International Journal of Population Data Science

Journal Website: www.ijpds.org

Understanding how to build a social licence for using novel linked datasets for planning and research in Kent, Surrey and Sussex: results of deliberative focus groups

Elizabeth Ford[1,*], Melanie Rees-Roberts[2], Kathryn Stanley[1], Katie Goddard[1], Sarah Giles[3], Jo Armes[4], Deborah Ikhile[1], Anotida Madzvamuse[5,6], Victoria Spencer-Hughes[7], Abraham George[8], Chris Farmer[2], and Jackie Cassell[1]

Submission History

Submitted:	16/09/2022
Accepted:	21/11/2022
Published:	30/01/2023

[1]Brighton and Sussex Medical School, Brighton, UK
[2]Centre for Health Services Studies, University of Kent, Canterbury, Kent, UK
[3]Digital Innovation Theme Public Advisor, NIHR ARC -KSS (Applied Research Collaboration Kent, Surrey, and Sussex)
[4]University of Surrey, Guildford, UK
[5]University of Sussex, Brighton, UK
[6]University of British Columbia, Canada
[7]East Sussex County Council, Lewes, Sussex, UK
[8]Kent County Council, Maidstone, Kent, UK

Abstract

Introduction
Digital programmes in the newly created NHS integrated care boards (ICBs) in the United Kingdom mean that curation and linkage of anonymised patient data is underway in many areas for the first time. In Kent, Surrey and Sussex (KSS), in Southeast England, public health teams want to use these datasets to answer strategic population health questions, but public expectations around use of patient data are unknown.

Objectives
We aimed to engage with citizens of KSS to gather their views and expectations of data linkage and re-use, through deliberative discussions.

Methods
We held five 3-hour deliberative focus groups with 79 citizens of KSS, presenting information about potential uses of data, safeguards, and mechanisms for public involvement in governance and decision making about datasets. After each presentation, participants discussed their views in facilitated small groups which were recorded, transcribed and analysed thematically.

Results
The focus groups generated 15 themes representing participants' views on the benefits, risks and values for safeguarding linked data. Participants largely supported use of patient data to improve health service efficiency and resource management, preventative services and out of hospital care, joined-up services and information flows. Most participants expressed concerns about data accuracy, breaches and hacking, and worried about commercial use of data. They suggested that transparency of data usage through audit trails and clear information about accountability, ensuring data re-use does not perpetuate stigma and discrimination, ongoing, inclusive and valued involvement of the public in dataset decision-making, and a commitment to building trust, would meet their expectations for responsible data use.

Conclusions
Participants were largely favourable about the proposed uses of patient linked datasets but expected a commitment to transparency and public involvement. Findings were mapped to previous tenets of social license and can be used to inform ICB digital programme teams on how to proceed with use of linked datasets in a trustworthy and socially acceptable way.

Keywords
data linkage; public engagement; focus groups; patient privacy, social license

*Corresponding Author:
Email Address: E.m.ford@bsms.ac.uk (Elizabeth Ford)

https://doi.org/10.23889/ijpds.v5i3.2114

Introduction

In England, the development of National Health Service (NHS) and Local Authority (LA) "Integrated Care Boards" (ICBs), has resulted in NHS providers and LAs across a geographical area, covering a population of 1-3 million people, coming together to form partnerships so that delivery of health and social care can be joined-up [1]. The main purposes of these ICBs are to improve outcomes in population health, tackle inequalities and enhance productivity and value for money in health and care services [2]. ICBs are mandated to make use of technology and service-user data to improve decision-making, planning and commissioning services [3]. This drive towards regional development of digital and data capabilities [4] has seen increased investment in infrastructure to take delivery of, and link data from LAs, general practitioners (GPs), acute, community and mental health NHS Trusts, adult social care, and other health and care settings. In the Kent, Surrey, and Sussex (KSS) region, funding for the ICB digital programmes has enabled data to be curated in this way for the first time, as previously no such data infrastructure existed.

Located in the Southeast corner of England, and despite general affluence, KSS contains some of the most socially deprived wards in England, mainly located in coastal towns. The region faces complex challenges for health and social care services with significant population growth while resources for service provision remain constrained. The region is characterised by an above-average number of looked-after children, and high levels of young people with emotional and mental health needs. The KSS population contains a greater proportion of older people than average for England and this older population is increasing quickly: growth of the older population (>65s) is four times greater than among younger populations (<65s) as rural and coastal locations are an attractive retirement prospect; this population growth will place significant demands on local health and social care services [5]. GP Practices in Medway, Swale and Thanet on the north east Kent coast, already have the highest ratio of people per GP in England [6]. As a result, health and social care services in KSS are increasingly supporting more frail people with complex health needs.

In the UK, electronic patient data, routinely collected or generated in the context of provision of health and social care, is collected, de-identified and linked together for secondary uses (such as service planning or research) without patient consent under the lawful basis of "task in the public interest" [7]. Using data to "*be aware of and consider what the health needs of its local population are*", is a statutory obligation for LAs [8] and NHS ICBs are expected to "*develop cross-system analytical capability to drive better decision making*" using data from their own services and systems [9]. A tension arises because patients and the public often feel that data generated during their care is theirs, it is about them, and they should have some ownership of it [10], and yet they are often unaware of secondary uses of their data or its governance [11]. Despite this lack of understanding of secondary uses of patient data, research through surveys and focus groups suggests the majority of the population are willing for their data to be used for public benefit purposes as long as patient privacy is preserved [11]. After a failed UK government plan to extract and use all English General Practice patient records for healthcare planning and research (known as care.data [12]) in 2014, the idea of a lack of social license was proposed to explain the scheme's downfall [13].

The concept of social license suggests that the public expect that organisations who manage potentially controversial schemes, for example, holding and using the public's data, will go beyond the requirements of formal regulation, and adhere to voluntary codes of trustworthy behaviour and transparency [14]. According to social licence theory, only when the public are satisfied that the motivations of the organisation are trustworthy, they may confer a "social license" to operate. Schemes which do not seek community involvement or approval, and are therefore non-transparent, will lose public trust, even if they meet legal requirements [15]. Carter et al., [13] proposed that conditions for a social license in health data sharing included reciprocity (two-way communication), non-exploitation and service of the public good. Stockdale et al., [11] in their review of studies on public views of data sharing, additionally fitted core research ethics principles as a lens for understanding public views, suggesting that respect for autonomy, contribution to the public good, fairness and justice, and prevention of harm were core values for health data sharing that were expected and shared by the public [11]. These two frameworks provide core tenets of social license in health data sharing and can be mapped together as shown in Table 1.

