Nuffield Council on Bioethics

The linking and use of biological and health data:open consultation

Response by: Pauline McCormack, Simon Woods, Jackie Leach Scully
Policy, Ethics & Life Sciences Research Centre (PEALS), Newcastle University, UK

Introduction

We welcome the opportunity to contribute to this consultation. Overall, we are of the view that established forms of ethical governance and oversight are currently being stretched by the use of new technologies that allow large amounts of data about individuals' health to be produced and shared. Bioinformatics and whole genome sequencing are rapidly becoming standard research tools in countries with developed healthcare, producing unprecedented volumes of data. Alongside the acknowledged uncertainty about how genuinely informative such data are, there is much debate about, but little agreement on, issues of identifying, handling and processing information that reveals unknown variants or unexpected results. As an example, the ongoing contention around the proposed European Reform of Data Protection Rules indicates the complexity and importance of this debate (1).
Below, we address some of the consultation's main questions in more detail. We have consciously avoided reiterating some of the more familiar points of discussion here, and concentrated on raising what seem to us to be more neglected or novel issues.

1 Do biomedical data have special significance?

That biomedical data have some special significance is reflected in the way these data are documented and managed via different regimes of ethical oversight in clinical care and healthcare research throughout the developed world. What remains more contentious is precisely what gives rise to this significance: whether, for example, it is intrinsic to the data because of their connection to individual identity and personhood, or to rights of autonomy and privacy, or whether special significance derives solely from the practical consequences of its misuse. Here, issues of potential discrimination through misuse of biomedical data are foremost. Notions of non-discrimination, confidentiality and trust are enshrined within clinical and research codes (eg: GMC 'Duties of a Doctor; Declarations of Geneva and of Helsinki), as is the notion that a physician's primary duty is to the care and best interests of the patient or research subject. From a normative perspective, the special significance of an individual's biological status lies in the possibility of harm or the thwarting of best interests if that status were to be known by third parties. The potential vulnerability of individuals is also reflected in UK legislation such as the Data Protection Act (1998). This and similar legislation within Europe regards biomedical data as 'sensitive personal information' demanding the highest level of security and prudence in access and use.
By contrast, aggregated and anonymised biomedical data are not regarded with equivalent concern nor are they subject to the same strict regimes of governance and regulation. This may be an indication that it is only when biomedical data are linked to identifiable individuals that it assumes 'special significance'. Even in the case of identifiable data it is not clear that it is the mere linkage that is significant. Some data are innocuous but this is often dependent upon the attitude of the subject (in the case of very private individuals most data will be regarded as 'special') or on the context in which data are used. Age, weight and height may be innocuous in most , but in the context of public health concerns about obesity or the insurance industry's interest in risk factors these same data take on very different connotations.
Person- respecting approaches have been fundamental to medical ethics and law as well as in broader social and political contexts. The idea that a person has a right to self-determination and autonomy is echoed in their right to privacy and to control over their personal information. These values are reinforced in practices that presume a right to privacy and confidentiality, and which normally require specific informed consent before such

personal information may be used by others. However, these rights are usually not regarded as absolute, since they may be trumped by more important concerns that outweigh personal interests. Examples include the public health interest in containing serious contagious diseases, the need to conduct public health research, and the requirement to audit health care delivery. Other erosions of these basic rights might also be justified on grounds of necessity and proportionality. In societies like the UK, where there is at least a tradition of a strong welfare state and national health care, a culture has evolved where a tradeoff of basic privacy rights against some important common goods is tolerated and indeed may be expected. This culture of expectation is at work in some of the contexts we discuss below. Some commentators now argue that such expectations can be extended to routine and extensive access to medical data with minimal consent, so called 'broad consent', or no consent at all (2). It is our view that this cultural shift is ripe for evaluation.

2 What are the new privacy issues?

General attitudes to privacy: Among the general public and from those who consent to participate in research, attitudes to sharing data are overwhelmingly positive, both as explicitly expressed in studies directly exploring motivation and also in the enactment of participation by consenting to research, use of their biomaterials and wide data sharing. Even where privacy and confidentiality concerns are expressed, they do not necessarily preclude participation. Studies show that participants are not overly concerned about privacy and confidentiality in terms of the sharing of their data, as long as:

- access to the data subject is not completely unrestricted;
- governance is mediated by an agency viewed as trustworthy;
- there is independent ethical oversight.

