

Chapter 7

Population research data
initiatives

Chapter 7 – Population research data initiatives

Chapter overview

This chapter describes research data initiatives in three clusters (biobanks, international collaborative research projects, and participant-driven research), identifies examples of good practice and draws lessons from some specific initiatives.

Biobanks are major resources of tissues and data that may be used for a variety of research purposes. UK Biobank is a large population biobank established to support the investigation of a range of common diseases occurring in the UK. Key features of this are the broad consent model, its Ethics and Governance Framework and the independent Ethics and Governance Council. Questions arise for which it is necessary to review the set of expectations underlying its operation, including feedback of findings and commercial access to the resource.

The UK10K Rare Genetic Variants in Health and Disease project confronts the problem of controlled disclosure of highly specific individual-level data among different groups of researchers working on distinct studies. It achieves this through a common ethical framework of policies and through guidelines that place considerable reliance on institutional sanctions and on the role of principal investigators.

International collaborative research initiatives such as the International Cancer Genome Consortium and the Psychiatric Genomics Consortium need to accommodate differing local practices and tackle complex consent issues to do with re-use and international transfer of data. The use of cloud-based storage and processing services is becoming increasingly important but it raises issues such as third party access (for example, by security services).

The wide availability of social networking platforms has facilitated participant-led research with norms and social dynamics that differ from more formal institutional research. They present distinct challenges of ensuring the protection of individual interests, of integration with institutional research, and of translation of findings into clinical products and practice.

A number of recommendations are made in relation to a greater role for subject participants, accounting for governance through explicit frameworks, and the use of institutional measures.

Introduction

- 7.1 Increased capability for linking databases, along with technological and IT innovations, have accelerated the pace of data acquisition for large-scale population studies. Typically, population research initiatives of this type collect and store information from identifiable individuals, the study participants. They are supported by the increasing pace and falling cost of automated collection and laboratory analysis of biomedical samples and information, including genetic analysis. Many of the research methodologies in use are well established and build on earlier developments in population epidemiology and data science, but life sciences research is turning towards increasingly large resource platforms for use by international teams of researchers studying a wide range of health-related conditions, and using data from many thousands of individuals. Another feature of these initiatives is the richness of the data available for each participant, made possible by linking existing medical databases, the inclusion of data derived from genomic analysis, as well as 'deep phenotyping' (see

chapter 1) including lifestyle and behavioural data, data from imaging technologies and data derived through sensing devices.

- 7.2 Biomedical research raises significant ethical and governance issues including recruitment of participants, how their morally relevant values and interests are respected (for example, through the choice of consent mechanisms), data security, decisions on feedback of medical information to participants, data access arrangements for researchers, data linking and movement trans-nationally, governance of the resource in the public interest, and strategies for disseminating the outcomes of research. In this chapter we consider a number of data initiatives in three broad clusters: biobanks, international collaborative research projects, and participant-driven research.

Biobanking

- 7.3 The term ‘biobank’ has become a catch-all phrase for many types of collection of biological samples and related data.⁴⁰⁴ Here we are concerned with collections that are established as prospective research resources comprising material and data from many participants. Our examples are two flagship British resources, UK Biobank and UK10K.

UK Biobank

- 7.4 The UK Biobank initiative is a major resource designed to support a range of research ‘to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses’ and the promotion of health throughout society.⁴⁰⁵ Through invitations sent to patients on geographically selected general practice lists, more than 500,000 people between the ages of 40 and 69 were recruited. The intention is to follow them for at least 25 years through their GP and NHS hospital records as well as through periodic collection of data directly from the subject participants themselves.⁴⁰⁶ In 2012 the resource opened for use by researchers worldwide. Applications to use the resource are screened to ensure that projects meet established ethical and scientific standards, and to consider how the research meets the criterion of being in the public interest. Researchers using the resource give undertakings to abide by certain conditions of use and to treat data confidentially. They pay a modest fee calculated on a cost recovery approach for providing the data or samples. UK Biobank carries out extensive work validating and cleaning data and preparing it for use but does not undertake the research itself.
- 7.5 An International Scientific Advisory Board advises on scientific and policy matters and the research community is invited to contribute to shaping the direction of future data

⁴⁰⁴ See: Kaye J, Gibbons SMC, Heeney C, and Parker M (2012) *Governing Biobanks: understanding the interplay between law and practice* (Oxford: Hart Publishing); Shaw DM, Elger BS, and Colledge F (2014) What is a biobank? Differing definitions among biobank stakeholders *Clinical Genetics* **85**(3): 223-7.

⁴⁰⁵ See: <http://www.ukbiobank.ac.uk/about-biobank-uk/>.

⁴⁰⁶ At the time of recruitment, participants give consent for UK Biobank to have long-term access to their existing and future NHS medical records and other health-related records. The consent form states: “I give permission for access to my medical and other health-related records, and for long-term storage and use of this and other information about me, for health-related research purposes (even after my incapacity or death).” Participants are able to withdraw from the study at any time on the basis of either no further contact, no further access to their data or no further use of their data. See http://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Consent_form.pdf and <http://www.ukbiobank.ac.uk/faqs/can-i-withdraw-from-uk-biobank/>.

acquisition. Right from the start it was recognised that such a large multi-purpose biobank resource, designed to collect data and samples prospectively to facilitate research, would raise ethical and governance challenges both at recruitment and in later decades. Decisions were taken to put in place two distinctive measures, namely, the UK Biobank Ethics and Governance Framework and an independent Ethics and Governance Council to advise the funders and UK Biobank.⁴⁰⁷

Recruitment

- 7.6 There is a long history of population studies in Britain, including the 1946 and 1958 Birth Cohorts studies up to the Life Study begun in 2012.⁴⁰⁸ It was recently estimated that 2.2 million people are currently taking part in large population studies in the UK (approximately one in thirty of the general population).⁴⁰⁹
- 7.7 UK Biobank sought to recruit as widely as possible across England, Wales and mainland Scotland (but not in Northern Ireland). Participants completed an automated questionnaire at local assessment centres and were interviewed about lifestyle, medical history and diet. In addition, basic assessments including weight, body mass index (BMI), heart function, blood pressure and bone density were made. Blood and urine samples were taken to be assessed for biomarkers, and DNA extracted for genomic analysis.⁴¹⁰
- 7.8 The overall UK Biobank volunteer rate was approximately 5.5 per cent of those approached.⁴¹¹ The locations of Assessment Centres, and the surrounding GP practices where the potential participants were registered, were selected with the aim of creating a generalisable population sample, 'so that research may ultimately benefit a wide diversity of people'.⁴¹² It is well known that it is harder (and more expensive) to recruit to population studies those with low income and of lower social status, those with poorer health and/or chronic conditions, and those from ethnic minority and rural communities.⁴¹³ This proved to be the case for the UK Biobank sample. It is unclear how far strategies to ensure greater representation from at least the larger UK ethnic minority groups were pursued, but it seems that a compromise was reached between the costs of ensuring participation from hard to reach communities and accomplishing timely recruitment of 500,000 people. Regardless, some researchers have been critical of the lack of representativeness of the sample for the UK population.⁴¹⁴ The UK Biobank recruitment strategies were justified on the grounds that nested case-control studies do not require a (near) representative sample. This may, however, limit the

⁴⁰⁷ UK Biobank has been funded (about £62 million by the time participant recruitment was completed) by the Medical Research Council, the Wellcome Trust and the Department of Health together with a number of other public and charitable resources. For the EGF and EGC, see: <http://www.ukbiobank.ac.uk/ethics/>.

⁴⁰⁸ See the Cohort and Longitudinal Studies Enhancement Resources (CLOSER) for details of British birth cohort studies: <http://www.closer.ac.uk/>.

⁴⁰⁹ Pell J, Valentine J and Inskip H (2014) One in 30 people in the UK take part in cohort studies *The Lancet* **383(9922)**: 1015-6 and <http://www.mrc.ac.uk/news-events/publications/maximising-the-value-of-uk-population-cohorts/>.

