Nuffield Council on Bioethics consultation on biological and health data: The collection, linking, use and exploitation of biological and health data: ethical issues

Submission prepared by the Medical Research Council (MRC) with contributions from the Economic and Social Research Council (ESRC) and the Biotechnology and Biological Sciences Research Council (BBSRC): January 2014

Introduction

1.   The Medical Research Council (MRC), Economic and Social Research Council (ESRC) and the Biotechnology and Biological Sciences Research Council (BBSRC) are UK-based non-governmental organisations funded by the UK tax payer via the Department for Business, Innovation and Skills.

2.   The research councils support excellent research, as judged by peer review, that has an impact on the growth, prosperity and wellbeing of the UK. To maintain the UK's global research position the research councils offer a diverse range of funding opportunities, foster international collaborations and provide access to the best facilities and infrastructure around the world. We also support the training and career development of researchers and work with them to inspire young people and engage the wider public with research.

3.   The mission of the MRC is to improve human health and support economic growth by supporting the delivery of world class medical research.   In 2012/13 the MRC spent £767 million on medical research. The MRC acts independently of Government in its choice of which research to support and aims to provide a broadly balanced yet responsive portfolio, ensuring that research reflects changing health needs. We fund across the biomedical research spectrum from fundamental laboratory-based science to clinical trials and population and public health research. Through rigorous peer review, we support the highest quality internationally competitive research. We work closely with the NHS and the UK Health Departments and other partners to deliver our mission.

4.   The MRC funds a wide range of activities relevant to biomedical data and informatics. As detailed below there is tremendous potential to advance understanding of health and disease by appropriate collection, linkage and use of such data. However, this must be done in a well governed, secure manner that assures public confidence and trust in such use with benefits being clear to the public and those from whom data are collected.

5.   The ESRC is the UK's largest organisation for funding research on economic and social issues. Supporting independent, high quality research which has an impact on business, the public sector and the third sector.   The ESRC

promote and support, by any means, high-quality basic, strategic and applied research and related postgraduate training in the social sciences; advance knowledge and provide trained social scientists who meet the needs of users and beneficiaries, thereby contributing to the economic competitiveness of the UK, the effectiveness of public services and policy, and the quality of life; provide advice on, disseminate knowledge of and promote public understanding of, the social sciences.

6. The ESRC has made a substantial investment in data resources and initiating research for a range of biosocial research areas that fall within our strategic priorities. As a research funder we have invested in the collection and analysis of biological data as part of our longitudinal studies.

7. BBSRC funds academic research in the non-clinical life sciences in universities and research institutes, together with related postgraduate training, knowledge exchange and innovation, and public engagement activities. The Council funds a significant portfolio of fundamental, underpinning bioscience research in areas relating to human health. The use of resources and data from cohort studies, biobanks and longitudinal monitoring, and the application of new technologies in areas such as genomics, bioinformatics and modelling, are therefore relevant to BBSRC.

8. This response focuses only on those questions or parts of questions relevant to RCUK or the individual Councils who have contributed to the response.

9. In order to further realise benefits from data use we recognise that it is critical to ensure the research directions are aligned with public support and expectations as to uses of data – in the context of potential and realised benefits. The research council's welcome this work by Nuffield Council on Bioethics which we hope will help describe and clarify ethical approaches to these important issues in the context of public expectations of biomedical and health researchers and funders.

10. In relation to support for this area, the MRC has five 'pillars' to support this area set out in Box 1:

Box 1: MRC approach to research involving biomedical data:

> 1. MRC investments in data and tissue collection, and data curation, infrastructure and tools, need to **be science driven** – i.e. tightly linked to well-developed visions for specific future scientific needs.
>
> 2. MRC support for **enabling infrastructures**, including enhanced computing power, data storage, data exchange and secure linkage systems. Investment is driven by research community needs, opportunities for partnership (including across disciplinary boundaries), effective and efficient delivery, and value for money.
>
> 3. MRC is **building capacity in strategically important skills** such as bioinformatics, biostatistics, population health sciences, methodology research and interdisciplinary social and biomedical sciences. These analytical skills will enable the development of new theories and methodologies for data integration, linkage and analysis which will critically underpin "big data" science now and in the future.

