The right to privacy: digital data

PHG FOUNDATION

making science work for health

Submitted to

House of Commons Science and Technology Select Committee

January 2022

About the PHG Foundation

1. The PHG Foundation is non-profit, independent policy think-tank and a linked exempt charity of the University of Cambridge. Our mission is making science work for health – providing multidisciplinary analysis of innovations and ideas in genomics and other biomedical technologies to inform health policy and practice. We have over twenty five years' experience in issues surrounding the responsible and effective use of genomics and data in health services, for public health and personalised prevention.

2. Our response to the inquiry draws on a range of recent research projects, reports and other activity that we have undertaken on this topic. These are freely available from our website. We would be happy to comment in greater depth or to provide oral evidence.

The importance of data sharing for health research and medicine

3. Data sharing is crucial for health research and modern medicine. Data sharing has enabled significant advances in scientific research, including in the genomic field where data from hundreds of thousands of individuals enables researchers to identify novel causes of disease.[1] In turn these insights are built on to develop treatments and therapies for both rare and common diseases.

4. For example, our work on data sharing to support UK clinical genetics and genomics services identified three key benefits for data sharing in genomic medicine: Increased, improved and faster diagnosis for patients; improved and more tailored treatments for patients, and; cost and efficiency savings for the NHS.[2] Our recent research on use of genomic data to combat COVID-19 identified that improvements to data access and sharing for research have

1 The 100,000 Genomes Project Pilot Investigators, "100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report." New England Journal of Medicine (2021) 385(20): 1868-1880.

2 PHG Foundation, *Data sharing to support UK clinical genetics & genomics* (Dec 2015) available from: services https://www.phgfoundation.org/report/data-sharing-to-support-uk-clinical-genetics-and-genomics-services

been integral to the pandemic response in England.[3]

5. Building insights from data requires data to flow between the various actors in the ecosystem (e.g. health professionals, health systems, researchers and commercial companies); but deriving insights may also require the integration of different types of data, e.g. clinical phenotypic data with genomic sequence data to better understand the role of a particular genetic variant, in causing disease.

6. The range of relevant data that are useful in health research and care has also expanded rapidly. This not only includes routinely collected health data such as individual health records, but also an increasing range of data from outside the healthcare setting. Such as data generated through the use of fitness trackers or home monitoring devices, passively generated data through environmental sensors, location data and online activity (so called citizen-generated data).[4]

7. This leads to a general paradox/challenge which is inherent in this space: the more informative the data, the greater the risks to privacy posed by sharing that data. Although these risks may be mitigated by infrastructural and technological safeguards (see below), residual risks may remain. This is one of the central challenges for data sharing but is by no means the only challenge or barrier.

Challenges for health data sharing and privacy

8. In our work over the last decade we have identified a range of challenges in reconciling privacy and data sharing for both health research and medical purposes. These can be categorised into technical challenges,[5] cultural challenges and ethical and legal challenges, and they are often heavily interrelated. The nature and scale of these challenges vary according to context and purposes, and they are not always easy to address.

9. A key challenge that we have repeatedly identified in our research has been significant uncertainty and disagreement about the requirements of the legal framework and what is required to safeguard privacy, confidentiality and data protection. In particular, in our Information Commissioner's Office (ICO) funded

3 PHG Foundation, *Regulation and use of confidential patient information for genomic and medical research during and post COVID-19* (2021) available from https://www.phgfoundation.org/report/control-of-patient-information

4 PHG Foundation, *Citizen generated data and health: predictive prevention of disease* (2020) available from https://www.phgfoundation.org/research/citizen-generated-data

5 For an example of these in the context of routine health data see the PHG Foundation report on *Linking and sharing routine health data for research* (2017) available from: https://www.phgfoundation.org/report/linking-and-sharing-routine-health-data-for-research

work on the impact of the General Data Protection Regulation (GDPR) on genomic healthcare and research, we identified that there is considerable divergence and uncertainty about which data count as 'personal data', whether this always includes 'pseudonymised' data and whether genetic or genomic data may ever be anonymised. We identified further uncertainty and disagreement about the nature and level of safeguards that must be applied to protect personal data in health research and what is required to comply with a range of data protection obligations.

10. These findings were echoed in relation to other parts of the legal framework in our recent DHSC and NIHR funded research on the regulation and use of confidential patient information for genomic and medical research during and post COVID-19.[6] In this research we identified confusion among health data custodians and researchers about the alignment or friction between data protection law and the common law of confidentiality. In particular, it is unclear whether the scope of confidential patient information matches the scope of personal data or whether pseudonymised data may be considered to remain personal data but not confidential patient information.