⁴¹⁰ This was followed by an enhanced baseline assessment including eye measure, additional blood collection for RNA analysis and saliva sample collection. See <http://www.ukbiobank.ac.uk/data-showcase-timeline>.

⁴¹¹ Allen N, Sudlow C, Downey P, et al. (2012) UK Biobank: current status and what it means for epidemiology *Health Policy and Technology* **1(3)**: 123-6; see also Swanson JM (2012) The UK Biobank and selection bias *The Lancet* **380(9837)**: 110.

⁴¹² See UK Biobank Ethics and Governance Framework, <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>, at page 4.

⁴¹³ Ridgeway JL, Han LC, Olson JE *et al.* (2013) Potential bias in the bank: what distinguishes refusers, nonresponders and participants in a clinic-based biobank? *Public Health Genomics* **16(3)**: 118-26; for ethnic minorities, issues include distrust of the medical profession, lack of awareness and economic burden. See Paskett ED, Reeves KW, McLaughlin JM *et al.* (2008) Recruitment of minority and underserved populations in the United States: the centers for population health and health disparities experience *Contemporary Clinical Trials* **29(6)**: 847-61.

⁴¹⁴ Swanson JM (2012) The UK Biobank and selection bias *The Lancet* **380(9837)**: 110.

range of research studies that use UK Biobank since those requiring a (near) representative sample may lack sufficient scientific validity. It also follows that the ambition for research outcomes to benefit a 'wide diversity of people' might fail to be achieved equally.

Consent and governance

7.9 Population-based biobanks generally create a resource for the use of researchers for future research that is unspecified at the time of collection. Thus, at the time of recruitment it is not possible to tell would-be participants in detail what research may be carried out with the data and samples they may donate. A 'broad consent' model is therefore adopted by many biobanks and was adopted by UK Biobank. Participants were given information about what data and samples would be collected and how the project would be governed but the scope of future research was defined only in general terms, as health-related research in the public interest. Participants are able to withdraw at any time from the project to different degrees: they may choose no further contact, no further access to their data or no further use of their data.⁴¹⁵ As we noted in chapter 4, the broad consent model is not uncontroversial and these controversies have found a focus in relation to biobanks.⁴¹⁶ As we argue in chapter 5, however, consent, as a way of respecting morally relevant individual values and interests, is only ever part of the story – it is the relationship between norms of privacy, the way of respecting individual preferences (in this case, through broad consent), and mode of governance regulating the public and private interests in play, that determines whether a data initiative can find an ethically acceptable form. While broad consent has been a practical solution to the difficulties of obtaining informed prospective consent for a large number of diverse research projects from thousands of participants, the need for complementary governance structures is generally accepted. For many population studies, an ethics oversight committee is established, but for UK Biobank an independent Ethics and Governance Council (EGC) acts as the guardian of a dedicated and detailed Ethics and Governance Framework (EGF) (see Box 7.1.)

Box 7.1: UK Biobank Ethics and Governance Council and Framework

The UK Biobank Ethics and Governance Council (EGC) is an advisory body with members appointed by the funders independently of UK Biobank. It has no formal regulatory role but rather advises UK Biobank in the manner of a 'critical friend'.⁴¹⁷

The Ethics and Governance Framework (EGF) sets out the relationship between UK

⁴¹⁵ See <http://www.ukbiobank.ac.uk/faqs/can-i-withdraw-from-uk-biobank/>. They are unable, however, to have their data expunged entirely from the biobank's systems. In fact, in June 2007, UKBB was obliged to amend its advice regarding the 'no further use' option offered as part of the right to withdraw as a consequence of a technical feature of the data archiving system, after some participants had been recruited. The advice was amended to indicate that while information would be made unavailable to researchers it would nevertheless be retained for archival and audit purposes. Furthermore, UK Biobank advises that while it would destroy all biological samples, 'it may not be possible to trace all distributed sample remnants' and data could not be removed from completed analyses. See: 'No further use' withdrawal option: February 2008' at <http://www.ukbiobank.ac.uk/resources/>.

⁴¹⁶ For criticism of broad consent as inadequately informed, see: Greely, HT (2007). The uneasy ethical and legal underpinnings of large-scale genomic biobanks *Annual Review of Genomics and Human Genetics* 8(1): 343-364; for a defence of broad consent for larger-scale research infrastructure projects, see: Knoppers, BM (2005) Consent revisited: points to consider *Health Law Review* 13(2-3): 33-8; Hansson MG, Dillner J, Bartram CR, *et al.* (2006) Should donors be allowed to give broad consent to future biobank research? *The Lancet Oncology* 7(3): 266-9. See also: Corrigan O, McMillan J, Liddell K, *et al.* (Editors) (2009) *The limits of consent* (Oxford: Oxford University Press).

⁴¹⁷ Laurie, G (2011) Reflexive governance in biobanking: on the value of policy led approaches and the need to recognise the limits of law *Human Genetics* 130(3): 347-56.

Biobank and participants, research communities, individual researchers and society. This twenty page document was widely circulated and discussed in the research and bioethics communities, and at public meetings, before a version was agreed by the funders during the planning stage of the project. The EGF may be seen as an instrument, legitimised through wide discussion, which serves to align the public interests in research and the privacy and other interests of participants, as well as engendering trust. It is a 'living' document that is intended to be responsive to changing needs and circumstances. The EGC is charged with monitoring and reporting publicly on the conformity of UK Biobank with the EGF and to advise more generally on the interests of research participants and the general public in relation to UK Biobank.

- 7.10 While participants cannot know in advance for exactly what purposes their data will be used, UK Biobank has attempted to keep them informed about the developments of the resource, the research being carried out using it and the results of that research. It does this via newsletters and information on the website, as has been the case with other biobanks.⁴¹⁸ It is a way of encouraging the continued support and motivation of participants on which the long-term value of the resource depends.⁴¹⁹ But it is also important in order to enable participants meaningfully to exercise the option to withdraw from the resource if they consider that morally salient features of its use, as approved under the governance system in place, depart from the scope of their expectations.⁴²⁰ Aside from this, however, these means of communication are mainly one-way, an opportunity for the biobank to deliver information, rather than a deliberate opportunity to hear the views of participants or involve them in the development of the resource.
- 7.11 The EGF states, among other things, that 'further consent will be sought for any proposed activity that does not fall within the existing consent'. As circumstances change it may not be clear whether new proposals raise issues that fall outwith the scope envisaged by the initial broad consent. However, relying on renewed consent may limit activities that are in the interests of participants and in the public interest. This is because seeking renewed consent from the whole body of participants may be impractical, very expensive and likely to damage the resource as many participants, especially those who are hard to reach, might drop out by default. Thus activities beyond the obvious scope of initial consent are likely to be avoided. Possible examples here include data linking to other biobanks or to administrative data that are not directly health related.
- 7.12 Revising the EGF is one potential response to these changing circumstances and the changing horizons of public and private interests. However, revising the EGF simply by agreement with the Board of Directors falls short of the standard of engagement set by UK Biobank's creation and legitimisation, which was characterised by wide-ranging and inclusive discussions, pursued through public meetings. Despite time and resource implications there is growing support for the view that long-term studies should attempt to involve participants in a meaningful way and some initiatives have moved in this

⁴¹⁸ McCarty CA, Garber A, Reeser JC, and Fost NC on behalf of ; the Personalized Medicine Research Project Community Advisory Group and Ethics and Security Advisory Board (2011) Study newsletters, community and ethics advisory boards, and focus group discussions provide ongoing feedback for a large biobank *American Journal of Medical Genetics* **155A(4)**: 737-41.

⁴¹⁹ Gottweis H, Gaskell, and Starkbaum J (2011) Connecting the public with biobank research: reciprocity matters *Nature Reviews Genetics* **12(11)**: 738-9.