> 4. We work with the Wellcome Trust, the UK Research Councils, NIHR, CRUK, other funders and government departments to **align sharing and open access policies**.
>
> **Continued**
>
> 5. We work with a wide range of stakeholders nationally and internationally **to secure a robust regulatory and ethical environment**, and to develop and implement good practices, that enable data and tissue to be shared ethically for excellent research.

11. Examples of funding initiatives include the partnership investment in eHealth research (joint funding of £19m (from a consortium of funders including the ESRC and MRC) for four eHealth centres); large cohort studies (such as UKBiobank –initial joint funding of £60m; birth and other cohorts – in which MRC invests a further £9.6m per annum) and support for genomic and other informatics platforms and related research. We are currently in the final stages of decision making on a UK platform for dementia research that will build on cohort and patient data relevant to the dementias with support of up to £12m.

12. ESRC have invested in biosocial data collection and analysis, in particular as part of our longitudinal studies. The UK has a unique, world-leading collection of longitudinal studies spanning 65 years which follow the life trajectories of families and individuals. New investments such as the Life Study will strengthen this series bringing together social, economic and biological measures to deepen our understanding of early childhood health and development. The enhancing of other studies, including the Millennium Cohort and Understanding Society, through the addition of biological samples and responsible linkage to administrative data sources will radically extend our capacity to understand the key influences that affect people's life chances and shape policy intervention to improve life outcomes.

13. MRC and ESRC have both invested in the Life Study which will be a truly interdisciplinary new birth cohort of over 80,000 babies. Life Study is the first UK-wide study to recruit in pregnancy and will thus be able to collect data and consider factors before birth. It will collect both socio-economic and biological data and is starting data collection in 2014.

14. Along with the National Institute for Health Research (NIHR), ESRC recently announced it is awarding £20m to six research projects which will significantly add to our understanding of dementia. Amongst other things, the research will look at how we can better prevent dementia, and improve the quality of life of those with dementia and their carers. One of these awards will develop a publicly available tool to help meet the future needs of dementia patients and their carers, and a. A model will be developed from this which will enable us to better predict the future costs of dementia.

15. In 2013, the MRC committed an additional £70m in health and biomedical informatics research to enable the UK to help tackle the "big data" challenges to enhance and the UK as world-leader and to deliver new benefits for UK science, health and the economy. This investment was allocated as follows:

a. Strengthening capability in electronic health records research (£20m capital which supports the Farr Institute linking the eHealth centres).
b. Medical bioinformatics - building capability, capacity and infrastructure (£50m - comprising £35m capital and £15m resource).
c. Key challenges in policy, regulation and ethics – MRC support for initiatives such as the Global Alliance for Genomics and Health; US Institute of Medicine project on clinical trial data sharing; Ethics and Governance of Data access (EAGDA).

16. The MRC approach to ensuring best practice in relation to ethics and governance is as detailed below in box 2:

Box 2: MRC approach to ethics and governance frameworks

- To build public confidence in medical research and researchers, in particular in relation to frameworks for consent, confidentiality and privacy of personal information.
- To help create a legal and regulatory framework that promotes and enables excellence and innovation in research while taking a proportionate approach to the risks to people and the research process.
- To promote a research culture that recognises and rewards 'sharing' of research information within appropriate and proportionate governance.
- To promote and enable collaboration for research across disciplines and between academia, government departments, industry and regulators.
- To strengthen transparency, so that sources of research data and tissues can be 'discovered,' assessed for their potential for new research and accessed efficiently through appropriately governed mechanisms.

Response to consultation questions

17. Please note, as the questions cover a range of overlapping areas where the MRC, ESRC and BBSRC have interests, responses to one question may also be relevant to another, we have not repeated these points in each section.

Consultation question 1: Do biomedical data have special significance?

18. There is tremendous scope for research involving biomedical data to provide advances in prevention, diagnosis and treatment of disease and promotion of health. As the remit of the MRC is to deliver health benefits through medical and health research it follows that the MRC considers biomedical data to be of particular value in this area. However, there is also considerable value in linkage to social and economic data to provide wider perspectives in impact of these factors on health and disease.