11. Another challenge is that consent operates differently and has different requirements depending on whether it is consent for the purposes of disclosing confidential information, consent as a legal basis for processing personal data, or consent as a requirement of the ethical framework for health research. These differences are difficult for professionals to understand, and are highly confusing to individual patients and research participants.

12. A perennial tension in the genetic/genomic context is that such data are highly identifying and sensitive but that sharing some genetic information can have clear clinical benefits for family members.[7] Despite legal attention to this challenge in the case of ABC v St George's NHS Trust (and other NHS defendants) there is still some uncertainty about when such data should be shared and how this aligns with confidentiality and data protection law.[8]

6 PHG Foundation, *Regulation and use of confidential patient information for genomic and medical research during and post COVID-19* (2021) available from https://www.phgfoundation.org/report/control-of-patient-information

7 PHG Foundation, *Data sharing to support UK clinical genetics & genomics* (Dec 2015) available from: https://www.phgfoundation.org/report/data-sharing-to-support-uk-clinical-genetics-and-genomics-services

8 Mitchell C, *'After ABC v St George's: a new duty to consider'* (PHG Foundation Blog, 3 March 2020) available from: https://www.phgfoundation.org/blog/abc-v-stgeorges-new-duty#:~:text=The%20ABC%20case,(and%20other%20NHS%20defendants).&text=ABC%20brought%20a%20claim%20of,her%20of%20the%20genetic%20risk.

Solutions, strategies and national policy

13. Despite the range and scale of these challenges, there have been some welcome developments in national level strategies and policies which may help to address them. The Government's National Data Strategy recognised the need to support the development and application of privacy enhancing technologies, a responsibility to ensure that there is a clear and predictable legal framework for uses of data in the public interest and the central importance of earning and retaining people's trust in the use of data.

14. The NHS data strategy, Data saves lives: reshaping health and social care with data builds on this with a range of proposals and commitments, including to clarify and simplify information governance and provide safe and secure analysis environments for research. The DCMS consultation on reform of data protection law, Data: a new direction, also contained a welcome focus on the importance of health research and challenges for researchers navigating the regulation. We think many of the barriers identified in the consultation and a number of the proposals in this area are sensible.

15. However, we have reservations about the emphasis of some of these proposals and elements that may be missing or under-emphasised in terms of potential solutions. In terms of the overall approach to data protection law, we question the strength of some of the claims made in the DCMS consultation about the role of law as a barrier. In our research, we have repeatedly identified that uncertainty and complexity were the key and overarching challenges for researchers. However, we recognised that both are inevitable in the application of a comprehensive and sector-agnostic new law.

16. Our work highlights that the challenge is not the wording of the law (although there are undoubtedly aspects that could be better drafted) but rather the inevitable scope for argument about its proper application in a specific context, such as genomic research. To change the law would be unlikely to significantly alter the scope for that argument. We recommend that as much resource and effort as possible is put into developing specific guidance, in consultation with relevant industries and sectors to address these challenges, rather than changes to the legislation.

17. Our strong view is that the challenge of developing, interpreting and applying appropriate standards for the protection of privacy, confidentiality and data protection in the health context would be best achieved through the co-development of specific guidance addressing particular topics and issues between regulatory authorities such as the ICO and specialists in health, health data

and genomics. We strongly refute the concept of a 'surfeit of guidance' (as it is put in the DCMS consultation) in this regard.

18. In our research we found considerable approval for the ICO's track record in producing user friendly and sensible guidance. More broadly, there are a range of regulatory and advisory bodies who are well placed to develop clarity and specific guidance about appropriate measures that should be put in place to facilitate ethical and privacy-preserving data sharing. These include the National Data Guardian and the Health Research Authority.

19. We are also concerned that the proposed changes to data protection law risk diverging sufficiently from the European Union's standards that the UK will be adjudged to offer a lower (and inadequate) level of protection for personal data. This is also a latent risk in the context of the pro-growth National Data Strategy, and it is crucial that the level of protection of privacy and data protection for UK citizens is not lowered in the name of prosperity. This would run counter to the global trend and would also jeopardise free flows of data between the UK and the EU, which are crucial to scientific and genomic research in particular.