Because anonymized and linked datasets are being developed in the KSS region for the first time, there is currently very little local evidence of public awareness of secondary uses of data, or evidence testing "reasonable expectations" of KSS patients and public around health and care data linkage for secondary purposes. Our aim was to initiate and establish a culture of public engagement and involvement in the region to grow alongside the development of the data assets. It was therefore important for us to involve local people in early discussions on data uses, rather than rely on knowledge on public views gained from other regions or national pieces of work. The set-up phase of these datasets in KSS allowed an opportunity to consult with citizens from the very beginning, to understand how a social license might be built, for the use of linked data. Publicly supported and informed governance and data access structures would ensure the questions asked of the data are appropriate, ethical, can be prioritized against public benefit and need, and that projects do not perpetuate stigma, discrimination or inequalities in health. In Kent and Medway, local research suggests that members of the public are enthusiastic about public health research and being involved in this work, to ensure research reflects community priorities [16], although this study did not investigate attitudes to secondary uses of linked health data.

We aimed, therefore, to engage with the citizens of Kent, Surrey and Sussex, to understand public views around integrated dataset use, identify ways of including public expectations in governance models, explore routes to public involvement in prioritisation and decision-making for data use, and gather public suggestions on how the use of data can be communicated transparently.

Methods

Study design

Given that so little is publicly known about the regional use of routinely collected health and administrative data for service planning, evaluation and research, we anticipated that participants would need to be informed and supported to understand the different issues before deliberating on their opinions and views. Therefore, we used a deliberative research design, which is an approach for gathering wider views for health issues when there are many complex issues to weigh against each other [17]. We chose the method of deliberative discussion focus groups outlined by Rothwell et al. [18]. These focus groups include a range of informative presentations, interspersed with facilitated plenary and breakout discussions. According to Rothwell et al., information presented should be balanced, comprehensive and incorporate the benefits and risks of the topic of interest, and the resulting discussion must be inclusive of all group members.

Materials and methods

We conducted five 3-hour focus groups, each covering three topics of information and discussion:

1) What integrated datasets could be used for; and what their potential benefits might be.
2) How patient data is linked, de-identified and protected; and what potential risks and concerns about linked datasets might be.
3) How the public could be involved in oversight and governance of datasets, and how information about datasets and their use could best be communicated with the public.

Each topic consisted of an expert presentation (10–15 minutes) followed by the opportunity for participants to ask questions for clarification. Presentations were written and delivered by authors EF, MRR and JA, who are health and data researchers based in the KSS region. Participants were then split into breakout groups of 4–6 participants and a facilitator followed a question guide. All participants completed an online pre-group demographics questionnaire and post-group feedback questionnaire. Both included a question on willingness to share their medical records to ensure a diverse range of views were represented, this was taken from the Wellcome Trust monitor report wave 3 [19].

Focus groups were conducted remotely using Zoom video-conferencing software and were recorded. Each group consisted of two sessions, one week apart, totalling 3 hours; at the beginning, participants were informed that both benefits and risks of data usage would be explored.

Recruitment and procedure

We advertised our study through a large range of health, social care and support organisations in Kent, Surrey and Sussex via an online flyer, giving a link to a website at which potential participants could express an interest. Expressions of interest were followed up by sending participants a link to a Qualtrics survey site which presented the full participant information sheet and consent form. Once consented, participants were booked into a focus group via email communication and were sent further online reading about issues to be discussed. Participants were sent a £ 50 voucher after completing the focus group.

Addressing digital exclusion

After consultation with public advisors on the materials and methods of our focus group, we recognised the possibility of digital exclusion arising from holding our focus groups online. We tackled this by:

1) Splitting the focus group into two shorter sections (rather than 3 hours in one go).
2) Inviting participants who were not proficient in English to attend with another household member who could help them with English.
3) Holding focus groups at different times of the day and in the evening.
4) Providing a university telephone number that could be called to express an interest for those who did not want to register via the website.
5) Offering to send study documentation and payment vouchers by post, and offering a phone number with which to join the Zoom call
6) Creating a step-by-step user guide for Zoom and offering each participant a pre-group accessibility check via a short Zoom call, to talk them through the functionality of the software.

Data collection and analysis

All focus groups were recorded using Zoom recording features which capture both audio and video recordings. The video was stripped from the recordings and audio files were sent to an external transcribing company for verbatim transcription with removal of all identifiers. Transcripts were then checked for quality and anonymity against the original audio files by the research team before analysis.

Questionnaire data from Qualtrics was aggregated across the groups and descriptive statistics reported. Qualitative

Table 1: Core tenets of a social license for secondary uses of health data

Carter et al., 2015 [13]	**Stockdale et al., 2019 [11]**
Reciprocity	Autonomy Fairness and Justice
Non-exploitation	Prevention of harm
Service of the public good	Contribution to the public good

data was analysed inductively according to 6-step thematic analysis principles outlined by Braun and Clark [20] using NVivo software. First, one researcher (KS) familiarised herself with the data by reading and re-reading transcripts and noting down initial ideas. The same researcher coded the data (KS), highlighted relevant features of the data and assigned them codes, then applied codes in a systematic fashion across the entire dataset. A second researcher (KG) checked the data, and coded additional data within the coding structure. Codes were then collected into potential themes, and themes were then discussed among all members of the research team and themes mapped and adjusted. Checking back against the data, the specifics of each theme were refined, and clear names for each theme were generated by two researchers (KS and EF). Finally, compelling extracts and quotes were selected to illustrate each theme. Because our focus groups were structured in the form of three discussions on separate topics, our themes were organised into three groups: 1) benefits, 2) risks and concerns, and 3) values for safeguarding data and including the public.

Results

Participant characteristics

79 participants took part in the focus groups and demographic characteristics are given in Table 2. Over 65% of the sample was female, with an age range of 19-83 years and a median age of 51, and most participants came from Sussex. Participants were largely in favour of sharing medical records data both before and after the focus groups, and there was slight shift overall to being more willing to share at the end of the focus groups compared to before (Figure 1).

Thematic analysis

The thematic analysis generated 15 themes organised into three groups according to the topics presented and discussed within the focus groups: 1) perceived benefits to data linkage, 2) risks and concerns about data linkage, and 3) values for safeguarding data and including the public. The theme names are depicted in Figure 2.

1) Perceived benefits of data linkage

Efficiency and resource management

Focus group participants were generally positive about data being linked together to improve services, particularly as they envisaged that joined-up data would help to improve the efficiency of services and would help manage resources.

> *"All this dataset analysis and helping to link them together will enable each department to do their job better. All of them are wasting so much time in doing the same thing again or doing the wrong thing because they haven't got the information. This will make everything so much more efficient for everybody." (FG2)*

Improving preventive services and out of hospital care

Participants described times when they had struggled to receive community-based care for long-term conditions and felt that connected data would help to connect services, ultimately keeping people who lived with long-term conditions out of hospital.