⁴²⁰ Melham K, Briceno Moraia L, Mitchell C, *et al.* (2014) The evolution of withdrawal: negotiating research relationships in biobanking *Life Sciences, Society and Policy* **10**:16, available at: <http://link.springer.com/article/10.1186/s40504-014-0016-5>.

direction.⁴²¹ However, whereas the UK Biobank EGF proposes a representative participants' panel, this has not been pursued to date.⁴²² Given its central role in the regulation of the relationship between UK Biobank and its participants, there is a strong argument both that the process of revision and development of the EGF would benefit significantly from including participants' views and interests, and, as we argue in chapter 5, there is a substantive moral reason for them to participate.⁴²³ Furthermore, the involvement of 'publics' (e.g. participants, policy makers, researchers, future beneficiaries and other publics) in open discussion, particularly in relation to issues where practical decisions are to be made, can help build, maintain and develop a trusted governance structure.⁴²⁴

Data security, access and linkage

- 7.13 Permanently de-identified data (which the data controller cannot link back to the individual case) is little use in the context of continuing longitudinal research. Data and samples are therefore assigned a pseudonym or code (see chapter 4) so that they can be linked back to the same index case over time. The obvious personal identifiers (names, addresses, etc.) are separated and stored in a protected file store. Coded data are then made available to researchers, who are usually bound by undertakings to not try to re-identify participants. UK Biobank has an Access Committee (a subcommittee of the UK Biobank Board) that takes decisions about research access in the light of advice from UK Biobank managers and external ethics advisors. Access requirements for users focus on the three areas described below in order to protect the confidentiality of participant data, as well as to promote the trustworthiness of the project.
- 7.14 First, researchers are checked to see if they are 'bona fide' (acting 'with good faith') and from recognised institutions (and so governed by ethical codes of practice). If there is a breach of use by a researcher, there are a number of ways in which penalties may be imposed.⁴²⁵ Although there may be no legal basis for UK Biobank to reprimand individuals, beyond refusing further access to the resource, there exist ways that those using data could be penalised by their institution and ways that some institutions can be penalised by funders if one of their staff breaks the rules, which may be given effect through contracts and legal agreements.⁴²⁶ Second, proposals must meet criteria set by the biobank for the use of the resources. Access committees need to review requests for samples to ensure that the research is scientifically valid and the use falls

⁴²¹ UK Biobank Ethics and Governance Council (2009) Workshop report: involving publics in biobank research and governance, available at: <http://www.egcukbiobank.org.uk/sites/default/files/meetings/EGCpercent20workshoppercent20report.pdf>.

⁴²² UK Biobank Ethics and Governance Framework, <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>, at page 8.

⁴²³ As it happens, some members of the EGC have also been participants. However, there is no requirement that there should be participants on the Council.

⁴²⁴ Burgess MM (2014) From 'trust us' to participatory governance: deliberative publics and science *policy Public Understanding of Science* **23**(1): 48-52. See also Armstrong V, Barnett J, Cooper H, et al. (2007) Public attitudes to research governance. On the Governance of biomedical research: a qualitative study in a deliberative context, available at: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtx038443.pdf.

⁴²⁵ Joly Y, Zeps N, and Knoppers BM (2011) Genomic databases access agreements: legal validity and possible sanctions. *Human Genetics* **130**(3): 441-9.

⁴²⁶ The 2008 Thomas and Walport Data Sharing Review Report points, with approval, to the application of legal penalties through the Statistics and Registration Service Act 2007: "The Board may extend access to researchers from various organisations, including academic institutions, public bodies and nongovernmental organisations. These researchers are then bound by a strict code, which prevents disclosure of any personal identifying information. Any deliberate or negligent breach of data security by the approved researcher would entail criminal liability and the prospect of a custodial sentence up to a maximum of two years." Thomas R and Walport M (2008) *Data sharing review report*, available at: <http://systems.hscic.gov.uk/infogov/links/datasharingreview.pdf/view>. See also our recommendation 5 above.

within the broad categories given in the participant consent. Third, researchers must agree to certain undertakings regarding the confidentiality of the data and handling of samples. Material and Data Transfer Agreements are used to bind researchers through their institutions.⁴²⁷ Researchers are required, for example, to provide a secure environment for the samples and data, to ensure that the data are not used other than for the agreed purposes and not to attempt to re-identify individuals from data. UK Biobank also asks researchers to return results of their work, which may be added to the resource, and for research results to be published, which promotes outcomes for public benefit and thereby demonstrates respect for the altruistic motivations of participants.⁴²⁸

7.15 One of the core aims of UK Biobank is to link data provided by participants with other health-related and administrative records in order to track the emergence and/or progression of disease and to collect data to support epidemiological research. The complex process of linking UK Biobank records to both NHS hospital and GP records is currently underway. A recent MRC strategy review of population cohort studies saw increasing opportunities for cohort studies not only to link to NHS records but to also to link more widely to cross-sector administrative and environmental information. It noted a number of initiatives that will improve secure access to data.⁴²⁹ However, the resource could potentially have much broader uses, some of which may challenge the initial health-related purposes or address them in unusual and unanticipated ways. Examples already exist of biobanks seeking to link their data with criminal convictions and cautions, as well as financial benefits, earnings and employment data.⁴³⁰

7.16 There are also opportunities to use commercial data (such as geospatial location data from mobile phones) in research. It has been suggested that purchasing data from supermarkets might be used to infer the effects of diet on the health of research participants. With researchers' growing interests in 'deeper' phenotyping (see chapter 1) the appetite for data from a wider range of sources is likely to increase. As we noted at the outset, when these are linked in the context of health-related research, this data may be informatively 'health-related' but such linkages may well fall outside the expectations participants had when they signed up to participate.⁴³¹ To determine this, it may be necessary to reason in relation not only to the nature of the research proposal within the developing field of science and the initial expectations of participants but also the norms that apply at the time.⁴³² As further opportunities emerge, enhanced participation may prove to be an important way of demonstrating respect for subject participants and undertaking the moral reasoning necessary to relate the scope of reasonable expectations to the scope of potential uses.

⁴²⁷ Fortin I, Pathmasiri S, Grintuch R, and Deschênes M (2011) 'Access arrangements' for biobanks: a fine line between facilitating and hindering collaboration *Public Health Genomics* **14(2)**:104-14.

⁴²⁸ UK Biobank Return of Results Data: Guidance Note for Approved Projects, available at: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Return-of-Results_Guidance-Note_v2.pdf. Where research results may be seen as controversial, UK Biobank can ask for sight of papers before publication.

⁴²⁹ Medical Research Council (2014) Maximising the value of UK population cohorts. MRC strategic review of the largest UK population cohort studies, available at: <http://www.mrc.ac.uk/news-events/publications/maximising-the-value-of-uk-population-cohorts/>.

⁴³⁰ See for example, the Avon Longitudinal Study of Parents and Children data linkage information. <http://www.bristol.ac.uk/alspac/researchers/resources-available/data-details/linkage/>.

⁴³¹ See footnote 407 above.

⁴³² See Laurie, G (2009) Role of the UK Biobank Ethics and Governance Council *The Lancet* **374(9702)**: 1676, available at: <http://www.thelancet.com/journals/lancet/article/PIIS0140-6736percent2809percent2961989-9/fulltext>. Whether proposed activities fall within the scope of the original consent "depends on what is proposed scientifically, expectations of participants, and social mores at the time of an application."

Overlap between research and medical care

- 7.17 UK Biobank states clearly that participants are not likely to benefit directly from participation; the intention is instead that research discoveries will benefit future generations and be in the broad public interest. This appeal to altruistic participation mirrors appeals to participate in blood donation programmes or early phase clinical trial research. But despite no lack of clarity in the message from UK Biobank there remains some ambiguity at its reception. For example, follow-up research in relation to other biobanks has shown that participants may regard the initial assessment as a 'health check'.⁴³³
- 7.18 This tension is particularly keen when looking at the potential to feed back health-related findings that may have significant benefit for individuals. At the initial assessment, UK Biobank participants are given results from some of the measurements taken, such as their blood pressure and weight. If staff notices abnormalities, such as elevated blood pressure or a suspicious mole, they may advise the participants to see their GP.⁴³⁴ However, participants are told at the time of recruitment that no further personal feedback would be offered either from analyses carried out by UK Biobank or by researchers using the resource. This blanket 'no feedback' policy was typical for many population and cohort studies that had been established prior to UK Biobank.⁴³⁵ Since then, however, such a policy has become more contentious in the light of discoveries that may come about through further data collection from participants or as a result of the analysis of data. It is now often argued that there is a moral obligation to consider feedback of health-related findings, including in the case of whole genome sequencing.⁴³⁶

Box 7.2: UK Biobank imaging study

UK Biobank is seeking to enhance the resource through the addition of an imaging study. In the pilot study, launched in May 2014, existing participants are invited for an integrated series of imaging studies of the brain, heart, abdomen, bones and carotid arteries.⁴³⁷ The aim is to use these data as part of an increasingly detailed ('deep') phenotyping of participants to enhance the potential to deliver research objectives of UK Biobank.

