2018;13(10):e0204417.
12. Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. *Expert Rev Mol Diagn*. 2014;14(8):1053-65.
13. Aitken M, De St Jorre J, Pagliari C, Jeps
13. Aitken M, De St Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC medical ethics*. 2016;17(1):73.
- n R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. *BMC medical ethics*.

2016;17(1):73.

14. Tully MP, Bernsten C, Aitken M, Vass C. Public preferences regarding data linkage for research: a discrete choice experiment comparing Scotland and Sweden. *Bmc Med Inform Decis*. 2020;20(1):109-.
15. Viberg Johansson J, Bentzen HB, Shah N, Haraldsdóttir E, Guðbjörg AJ, Kaye J, et al. Publics' preferences for sharing health data: a discrete choice experiment. *JMIR Medical Informatics* 2021;in press.
16. Hinds PS, Vogel RJ, Clarke-Steffen L. The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *J Qualitative health research*. 1997;7(3):408-24.
17. Musschenga AW. Empirical ethics, context-sensitivity, and contextualism. *J Med Philos*. 2005;30(5):467-90.
18. Bogner A, Menz W. The theory-generating expert interview: epistemological interest, forms of knowledge, interaction. *Interviewing experts*: Springer; 2009. p. 43-80.
19. Bryman A. *Social research methods*. Fifth ed. Oxford: Oxford University Press; 2016.
20. Kvale S, Brinkmann S. *Interviews: Learning the craft of qualitative research interviewing*: sage; 2009.
21. Krueger RA, Casey MA. *Focus groups: A practical guide for applied research*: Sage publications; 2014.
22. ATLAS.ti 8. The Qualitative Data Analysis & Research Software 2019 [Available from: <https://atlasti.com/>].
23. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analysing and presenting qualitative data. *Br Dent J*. 2008;204(8):429-32.
24. General Data Protection Regulation. General Data Protection Regulation (GDPR) 2016 [Available from: <https://eugdpr.org/>].
25. Rivas Velarde MC, Tsantoulis P, Burton-Jeangros C, Aceti M, Chappuis P, Hurst-Majno S. Citizens' views on sharing their health data: the role of competence, reliability and pursuing the common good. *BMC Medical Ethics*. 2021;22(1):62.
26. Nicol D, Eckstein L, Bentzen HB, Borry P, Burgess M, Burke W, et al. Consent insufficient for data release. 2019;364(6439):445-6.
27. Peterson M. What Is a de Minimis Risk? *Risk Management*. 2002;4(2):47-55.
28. Nemerofsky J. What Is Trifle Anyway. *J Gonz L Rev*. 2001;37:315.
29. Lundgren B, Stefánsson HO. Against the de minimis principle. *J Risk Analysis*. 2020;40(5):908-14.

30. O'Doherty KC, Christofides E, Yen J, Bentzen HB, Burke W, Hallowell N, et al. If you build it, they will come: unintended future uses of organised health data collections. *BMC Medical Ethics*. 2016;17(1):54.
31. Heikkilä M, Cerulus L. Hacker seeks to extort Finnish mental health patients after data breach: POLITICO; 2020 [Available from: <https://www.politico.eu/article/cybercriminal-extorts-finnish-therapy-patients-in-shocking-attack-ransomware-blackmail-vastaamo/>].
32. Parkins D. The world's most valuable resource is no longer oil, but data. *J The economist*. 2017;6.
33. Graeme Laurie, Kerina H. Jones, Leslie' Stevens, Christine Dobbs. A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data: mi mason institute for medicin, life sciences and the law; 2014 [Available from: <https://www.nuffieldbioethics.org/wp-content/uploads/FINAL-Report-on-Harms-Arising-from-Use-of-Health-and-Biomedical-Data-30-JUNE-2014.pdf>].
34. Braun M. Represent me: please! towards an ethics of digital twins in medicine. *J Med Ethics*. 2021;47(6):394-400.
35. Loh J. 'Represent me: please! Towards an ethics of digital twins in medicine': Commentary. *J Journal of Medical Ethics*. 2021;47(6):412-.
36. BBC News. Grindr and Twitter face 'out of control' complaint: BBC News; 2020 [Available from: <https://www.bbc.com/news/technology-51106357>].
37. Wee S-L, Mozur P. China Uses DNA to Map Faces, With Help From the West: *The New York Times*; 2019 [
38. The Guardian. Pentagon awards \$10bn cloud computing deal to Microsoft, snubbing Amazon: *The Guardian*; 2019 [Available from: <https://www.theguardian.com/global/2019/oct/26/pentagon-awards-10bn-cloud-computing-deal-to-microsoft-snubbing-amazon>].
39. Annoni A, Benczur P, Bertoldi P, Delipetrev B, De Prato G, Feijoo C, et al. Artificial Intelligence: A European Perspective. Joint Research Centre (Seville site); 2018.
40. Brierley H. When should private companies be allowed to use NHS data? Take part in a Uni of Manchester opinion study and give your views. Reimbursement offered. More info here: <https://connectedhealthcities.org/research-projects/qstudy/> #HealthDataResearch. 2019.
41. Statista. Leading medical technology companies worldwide based on revenue in 2019 2019 [Available from: <https://www.statista.com/statistics/281544/revenue-of-global-top-medical-technology-companies/>].
42. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *J Scientific data*. 2016;3(1):1-9.

43. Boeckhout M, Zielhuis GA, Bredenoord AL. The FAIR guiding principles for data stewardship: fair enough? *J European journal of human genetics*. 2018;26(7):931-6.
44. Warren SD, Brandeis LD. The Right to Privacy: *Harvard Law Review*; 1890 [downloaded 18 May 1996 from an internet site hosted by Stephen R. Laniel; and reformatted from a site hosted by Lawrence University, 1999. Available from: https://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html.
45. Nissenbaum H. Privacy as contextual integrity. *J Wash L Rev*. 2004;79:119.
46. Guillemin M, Gillam L. Ethics, reflexivity, and “ethically important moments” in research. *J Qualitative inquiry*. 2004;10(2):261-80.
47. Wiese Scharthum D. Making privacy by design operative. *J International Journal of Law Information Technology*. 2016;24(2):151-75.
48. Bygrave LA. Data protection by design and by default: deciphering the EU’s legislative requirements. *J Oslo Law Review*. 2017;4(02):105-20.
49. Bygrave LA. Article 25 data protection by design and by default. *The EU General Data Protection Regulation (GDPR)*: Oxford University Press; 2020.
50. Brey P, Dainow B. Ethics by Design and Ethics of Use in AI and Robotics: SIENNA; 2020 [Available from: https://sienna-project.eu/digitalAssets/915/c_915554-l_1-k_sienna-ethics-by-design-and-ethics-of-use.pdf].
51. Rességuier A, Rodrigues R. AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Sociaty* 2020;7(2):2053951720942541.