19. From the economic and social science perspective, it is difficult and not always helpful to tightly define biomedical data. It is of a heterogeneous nature, but often default definitions may lead to an assumption that it is solely data collected for medical or health purposes, where the underlying reality for social research may be different (e.g. understanding antecedents to poverty). So data collected in other contexts may also be of relevance to understanding health and disease.

20. Biomedical data are stored and used in a wide range of research studies and have provided significant insights into variation and health – classic examples are drawn from epidemiology and population heath and include linkage between smoking and cancer and more recent findings from cohorts of diabetic patients indicating potential benefits from antiglycaemic agents and cancer. We believe that biomedical data occupies a unique position in the larger scheme of data intelligence and informatics. Its potential to provide insights into health in relation to risk prevalence is matched by its emerging existence as an explanatory variable in a social and behavioural context.

21. Beyond such population studies almost all medical and health research involving people requires collection, storage and analysis of data relating to those people and their health, for example clinical trials and experimental medicine studies. This data may be identifiable or anonymised – a key principle is that data seen by research teams should always be as anonymised as possible in the context of the research – for many projects this will mean that no identifiable data is required by the research team.

22. The change in scope referred to in the question would seem to relate to the relatively recent exponential increase in data and linked information available – due to the power of genetic sequencing and informatics capability to link many sources of data, either from routine health records, research derived data or other socio-economic data held in the public or private sector. Thus the collection, storage and use of biomedical data is not novel, it is the scale and interoperability of datasets that has changed, as well as the breadth and implications of some of the data (for example genetic).

23. The increased potential for data generation; discovery and linkage referred to above has tremendous potential for benefits to individual and population health through areas such as population health and modelling, identification of biomarkers for diagnosis and prognosis and development of, stratified (personalised) medicine approaches. The scope and richness of information provided by extensive data linkage and easier access to big data can only be beneficial to research leading to improved understanding of health and well-being for society and for individuals , provided that privacy concerns are met.

24. The NHS is a unique resource in the UK that contains a wealth of longitudinal health data from cradle to grave on the UK's population of 63m people. The ability to utilise NHS electronic health records (eHealth) to identify effective treatments, monitor drug safety, identify public health risks and provide insights into the cause and development of diseases offers the UK unrivalled opportunities in health research on the global stage with the potential to accelerate and deliver unique advances for patient and population benefit.

25. The increase in scale and scope of research involving data builds on the existing symbiotic relationship between research and improving health and healthcare. Maximising the value of data (in a responsible way through strong governance of access and linkage), particularly that produced through the public purse, is critical to realising the potential for significant advances in patient care. Commercial partners are and will be critical to this relationship in order to realise the potential for novel treatments and interventions. This does not describe a change in practice, but rather in scale and sophistication of approaches.

26.   In addition to NHS records, valuable health related population data are also held across a spectrum of public sector bodies, including Government Departments and agencies. The ESRC led Administrative Data Taskforce (ADT) with MRC and Wellcome Trust to consider how such data might be accessed and linked for research. As part of their deliberations the ADT took into consideration advances already made on access to NHS records for research. The ADT recommendations proposed a UK Administrative Data Research Network that would be responsible for linking data between government departments. ESRC recently commissioned the ADRN in the first phase of our £64 million funding of investment in Big Data. The network will make routinely collected administrative data accessible for research in ways that prevent the identification of individuals, while providing a sound evidence base to inform research, and policy development, implementation and evaluation.

27.   Biomedical data are part of a wide range of data relating to individuals that may be stored and used for research. In a sense, the actual data and form in which it is held are no different from other information about individuals, such as social or economic data. However, biomedical data may contain information that individuals and/or society consider particularly sensitive or private; it may also have relevance to current or future disease risks which may have significant implications for individuals and their families.

28.   The legal framework (through the Data Protection Act 1998) currently recognises personal data and sensitive personal data – and clearly individuals and society consider different types of data as being of differing sensitivity. This will relate to both the nature of the data and the context in which it was obtained e.g. through health consultation; public; self-disclosure on social media etc. It is important that society determines how these differing levels should be dealt with – in the context of effective public engagement as to the potential benefits of research using such data and the options for governance of these.

