Nuffield Council on Bioethics: The linking and use of biological and health data

Response by the Wellcome Trust

January 2014

Key Points

- The linking and use of biological and health data is central to a wide range of biomedical research. There is a strong need to ensure a legal and regulatory environment in which these data can be used and linked for the purposes of research that aims towards improving health.

- Rapid advances in biomedical technology, data storage, interpretation and analysis are creating a pressing need for open dialogue about the potential benefits of such research for individual and public health, and the ethical, legal and social implications of these innovations on individuals and society. We welcome this consultation as an opportunity to inform this dialogue.

- There are legitimate ethical concerns about the privacy implications of the use and linking of such data and future developments in using these data cannot necessarily be anticipated. Participants' wishes are paramount, and there is a need for strong governance mechanisms for the use of participant data in research that can appropriately protect participant confidentiality.

Introduction

1. The Wellcome Trust is a major funder of biomedical research in the UK and abroad. We support several large cohort studies, which collect data from large numbers of individuals over time, and which bring together biomedical, health and other types of data to enhance our understanding of health and disease. We are also a major funder of medical humanities research which seeks to explore the historical, ethical and social context of advances in biomedicine. We support a broad range of work to engage the public on these issues and promote discussion and debate.

2. As a research charity dedicated to achieving extraordinary improvements in human and animal health, we are committed to ensuring that the data outputs generated by the research we support can be accessed and used in a way that maximises their value in progressing research and its application for health benefit, whilst protecting the privacy and confidentiality of research participants in line with the consents given.

3. We are taking forward a broad range of activities in partnership with other organisations to enhance the availability of research data and address the associated infrastructural, cultural and ethical issues. Key initiatives include:
   - The Expert Advisory Group on Data Access (EAGDA) [1] – which was established in 2012 by the Medical Research Council, Economic and Social Research Council, Cancer Research UK and the Wellcome Trust to provide strategic advice on emerging scientific, ethical and legal issues in relation to data access for studies

[1] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/EAGDA/index.htm

across genetics, epidemiology and the social sciences. EAGDA is also exploring models of good governance for data access (see paragraphs 13-14, 21, 45).

- The Public Health Research Data Forum – which brings together funders of global health research to increase the availability of health research data in ways that are equitable, ethical and efficient.

- Clinical trial data – we are engaged actively in discussions with other stakeholders aimed at increasing the transparency of datasets resulting from clinical trials.

4. In this response, we consider patient health information and clinical trials data to be included in the category of 'biomedical' data.

Q1. Do biomedical data have special significance?

5. We do not believe it is necessarily appropriate or useful to treat 'biomedical data' as a distinct class of data. It is certainly true that some types of biomedical data are highly sensitive, and need to be subject to robust controls to protect confidentiality and privacy. But the same is true for many other types of personal data.

6. As with other types of personal data, the manner in which biomedical data are regulated and governed needs to be proportionate to the level of risk to data subjects, and this is highly context dependent. There are cases in which biomedical data may be shared openly (for example, where fully-anonymised, non-identifiable and reported in aggregate) and others where strict security and access controls are required, for example where datasets contain strong phenotype information that could be linked to identifiable individuals. We would argue that simply falling into the category biomedical data does not in itself confer a special significance.

7. We do understand, however, that there is a perception among some elements of the public that biomedical data, particularly genomic data, can be different to or more sensitive than other kinds of personal data, for example, financial or consumer data. Commissioned research and stakeholder dialogues supported by the Wellcome Trust during 2013 explored public attitudes towards different types of personal data and revealed a complex picture of varying perceptions, levels of trust in research, and awareness of how biomedical data can be used in research (see also paragraph 39).[2] A key concern evident from this research was the potential for commercialisation, with a strong sense from participants that companies should not profit from buying and selling personal data, of all kinds. It is imperative that further work is done to engage the public on these issues if trust in biomedical research and the use of participant data is to be maintained.