In discussion with the EGC, the UK Biobank International Scientific Advisory Board, the project's funders and others (though not subject participants), it was agreed that during the pilot phase participants and their GPs would receive feedback on any serious health-related finding that might be observed during the collection of the imaging data.⁴³⁸ Those receiving this feedback will be followed up so the consequences of being offered such

⁴³³ Halverson CME and Ross LF (2012) Incidental findings of therapeutic misconception in biobank-based research *Genetics in Medicine* **14**(6): 611-5, available at: <http://www.nature.com/gim/journal/v14/n6/abs/gim201150a.html>.

⁴³⁴ See UK Biobank Ethics and Governance Framework, <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf>, at page 7.

⁴³⁵ Wallace SE and Kent A (2011) Population biobanks and returning individual research results: mission impossible or new directions? *Human Genetics* **130**(3): 393-401.

⁴³⁶ Wolf SM, Crock BN, Van Ness B, *et al.* (2012) Managing incidental findings and research results in genomic research involving biobanks and archived datasets *Genetics in Medicine* **14**(4): 361-84. See also Medical Research Council and Wellcome Trust (2014) Framework on the feedback of health-related findings in research, available at: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp056059.pdf.

⁴³⁷ See <http://www.ukbiobank.ac.uk/2014/05/uk-biobank-imaging-study-launched/>.

⁴³⁸ Some participants were invited to comment on the information materials for the imaging study but the direction of the study was set, effectively, by the groups mentioned.

feedback can be assessed. This will provide vital evidence of the implications and consequences of receiving feedback, an area that remains very much debated worldwide.⁴³⁹

- 7.19 One of the assumptions behind offering findings to participants is that it is likely to prove of benefit through, for example, earlier diagnosis and therefore more effective treatment of a condition. There is also an acknowledgement that some people may be harmed, for example by anxiety caused by notification about a serious abnormality that may be revealed later to be of little or no consequence to their well-being. However, given the lack of evidence about the consequences of providing such feedback it is unclear what the benefits or disadvantages may be. Some have argued that it is not the right time for biobanks to institute such a policy.⁴⁴⁰ However, if a project considers revising a current, general 'no feedback' policy, this may be considered as outside the terms of the initial consent given by participants.⁴⁴¹

Commercialisation

- 7.20 One specific use of UK Biobank data that has stimulated discussion is the use by commercial entities. While research suggests that academic researchers are generally trusted by the public, industry is viewed with more suspicion owing to its supposedly more mixed motives (as we note in chapter 5).⁴⁴² The debate focuses on whether the outcomes of the research will be shared and benefits returned to the public domain instead of boosting profits of commercial entities. There may also be fears that private interests will restrict the benefits available to the public, through, for example, commercial pricing of products.⁴⁴³ It is possible that public discussion about the commercial use of the resource may have led to some people choosing not to join the study and so depressing the uptake rate for the project.

An ethical framework

- 7.21 A review of the conditions under which UK Biobank has been established suggests that it does provide a secure moral basis for the proposed uses in most respects. There was initial consultation with the public and other stakeholders in the development of the plans for the project and its governance. In order to respect the range of interests involved while at the same time acknowledge uncertainty about the specific future uses of the resource it uses a model of broad consent, together with a governance framework for the use of data that includes a criterion of public interest to ensure conformity with the expectations of the stakeholders. The EGF provides an explicit

⁴³⁹ See Johnson KJ and Gehlert S (2014) Return of results from genomic sequencing: A policy discussion of secondary findings for cancer predisposition *Journal of Cancer Policy* **2(3)**: 75-80; Hallowell N, Hall A, Alberg C and Zimmern R (2014) Revealing the results of whole-genome sequencing and whole-exome sequencing in research and clinical investigations: some ethical issues *Journal of Medical Ethics* (Online First), available at: <http://jme.bmj.com/content/early/2014/07/18/medethics-2013-101996.long#responses>.

⁴⁴⁰ Viberg J, Hansson MG, Langenskiöld S and Segerdahl P (2014) Incidental findings: the time is not yet ripe for a policy for biobanks *European Journal of Human Genetics* **22(4)**: 437-41, available at: <http://www.nature.com/ejhg/journal/v22/n4/abs/ejhg2013217a.html>.

⁴⁴¹ Trinidad SB, Fullerton SM, Ludman EJ *et al.* (2011) Research practice and participant preferences: the growing gulf *Science* **331(6015)**: 287-8, available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3044500/>; Stjernschantz Forsberg J, Hansson MG, and Eriksson S (2011) The risks and benefits of re-consent *Science* **332(6027)**: 306.

⁴⁴² Clemence M, Gilby N, Shah J, *et al.* (2013) *Wellcome Trust monitor wave 2: tracking public views on science, research and science education*, available at: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_grants/documents/web_document/wtp053113.pdf.

⁴⁴³ Huzair F and Papaioannou T (2012) UK Biobank: consequences for commons and innovation *Science and Public Policy* **39(4)**: 500-12; Caulfield T, Borry P, and Gottweis H (2014) Industry involvement in publicly funded biobanks *Nature Reviews Genetics* **15**: 220.

correlative for the expectations of participants and, as a 'living' element of governance, provides a focus for reflective governance. The challenge to UK Biobank is to ensure that this ambition is realised in practice, in response to developments, such as developments in the potential uses of data we describe in this report.

7.22 Because of the developments of the initiative (such as the introduction of imaging) and new possibilities for using the data, as well as increasing knowledge (for example, in human genetics), this process of reflection and engagement with participants, should be maintained throughout the life of the initiative. The norms and expectations, such as 'no feedback', on which initial consents were premised, and corresponding views about the level of duty of care owed to participants, may alter in relation to new information. The EGC has responded to these but there is no embedded process through which others' views can be engaged in relation to such matters as part of the revision of the ethics and governance framework and the development of practice and governance more generally.⁴⁴⁴ There is scope for bringing the views of the subject participants, the research users of the resource and the greater research community into this reflection so that both the promotion of research in the public interest and the privacy and other interests of all participants in the process are enhanced.

UK10K Rare Genetic Variants in Health and Disease

7.23 The UK10K project, established in 2010, was concerned with using genome sequencing to illuminate the genetic contribution to disease in a research culture where open access to data had been the norm. The premise of open access to genetic sequence data was established by the Human Genome Project (1999-2004) and the key principles were set out in a series of international agreements.⁴⁴⁵ This has been widely accepted and endorsed by the research community and open access with DNA sequence data deposited online became the accepted practice. This was based on the assumption that there would be no risk of re-identification of research participants who had given biological samples for sequencing. However, this assumption was overturned by a study showing that data from individuals could be distinguished in genome-wide association study (GWAS) data using only summary statistics.⁴⁴⁶ A later study demonstrated that male participants could be re-identified by linking individual mutations (single nucleotide polymorphisms, or SNPs) on the Y chromosome with data found in publicly available datasets on the Internet (see chapter 4).⁴⁴⁷ Policies then changed: some datasets have been removed from the web, and models of managed or conditional access to data have developed, of which the UK10K project is one example. However, some initiatives in the spirit of citizen science, for example the Personal Genome Project, have continued to offer open access (see paragraph 7.41ff.)