29.   One sensitivity in relation to biomedical data is the potential identification of individuals and the health implications of data held – which may not be known to the individual. Disclosure of identifiable data may lead to harm to an individual – whether financial, reputational or in relationships and this is the basis of the common law expectations of confidentiality. The MRC and Wellcome Trust will shortly publish a framework for feedback of heath related findings from research, which provides context for researchers to consider what information individuals should be made aware of.

30.   In relation to identifiability and health implications, genomic data has been seen as particularly sensitive – although information can rarely be derived from such data without further information about an individual. This has been challenged through a recent study indicating individuals' surnames may be determined from parts of genetic data. It is important to consider that such identification could only occur deliberately, mitigation could therefore be through stringent governance and greater understanding that such deliberate identification is deemed illegal and / or professional misconduct. The ICO has stated that '*The ICO will treat deliberate re-identification activity with the utmost seriousness'.* The full text of this statement helpfully clarifies not only that the act of re-identifying anonymised subjects without their consent is in itself potentially subject to penalties, but also that the use of such data to cause harm to data subjects (e.g. by discriminating against them) may also

be subject to penalties even if the actual re-identification occurred outside UK jurisdiction. This may serve to reassure both research subjects and researchers that, although re-identification is becoming technically feasible, it is not an acceptable practice within the UK and may be subject to legal action.

31. Issues such as sensitivity and possible re-identification of research participants are not unique challenges for biomedical data and in fact social science in the UK, which has had 'open access' approach to data findings for many years, has grappled with these challenges for some time.

32. An alternative route to mitigate the risks of identification is through providing information and obtaining participant consent on the basis that such identification could occur. The approaches of consent and robust governance should be complementary and not seen as mutually exclusive.

33. Genomic data does pose a particular ethical challenge as there may be an implied disclosure i.e. learning something about an individual's genetic make-up may tell the investigator something about other individuals. There are two issues – potential breach in confidentiality to the proband and a putative obligation to inform other individuals about any potential risks to them. Recent work and recommendations by the Expert Advisory Group on Data Access (EAGDA)[1] to research funders in this regard has been of assistance in this area (see further below)

Consultation question 2: What are the new privacy issues?

34. As above, the issues relating to biomedical data relate to new innovations providing different types of data (e.g. genomic; metabolomics) and to new technologies and informatics capabilities expanding opportunities and scale of data linkage. These issues arise in the context of a rapidly evolving landscape of data generation, not just in the healthcare and research environments but in personal and social provision of data (e.g. through Facebook; Instagram; consumer outlets etc. Another recent advance is the increasing sophistication of tools that perform 'black box analyses' – so no identifiable data is seen in outputs. The research councils are aware of evolving issues around intellectual property and patient control: such as, who actually owns/governs a consolidated dataset from several sources/many different patients?

35. The implications for linkage across multiple datasets need to be understood – and in particular whether there are risks of disclosure of identifiable information. It is important that such identification and disclosure could be inadvertent or could be deliberately effected. This could occur with or without individual knowledge and/or consent. Steps to address each of these situations will differ and should not be conflated. One approach is to accept these risks and focus on researcher responsibility to ensure harmful identification and disclosure does not occur, or to involve participants on the basis that such risks are present (recent genomic project).

36. EAGDA recently considered this issue, highlighting that every reasonable effort must be made to protect the privacy of research participants in consultation with experts in the field. It concluded that large datasets,

---

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

particularly those including extensive genomic information, cannot be completely safe from inferential exploitation, including subject re-identification. Although the likelihood of such re-identification may currently be low for most types of study, it is likely to increase in the future as:

- Research datasets become richer, more complex and more readily accessible;
- Methods of analysis and interpretation increase in sophistication and reduce in time and cost;
- Improved and wider use of demographic and administrative data become possible; and
- Individuals release more information about themselves into the public domain e.g., through recreational genomic and social networking web sites.

37.   There is clearly a balance to be achieved between maximising the value and the benefits of data produced by the researchers we fund, whilst safeguarding the confidentiality of research participants. These risks can be mitigated by the steps described in the response to the previous question, such as clear information and consent pathways; robust governance with clarity that deliberate re-identification without appropriate approval and governance will be viewed as illegal and/or misconduct.