8. As a funder of genomic research, we are aware that genomic information is often perceived to be more deterministic, and therefore more revealing of personal and health characteristics, than it actually is. In addition, although genomic data are intrinsically unique to each person, it does not follow that individuals are straightforwardly identifiable from those data: substantial levels of technical expertise, time, resources and the availability of other relevant data would be required to identify an individual using their genomic data. We believe that *in principle*, genomic data ought to be categorised and treated in the same way as other types of sensitive personal data. However, we recognise that to ensure essential public confidence and

[2] Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_grants/documents/web_document/wtp053205.pdf

trust in research that handles genomic data, public sensitivities ought to be taken into account in the development of governance systems for such research.

9. Traditional models of one of the most prominent ethical issues in research, informed consent, have been challenged by some types of biomedical research, as in particular cases it may be possible to draw inferences about individuals' relatives from their medical information, for example, relating to disease risk. This necessitates a modification of consent processes to acknowledge and understand the possible implications for others of using and analysing an individual's biomedical data. Whereas in the past this issue has been largely restricted to rare diseases, the use of genomic data means that inferring information that has implications for relatives and future relatives of the individual may be possible much more frequently.[3] This has long been recognised as a challenge for research and clinical practice involving genetic testing. It is standard practice to inform participants of these potential implications during the consent process for clinical genetics and research involving genomic data. Consent processes must be upfront about these potential implications, and acknowledge that there may be currently unknown implications for the individual and his/her relatives arising from their genomic data. We do not therefore consider that genomic data present qualitatively different ethical challenges for research, but rather that they are greater in scale and extent (see response to question 2).

10. At present, genomic data are mainly gathered in the context of research projects and specific clinical settings, for example in assessing inheritance risk for some rare genetic diseases. In these situations, participants and patients explicitly opt-in and consent to their genomic data being analysed and used. In the future, as the collection and use of genomic data becomes more mainstream and integrated into the NHS, a substantial amount of work will need to be undertaken to establish what the appropriate consent mechanisms ought to be. These should be informed by public engagement activities examining views on the sensitivity of genomic data (see also paragraph 63). The 100,000 Genomes Project being developed by Genomics England, which will use active, opt-in consent, could provide an excellent opportunity to inform this process.

Q2. What are the new privacy issues?

11. We consider that the privacy issues raised by the linking and use of biological and health data are different in scale and frequency, but not different in kind, to those already widely acknowledged in biomedical research. As a greater number of data sources are analysed and linked in more and different ways, it is likely that there will be increased scope for re-identification of data subjects in some cases. We are not aware of any evidence to suggest that qualitatively different privacy concerns are being raised as a result.

12. Particular concerns have arisen regarding the possibility of re-identifying individuals from genomic and other datasets. In 2013 a paper in *Science* indicated that anonymised genomic data could be combined with other publicly available information to yield participant identities.[4] It should be noted that the risk of breaching participant privacy is not unique to genomic studies, and the interpretation of genomic data requires substantial expertise and resources. However, technologies are developing

[3] It may also be possible in some circumstances to draw inferences about individuals' relatives from other information such as postcodes/addresses: genomic data are not qualitatively different from other kinds of data related to an individual in this respect.
[4] http://www.sciencemag.org/content/339/6117/321.short

rapidly: both the techniques being devised to hack or misuse data and those to mitigate the risk of data security breaches are increasing in sophistication. It is imperative that data governance bodies and those responsible for managing genomic data maintain vigilance and awareness of these techniques as they emerge.