⁴⁴⁴ See, for example, Bjugn R and Casati B (2012) Stakeholder Analysis: A Useful Tool for Biobank Planning *Biopreservation and Biobanking* **10(3)**: 239-44; Lemke AA, Wu JT, Waudby C *et al.* (2010) Community engagement in biobanking: experiences from the eMERGE Network *Genomics, Society, and Policy* **6(3)**: 35-52, available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3434453/>.

⁴⁴⁵ Muddyman D, Smee C, Griffin H, Kaye J, and UK 10K Project (2013) Implementing a successful data management framework: the UK10K managed access model *Genome Medicine* **5(11)**: 100, available at: <http://link.springer.com/article/10.1186/gm504>.

⁴⁴⁶ Homer N, Szlinger S, Redman M *et al.* (2008) Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays *PLoS Genet* **4(8)**: e1000167, available at: <http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1000167>.

⁴⁴⁷ Gymrek M, McGuire A, Golan D, Halperin E and Erlich Y (2013) Identifying personal genomes by surname inference *Science* **339(6117)**: 321-4.

- 7.24 The objectives of the UK10K project are to apply genome-wide sequencing to existing research collections of patients from the UK and abroad with specific diseases (some 5,500 individuals) using comparisons with some 4,000 deeply phenotyped participants from the Twins UK and ALSPAC longitudinal cohort studies. This allows the identification of genetic sequence variants that may be associated with specific (and usually rare) diseases recorded in the phenotype data with the aim of characterising the genetic basis of diseases. In addition, a long-lasting research resource for UK and global genetic research is being established through rapid data release to the European Genome-phenome Archive, which is held by the European Bioinformatics Institute.⁴⁴⁸
- 7.25 The key instrument of governance for the initiative is, like UK Biobank, an Ethical Governance Framework (EGF),⁴⁴⁹ based on informed consent and approval from an appropriate ethics committee. This was drafted by the UK10K Ethical Advisory Board with independent advice and international (but not public) review. The Board includes members representing the interests of some patients through the patient interest group Genetic Alliance UK and cohort study participants.
- 7.26 The aim of the UK10K EGF is to enable the UK10K project to operate as a federated system. This means that the projects can work together under a common ethical framework, which can acknowledge the nuances of particular studies while still allowing them to be part of a common endeavour. At the same time, it strives to ensure that there can be sufficient harmonisation so that these very different studies can participate in an ethically-coherent project that maximises the research benefit, while acknowledging the responsibilities and obligations that are owed to research participants.
- 7.27 An important precept is acknowledging and respecting the role of the principal investigator (PI) of each collaborating study, the person (usually a senior researcher) who has management responsibility for each study. Many of the principal investigators have collected samples from research participants themselves and may have a continuing clinical and/or research relationship with research participants. This means that they may be in a good position to develop an understanding of the interests of research participants, through consultation with participants themselves and others.
- 7.28 The EGF describes both policies, by which all project members agree to abide, and guidelines, which represent best practice as it is currently understood. It deals with aspects of the project including the feedback to patients of pertinent and incidental clinically significant findings, and management pathways. Feedback may take place when a clinician believes it to be appropriate, the patient has consented and the finding has been validated to clinical standards.⁴⁵⁰ There is a diagram illustrating the process of data flow through the project and the necessary approvals.⁴⁵¹ Data access policy is described in a Data Sharing Policy Document.⁴⁵² All this information is publicly available. Access for the research community to sequence data held in the European Genome-phenome Archive is overseen by an independent Access Committee, which

⁴⁴⁸ See <https://www.ebi.ac.uk/ega/dataproviders/EGAO00000000079>.

⁴⁴⁹ See <http://www.uk10k.org/ethics.html>.

⁴⁵⁰ Kaye J, Hurler M, Griffin H, et al. (2014) Managing clinically significant findings in research: the UK10K example *European Journal of Human Genetics* **22**(9): 1100-4, available at: <http://www.nature.com/ejhg/journal/v22/n9/full/ejhg2013290a.html>.

⁴⁵¹ Muddyman D, Smee C, Griffin H, Kaye J, and UK 10K Project (2013) Implementing a successful data management framework: the UK10K managed access model *Genome Medicine* **5**(11): 100, available at: <http://link.springer.com/article/10.1186/gm504>, at page 3.

⁴⁵² See http://www.uk10k.org/data_access.html.

will only approve applications from ‘appropriately qualified’ researchers who sign a legally binding agreement, making a number of undertakings that include protecting data confidentiality, providing appropriate data security and not attempting to identify individual participants.

- 7.29 The UK10K project is more modest in resources and limited in purposes and research methods than UK Biobank, and aims to leverage existing resources with genomic sequencing to generate new knowledge. Nevertheless it shares many of its governance principles with UK Biobank. Both make use of explicit Ethics and Governance Frameworks and place considerable reliance on institutional academic regulation to ensure the probity of individual researchers. Both foreground the role of consent and recognise the challenges of interpreting it in different and changing circumstances (UK Biobank through the EGC, UK10K placing significant emphasis on the role of the principal investigator to interpret the interests and expectations of participants). Reflecting on these different approaches we make a number of recommendations below with regard to governance that are relevant to biobanks.

Recommendation 10

We recommend that appropriate mechanisms should be put in place to allow governance arrangements to evolve during the life of an initiative, through deliberation with morally relevant stakeholders including participants, the public, funders and the research community. Arrangements may include, e.g., representation of relevant stakeholder groups in the governance of the biobank; regular review of a public ethics and governance framework document legitimated through deliberation with interested parties that sets out the relationships of a biobank with participants, the research community, individual researchers, funders and the wider society. This may serve as an instrument to maintain alignment of the public interest in research with the privacy and other interests of the participants. Governance arrangements should, among other matters, outline policies for maintaining data security, the feedback of health-related findings to participants and for research access to the resource. In large scale and complex initiatives detailed diagrams of data flows should be available to support good governance. The responsibility to ensure appropriate governance arrangements are in place rests with funders.

Recommendation 11

Where broad consent is sought for the use of data additional, adaptive safeguards should be in place to secure the interests of participants over the life of a project. A possible model is provided by a publicly articulated, ‘living’ ethics and governance framework that reflects the expectations of participants and is subject to review and revision through mechanisms that involve representatives of the full range of interests of participants in the initiative.

Recommendation 12

We recommend that researchers should operate demonstrably within a local governance framework able to maintain reasonable surveillance in order to

identify inappropriate data use and administer sanctions for misuse. Researchers should be members of a recognised research environment with appropriate arrangements in place to ensure their research meets ethical standards. They should provide undertakings regarding the confidential and secure use of data and that they will refrain from any attempt to identify participants from whom data may have been derived.

International collaborative research

7.30 As noted earlier, science is becoming an increasingly global enterprise, and as the ease with which research groups can communicate, share knowledge and carry out research collectively increases, more international collaborations will be formed. International collaborative initiatives can allow the sharing of knowledge and best practice and spread research expertise and funding across both well- and less-well-supported countries.⁴⁵³ While the benefits can be great, there are significant difficulties to be faced. For example, scientists in one country cannot ‘police’ the activities of those in another country as there may be differing national laws and governance frameworks that may prevent single policies being imposed.⁴⁵⁴ This requires conducting science in a way that provides accountability both at the local and consortium level, while respecting local legal, ethical and cultural norms.

Box 7.3: Examples of international collaborative research involving genetic data

International Cancer Genome Consortium

The International Cancer Genome Consortium (ICGC) coordinates large-scale cancer genome studies in tumours from 50 different cancer types and/or subtypes that are of clinical and societal importance across the globe.⁴⁵⁵ As of May 2014, 74 projects representing over 17 countries and jurisdictions had sequenced over 25,000 cancer tumour genomes. Samples are held by each member project, while data is deposited in a central repository located in Toronto, Ontario. The project distinguishes two ‘types’ of data. Open access data, which does not contain obvious personal identifiers, is available from the ICGC Data Portal.⁴⁵⁶ Controlled access data, which is more readily identifying, is available to authorised researchers for approved research through the ICGC Data Compliance Office.⁴⁵⁷ After approval the researcher is able to download the data onto their own system for analysis.