> *"I think it's about prevention and being able to put in that early help and connecting the services and using the data to know where that is needed." (FG1)*

They saw connection of data as an additional way of improving the management of health care resource, as keeping people healthier would prove cheaper in the longer term:

> *"Getting older when you're healthier is cheaper for everybody." (FG2)*

Improved communication between services

Participants often discussed the hopes they had that linking data would improve their direct experience of care, with the expectation that data flows for live patient records would also be improved through a data linkage programme. They referenced previous experiences when clinicians had had a lack of information about the patient in front of them and this had reduced the effectiveness of care provided.

> *"The specialist hadn't even read [her notes]. She had to go over everything. Whereas if the data is all linked up and they actually read it before they speak to you, that would save so much time for everybody I would have thought." (FG3)*

Benefits across services in the community

Participants supported the idea that services provided in the public sector would be improved with a better understanding of what other services were doing and with increased data available on service users. They supported the idea that linkage should cross boundaries, out of just linked health data and into other services provided by LAs, for example:

> *"We cannot look at all these different services as being very individual. It might seem a great, vast difference between the housing department and the community mental health service, but actually they do need to be linked together and I think this is a good way, with integration in being able to do that." (FG4)*

2) Risks and concerns about data linkage

Accuracy and missing data

Participants were concerned about whether the data recorded in their health records would be a good representation of the ground truth of their health or illness or lived experience. Participants worried that important data might be missing, and that data would not be accurate enough to plan appropriate services. They acknowledged that data is created by "*humans*" in the course of a clinical encounter, and so could have errors in it.

Table 2: Participant characteristics

Demographic characteristic	N (Total = 79)	%
Sex: Female/ Non-binary*	53	67
Male	26	34
Median Age	51	Range 19–83
Ethnic background:		
Any white background	63	80
Asian, Asian British, Asian Mixed, Indian, Chinese or Pakistani*	11	14
Black British, Black African or Black Caribbean*	5	6
County of residence:		
Sussex	48	61
Surrey	20	25
Kent	11	14
Willingness to share data, before focus groups:		
Very willing	31	39
Fairly willing	32	41
Fairly unwilling	9	11
Very unwilling	6	8
Don't know	1	1
Willingness to share data, after focus groups:		
Very willing	35	44
Fairly willing	30	38
Fairly unwilling	7	9
Very unwilling	0	0
Don't know	7	9
Education level		
GCSE/O-Level (up to age 16)	5	6
A Level (up to age 18)	18	23
University degree	30	38
Postgraduate degree	13	17
Not disclosed	13	17
Employment status		
Full time employed	20	25
Part time employed	11	14
Homemaker, carer	3	4
Retired	14	18
Self-employed	7	9
Student	4	5
Unemployed/Job seeking.	2	3
Other/not disclosed	18	23
Disability		
Yes	17	22
No	46	58
Not disclosed	16	20

*Groups combined to support anonymity.

> *"Missing data would mean that you plan incorrectly, and you have to use everybody's data not just little groups of data." (FG2)*

> *"How do we ensure the cleanliness of the data and make sure that the data is accurate enough to build the right picture?" (FG1)*

> *"Data is only as good as who's inputting it, so you've still got a human, haven't you? And errors do happen." (FG5)*

Data breaches and hacking

Participants were worried that despite controls over the data, and infrastructure to ensure its security, the data might be hacked by unknown adversaries.

> *"I don't trust personally that all of this sharing of information isn't going to get hacked. I don't trust the fact that it's all going to be anonymized, there's human error that's going to come in here" (FG1)*

Figure 1: Willingness of participants to share their medical data for a research study

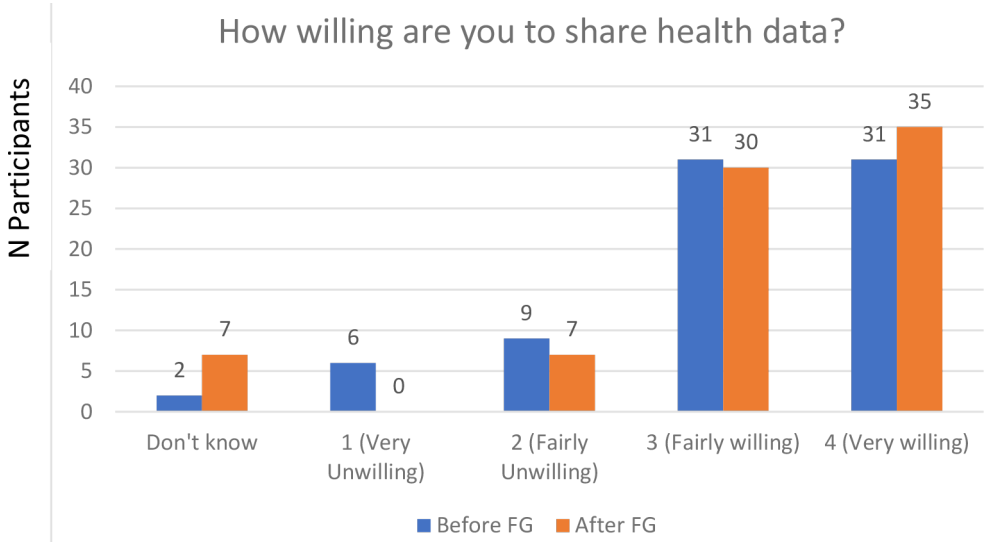
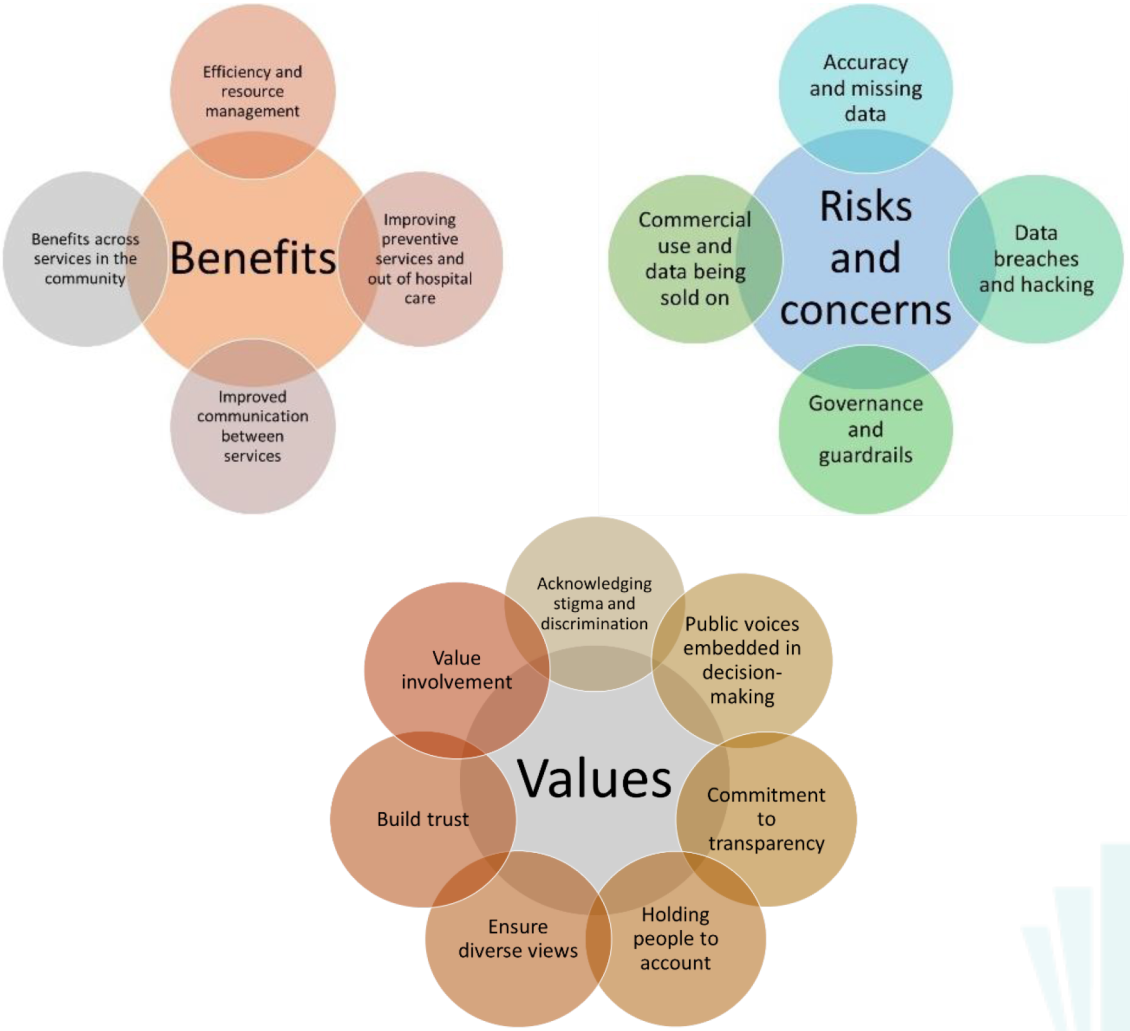