13. EAGDA has examined the risks of re-identification in the UK from genomic studies, consulting with experts in the field of genetics, computer security and statistical disclosure risk. It has consulted with the Information Commissioner's Office to produce a statement on identifiability that addresses the risks and provides several recommendations for funders, concerning:
    - Modifications to participant consent for genomic studies. As the technological feasibility of re-identifying research participants from anonymised data is increasing, anonymity cannot in some cases be fully guaranteed and consent processes should acknowledge the technical risk of re-identification;
    - Reviewed assessments of the risk of re-identification based on actual (rather than hypothetical) risks – taking into account the likelihood and consequences of re-identification under current conditions. Risks will depend on the type of data used in the study and the possibilities of linking to other potentially identifying data;
    - Flexible control of data access mechanisms in light of the changing potential for anonymised data to be rendered identifiable;
    - Clarity on the availability and use of legal and funder sanctions against attempts to re-identify individuals without their consent.

14. Informed consent has typically been used as the mechanism by which individuals can maintain control over their privacy, determining the conditions under which their data can be collected and used. However, EAGDA considers that good data governance can enable broader, more flexible consent to be given (see response to question 7).

15. It is our view that in every circumstance participant re-identification may not always be preventable, but that the risks of re-identification can be well-managed. These risks can be reduced through good data governance, controls on access to potentially identifiable data, clear protocols for data handling and strong sanctions against those who attempt to re-identify individuals.

16. To understand and address the risks of re-identification, there is a pressing need to conduct research to establish what the actual harms resulting from the re-identification of research participants from genomic or other biological or health data would be, and discussions are underway as to how best to undertake this complex research. In our view, the re-identification of individuals is a breach of confidentiality in itself, but it is an important to consider what kinds of harms could accrue in order to develop strategies to mitigate these in case of a data breach.

Q3. What is the impact of developments in data science and information technology?

17. Developments in data science and information technology are having a significant impact on the way biomedical research is carried out, and data are analysed, shared and disseminated. Much data is currently in silos, divided by disease type, institution, platform of analysis or method of collection. In short, many research datasets are largely untapped resources. We consider that developments in data science and technology have the potential to enable greater linkage between these datasets, and this has tremendous potential to lead to new insights in the understanding of disease and human health (see **Box 1**).

Research design and funding

18. Many funders now have explicit expectations that data generated by the research they support should be shared in a timely and responsible manner, with as few restrictions as possible whilst being strictly in line with participants' consents. The Wellcome Trust, in common with other funders, requires that researchers generating datasets of value submit a data management and sharing plan as part of grant applications – and we commit to meet the costs of implementation (see also paragraphs 32-33).

19. In addition to supporting data sharing as an integral component of the research we support, funders are also making significant investments to build key data resources and enable data re-use. For example:
    - A consortium of funders, led by MRC, has funded the Farr Institute of Health Informatics Research, a network of e-Health Information Research Centres (HIRCS). The Institute will provide the physical and electronic infrastructure to support collaboration across the HIRCs. These will build on existing strengths in health informatics to: enable linkage and analysis of anonymised health and health-related datasets; develop and refine research methodologies; and build research capacity in UK health informatics research
    - Our recent funding calls for "Sustaining Health" encouraged proposals with the potential to unlock the power of data by making it more relevant, available, accessible and useful (e.g. by formatting to permit inter-linkage).

Barriers to developments

20. Significant barriers do exist to development and innovation in using and linking biomedical data more widely. Key constraints include:
    - Infrastructural issues – building and sustaining the databases and tools needed to store, analyse and make data available;
    - Cultural issues – developing a culture in the research community that supports ethical data sharing, underpinned by appropriate incentives;
    - Technical issues – developing the data standards, metadata formats and platforms needed to allow data to be used and linked effectively;
    - Professional issues – providing the skills, training and career structures needed to equip researchers and key support staff to manage and share data;
    - Ethical and legal issues – setting in place the safeguards and governance systems needed to protect research participants, and enabling data sharing across borders;
    - Engagement – much needs to be done to engage the public, policy makers, law makers and health professionals, particularly GPs, on why researchers (both academic and commercial) want to use and link data. It is also important to openly discuss the safeguards, governance and sanctions that are in place to ensure the security of the data being used in research (see paragraphs 37-39; 61-63).