38.   The potential 'public interest and benefit' from medical research lies in the significant potential and actual use of biomedical data to provide insights into health and disease leading to more effective targeted treatments and interventions – with associated reduction in overall cost and reduced morbidity and mortality. As detailed in the first section, this includes population health but also, for example, reuse of clinical trial data to understand stratification in responses to treatment.

39.   Many issues in this question relate to public and social expectations around privacy which are not for research funders to determine. The MRC and ESRC aim, however, to ensure active engagement in these discussions to ensure that the evolving research landscape is described in these discussions and also to ensure that governance reflects evolving public expectations, legal requirements and best practice in ethics

Consultation question 3:   What is the impact of developments in data science and information technology?

40.   As described in the introduction the impact of these developments is reflected in the significant research council investments – which align with similar initiatives from many other public, charity and private sector organisations.

41.   The need for coordination of approaches and ensuring robust governance for these developments is also very clear. For example, to add value to the eHIRCs initiative, funding was made available to support the Farr Institute to link the four centres and draw in informatics expertise from across the UK research community. The aims of the network are to:

- Coordinate training and career development;
- Share good practice and explore novel linkages between datasets;

- Engage with the public to promote better understanding of the benefits and risks of e-health records research;
- Provide an interface for collaborations between academia, industry and the NHS.

42. The wider arena of biomedical informatics (which is not limited to data from electronic health record data) covers a broad spectrum of capabilities, including bioinformatics, computational biology, statistical genomics/genetics, stochastic systems biology and other quantitative methods. High-throughput technologies for analysing genomes and biological processes in health and disease are transforming medical R&D and will revolutionise individual health care. In parallel to low cost whole-genome sequencing, large-scale metabolic profiling, imaging, and immunological techniques are providing vast depths of information at cell/tissue/organ levels. Future biomedical R&D in industry and academia will involve much greater emphasis on manipulating and drawing evidence from immensely large and complex biomedical data sets.

43. The MRC and BBSRC support a large and diverse portfolio of investment in medical and biological bioinformatics ranging from investigator-led research projects to large-scale infrastructure support through cohort studies and intramural programmes, training at multiple career stages, and strategic partnerships (e.g. EBI/ELIXIR). Some examples of the developments and their impact are described below:

- Improvements in cohort selection, population meta-analysis, ability to find associations between diverse attributes. Power of using large datasets – e.g. may spot patterns across large geographic areas that would otherwise be missed.
- Drivers include technical advances e.g. sophisticated analysis tools, ability to store mass data cheaply, increased ability to detect and record, wider use of electronic systems, increased coding/machine language processing.
- Effect on research priorities includes more database-driven, opportunity to have large, combined collections of data (thus funders will increasingly look for research with maximally reusable outputs through appropriate data access and sharing).
- New developments such as use of tiered research queries for selecting people e.g. for trials (can do automatic preliminary investigation on research dataset to see if there is evidence to proceed).
- Evolution of informatics and analysis, as well as data available allows opportunities for increasing understanding of stratification of diseases; ability to elucidate molecular signatures of disease across genomics; proteomics etc. understanding risks of disease and predisposing factors allowing more effective prevention strategies.

44. The UK is home to the largest and longest-running longitudinal studies in the world. The ESRC and MRC jointly invested in the Cohort and Longitudinal Studies Enhancement Resources (CLOSER); a programme, which aims to maximise the use, value and impact of these studies both within the UK and abroad. One of the workstreams is developing a uniform search platform (USP) to provide a portal to hundreds of thousands of variables, questions, and data collection instruments from across the 9 CLOSER studies. In order to achieve this, substantial work will be undertaken to enhance the metadata available for all of the CLOSER studies. The research councils also support access to cohort studies through a range of other funding mechanisms. For

example, under the auspices of the cross-Council Lifelong Health and Wellbeing Programme, MRC and BBSRC fund the Centre for Cognitive Ageing and Cognitive Epidemiology hosted at the University of Edinburgh, which utilises the Lothian Birth Cohorts in its research programmes.