Figure 2: Depiction of themes across the 3 topics covered in the focus groups



Participants wanted to know how appropriate use of the data would be enforced, and whether or how they could trust the systems put in place for data security. To some extent they accepted that guarantees could not be given but were nevertheless concerned that hacking events could occur even with the best privacy systems.

> *"If we're putting forward all these concerns, are they going to be actually addressed? How can*

they address it though, because nobody's going to be able to guarantee there's not going to be a hacking event, nobody's going to be able to guarantee dishonest people, looking at the data? Nobody can guarantee any of those things." (FG1)

"I think one of my concerns would be even when you have an organisation which has been vetted and therefore they've been given access to the data set; it's how you ensure that they ensure that the people who access the data are appropriate, not somebody who thinks I've got five minutes, I'll just have a look." (FG1)

Governance and "guard-rails"

In the information presentation, a diagram was presented showing the proposed governance framework for the Sussex linked dataset, whereby analysts wishing to access the data would need to come from approved organisations, completed a data access request which would be checked against a prescribed set of rules known as "guard-rails" (shown in Figure 3). Analysts would then get to log into a data safe haven to access only the data needed for their approved project. When their analysis was complete, their outputs would again be assessed against the guard-rails and only aggregated outputs would be allowed to leave the safe haven. This system largely met with approval from the participants, with one of them saying:

"*I think most of my concerns were helped by the diagram of how the system would work.*" *(FG1)*

However, participants discussed the need for transparency over what would happen if someone broke the rules set up in the framework, and the importance of transparency over accountability for following the rules.

"It is absolutely important that we have governance, assurance and accountability because when that system is not being used, individuals or the organisation should be able to go to that accountable person and say your systems have failed, you should now do something about it, or you should be held accountable to it. I think that's important." (FG1)

Commercial use and data being sold on

Lastly, participants were worried about the motivations behind the use of the data. They were particularly concerned that if profit-making companies were to gain access to the data, they might "*bury*" any research findings which were not in their favour.

"Researchers are funded by people, organisations. Who they are funded by and what is their ultimate goal? We need to be very careful who has access to these data sets. For example, drug companies fund a huge amount of research, but they will not publish those research results if it's not advantageous to them. They will not advertise it; they will lose it." (FG2)

Participants offered a suggestion, that increased transparency about all projects conducted using the data and their results, might help to increase trust:

"*Maybe we need to add in that when these researchers have access to the data, they must broadcast their results regardless of what the results show." (FG2)*

3) Values for safeguarding data and including the public

Participants were asked what they would like to see in terms of safeguards for the data, and how the public should be involved in oversight and decision-making for using the linked datasets. Participants responded with a number of values that they would like to see dataset custodians and users adhere to, and which might help to engender a social license.

Acknowledging stigma and discrimination

Participants wanted a commitment to acknowledging people's lived experiences of stigma and discrimination and making sure further uses of their data did not perpetuate either.

"I just think in terms of some of the things, like mental health and sexual health, can be a bit sensitive for people and things like whether people are on benefits, crime issues, domestic abuse, lots of that gets linked in already now. It's quite stigmatising and it's very sensitive and they need to bear that in mind if they're linking certain things." (FG1, pt1)

"*Stigma is such a big thing, and it's such an important thing to bear in mind, to have at the forefront of any decisions of moving forward with the research." (FG1, pt3)*

Public voices embedded in decision-making

One of the suggested ways to make sure that data linkage and uses would not perpetuate stigma, was to make sure public voices are heard throughout planned uses of the data. Participants equated involvement of people with lived experience as "*good working practice*" and felt that public involvement would generate more trust in the use of the data.