21. The Trust is working in partnership with other funders to address these challenges, and to harmonise policies both across studies and jurisdictions. A key challenge for data linkage at present is the divergence of data access procedures across studies, and EAGDA is currently working to develop recommendations to ensure consistency and interoperability. EAGDA is also taking forward pieces of work to explore the need for new incentive structures within the research community to enable data sharing.

22. Data sharing across borders also creates substantial issues for research governance and management, as there are differing privacy and confidentiality rules, and safeguards on data sharing between jurisdictions. Initiatives such as the P3G consortium have been set up in part to address these issues on an international scale and promote harmonisation of standards for data sharing in genomics.[5]

23. In the past few years there has been increased political attention focused on increasing the research value of datasets and the potential gains to be made from linking datasets on a large scale: so-called "big data" initiatives. We welcome these developments. Of course, we need to be aware that there is a risk of over-hyping the potential of these initiatives and that this could lead to some unrealistic expectations. However, at the current time, we believe that the benefits in terms of unlocking the value contained in data for health outweigh these potential concerns.

Significant developments

24. There are significant opportunities for research, as well as perceived drawbacks, to the recent moves by Government towards improving accessibility to administrative data (through the new Administrative Data Research Centres) and also creating a centralised collection of patient records and health information (the Health and Social Care Information Centre). These data have the potential to be linked in novel and valuable ways, for example, in identifying trends in the relationship between socio-economic and health conditions. It would not be possible to identify such trends and relationships without sharing different types of data on a national and international scale. Such linkage already occurs in parts of Scandinavia, Australia and Canada.

25. In the field of genomics, a new "Global Alliance for Genomics and Health" has been formed to develop the data standards needed to link genomic and associated clinical data internationally and to consider how to approach the ethical, legal and social issues associated with doing this.[6] The Alliance has five working groups, four of which are examining key issues in the development of data sharing infrastructure: security, data standards, clinical interfaces and regulation. It hopes to create a basis for collaborative working across jurisdictions, through international codes of conduct and/or flexible standards on issues such as consent, data access and privacy.

26. Recently, discussions about data sharing have focused on the need to open up access to clinical trial data, both to build public confidence and to facilitate research. We held a cross-sector workshop in April 2013 to discuss issues around clinical trial transparency. As a result, we are exploring the potential of establishing an international consortium, to facilitate appropriate access to data from both commercial and academic trials.

Q4. What are the opportunities for, and the impacts of, the use of linked biomedical data in research?

Opportunities and concerns

27. We believe that large-scale, international sharing and linking of biomedical data will create substantial opportunities to further understand the causes, mechanisms and possible treatment routes for many diseases. It may also help avoid unnecessary duplication of research. As a funder, we believe that enabling access to research and

[5] http://www.p3g.org/about-p3g
[6] http://oicr.on.ca/node/11861/

other datasets in a safe and secure manner will result in: faster progress in improving health; better value for money; and higher quality science.[7]

28. Indeed, we would argue that there is a strong case for ethical data sharing. On one level, it is critical to the scientific enterprise that data underpinning research findings can be accessed and scrutinised. There is also a strong argument that there is an obligation to maximise the value of the outputs of research that has been supported by public or charitable funding. Finally, it could also be argued that there is an obligation to research participants that maximum value is derived from the data and samples they provide for research.

29. The ethical case for data sharing includes the need to safeguard the privacy and confidentiality of research participants and ensure that data is used in line with the consent provided. Data sharing should always be equitable, ethical and efficient.