45.   Data are only of value and relevance once analysed to provide information that can then be interpreted into research findings. As described in the Introduction, the MRC invested £50m (£35m capital and £15m resource) in 2013 in medical bioinformatics to support improved linkage and analysis of large scale genomic information, complex phenotypic data and electronic health records. This support strongly complements the additional investment in the Farr Institute (£20m capital, see above) and will help strengthen critical mass by building flexibly upon existing research excellence in Universities and other research establishments. Support will be targeted at developing new infrastructure and tools to improve linkage within and across diverse heterogeneous datasets ranging from the cellular level through to the whole disease phenotype, and new analytical approaches to interrogate and share this data to improve our understanding of human disease. Importantly, the initiative will also invest in research skills and long-term capability by supporting career development of data scientists and new research leaders in medical bioinformatics.

Consultation question 4: What are the opportunities for, and the impacts of, use of linked biomedical data in research?

46.   The scale and opportunities for medical research are as described in previous sections – as well as current research council initiatives relating to these.

47.   In order to further realise benefits we recognise that it is critical to ensure that research approaches are aligned with public support and expectations as to uses of data and mechanisms for linkage – in the context of potential and realised benefits. The Councils welcome this work by Nuffield Council on Bioethics which we hope will help describe and clarify ethical approaches to these important issues in the context of public expectations from researchers and funders.

48.   There is a critical need to ensure public confidence in approaches to robust governance to protect privacy and trust for use and reuse of data – based on careful information but often requiring a broad model of consent linked to trust and continued engagement (e.g. birth cohorts and UKBiobank). There are significant risks if consent is seen as only route to protect privacy, as exemplified in discussions around draft EU Regulations. These risks are outlined in a joint statement from a number of scientific research organisations on the implications of the proposals for the draft EU Data Protection regulation[2] (BBSRC, MRC and ESRC are co-signatories) and paper from Baroness Onora O'Neill (at Annex 1).

49.   The research councils have agreed common principles on making research data available[3] and are committed to transparency and to a coherent

---

[2] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Personal-information/Data-protection-legislation/index.htm
[3] http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx

approach across the research base. Individual Research Councils, including BBSRC, ESRC and MRC have data sharing policies[4] and a clear expectation that funded data will be shared. . We are engaged in further discussion as to the mechanisms for sharing – for example through collaboration and safe haven approaches.

50. In order to further incentivise data sharing under these governance frameworks there needs to be recognition of production and sharing of datasets, for example through developments in the REF assessment to ensure that there is appropriate recognition of data sets and their sharing as an output.

Consultation question 5: What are the opportunities for, and the impacts of, data linking in medical practice?

51. Delivery of medical practice and clinical care are beyond the scope of the MRC and ESRC. However, we consider it important to recognise the value of transition of findings and interventions from research into clinical practice – e.g. the generation and interpretation of genomic or proteomic data – a small amount of information provided by such technologies may be relevant to immediate care but other data may be of significant value in future care or for research initiatives. This is key for advancing molecular pathology.

52. It is very important that information on collection, linkage and use of data in all contexts is provided in an accessible and understandable way to patients so that they can understand how data might be used and raise queries or objections. The benefits of approaches should be clearly articulated as well as provision and description of robust governance measures to protect confidentiality.

Consultation question 6: What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?

Biosocial Research opportunities

53. On the biomedical social science interface, biosocial research has the potential to enhance the depth and breadth of insights drawn from research, thereby improving the positive impacts of research on policy and society. For the ESRC as the leading UK funder of social and economic research, it is important that biosocial research informs understanding of the determinants and consequences of human experiences and behaviours and clarifies the pathways and mechanisms involved in shaping lasting consequences, both under and outside the skin.

54. As a research council, ESRC is committed to enabling the growth of exciting, innovative research, including that in the biosocial research field. Engagement in biosocial research that links the social sciences to biological and biomedical sciences is a growing priority.

---

[4] http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/index.htm
http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx
http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx

55.   As a research funder ESRC have invested in the collection and analysis of biological data as part of our Longitudinal studies and are committed to continuing investment in this area over many years, working in close collaboration with other research funders.

56.   For example, Understanding Society is a major household panel study that provides valuable new evidence to inform research on issues of importance to a wide scientific and policy community of interest. The study collects biomedical measures and samples to enable new research on the social determinants and impacts of health in a household context. This opens up exciting prospects for advances at the interface between social science and biomedical research.