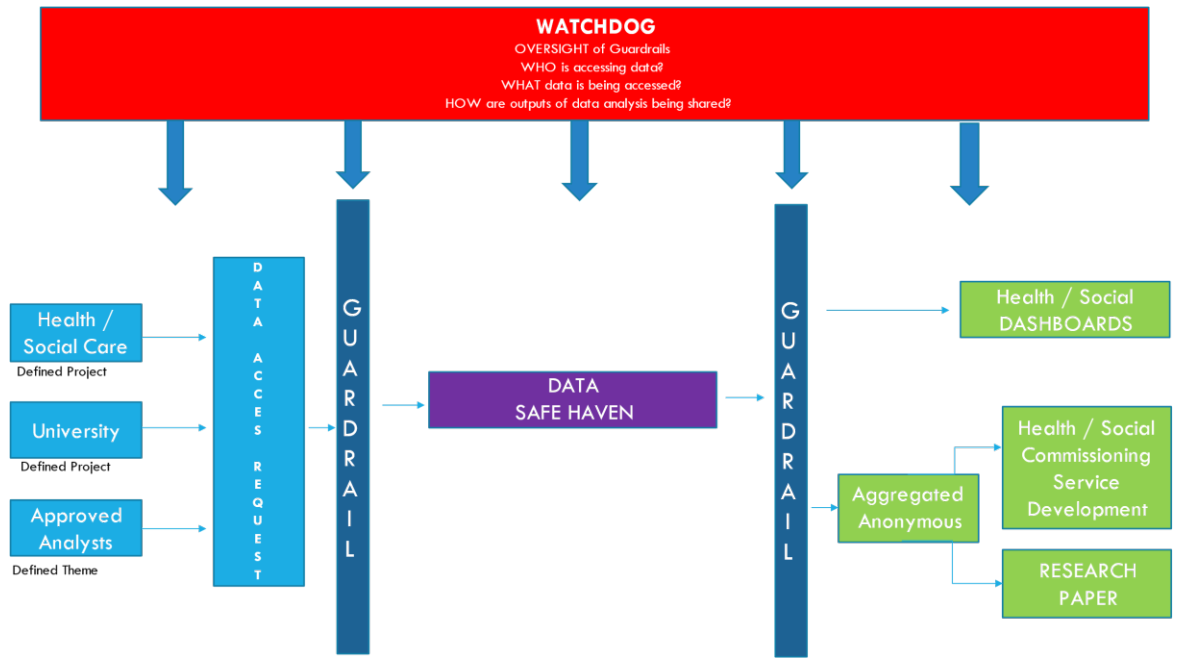
"I think the key to understanding our desired outcomes, how we're going to achieve those, not just through legislation but good working practice: that good working practice is people with lived experience." (FG1)

"I do think the public need to be involved because I also feel that then you've also got more level of trust" (FG3)

Commitment to transparency

Participants wanted a commitment to transparency about who was using the data and for what purposes, and accountability for anyone who went outside their permitted use of the data. One suggestion was that the system within which data was housed and analysed could keep a trail of all uses of the data:

Figure 3: A depiction of how data would be kept safe and accessed by trusted users as proposed by Sussex ICB



"As everyone else has said, it's about who has access to it and what's the data trail on them. If it is a log in, so that you can see. It's a bit like you swipe a key card in places, you can see who's come in and out of the building, who has gone to different rooms at different times." (FG1)

Holding people to account

A theme which followed from the discussions on risk and concerns was that participants wanted clarity on how people would be held accountable if they broke the rules while accessing the data. Participants wanted to understand what penalties would happen to rule-breakers, and how these would be enforced.

"This holding people who violate the rules and ethics and what have you, holding them to account, how do we do that? Is it a crime, do we take them to court, put them through the whole legal system and then they go to jail for it? What is the way to uphold people to account? Is it headline news and embarrass them, or do we fine these companies? Please think about that." (FG2)

Ensure diverse views

Participants were asked what sort of involvement they would like to see in the KSS datasets and how the public voice could be represented in decision-making. Participants were clear that they thought all sectors of society should be represented in any public panel or lay representatives in decision-making groups.

"Make sure that different areas of society's views are represented on it, because if we're going to go with implied consent, we need to make sure that people from different backgrounds and their views are still being considered." (FG4)

"In order to increase involvement, we need to become more diverse and more inclusive." (FG1)

Build trust

Participants suggested ways to build trust between the public and the dataset custodians and users. These centred on personal relationships and engagement and taking time to build trust. Participants suggested that building trust and involvement was *"not a one-off exercise, it's a journey and it's a developing journey" (FG1)*. By reaching out to "*community champions*" (FG1) it was suggested that community members could then involve seldom-heard groups and act to build further trust.

"The hardest thing to do is personal engagement with people, to gain their trust and then they will talk to their friends and colleagues, but you need to do a lot of that." (FG2)

"If you really want true engagement with a community, whatever that community looks like is something that you need to build up over time, because that's how you build up trust as well. It's about, the solution is often in the community." (FG5)

Value involvement

Participants were clear that public involvement should be valued in the same way that professional roles were valued in decision-making and advisory boards for the governance and uses of the dataset. Lay or public representatives should share equal power with other stakeholders and their roles should be paid.

> *"You've got to work together in a way that values all contributions and that builds and sustains mutually respectful and productive relationships." (FG4)*
>
> *"If you're looking at panels, consumer groups, etc, work has got to be valued." (FG1)*

Participants were worried that any consultations or involvement roles may be treated as "*tokenistic*" by decision-making boards when deciding on uses of the datasets. They wanted to see a commitment to involvement of the public from the outset and also a commitment to acting upon the advice given to them by public contributors. Participants were worried that a public panel might be set up but that no-one would act on the advice coming from the panel.

> *"I think it's so important not to invite the public to get involved when the decisions already been made. You need involvement at the beginning." (FG1)*
>
> *"What clout, what teeth has this committee got? Is it the law, this committee?" (FG2)*

Discussion

We have shown through deliberative focus groups that citizens of Kent, Surrey and Sussex were largely in support of the linkage of health care records for re-use for audit, planning health services and research. Their main motivations for support were the expectations that linkage would lead to improvements in experience of care, information flows and efficiency within the health service. They had a number of concerns about the quality and content of the data, how the data would be protected from hacking or misuse, and what purposes the data would be used for.

These findings are very similar to views expressed by other groups asked similar questions in the UK and Ireland [11]. Several reviews [10, 11, 21] have shown a high level of willingness among the public to share data for the public good, often with around 70–85% of people agreeing that their data should be shared in this way. Previous studies have also captured the reasons that patients give for supporting data sharing; these include improving clinical decision making and patient care [21]. Reviews have found that patients with particular conditions such as rheumatoid arthritis or Parkinson's are most likely to give these reasons [21]. Additionally, reviews have found that patients see more efficient use of resources as a benefit from data linkage [22]. However, the perceived benefit of data linkage for better provision of public services outside of healthcare has been less well explored. One study examined parents' views of data linkage for provision of early family intervention services [23], and found overall high levels of support, however support was much lower among marginalised groups such as black and lone parents.