International collaboration

30. Sharing data in an international context does raise potential concerns around equity and fairness, particularly where there is a disparity in terms of the capacity of partners to use the data, and variations in the laws and conventions on data privacy and confidentiality. This is a particular concern in the context of global health research involving collaborations with low and middle income countries, where an obligation upon researchers in these settings to share data may place them at a disadvantage relative to better-resourced researchers overseas. The Human Heredity and Health in Africa (H3Africa) initiative, supported by the Wellcome Trust and the National Institutes of Health, has been set up in part to address the issue of capacity building in genomics research within Africa and to consider how genomic and clinical data can be shared more equitably.[8]

31. In addition, there is a current lack of understanding and evidence about what the expectations of those participating in research are about how their data are used – particularly in low and middle income settings. An empirical research project funded by the Wellcome Trust and led by the Ethox Centre is examining the views of research participants and stakeholders (including ethics committees) towards the value of sharing data and protecting their privacy, and exploring whether these interests are in fact in tension.

Access to data

32. As noted above, for research we fund we require that datasets of wider value should be preserved and made accessible to users (including academics and researchers at commercial organisations). We recognise that, while the default position should be that researchers make data available for re-use as widely as possible, there will be legitimate cases where data either cannot be made available or strict controls need to be applied.

33. In cases where access to data is controlled, we would normally expect applications from commercial researchers to be considered on an equivalent basis to those from other users unless the consent does not permit this (e.g., 1958 Birth Cohort). For example, UK Biobank adopts a policy of allowing both academic and commercial uses

[7] Sharing research data to improve public health: joint statement by funders of health research: http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm
[8] http://h3africa.org/

for its data and the UK Data Archive also allows commercial use under special licenses. Commercial users play a critical role in the research enterprise, often enabling translation from the discovery of drugs or early stage development of technologies to developing them into effective pharmaceuticals or healthcare treatments. Enabling access to commercial users will often be a vital element in maximising the benefits of data sharing and bridging the gap from bench to bedside.

34. We recognise that there are concerns among the public about commercialisation and the possibility of companies deriving profit from research data. We acknowledge that more work needs to be done to communicate the role of commercialisation in translating scientific discoveries into technologies to improve healthcare and in promoting dialogue on these issues.

Q5. What are the opportunities for, and the impacts of, data linking in medical practice?

Use of linked electronic records

35. As with data generated in the course of research, we believe that using patient data, appropriately and subject to rigorous governance, can be invaluable in contributing to research that aims to improve health. Correlations between health-influencing factors, causes or predictors of disease, protective and risk factors may all be discovered through the use and linking of these data and it is only through large scale studies that these factors could be established. Patient records also have a key role in enabling the identification of suitable participants to be invited to take part in clinical trials. We strongly support efforts to improve access to electronic records and patient data for legitimate research purposes, where these uses are controlled by strict and robust codes of conduct and governance arrangements (such as through the Clinical Practice Research Datalink)[9].

36. In the healthcare arena, there are several good case studies illustrating the benefits that can be gained from a integrating and linking biomedical and health data. See **Box 1** and www.patientrecords.org.uk for some examples.

Patient and public perceptions

37. The public need to have confidence in how their patient records and health information are used for research, and this will require both openness and transparency. Research has shown that the public are generally supportive of research using health information, but there is little understanding of what this actually means in practice. A summary of recent surveys exploring public attitudes towards the use and sharing of health data can be found on our website.[10]

38. We therefore welcome the current efforts by NHS England to provide the public with more information about how their patients records may be used, and their right to object (see paragraphs 61-63). In conjunction with over forty other medical research charities we are currently conducting an advertising campaign to raise awareness of the benefits of using health records in research, and to showcase a range of case

[9] http://www.cprd.com
[10] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Personal-information/Public-engagement/index.htm

studies.[11] The main driver for this campaign is to ensure that people to be able to make an informed decision about the way their data are used.