The Psychiatric Genomics Consortium

The Psychiatric Genomics Consortium (PGC) is an international initiative with over 500 investigators from over 80 institutions in 25 countries. Its purpose is to conduct mega-analyses (individual-level data meta-studies) of genome-wide genetic data for

⁴⁵³ For example, the Global Alliance for Health and Genomics has created a Framework for Responsible Sharing of Genomic and Health-Related Data, <http://genomicsandhealth.org/>; the Human Heredity and Health in Africa (H3Africa) Initiative is a cross-continental initiative to support African researchers and improve African health, <http://h3africa.org/>.

⁴⁵⁴ Romeo-Casabona C, Nicolás P, Knoppers BM, *et al.* (2012) Legal aspects of genetic databases for international biomedical research: the example of the International Cancer Genome Consortium (ICGC) *Revista de Derecho y Genoma Humano/Law and the Human Genome Review* **37**: 15-34.

⁴⁵⁵ The International Cancer Genome Consortium (2010) International network of cancer genome projects *Nature* **464(7291)**: 993-8. See: <https://icgc.org/>.

⁴⁵⁶ Zhang J, Baran J, Cros A *et al.* (2011) International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data *Database* **2011**: bar026, available at: <http://database.oxfordjournals.org/content/2011/bar026.long>.

⁴⁵⁷ Joly Y, Dove ES, Knoppers BM, Bobrow M and Chalmers B (2012) Data sharing in the post-genomic world: the experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO) *PLoS Computational Biology* **8(7)**: e1002549, available at: <http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1002549>.

psychiatric disorders. It is the largest biological experiment in the history of psychiatry.⁴⁵⁸ The PGC data repository is located in the Netherlands. All phenotype and genotype data is stored there and all analyses of the data are carried out on its Genetic Cluster Computer.⁴⁵⁹

- 7.31 Similarly to biobanks, because of the developing knowledge environment, flexible and continually reviewed governance mechanisms are needed to guide the science while protecting the interests of participants. However, because of the diversity of their membership such consortia have to rely on agreements among members, peer pressure and the limited sanctions that can be imposed at the local level, such as cessation of funding. In the early days, ICGC members agreed to a set of overarching policies, together with flexible guidance that could be followed if desired.⁴⁶⁰ It was agreed that projects, which must obtain participants' consent for whole genome sequencing, could be flexible regarding the return of individual health-related findings.⁴⁶¹ This contrasts with the more formal ethics and governance frameworks previously discussed and highlights the reluctance of researchers to impose a single governance structure over many different national and regional groups, a reluctance not shown in other sectors (finance, for example).
- 7.32 While ICGC was, for many, a prospective study, the PGC Schizophrenia Working Group uses existing data from a number of previous studies that were carried out in different countries to identify the genetic variants which may confer genetic risk for individuals.⁴⁶² Retrospective studies such as this present an obvious challenge to conventional research governance arrangements because the data were originally collected from participants in a number of different countries, and in varying circumstances. Furthermore, the extensive repurposing, data linking and analysis are carried out by a research collective whose members are themselves based in institutions in a number of countries. Gaining consent for international extensions of data access, or even disclosure outside a single institution, is actually a relatively recent circumstance for researchers to consider, let alone subject participants, and it is unclear whether the original consents would have been informed by foresight of such extensive re-use.
- 7.33 It is especially difficult to elicit participants' expectations, because deliberative engagement is extremely difficult at the international consortium level. Even if such procedures were undertaken for each local project, it would be very difficult to generalise results across sites. Yet it is important that participants and publics are able to obtain information about what is being done with the data they have provided. Specific investigations are therefore needed to assess the impact of cross-border research and how best to verify consent for extended uses, as well as how to disseminate the results of research in as transparent and accessible but secure a way as possible. Similarly, policy bodies and funding agencies need to be clear under what circumstances data can be re-used.

⁴⁵⁸ <http://www.med.unc.edu/pgc>.

⁴⁵⁹ <http://www.med.unc.edu/pgc/documents>.

⁴⁶⁰ See <https://icgc.org/icgc/goals-structure-policies-guidelines>.

⁴⁶¹ Wallace SE and Knoppers BM (2011) Harmonised consent in international research consortia: an impossible dream? *Genomics, Society and Policy* 7: 35-46, available at: <http://www.lssjournal.com/content/pdf/1746-5354-7-1-35.pdf>.

⁴⁶² Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci *Nature* 511(7510): 421-7, available at: <http://www.nature.com/nature/journal/v511/n7510/full/nature13595.html>.

7.34 Another concern is the security of data, as data protection regimes are not the same across all countries.⁴⁶³ Global commerce has been dealing with this issue for many years, and it has been suggested that the research and clinical communities can learn from their experience.⁴⁶⁴ For example, the adequacy test asks if the level of protection in the jurisdiction that receives the data is comparable to that of the origin of the data with the implication that if it is, personal data may be transferred with confidence.⁴⁶⁵

Cloud storage and computing

7.35 As more and more data become available, larger data analysis projects are being undertaken. These use the Internet to access appropriate technologies and operating power to transfer files, provide storage and drive analyses. Although commonly used in corporate settings, cloud computing is still a relatively new and potentially confusing concept to many in the research setting. It raises fears of loss of privacy due, for example, to a lack of clarity about responsibilities for data protection. Cloud services can be layered, with one provider being responsible for software while another is responsible for infrastructure.⁴⁶⁶ With competing responsibilities there is a fear that adequate protections may not be in place for secure data processing. As already mentioned, if a provider is located in a different country, the data may be subject to a different data protection scheme, possibly one of lesser stringency.

7.36 The location of data repositories is now an issue of potential concern. This led, for example, to some countries declining to participate in the Type 1 Diabetes Genome Consortium, owing to study requirements for the processing of samples at network laboratories and/or final deposition of samples in US-based Central Repositories.⁴⁶⁷ Of particular concern is that many companies offering cloud computing facilities are based in the USA and therefore all data is subject to the homeland security legislation that allows access to data by the NSA (see chapter 2 above). The use of cloud computing in research highlights the difficulties of balancing the desire to analyse large datasets from around the world to advance scientific discovery with the risks of the potential loss of the confidentiality of data.

Box 7.4: International Cancer Genome Consortium PanCancer Analyses of Whole Genomes (ICGC PCAWG)

The ICGC PCAWG will study the whole genome sequence from tumours and matched samples (usually blood) from an estimated 2,000 patients internationally who have been recruited to ICGC member projects. Demographic, clinical and pathology data will be available for all 2,000 matched samples. The dataset on which the analyses will run is expected to exceed one petabyte of data. Examining and comparing data across cancers internationally is now possible due to the large number of cancer genome

⁴⁶³ See, for example, The Organisation for Economic Co-operation and Development (2013) *The OECD privacy framework*, available at: http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf.

⁴⁶⁴ Kosseim P, Dove E, Baggaley C *et al.* (2014) Building a data sharing model for global genomic research *Genome Biology* **15(8)**: 430, available at: <http://genomebiology.com/2014/15/8/430>.

⁴⁶⁵ For this purpose the US is deemed to be a 'safe harbour' by the EU as US organisations have voluntarily self-certified that they will comply with mutually agreed-upon data protection principles. See: http://ec.europa.eu/justice/policies/privacy/thirdcountries/adequacy-faq1_en.htm. However, see chapters 2 and 4 regarding concerns and evolving case law about jurisdictional differences.

⁴⁶⁶ Information Commissioner's Office (2012) *Guidance on the use of cloud computing*, available at: http://ico.org.uk/for_organisations/guidance_index/~media/documents/library/Data_Protection/Practical_application/cloud_computing_guidance_for_organisations.ashx.

⁴⁶⁷ Hilner JE, Perdue LH, Sides EG *et al.* (2010) Designing and implementing sample and data collection for an international genetics study: the Type 1 Diabetes Genetics Consortium (T1DGC) *Clinical Trials* **7(1 suppl)**: S5-S32, available at: http://ctj.sagepub.com/content/7/1_suppl/S5.short.

tumours that have been sequenced and made available in accessible form. As no one site will have sufficient capacity to host the project, using a cloud environment is being explored, if it can be shown to be consistent with the ethical and legal requirements of the ICGC.