A key finding in our focus groups was how often participants referred to benefits they expected to see in patient-facing services, such as specialists having better information to hand when patients were seen in clinic. This was despite us emphasising that we were discussing secondary uses of anonymised patient data. It may be important in the future to start discussions at the place where participants want to start, which in our study, we found to be their experience and expectations of direct care. It is important, during deliberative and reciprocal dialogues, that researchers should expect to explain the difference between data used for direct care and data used for secondary purposes, and to reiterate the different uses of data throughout discussions. They should also be mindful that public contributors must be given space, and control, over the content of discussions so that they can give insights into topics on which they are experts by their lived experience. By having open, two-way dialogue in this way, researchers can achieve the spirit of reciprocity outlined in social license theory.

The concerns expressed by KSS participants also match closely those mentioned in previous studies. Reviews of public concerns about patient data sharing show that patients worry that their data would not be secure, would be misused, or would fall into the wrong hands, and that patient privacy was not guaranteed [21]. Previous citizen engagement has also highlighted the worry that data might be sold on or used for commercial purposes [24, 25]. On the other hand, few previous studies have presented specific governance models to participants and asked whether they are supported and appropriate. In addition, participant worry about data quality has not regularly been reported in surveys and interviews, although this has been mentioned in more in-depth deliberative research such as citizens' juries [17]. For example, participants in a citizens' jury exploring views of sharing data from patient clinic notes and letters, gave the view that "*patient data could contain information about other patients, judgements, offhand comments and other data requiring interpretation, and could be misinterpreted by researchers*" [17]. In historical work, 25–33% of patients found incorrect data in their record, or felt their data was incomplete, when they could see it for themselves [26, 27], suggesting that when patients have the opportunity to view their health record, a substantial minority finds that it does not represent their ground truth of health or illness. This could lead to public concern about accuracy of large bodies of patient data when datasets are used for planning or commissioning services, especially as more patients are getting access to their own record.

When asked about safeguards and values for data-sharing, participants in this study were keen to see that usage of data generated "*data trails*" so that usage could be audited, and also that there was a transparent mechanism by which data users would be held to account if they did not adhere to governance rules. While few previous studies have highlighted the public desire for reassurance on accountability, previous reports have identified that for a "social license" to be granted, organisations must commit to transparency; and transparency is a principle of most data protection laws such as the European GDPR [28]. Work by the Data Futures Partnership in New Zealand, which has made recommendations for organisations using public data, has suggested that organisations should make transparent "*the access rules and protocols in place and the consequences for staff who break them*" as well as "*tell people what you will do if there is a data breach*" [29]. One could consider that the more *transparent* the system of data protection and accountability is, the less the social license has to rely on the public *trusting*

the organisation. Trust is a firm belief in the reliability or truth of something, which may rest on limited information. The more information about data protection mechanisms which is available, the more the public understand about the safety of their data, and the less they have to rely on "*trusting*" data users as beneficent agents.

The final part of the focus groups encouraged discussion on how communication strategies could achieve an acceptable level of transparency; this is a question which has not yet been well answered in academic literature. In our analysis we focussed on shared values which would make the data holders and users more trustworthy in their use of data but also which, if operationalised, would lead to more trustworthy communication about data uses. The values advocated by participants were public involvement and diversity, building relationships, valuing contributions and acknowledging potential for discrimination and stigma. These loosely map onto the research ethics principles of autonomy, justice, and non-maleficence, as citizens wanted all sectors of society to have the opportunity to be involved, wanted involvement to create change for the good of diverse sectors, and wanted involvement of citizens to help with avoidance of harm from perpetuating discrimination. Participants wanted dataset teams to actively commit to reaching out to communities, especially those which are seldom heard in research, and to work together with community champions to communicate information about datasets and their uses. They also wanted a commitment that individuals from these seldom heard communities would have the opportunity to have their voices heard, and commitment to act on advice given by public groups, rather than public involvement being tokenistic. These expectations align with the proposed best practice in public involvement in data intensive health research [30].

Strengths and limitations

This was the first time that citizens of Kent, Surrey and Sussex have been asked specifically about local data being used to improve local health services and local research. We achieved a large sample size across five focus groups, and our findings map onto a number of concepts established in the literature. However, we acknowledge several limitations in the generalisability of our findings. Firstly, we did not recruit equally from all three counties, with the majority of our sample coming from Sussex. Therefore, we may not represent views of citizens from all parts of Kent and Surrey. Kent especially is a large and diverse county, with pockets of substantial deprivation in coastal and Thames estuary towns. We may not have captured views from citizens living in these areas. We also had a highly educated sample, with very few participants who left education after GCSEs (taken at age 16). The structured data we collected showed that most participants were fairly or very willing to share their medical records for research at the beginning of the focus groups, and this may have reflected the way the study was advertised via community, health and patient groups, thus targeting individuals who were already potentially interested in advocating for health service improvements. While we opened up the research to and welcomed interests from participants who were unwilling to share data, often these people did not, in the end, take part in the research, which means the dissenting view may not be well-represented in our findings. Lastly, because of recruiting and hosting focus groups online, we may have excluded people who do not use computers or who could not afford internet data costs for joining video conferencing calls. We tried to mitigate this with a digital accessibility strategy but acknowledge that this may have been a challenge for some potential participants. We also acknowledge that online spaces are not necessarily the most comfortable for deliberative processes, with online discussions sometimes struggling to meet criteria for quality deliberation [31], although some research has found that online groups express more candid and direct opinions than face-to-face groups [32]. We aimed to maintain the deliberative approach online via personalised pre-group introduction sessions, breaking into small groups ($N \leq 6$) for discussion, conducting discussions in two shorter meetings rather than one long meeting, and having a dedicated technology facilitator separate from the group host facilitator to tackle any technology issues. We encouraged participants to keep their cameras on and facilitated ease of contribution through some icebreaker tasks early in the meeting. We also sent out preparatory materials before the group discussion and checked for comprehension of materials prior to and at the outset of the groups. These are all suggested elements of good practice for conducting deliberative discussions online [31].

Implications and recommendations

It is encouraging that KSS participants were largely in favour of data being linked and used for improving service planning and potentially for research. The risks they perceived were largely as expected in the literature and discussions allowed us to identify values and safeguards to enable trustworthy use of data and, ideally, establish a social license for the use of these linked datasets going forward. By mapping together the tenets of social license by Stockdale et al. and Carter et al. [11, 13] and findings from this research, we have distilled out three recommended principles for ICB teams working on linked dataset development, governance, communication and public engagement. Work is ongoing in Kent and Sussex to implement these recommendations, and establish a standing "patient advisory panel" alongside the development of a secure data environment for health and care data.