39. In 2013 research we commissioned indicated that there is an expectation that health data are shared within the NHS to support individuals' direct care, and the majority of people are supportive of the use of anonymised data for research.[12]

Box 1: Examples of data linking in medical research and practice

- Sharing of clinical data on the outcomes of diabetic patients in Tayside has resulted in a 40% reduction in amputation and a 43% reduction in the number needing treatment for eye disease.
- Espoo, Finland, has adopted an eHealth system that supports integrated and self-care. This has led to improvements in patient care and savings in healthcare costs. For example, the proportion of diabetes patients who have well-controlled glucose levels has increased from 72 to 82% over four years – generating cost savings due to reductions in surgery and hospital visits.
- There have been concerns that orlistat – a drug prescribed to treat obesity – increases the risk of liver problems. Researchers linked data from the Clinical Practice Research Datalink and Hospital Episode Statistics to show that this isn't the case and that the drug is safe to use.
- A study in Scotland linked health and education data to show that the risk of a child developing special educational needs increases steadily with increasing prematurity, suggesting that elective early delivery is not a risk-free choice.
- A study was carried out using data from the Scottish Morbidity Record (SMR2) to show that the introduction of smoke-free legislation ban was associated with a drop in the number of preterm deliveries and low-birth weight infants.
- Researchers from Wales linked data from Cervical Screening Wales and the National Community Child Health Database to see whether the treatment of precancerous changes to the cervix is associated with preterm birth and low birth weight. They found that surgical treatment did not affect these outcomes.
- The Avon Longitudinal Study on Parents and Children (ALSPAC) has been in discussion with HMRC and DWP with regard to linking their research datasets with administrative datasets. This linking could be enormously valuable. For example, economic data on tax and income would greatly enrich the study, with such factors being considered as both exposures and outcomes.
- In contrast, in a study on the use of erythropoietin in renal patients who are anaemic, data were not released and linked, and the harmful effects of high-dose erythropoietin were overlooked: an effect that could have been detected earlier if data could have been linked.

Q6. What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?

40. We are aware that biological and health data have potential further utility, for example, by insurance companies, employers or marketers. We recognise that potential use by such groups is a particularly sensitive matter of public concern, and consider that extremely careful thought must be given to how to manage access to data for use outside biomedical research and healthcare.

41. In terms of using biological and health datasets for purposes other than research, gaining access to these typically requires the user to sign a Material Transfer or Data Access Agreement. These specify that the user is legally bound to use data only for

[11] http://www.patientrecords.org.uk
[12] http://www.wellcome.ac.uk/About-us/Publications/Reports/Public-engagement/WTP053206.htm

specific stated research purposes and to not pass it on to any third party. Although breaches may be possible, we are not aware of any instances of abuse of these agreements for purposes beyond biomedical research.

42. In its final report in 2012, the Human Genetics Commission recommended that its successor, the Emerging Science and Bioethics Advisory Committee (ESBAC), work with Government and insurance companies to resolve concerns over the use of genetic data in insurance, and we fully endorse these efforts.

43. There is a moratorium in place until November 2017 on the use of predictive genetic test results for insurance purposes, agreed by the Government and the Association of British Insurers in 2011.[13] The review of the Concordat, due to be undertaken this year, will need to take into account the rapid developments in genomic and data technologies since the agreement was made.

Q7. What legal and governance mechanisms might support the ethical linking of biomedical data?

44. It is our view that the potential for societal benefit, such as the development of treatments or preventative measures for health conditions, is heavily in the public interest. All considerations about appropriate legal and governance mechanisms to support the ethical linking of biomedical data ought to take this benefit into account.

45. EAGDA was set up precisely to proactively consider how ethically rigorous and practical governance models can be developed in light of the increasing possibilities for linking biomedical data. An advisory group such as EAGDA is able to address emerging ethical issues in the rapidly developing field of biomedicine quickly and efficiently and promote good models of data governance among researchers more effectively than a legislative vehicle.

46. The linking and use of biomedical and health data do not require distinct governance mechanisms compared to other uses of personal data. The primary considerations in determining appropriate governance arrangements are that they are proportionate, commensurate with risks, fit for purpose and respect the wishes of the data subject. This means that they are flexible enough to accommodate novel proposals for the use of data and reduce the barriers to researchers wishing to access and link datasets, while at the same time, most importantly, being sufficient to protect participant confidentiality.