Several cloud providers, rather than one, may be used, based in different countries in North America, Europe and Asia. Annai Systems, an academic cloud based in the US, was approved for use by the ICGC Executive in November 2014 and ICGC data is already being mirrored, so far for five projects in its cloud. Any provider's Terms of Service will be reviewed and agreed by the ICGC. A small number of ICGC Portal staff and PCAWG working group members will align and annotate the data to create a uniform dataset. Only ICGC PCAWG team members who have received approval will be allowed access to the dataset. The dataset will be removed from the cloud providers after the analysis is completed and archived at the EGA.

- 7.37 At least four benefits to using cloud services have been identified for international collaborative biomedical research: lower costs, as one 'rents' space rather than purchases it; better data security, as such providers have the money to invest in state-of-the-art security mechanisms; increased data storage capacity; and lower environmental impact, as a resource is being reused rather than newly constructed.⁴⁶⁸ However, the terms of service of many of these providers have not necessarily been developed with specific attention to the needs and sensitivities of biomedical research.⁴⁶⁹ There may therefore be a gap to close through the research community (including participants) working with providers to agree control of the data, security measures (such as appropriate encryption), and access to the data.
- 7.38 Cross-border data access and transmission and the use of cloud services should provoke research studies to review the ethical and legal implications, particularly where they are introduced to existing projects. For example, using cloud providers was not considered at the beginning of the ICGC and is not included specifically in consent documents. Seeking specific consent for this use from the 2,000 participants from multiple countries would be unfeasible. Through its oversight committees, the ICGC has approached this problem by working with cloud suppliers who will design systems that will provide for the needs of the scientific community.
- 7.39 One US-based market intelligence firm has predicted that by 2020 80 per cent of all health care data will pass through a cloud provider at some point and that cloud-based products will increasingly be used to manage costs and enable the analysis of the increasing amounts of health-related data becoming available.⁴⁷⁰ This could be simply one more standard technology that will be commonly used. However, it is not clear that there is a high level of understanding amongst the general public, and indeed researchers and health care administrators, of the implications and, therefore, the moral relevance of cloud technologies.
- 7.40 Details of how data initiatives use cloud systems need to be disseminated and discussed in the public arena, allowing any misconceptions to be explored and facts

⁴⁶⁸ Dove Es, Joly Y, Tassé A, et al. (2014) Genomic cloud computing: legal and ethical points to consider *European Journal of Human Genetics* (advance online publication), doi:10.1038/ejhg.2014.196.

⁴⁶⁹ Ibid.

⁴⁷⁰ <http://www.idc.com/getdoc.jsp?containerId=prUS25262514>.

explained so that this technology can be used transparently and with appropriate safeguards. Any research-based data initiative seeking to use such technologies should discuss this with study partners and, if possible, potential participants for acceptability. This would allow prospective initiatives to include details of cloud use in consent materials as well as the governance framework. Any agreements with providers will need to be tailored to ensure that data will be kept secure from breaches of privacy and reviewed regularly. As norms in research practice change, it may be that the use of cloud providers will no longer be seen as contentious. But this will only happen with detailed examination of the issues and public debate, which will help us recognise morally reasonable expectations and formulate appropriate governance and oversight mechanisms.

Recommendation 13

We recommend that all international collaborative data research initiatives should operate within an explicit, public ethics and governance framework that has agreement from the initiative's constituent partners. International collaborators should be able to demonstrate that they can fulfil recommendation 12 by applying equivalently strong governance standards (using legal and other mechanisms available in their national jurisdiction).

Recommendation 14

We recommend that all partners in international collaborations integrate the provisions of the ethics and governance framework (EGF) agreed by the initiative as far as possible at their local research site. The partner should ensure that they adhere to the EGF, for example by ensuring participants have given appropriate consent for the use of data and samples in the initiative and that they are informed of potential transfer across borders.

Recommendation 15

We recommend that national bodies publish their policies on the use of cloud services in health data settings so that data initiatives can include this in their decision making and interactions with publics and participants.

Open data

7.41 There are a number of projects that involve uploading individual genomic data and other data to the world wide web so that it becomes freely available to anyone to use for any purpose. The best known such initiative is the Personal Genome Project. This was initiated by the prominent Harvard University genomics researcher, George M. Church in 2005.⁴⁷¹ The Personal Genome Project is a long-term cohort resource that aims to publish the genome sequence, medical records and various other measures

⁴⁷¹ Church GM (2005) The Personal Genome Project *Molecular Systems Biology* 1: 2005.0030, available at: <http://msb.embopress.org/content/1/1/2005.0030>; Lunshof JE, Bobe J, Aach J, et al. (2010) Personal genomes in progress: from the Human Genome Project to the Personal Genome Project *Dialogues in Clinical Neuroscience* 12(1): 47-60, available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181947/>.

such as MRI images of 100,000 volunteers so that the data are freely available to anyone who chooses to use them and to enable citizen science. George Church himself, together with other prominent figures in the biotechnology industry, genome science, and science policy made up the initial 10 participants, the 'PGP-10'. (The Harvard University Medical School Review Board that considered the project had requested that the first group of volunteers included Church himself and other stakeholders in genomic science.) Today more than 3,500 volunteers have joined the USA study and additional studies have been established in Canada (2012) and UK (2013), with others planned. There is a long waiting list of potential volunteers, with over 10,000 people registered in the UK within three months of the project launch, although there is a significant lag, due to funding constraints, in generating and uploading sequence data.

7.42 Participants in the PGP go through a different recruitment process to many conventional research projects. Firstly, there is no promise that the identity of individuals will remain anonymous as the whole purpose of the project is to make sequence data freely available. To be accepted, volunteers must be over 21 and pass an examination to test whether they are aware of the potential risks to participation – including possible discrimination by insurers and employers. If accepted for inclusion in the study, participants are required to contribute a sample from which a genome sequence will be produced and encouraged to upload other kinds of medical information. Before the sequence information is deposited on the project website, they have thirty days to review the data and make a decision whether they want it to be made public. If they decide to withdraw from the project during this period their data will not be publicly released. However, if data are put on the web and participants later decide to withdraw from the study, already released information will remain publicly available and only future information will not be released. Participants are asked to report any discrimination or harm that they experience as a result of participation in the project. There is a continuing relationship and engagement between participants and the project. Participants in PGP (USA) were required initially to do this on a quarterly basis, but this was reduced to six-monthly as it was felt to be too onerous for participants who had nothing to report. Participants in the PGP may be viewed as 'information altruists' who are prepared to allow their genome sequence to be made public.⁴⁷² Although this level of openness is not for everyone, the positive response to the launch of the UK and the Canadian arms of the project suggests that such projects do have public appeal.

7.43 The open publication of data, as exemplified by the Personal Genome Project, is a limit case for the governance of data for research. Nevertheless, it is not meaningless to ask what the morally reasonable expectations of participants may be. In terms of what they may expect the limits to data use to be, the answer will depend on public norms rather than those maintained in the context of a specific data initiative, and on governance by law and the conventions of public morality. But while the expectations may not be bounded (indeed, participants are urged to contemplate the worst that can reasonably be imagined) they may nevertheless have some positive content. Subject participants may expect, for example, that their supposedly altruistic gesture should be answered by a commitment on the part of the PGP organisation actively to secure the best use of the data to advance scientific knowledge (for example, by ensuring the

⁴⁷² See also Kohane IS and Altman RB (2005) Health-information altruists – a potentially critical resource *New England Journal of Medicine* **353**: 2074-77.

quality, accessibility and interoperability of the data published). Beyond that, the PGP implicitly poses a challenge to societies to affirm as a norm that the relevant rights of altruistic subject participants will be protected. To do this is to abandon the 'arms race' of developing ever stronger data security measures and rely instead on regulating the conduct of data users, not purely within the context of 'bona fide' research, subject to institutional codes and penalties, but generally, under public morality and the rule of law.⁴⁷³

Citizen science and participant-driven research

7.44 Increasing access to digital technologies and the rise of online social networks has facilitated the formation of communities of people engaged in establishing and conducting health research including self experimentation, self surveillance, analysis of genomic data and genome-wide association studies.⁴⁷⁴

Box 7.5: PatientsLikeMe

Founded in 2004, the largest participant-driven research network, PatientsLikeMe (PLM) has more than a quarter of a million members representing over 2,000 health conditions. Through this company ('Live better, together') people connect with others who may have the same disease or condition, and track and share their own experiences. In doing so they generate data about the real world nature of disease that can help researchers, pharmaceutical companies, regulators and health providers develop more effective products, services and care. PLM allows members to contribute their own data about their conditions (treatment, history, side effects, hospital episodes, symptoms, function scores, weight, mood, quality of life, etc.) on a continuing basis. The resulting longitudinal record is organised into charts and graphs that allow members to identify patterns, gain insight and place their experiences in context, as well as to see what treatments may have helped other patients like themselves. The website also gives members lists of relevant clinical trials and they can search the site for trials for which they may be eligible. The company also offers a commercial service to actively message potential participants for specific clinical trials.