Reciprocity, Autonomy, and Justice

To fulfil this tenet, dataset teams must commit to ongoing and meaningful two-way communication with the public in their jurisdiction. To achieve this, they need a robust strategy by which diverse public voices can contribute to agenda-setting for the use of these datasets in local health systems and for research. Our research has shown such a strategy must outline that public contributors will be valued and paid, and that there is a clear mechanism by which contributions from the public will be acted upon. Public involvement must start early and be continuous, and there should be investment in building relationships with diverse community members over time. To foster autonomy, a communications strategy must be developed which actively reaches out to

diverse publics, informs them of uses of data, and directs them to data opt-out mechanisms, where they exist. Through these activities and strategies, the aim must be to build trust through two-way relationships, responsiveness to the community, and transparency. This strategy should also aim to actively approach groups who were not well represented in our sample, such as those who left education at age 16, and residents of Kent's coastal and deprived areas.

Non-exploitation and prevention of harm

For this principle, our results suggest three things must be considered: protecting patients' data and privacy; not using the data for commercial purposes; and avoiding perpetuating stigma or discrimination. While participants were largely happy with the proposed way that data would be protected in KSS integrated datasets, they valued communication about this, and wanted to make sure an accountability system was in place for misuse. These systems should be communicated clearly to the public to ensure systems for protecting datasets are transparent. This communication strategy should also include information on how decisions are made on what purposes data will be used for (e.g. public benefit) and will not be used for (e.g. commercial gain) and how and who makes decisions on this, as well as a clear indication of what sanctions would be in place for anyone who does not follow data security rules. Lastly, outreach work to engage with diverse community members (e.g. those who left education at the age of 16) and seek their contributions must be properly funded and valued, as by hearing public voices and embedding lived experience in data projects, unintended harm by use of datasets can, to some extent, be avoided.

Service of the public good

Dataset teams in KSS are already fully committed to using data only in service to the public good and our findings indicate what the public hopes for in terms of benefit: improvements to services, more efficient use of public resources, better data flows and communication between health services, and wider consideration of mental and physical health issues in other public services. Dataset teams need to focus on clear and active communication about allowable uses of data, showing how data uses are aligned to public priorities. There should be a commitment to publicising the results of data projects (in accessible formats) where they have contributed to improved services or health outcomes; an end of project reporting requirement could be written into the data usage lifecycle as part of the dataset commitment to transparency on how data uses have contributed to the public good.

Future research recommendations

While public views on data sharing, including public concerns and suggested safeguards, are now fairly well established, there is not much research on how to transform public expectations and preferences into a testable "social license to operate". Ultimately, teams will never be able to know if they have achieved a social license unless there is a framework for how to achieve such a thing and how to measure or test whether it has been achieved. Muller et al. [33] argue that dataset teams and researchers using health data should move towards a shared or reflexive data governance model, which focuses on mutual learning, communication and deliberation, via establishing "extended fora" in which all stakeholders share power and are consulted [33]. This also identifies that achieving a social license is not a one-off event, but that it requires continuous maintenance. Researchers should consider developing methods to evaluate attempts to establish and maintain a social license to articulate steps or activities which best achieve this. It is also important to make sure that researcher and public views on what constitutes "public benefit" from data intensive research are aligned; research on what "public benefit" means to the public is ongoing [34, 35].

Conclusions

After deliberative focus groups with Kent, Surrey and Sussex citizens about their views on new linked datasets of health and care records emerging in their region, we have shown that their views largely map onto previous studies elucidating public views on secondary uses of health data. By mapping themes onto existing tenets of social licence and research ethics principles, we have made three recommendations on how digital and dataset programmes in the KSS integrated care systems can involve and communicate with the public in a transparent, inclusive and trustworthy way, thereby starting to build the foundations of a social license for the re-use of health data in the region. This in turn will allow the health data infrastructure in the KSS region to develop, accelerating the applied health research economy and supporting improvements in the health and health care of KSS citizens. These findings can equally be applied in any UK region or internationally, particularly in areas where health data infrastructure is in its infancy. We have shown that the public expect that citizen involvement with health data linkage and re-use should start early, be embedded throughout the data usage pipeline and be highly valued.

Acknowledgments

This project was funded by a grant from the NIHR public health research programme (NIHR133761), and some researcher time (MRR, JA and DI) and public engagement throughout the project was supported by the NIHR Applied Research Collaboration Kent Surrey and Sussex (grant number NIHR200179). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Statement on conflicts of interest

The authors declare no conflict of interest.

Ethics statement

This study was approved by the Brighton and Sussex Medical School Research Governance and Ethics Committee (ref ER/BSMS2730/8). All participants gave informed consent before participating in the study.