47. The prospect of linking data in itself does not create particular issues for governance or increase the risk of confidentiality being breached, because the risk depends on what data are actually being linked. Governance mechanisms need to be sensitive and flexible in their approach to ensure that access controls are commensurate to the risk posed by the linking of the data being used in each circumstance.

Consent and participant involvement

48. One of the key issues for data-driven research is that any consent given should be clear and flexible enough to allow a range of future uses, as the specific uses to which the data will or could be put are not necessarily known at the time of consent. The use of broad consent is becoming more widespread and under certain circumstances,

[13] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216821/Concordat-and-Moratorium-on-Genetics-and-Insurance-20111.pdf

where supporting governance mechanisms are in place, can be conducive to conducting ethically rigorous data-driven research.

49. We welcome the development of new models of consent that allow flexibility in consent procedures depending on the preferences of individual participants. With "dynamic" consent, participants can opt to provide broad consent for future uses of their data, or request to re-consent for fixed time periods or particular uses of their data. We recognise that there are likely to be a wide range of attitudes towards giving consent over time and for different research purposes: some participants may wish to have their consent sought for uses of their data on an on-going basis, whilst others may wish to not be repeatedly contacted for consent over time.

50. Whichever type of consent is sought, it ought to be clear to the participant how their data can be used, including: whether it would be accessible to commercial organisations; whether it will be possible to withdraw consent once data have been made available for access to researchers; and whether or not they will be informed of any health-related findings resulting from the processing of their data.

51. Legal and governance arrangements need to accommodate the fact that studies vary substantially, both in terms of their methodologies and in the potential for continuing involvement from participants. Longitudinal studies such as UK Biobank and birth cohorts such as ALSPAC extensively involve their participants on an on-going basis, providing information about the way their data are being used and might be used. ALSPAC includes participant members on its own ethics committee that reviews proposals for data access and linking, but this level of involvement would not be feasible for a study the size of UK Biobank. We thus consider that the reasonable level of continuing involvement for participants will depend on the type of study, what data are collected and the individual preferences of the participants. Whatever governance arrangements are in place, they need to be communicated clearly to participants to inform their consent.

52. We consider that good data governance mechanisms can provide an ethically rigorous set of safeguards for research participants, enabling them to provide broad consent for further use of their data. Such governance mechanisms ought to include an appropriately constituted committee to decide what requests for access and linking should be granted. Establishing a robust system of data governance will, we believe, help to ensure that the need for explicit and specific consent for research uses of health and biological data will be reduced.

Legislative and regulatory framework

53. Many research projects handling data from participants have well established governance mechanisms for the ethical use and linking of data in place, and these often cover many types of personal data other than biomedical data. Large cohort studies in particular usually have a Data Access Committee (DAC) to make decisions on access requests and ensure that all uses of participant data, whether identifiable or anonymised, are consistent with the terms of participants' consent.

54. Appropriate safeguards when using potentially identifiable information could include:
    - Technical safeguards:
        - Using electronic technologies to ensure security
        - Ensuring the linkage of identifiable (or potentially identifiable) clinical data for any purpose other than direct care occurs only in specialist, well-governed,

independently scrutinised and accredited environments called 'accredited safe havens', such as England's new Clinical Practice Research Datalink;

- Researcher level safeguards:
  - Introducing a system of 'approved researchers', as recommended in the Data Sharing Review[14], bound by the "same duty of confidentiality as the clinical team", and with "criminal sanctions in case of breach of confidentiality";
  - Contractual data sharing agreements and professional standards for researchers that prohibit re-identification.
  - Clarity in the UK law on appropriate sanctions or penalties (including criminal charges where applicable) that are in place for anyone who is found guilty of misusing these data.