PLM describe their four core values as follows⁴⁷⁵:

- Honour the trust patients put in us – patients trust the company to protect their health data and to use it to advance knowledge of their disease.
- Transparency. The company aims for 'no surprises'. It discloses its business partnerships, what it does with patient's data and how the company makes money.
- Openness. The company believes that sharing health information openly has potential to benefit patients.
- Create 'wow'. This is a goal for what patients should feel when they visit the website.

The company has a team of in-house researchers who produce many (peer reviewed) papers and also a number of collaborative partnerships with academic research

⁴⁷³ There are two routes to this: general data protection legislation and anti-discrimination legislation. A number of legal instruments give protection against discrimination and the existence of measures comprising the 'welfare state' offers some practical insurance against the effects of discrimination. Although the Equality Act 2010 prohibits discrimination on the ground of health status, it does not explicitly include genetic status as a 'protected characteristic'; in response to the green paper that foreshadowed the Act, many, including the UK Human Genetics Commission, argued that it should. See <http://webarchive.nationalarchives.gov.uk/20100419143351/http://hgc.gov.uk/client/document.asp?DocId=134&CAtegorId=4>.

⁴⁷⁴ Vayena E and Tasioulas J (2013) Adapting standards: ethical oversight of participant-led health research *PLoS Medicine* **10(3)**: e1001402, available at: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001402>.

⁴⁷⁵ <https://support.patientslikeme.com/hc/en-us/articles/201245710-What-are-the-company-s-core-values->.

groups.⁴⁷⁶ It is run as a for-profit company that makes money by selling data uploaded by patients to other companies.

- 7.45 An early and influential example of PLM's research was their amyotrophic lateral sclerosis (ALS) lithium study. ALS is a progressive and incurable disease. In 2008 a small Italian study suggested that lithium carbonate could slow the progression of ALS. In response to this, many members of PLM began taking the drug. Two members with advanced stage ALS (from Brazil and the USA) initiated a study using self-generated data from members on the platform to test these findings. (Both died before the study was completed). The nine-month study indicated that lithium did not slow the progression of the disease, a result that was later confirmed in four randomised controlled trials.⁴⁷⁷
- 7.46 Patient-led and participant-driven research (PLR and PDR) is gaining wider recognition as a potential source of generalisable health knowledge that benefits both participants and society more widely, and that can realise the values of solidarity among communities of patients suffering from a common disease. It can complement conventional research on conditions, or on aspects of them, that may have been neglected. The researchers involved have claimed that it can speed up clinical discovery, and could potentially maximise it, setting a stage for better trials with more engaged participants.⁴⁷⁸ However, this may require new governance arrangements.
- 7.47 Like any clinical research, PDR can involve the risk of harms to participants or their relatives, including children. Self-experimentation can lead to participants taking excessive risks. Furthermore, the existence of a strongly solidaristic patient community may create or allow undue peer pressure or even exploitation. Conventional research has both scientific and ethical oversight, which facilitates the production of generalisable health knowledge that can be used by participants and society more widely. Research conducted outside the conventional academic and commercial institutions may not be subject to such oversight and study reports may not meet the basic acceptance criteria for peer-reviewed journals. While some participant-driven research may involve collaborators within the conventional system, thus bringing it within its ambit, much does not. However, trying to force this research into the conventional mould may stifle the very features that could make it so valuable.
- 7.48 While all forms of scientific research involving human participants should be subject to ethical as well as scientific appraisal, the appropriate standards for ethical oversight need to be adapted to the distinctive features of PDR. There have been calls for a broad dialogue to address the issues and to generate consensus on best practice as well as warnings that a failure to do this may pose threats of harms to participants,

⁴⁷⁶ <http://www.patientslikeme.com/research/publications>.

⁴⁷⁷ Wicks P, Vaughan TE, Massagli MP and Heywood J (2011) Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm *Nature Biotechnology* **29(5)**: 411-4, available at: <http://www.nature.com/nbt/journal/v29/n5/abs/nbt.1837.html>; Wicks P, Vaughan TE, and Heywood J (2014) Subjects no more: what happens when trial participants realize they hold the power? *British Medical Journal* **348**: g368, available at: <http://www.bmj.com/content/348/bmj.g368>.

⁴⁷⁸ Wicks P, Vaughan TE, and Heywood J (2014) Subjects no more: what happens when trial participants realize they hold the power? *British Medical Journal* **348**: g368, available at: <http://www.bmj.com/content/348/bmj.g368>, at page 2.

risks of undermining the credibility of PDR, and may provoke a backlash of over-regulation.⁴⁷⁹

Recommendation 16

We recommend that biomedical researchers give consideration to arrangements that will maximise the potential of participant-driven research to contribute to generalisable health knowledge and secure public benefits while providing adequate protection of those involved through continuing ethical and scientific appraisal. Key stakeholders are citizen patient researchers, biomedical research bodies, research funders and journal publishers. All stakeholders should encourage optimal use of human studies for improved health outcomes.

Conclusion

7.49 In view of the rapidly increasing importance to the research community of extending access to data, and the benefits that such research can bring to the public at large, developing best practice for the collection, governance and use, and extension of access to data in biological research and health care should be a very high priority across both research and clinical settings. Some work has been done and is continuing by international organisations such as the Public Population Project in Genomics and Society (P³G) which have brought together best practice regarding population research and the Global Alliance for Genomics and Health which has created a Framework for Responsible Sharing of Genomic and Health-Related Data.⁴⁸⁰ But more needs to be done at the level of individual patients and research participants and respect for their circumstances and protection of privacy must be at the centre of such systems. There are many stakeholders in the collective enterprise of health promotion and medical treatment but to marginalise those individuals who provide data for research will be to risk the trust of current and future generations.

Recommendation 17

We recommend that the research community, including all academic and commercial partners, data controllers and custodians, public services and government agencies, actively foster opportunities to create a more explicit and reflective foundation for extending data access in the public interest. We urge all stakeholders in the medical research enterprise to continue to develop robust and comprehensive, yet efficient privacy protecting rules, guidelines and measures. Among other things these should aim at:

- Providing greater clarity for members of the public about ways that their biomedical data are, and may be used in the future, along with a realistic acknowledgement that no system can guarantee privacy and confidentiality in all circumstances.
- Securing commitments from data controllers to a responsible approach to the extension of data access as part of their core mission statement; they must publish information about their approach to data access, transparency and accountability, and whether, and on what terms, they will consider extending access to data.

⁴⁷⁹ Vayena E and Tasioulas J (2013) Adapting standards: ethical oversight of participant-led health research *PLoS Medicine* **10(3)**: e1001402, available at: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001402>.

⁴⁸⁰ www.p3g.org; <http://genomicsandhealth.org/about-the-global-alliance/key-documents/framework-responsible-sharing-genomic-and-health-related-data>.

- Demonstrable and continual improvement of collection, storage and data access procedures against explicit standards for accuracy, reliability and security.