20. Finally, we believe there is a key missing element in the draft health data strategy which, based on recent experience and our research, is the requirement to engage with publics and patients about how they think data should be used, and build these considerations in a meaningful way into future policy development. This requirement is based on two related principles: that reasonable expectations around the use of data should guide the legitimacy of how confidential patient information is disclosed and used; and that meaningful engagement builds trustworthiness and public trust more generally.

21. Thus, it is not simply enough to explain what is being done with data, but there must be genuine engagement and the potential for public and patient attitudes and preferences to influence the rules and standards that are applied. This includes ensuring that those consulted are sufficiently representative of patients, participants and publics who will be affected. In the genomics context, we suggest that a key area where the expectations of publics and government should be aligned is around the uses of anonymised data, standards for pseudonymised data to mitigate against erroneous expectations that cannot be met. Without genuine engagement, policy developments and implementation measures run the risk of damaging public trust and confidence.

The ethical framework and effectiveness of existing governance

22. Meaningful engagement relies on an ethical assumption that individuals should have some control over data concerning them. There are a range of views on what the limits to this control might be. In our COPI project, PHG Foundation (working with engagement specialists Traverse Ltd) undertook a focus group to explore patient attitudes to the use of health data for medical research and the impact of the COPI notices more specifically on genetic and genomic research. A key finding was that some participants had a continuing expectation that they would continue to have an interest in their data even if it were de-identified or anonymised. Our work on GDPR and genomic data found that there was also a lack of clarity as to when data could be considered 'effectively anonymised'. Therefore a continuing challenge is how to resolve the potential lack of alignment between publics and governments around the uses of data that has been partially de-identified, which in some circumstances might be individually identifying (pseudonymised data) or data that originated from an identifiable individual but is no longer identifiable data (anonymised data).

23. Genuine engagement is only possible if there is transparency about the uses of data, clarity about the legal basis for processing those data, and a clear justification for using individuals' health and care data. Ethical frameworks that explore the use and sharing of individuals' data in health and care contexts typically require a balancing exercise to judge the potential benefits of accessing those data against the potential harms. [For example, this is the case where a health care professional considers whether to disclose confidential patient information from one individual, even where consent to this has been sought and refused, for the benefit of others. In genetics/genomics, there may be clear clinical benefits in sharing data from one individual within a family for the benefit of other family members despite the ongoing policy debate.[9]

24. Transparency is also important, in order to build trust and confidence in the use of data. Although past literature has highlighted the importance of fostering public 'trust', bestowing trust is an individual choice - it cannot easily be won, but can quickly be lost. Our COPI project explored this topic in detail. Our work highlighted that a failure of patient and public confidence in proposed reforms can critically delay or completely terminate plans for improvements to data access, linkage and sharing for important health purposes. Building on the work of Baroness Onora O'Neill, we concluded that it is less appropriate to talk about building and maintaining trust than it is to consider how to make systems, people and institutions worthy of trust. The

9 PHG Foundation, *Data sharing to support UK clinical genetics & genomics* (Dec 2015) available from: https://www.phgfoundation.org/report/data-sharing-to-support-uk-clinical-genetics-and-genomics-services

challenge for demonstrating trustworthiness in data reforms at a national scale is to show that the inclusion of government or commercial organisations does not undermine the higher trust in the NHS and in doctors. We concluded that investing time and resources into promoting characteristics of trustworthiness - such as transparency and public engagement - may help generate collaborative agreement on how data should be used and for what purposes.[10]

Safeguards and privacy

25. There is growing awareness of the different threats to privacy associated with the usage and sharing of data relating to individuals. These can be caused by many factors including inadvertent or deliberate data disclosure; inadequate controls on the use of data resulting in illegitimate uses; data mining to generate personal data from aggregated or anonymous data or from lifestyle data, and many more. The appropriateness of the safeguards adopted, depend on which type of threat is being addressed.

26. Within the NHS, there are safeguards relating to data protection and to the use of confidential patient information for direct patient care. For example, the UK GDPR has specific requirements for processing health data [GDPR and genomic data report]; and the National Data Guardian has described a set of principles underpinning the use and sharing of confidential patient information.[11] These were amended in December 2020 to emphasise that the duty to share information for individual care is as important as the requirement to keep information confidential (Principle 7), and to clarify that patients and service users should be informed how their confidential patient information is used.