57.   Along with the National Institute for Health Research (NIHR), ESRC recently announced an award if £20m to six research projects which will significantly add to our understanding of dementia. Amongst other things, the research will look at how we can better prevent dementia, and improve the quality of life of those with dementia and their carers. One of these awards will develop a publicly available tool to help meet the future needs of dementia patients and their carers, and a model will be developed from this which will enable us to better predict the future costs of dementia.

58.   There is significant opportunity for social science at the frontier of social and biomedical science, in areas such as epigenetics. Working with other research funders creates opportunities for exploring together rich sources of longitudinal data and for researchers to facilitate significant advances in this frontier science.

59.   As a general principle, the research councils consider it important to ensure people are aware of any potential uses of their data and that these only occur within a recognised legal and ethical framework. This is particularly important in order to maintain trust in provision of data for research where robust governance is required and implemented.

Consultation question 7:   What legal and governance mechanisms might support the ethical linking and use of biomedical data?

60.   Please refer to previous comments on current draft EU Regulations under question 5.

61.   We recognise the importance of evolving models of consent and governance and also the often on-going nature of the relationship between those whose data may be used and those who collect and use that data.

62.   As highlighted above, consent is not the only approach and considerations of trust and robust governance are also key frameworks to work within. At the heart of each of these approaches is the need for communication and understanding such that people can be aware of potential uses and outcomes from data linkage; risks of disclosure and measures to prevent these.

63.   The on-going relationship also means that, wherever possible participants in biomedical research should be able to have feedback, where desired, on the benefits e.g. research outcomes and changes to clinical practice.

64. When considering mechanisms there must be an appropriate balance between each individual and society. Again, this is not distinct from other domains of research or data use, but may be more challenging and emotive as biomedical data are often considered more sensitive (Q1). It is a policy and political decision about where the balance should be - however the debate would benefit from being more open as at present often extreme, advocacy positions are taken and reported in the media.

65. Biomedical data are not necessarily any more sensitive to disclosure than rich socioeconomic data and the criteria used should depend upon levels of sensitivity. There are issues around sensitivity and potential disclosure linked to many different types of data. It is not clear that linked biomedical data requires distinct governance arrangements compared to the use of other personal data. Data that include personal identification should probably only be used in safe settings, but once de-identified the most crucial element is to avoid any attempt to re-identify – the more sensitive the data the more constraints will be required, but researcher trustworthiness is crucial.

66. A good example of the promotion of data sharing and data access is provided by the ESRC's UK Data Service – this achieves appropriate levels of access and deals with privacy/ disclosure concerns in a consistent way; it also facilitates identification of appropriate research data sets, and ensures long-term retention of data.

67. There are currently serious inconsistencies in guidance and implementation of regulations in relation to biomedical data collection, linkage and use. The MRC Regulatory Support Centre has taken steps to address this through provision of a Data and Tissues toolkit[5] and training provision. Inconsistencies include, for example professional guidance and codes of conduct, conflicting legal obligations, individual rights (e.g. privacy) contrasted with societal roles. There is a compelling need for further analysis to set out each position in detail, promote dialogue between advocates of each position, effective public consultation (with explanation of the risks and benefits of each potential balanced outcome), and a clear framework to enable decisions about which should be adopted.

---

[5] http://www.dt-toolkit.ac.uk/home.cfm

ANNEX 1: Securing workable, ethically robust protection for personal data in biomedical and research contexts.

1. **Introduction**. Data protection is a distinctive approach to protecting aspects of personal privacy that has been devised and elaborated in the EU across some decades. Current proposals for a new *Data Protection Regulation* that will strengthen protection of privacy by requiring more explicit and more specific consent for any reuse of lawfully held personal data reflect a concern that new and more powerful ways of capturing, linking and mining data could be used to undermine privacy protection. This note sets out reasons for thinking that, while this diagnosis of a problem is correct, the remedies proposed in the draft *Regulation* for strengthening data protection may be neither workable nor ethically adequate, and that other approaches would be more robust.