References

1. NHS England. Integrated Care Boards 2022 [Available from: https://digital.nhs.uk/services/organisation-data-service/integrated-care-boards]
2. The King's Fund. Integrated care systems explained: making sense of systems, places and neighbourhoods 2021 [Available from: https://www.kingsfund.org.uk/publications/integrated-care-systems-explained]
3. UK Government. Policy paper: Integration and innovation: working together to improve health and social care for all (HTML version) 2021 [Available from: https://www.gov.uk/government/publications/working-together-to-improve-health-and-social-care-for-all/integration-and-innovation-working-together-to-improve-health-and-social-care-for-all-html-version]
4. digitalhealth.net. ICSs required to have 'smart digital and data' platforms by 2022 2021 [Available from: https://www.digitalhealth.net/2021/06/icss-required-to-have-smart-digital-and-data-platforms-by-2022/]
5. East Sussex County Council. Demographic projections. . . in brief 2022 [Available from: http://www.eastsussexjsna.org.uk/JsnaSiteAspx/media/jsna-media/documents/localbriefings/ESiF/Population-Projections-May-2022.pdf]
6. Healthwatch Kent. NHS announces plans to recruit more GPs in Medway, Swale and Thanet 2022 [Available from: https://www.healthwatchkent.co.uk/news/2022-03-28/nhs-announces-plans-recruit-more-gps-medway-swale-and-thanet#:~:text=The%20national%20average%20in%20September,would%20support%20employing%2020%20doctors.]
7. NHS Digital. Protecting patient data 2021 [Available from: https://digital.nhs.uk/services/national-data-opt-out/understanding-the-national-data-opt-out/protecting-patient-data]
8. Department of Health. The new public health role of local authorities London UK2012 [Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213009/Public-health-role-of-local-authorities-factsheet.pdf]
9. NHS Providers. BUILDING A DIGITAL STRATEGY: UNDERSTAND THE SYSTEM CONTEXT 2022 [Available from: https://nhsproviders.org/building-a-digital-strategy/understand-the-system-context]
10. 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;17:73. https://doi.org/10.1186/s12910-016-0153-x [Available from: https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0153-x].
11. Stockdale J, Cassell J, Ford E. "Giving something back": A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Res*. 2019;3:6-. https://doi.org/10.12688/welcomeopenres.13531.2 [Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6402072/]
12. NHS England. NHS England sets out the next steps of public awareness about care.data 2013 [Available from: https://www.england.nhs.uk/2013/10/care-data/]
13. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care. data ran into trouble. *Journal of Medical Ethics*. 2015;41(5):404–9. https://doi.org/10.1136/medethics-2014-102374 [Available from: https://jme.bmj.com/content/41/5/404.short].
14. Gunningham N, Kagan RA, Thornton D. Social License and Environmental Protection: Why Businesses Go Beyond Compliance. *Law & Social Inquiry*. 2004;29(2):307–41. https://doi.org/10.1111/j.1747-4469.2004.tb00338.x [Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1747-4469.2004.tb00338.x]
15. Gille F, Smith S, Mays N. Evidence-based guiding principles to build public trust in personal data use in health systems. *Digital Health*. 2022;8:20552076221111947. https://doi.org/10.1177/20552076221111947 [Available from https://journals.sagepub.com/doi/full/10.1177/20552076221111947]
16. Forbes L, Bates A, Rees-Roberts M, Hotham S, Kendall S. Developing a Public Health Research System to Support Local Government in Kent. Kent and Medway NIHR; 2020. [Available from: https://fundingawards.nihr.ac.uk/award/NIHR131931]
17. Ford E, Oswald M, Hassan L, Bozentko K, Nenadic G, Cassell J. Should free-text data in electronic medical records be shared for research? A citizens' jury study in the UK. *Journal of Medical Ethics*. 2020;46(6):367-77. https://doi.org/10.1136/medethics-2019-105472 [Available from: https://jme.bmj.com/content/46/6/367.abstract]
18. Rothwell E, Anderson R, Botkin JR. Deliberative Discussion Focus Groups. *Qual Health Res*. 2016;26(6):734–40. https://doi.org/10.1177/1049732315591150 [Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5847289/]
19. Ipsos Mori, Wellcome Trust. Wellcome Trust monitor report wave 3: Tracking public views on science and biomedical research: Ipsos Mori, Wellcome Trust; 2016 [Available from: https://wellcome.ac.uk/sites/default/files/monitor-wave3-full-wellcome-apr16.pdf]

20. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77-101. [Available from: https://www.tandfonline.com/doi/abs/10.1191/1478088706QP063OA]

21. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of medical ethics*. 2022;48(1):3–13. https://doi.org/10.1136/medethics-2019-105651 [Available from: https://jme.bmj.com/content/48/1/3.abstract].

22. Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. *Expert review of molecular diagnostics*. 2014;14(8):1053-65. https://doi.org/10.1586/14737159.2014.961917 [Available from: https://pubmed.ncbi.nlm.nih.gov/25260013/].

23. Edwards R, Gillies V, Gorin S. Data linkage for early intervention in the UK: Parental social license and social divisions. *Data & Policy*. 2021;3. https://doi.org/10.1017/dap.2021.34 [Available from: https://www.cambridge.org/core/journals/data-and-policy/article/data-linkage-for-early-intervention-in-the-uk-parental-social-license-and-social-divisions/2FE85AD3BF7925609251D3940E184BC8].

24. Hill EM, Turner EL, Martin RM, Donovan JL. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. *BMC Med Res Methodol*. 2013;13:72. https://doi.org/10.1186/1471-2288-13-72. [Available from: https://link.springer.com/article/10.1186/1471-2288-13-72].

25. NICE Citizens Council. What ethical and practical issues need to be considered in the use of Anonymised information derived from personal care records as part of the evaluation of treatments and delivery of care? . London: National Institute for Health and Care Excellence (NICE); 2015. [Available from: https://pubmed.ncbi.nlm.nih.gov/28230945/]

26. Jones R, McGhee S, McGhee D. Patient on-line access to medical records in general practice. *Health bulletin*. 1992;50(2):143-50. [Available from: https://europepmc.org/article/med/1517087]

27. Hassol A, Walker JM, Kidder D, Rokita K, Young D, Pierdon S, et al. Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging. *Journal of the American Medical Informatics Association*. 2004;11(6):505-13. https://doi.org/10.1197/jamia.M1593 [Available from: https://academic.oup.com/jamia/article/11/6/505/698041]

28. Information Commissioner's Office. Guide to Data Protection: Principle (a): Lawfulness, fairness and transparency [Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/]

29. Data Futures Partnership. A Path to Social Licence: Guidelines for Trusted Data Use: Data Futures Partnership; 2017 [Available from: https://aisp.upenn.edu/wp-content/uploads/2019/08/Trusted-Data-Use_2017.pdf]

30. Aitken M, Tully MP, Porteous C, Denegri S, Cunningham-Burley S, Banner N, et al. Consensus statement on public involvement and engagement with data-intensive health research. *International Journal of Population Data Science*. 2019;4(1). https://doi.org/10.23889/ijpds.v4i1.586. [Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8142968/]

31. Willis R, Yuille A, Bryant P, McLaren D, Markusson N. Taking deliberative research online: Lessons from four case studies. *Qualitative Research*. 2021;online first https://doi.org/10.1177/14687941211063483 [Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8142968/].

32. Min S-J. Online vs. face-to-face deliberation: Effects on civic engagement. *Journal of Computer-Mediated Communication*. 2007;12(4):1369-87. [Available from: https://doi.org/10.1111/j.1083-6101.2007.00377.x]

33. Muller SHA, Kalkman S, van Thiel GJMW, Mostert M, van Delden JJM. The social licence for data-intensive health research: towards co-creation, public value and trust. *BMC Medical Ethics*. 2021;22(1):110. https://doi.org/10.1186/s12910-021-00677-5 [Available from: https://link.springer.com/article/10.1186/s12910-021-00677-5]

34. Harkness F, Rijneveld C, Liu Y, Kashef S, Cowan M. A UK-wide public dialogue exploring what the public perceive as 'public good' use of data for research and statistics: Office for Statistics Regulation and Administrative Data Research UK, 2022 [Available from: https://osr.statisticsauthority.gov.uk/publication/a-uk-wide-public-dialogue-exploring-what-the-public-perceive-as-public-good-use-of-data-for-research-and-statistics/]

35. Aitken M, Porteous C, Creamer E, Cunningham-Burley S. Who benefits and how? Public expectations of public benefits from data-intensive health research. *Big Data & Society*. 2018;5(2):2053951718816724. https://doi.org/10.1177/2053951718816724 [Available from: https://journals.sagepub.com/doi/full/10.1177/2053951718816724]