27. In the genetics/genomics context, some of the most promising advances are from the integration of multiple health datasets to support health and care. This includes the prospect of combining genetic/genomic datasets with clinical records, and sometimes other lifestyle data to develop personalised interventions to predict future ill health and to develop targeted treatment. However, combining data in this way has potential to compromise patient privacy, if individuals are identified at risk of future disease without their knowledge or consent.

10 PHG Foundation, *Regulation and use of confidential patient information for genomic and medical research during and post COVID-19* (2021) available from https://www.phgfoundation.org/report/control-of-patient-information

11 Caldicott Principles: https://www.gov.uk/government/publications/the-caldicott-principles. For further information see PHG Foundation, *Regulation and use of confidential patient information for genomic and medical research during and post COVID-19* (2021) available from https://www.phgfoundation.org/report/control-of-patient-information

28. In the medical research environment, there have been a variety of responses to this: these include formalising commitments to develop technical infrastructure, in particular harmonising working and developing standards. In our report on the GDPR and Genomic Data, we highlighted the potential for sector specific codes of practice to be developed both to promote consistent approaches, but also to aid transparency.

29. These risks are potentially exacerbated by the increasing use of hypothesis-agnostic research which mines data for clinically relevant findings, especially if these utilise methods which are themselves opaque (such as artificial intelligence or machine learning). Our reports on Black box medicine and transparency addressed this problem, pointing out the need for such tools to be explainable, both as an aid to building trust, but also as a vital means of promoting patient safety.[12] This also requires resources to ensure that regulators and other key actors are in a position to develop a world-leading framework for AI and software as a medical device with input from the public, developers and the range of relevant stakeholders.

30. A further challenge in the context of genetics/genomics is the burgeoning use of research cohorts which collect successive waves of data and samples, supported by a range of research interventions, linked to lifetime clinical health records. Examples include Our Future Health, UK Biobank (healthy participants); and 100,000 Genomes Project, GRAIL (primarily participants who are patients). This ongoing data collection promises many benefits but also poses potential privacy risks. It is worth noting that many of these projects have developed pioneering Trusted Research Environments (TREs), and novel methods of data interrogation (e.g. BEACON) by which data can be interrogated by bona fide researchers without the data being shared or transferred in order to minimise potential privacy concerns. For these applications, it is also vital to develop robust methods for informing participants about the risks and benefits of participation, so that consent is fully informed; and to ensure that there is transparency about the potential uses to which the data will be put and the collaborations involved.

31. A combined approach which utilises the best available technologies (for example TREs and privacy enhancing technologies) in parallel with a commitment towards placing participants at the heart of governance e.g. in data access committees and scientific advisory committees) should be a priority.

12 Available from: https://www.phgfoundation.org/report/black-box-medicine-and-transparency

Concluding remarks

32. Balancing streamlining data access and analysis with privacy and other ethical norms is not easy. However, it is crucial that every effort is made to strike the right balance in the health context. A significant imbalance in either direction has the potential to impair research and healthcare activities with real world implications. This applies as much to an overly liberal approach to data access/sharing as it does to an overly restrictive one because any approach that diverges significantly from the reasonable expectations of patients and publics is liable to result in a loss of trust and confidence with profound potential consequences.

33. It is equally important not to adopt a narrow view that privacy (in particular individual privacy) is the only interest at stake and that there is no need to take care with sharing and uses of 'anonymised' information. In part, this is misguided because as our report on GDPR and genomics data suggests, it is increasingly difficult to categorise data as 'anonymised' in the health context. And there are wider group interests at stake, including an interest in ensuring that fair value is received in return for NHS data. These interests are as important as privacy in ensuring the continued trust and confidence of the public in the use of health data.

34. These considerations extend beyond the use of data in the public sector. Data collected from devices, wearables and social media are increasingly relevant for health research and potentially for care despite being considered 'lifestyle' or non-medical at present. In this context, it is arguably the case that more needs to be done to ensure privacy and the ethical use of such data while enabling innovation. Our view, in relation to this and other aspects of the overall framework, is that the way forward is rarely an overhaul of the legal framework.

35. Our work has suggested that the best way of striking the correct balance in most specific contexts is in interpreting and applying the law through authoritative subject-specific guidance, codes or policies. This requires regulators to work together, and to consult and engage with sectors and sub sectors such as genomics researchers, to develop appropriate standards on key topics. This has begun to take place in relation to artificial intelligence technology in the form of a multi-agency advisory service and it is crucial that regulators are well resourced and supported to perform these functions in the face of an ever-increasingly complex data landscape.

Contact: intelligence@phgfoundation.org