55. We are concerned that the amendments to the EU Data Protection Regulation adopted by the Civil Liberties and Home Affairs committee of the European Parliament include disproportionate provisions that could prevent much research that relies on the appropriate sharing of personal biological and health data and have a serious adverse effect on the European capability for biomedical research. These barriers at UK and EU level need urgent resolution.

56. In particular, the amendments to Articles 81 and 83 have the potential to block much research that relies on the appropriate sharing of personal biological and health data. We do not consider that the requirements stipulated in the amendments to obtain "specific" and "explicit" consent for all uses of personal data reflect an appropriate approach to data use for research. Furthermore, we believe that the over-reliance on consent as the primary means through which to assure privacy is misguided, as ethically rigorous and proportionate governance mechanisms can serve to protect participant confidentiality in line with their wishes where it is unfeasible to obtain re-consent for a future unanticipated use of their data.

57. In a global context, the relationships between different legislative and regulatory frameworks are highly complex, and this creates barriers to international collaboration. A harmonised approach to issues such as consent, data and sample sharing, data security and governance would help assist in the ethical linking of data, but we recognise that these issues present enormous challenges.

58. The current regulatory and governance framework for the use of health information in research in the UK has been perceived as complex and confusing, with requirements under the Data Protection Act overlapping with other requirements – such as Caldicott guardians and the common law of confidentiality, and associated section 251 exemptions under the NHS Act for the use of identifiable data in research. Researchers have faced a lack of consistency about the processes that should be used when information from patient records is required for research.

59. Terminological inconsistencies are also problematic in this area: often the terms "confidential data", "personal data", "identifiable data" are used interchangeably, as are "pseudonymised" and "anonymised linked" data. These inconsistencies generate significant problems for researchers in interpreting data protection law, ethical guidance, codes of practice and funder policies.

60. Recent attempts to clarify the framework have been made through:

[14] Data Sharing Review (2008), co-chaired by Mark Walport (former Director of the Wellcome Trust) and Richard Thomas (Information Commissioner)
http://www.connectingforhealth.nhs.uk/systemsandservices/infogov/links/datasharingreview.pdf

- Amendments to the NHS Constitution in 2012, which clarified a default assumption that non-identifiable patient data can be used for approved research, and patients can be approached about taking part in research studies. We supported these amendments.

- The second Caldicott Review published in April 2013, which contained the recommendation that where someone is concerned about their information being shared, they should have the right to make their objection heard. We agree with this and other recommendations of the Review.
  - The Academy of Medical Sciences is currently planning to host a workshop in collaboration with MRC and the Wellcome Trust, on the definition of 'safe havens', in response to the recommendations of the Caldicott Review.

61. We consider that an opt-out system for the use of patient records and NHS data will have important public benefits and lead to better research. These benefits would be more limited by an opt-in system because few people would take the steps required to actively opt-in, and there is a risk that a skewed dataset would be created as a result. This would lead to a potentially dangerous bias in research results[15].

62. With an opt-out system, it is very important that people understand their options and the right to object to the sharing of their patient records. We consider that the mechanisms for opting out should be straightforward, clearly explained and actively communicated.

Public engagement

63. We recognise that the benefits of using data may not be fully realised if there are public concerns about privacy and distrust of the bodies responsible for managing these data. It is critically important to conduct a wide range of public engagement activities in relation to the use and linking of data, particularly when this involves data routinely collected in the NHS. Such activities ought to be conducted by bodies that are widely trusted and supported, so that differing views can be openly aired and explored. An evidence base needs to be developed to capture the complexities of public attitudes, and public opinion must be taken into account when determining what governance arrangements will be appropriate for a dataset. Transparency in governance and decision-making is crucial to ensuring public trust in research in this area.

[15] See, for example, a case study on breast cancer care documented in the Academy of Medical Sciences' 2011 review of research governance, Box 6.1: http://issuu.com/acmedsci/docs/newpathw