2. **Two Underlying Problems with Data Protection**. The basic idea behind data protection is that the right to privacy can be safeguarded by prohibiting uses of personal data without consent from those to whom the data pertain—the *data subjects*—whereas non personal data, being nobody's data, may be used and reused without consent, and may (for example) be subject to transparency requirements or freedom of information requests. This basic approach has repeatedly faced two sorts of difficulty. First, the distinction between personal and non-personal data is insecure (Example: a home address may be treated as personal when included in a medical record, but as non-personal when included in an electoral register); this recalcitrant problem is set aside for present purposes. Second, personal information can often been inferred from non-personal information—as anybody who reads detective fiction knows! The second problem has recently become more prominent because inferential identification of data subjects from pseudonymised personal data—sometimes called 'inferential disclosure'—has become feasible by methods that depend on capturing, linking and mining data. Inferential disclosure is not new, but can be secured in new ways.

3. **Some Limits of Consent Safeguards**. Introducing supposedly stronger consent requirements for the reuse of personal data will not improve protection of personal privacy in medical and research contexts. Consent safeguards for personal data are currently reasonably strong in these areas, although often minimal in commercial contexts, where 'ticking and clicking' without reading, let alone understanding, the terms or conditions or other content to which consent is ostensibly given, is seen as achieving an adequate standard. By contrast, the use of personal data in medical and research contexts is already subject to reasonably robust informed consent requirements (cf. WMA *Declaration of Helsinki*).

   However, informed consent requirements in medical and research contexts have their limits. Data subjects—whether patients or research participants— cannot be expected to understand large amounts of medical or other technical information of high complexity that is relevant to their own medical treatment, or to grasp all of the ways in which data that pertain to them could be reused (including reuses of data that matter for public health; epidemiology; clinical improvements; other medical research). Where they cannot understand the data, or proposed reuses of data, they cannot give informed consent to the retention or the reuse of those data. In particular they cannot give fully explicit or specific consent to material uses of material that they do not understand. The thought that more exacting consent requirements, or in particular requirements for (more) explicit and specific consent, can carry the entire burden of ethically adequate regulation of the use and reuse of lawfully held personal data is an illusion. It is an example of a tendency to embrace

forms of 'consent fetishism' (a term first used by Prof Roger Brownsword), in which the entire burden of ethically adequate regulation is placed on the consent of individuals. However, the currently proposed ways of strengthening data protection by requiring more exacting forms of consent from individuals are not workable.

4. **Ethically Robust Governance of Personal Data.** Currently proposed ways of strengthening data protection by introducing 'higher' (but unworkable) standards for consent are also ethically inadequate. They assume that the consent of data subjects will be *sufficient* to ensure protection for personal privacy. This is not the case, for two quite specific reasons.

In the first place where personal data are shared by individuals, no individual's consent to the provision or reuse of data cannot be determinative, since it would amount to consent to provide or reuse others' personal information. For example an individual, who consents to the recording or reuse of his or her genetic information, thereby consents to the recording or reuse of information that also pertains to, or is likely to pertain to, related individuals. Consent procedures are by definition strictly individual, but data are not. This problem has been much discussed, but resolutions are not obvious. Informed consent approaches assume that individual choice can be determinative and sufficient, but offer no solution where this is not the case.

Secondly, since genuine consent for the use or reuse of highly complex data, sometimes for highly complex purposes, is unworkable, other ethically robust ways of safeguarding the privacy of personal information are needed. Various alternative structures for ethically robust data governance have been developed. These structures require only *broad* consent to the initial collection and retention of data, but   strengthen the protection of individual privacy by instituting   a range of further safeguards that bear on the retention and further use of the relevant data. Ethically robust structures for data governance typically provide not only that personal data will be securely held, but that they will be made available only for defined purposes within a specified range, to those meeting and agreeing to specified professional standards, often subject to the prior agreement of an independent ethical review committee, and to penalties for anybody who violates conditions set for data access, including penalties for attempts to   identify data subjects. Examples of well developed approaches to ethically robust data governance include the development of Biobanks and safe havens.

Any ethically robust approach to securing personal privacy by setting conditions on the use or reuse of personal data should do more than gesture towards standards for informed consent that cannot genuinely be met. It should   require feasible, ethically robust standards of data governance that couple broad consent with approaches to data governance that meet robust rather than merely aspirational ethical standards.

Onora O'Neill
November 2013