Participants have been reported as viewing the wider sharing of data as positive or as a practical benefit in terms of utility of their sample, and some have said that wider sharing would actively encourage them to 'sign up' (3-6). For example, Levy examined 9,000 consents between 2002-2009 and found that levels of consent to "providing a DNA sample, creating an immortalized cell line, conducting research on various genetic conditions including those that might be considered sensitive, and for notifying participants of clinically significant genetic findings were above 95%" (7 p1250).

Attitudes of specific groups: Particular communities which can expect to benefit from resulting improvements in clinical care and research are also often in favour of data and biomaterial sharing. This is particularly the case for rare diseases, a heterogeneous group with a membership that is fragmented and therefore often politically weak and neglected in debates about genomic information. In the rare disease community, advocates tend to see benefits in sharing data across different disease communities – for example, where specific therapies as well as particular approaches to one disease may be more widely applicable. This was the case with enzyme replacement therapy that was first tested in Gaucher disease (8) but is now widely used across several lysosomal storage disorders, and for exon skipping, initially developed for Duchenne Muscular Dystrophy but subsequently investigated for several other rare conditions (9).

Positive attitudes towards data sharing by some groups, however, do not necessarily mean that genuine threats to privacy should be ignored. Individuals, or the parents of children, with progressive and/or life threatening illness may be willing to share their data beyond the normative expectations of science and society, especially where perceived benefits strongly outweigh apparent concern about risks (10). A particular issue here is that the trend towards less expensive and more widely available technology makes it increasingly easy for small groups of people to start disease registries and/or biobanks of their own, outside the traditional academic or healthcare environments and therefore outside the established forms of ethical oversight. Such oversight needs to be developed in order to encourage as well as

manage these developments, addressing in particular the motivation for interest groups with unmet medical needs, eg: in ultra-rare diseases, to self organise.

New routes to identification: Although many research systems de-identify data before they are shared outside the originating institution, several reports have demonstrated how linking several sets of data, or new electronic algorithms, can lead to the re-identification of an individual (11, 12). Processes for de-identifying data need to take into account the possibility for re-identification as linking modes become more sophisticated and widely applied. In tandem with that, informed consent procedures need to be crafted to address the risks of linking and possibly reaggregating data.

Privacy is also potentially threatened, and public trust undermined, by the development of surveillance and other security technologies as highlighted in recent controversies such as the NSA's Dragnet operation in the USA (13). We have anecdotal evidence from recent work with members of rare disease groups that these high-profile developments can make group members query the security of data being exchanged with organisations based in the USA, and elsewhere. These concerns indicate that there are limits to how far 'public good' justifications for such intrusions can be pressed.

In some other instances, people may not be explicitly aware that their data are being shared, linked or utilised in existing healthcare processes. For example, the NHS has two national systems of linking individuals' care data – the Summary Care Record and Health & Social Care Information – both of which are opt out systems and therefore link and share data without consent. It is not known how aware members of the public are of these databases. Therefore continuing and appropriately targeted education of the lay public in general and of interested communities in particular is a crucial element in retaining public support for the sharing and linking of biomedical and health data (14).

In general, the benefits and risks of 'big data' sharing have not been fully examined. There is therefore a pressing need for research to establish the kinds of risk that are likely to arise and the contexts in which particular risks are more probable. Without a proper assessment of the risks, and equally importantly the ability to communicate the risks in accessible ways, the process of informed consent for data sharing becomes meaningless.

- Would it be helpful to treat biomedical data as property?

The first issue to recognise is that biomedical data are routinely regarded as property, but rarely the property of the individual. Pharmaceutical companies, academic researchers, medical practitioners and charities have all established databases of biomedical information and these physical registries and databases are regarded as a species of private property. This has been demonstrated in cases of bankruptcy or commercial takeover where such databases are regarded as assets. There is of course a complex conceptual problem as to the nature of such property as, unlike human tissue, participating in a registry or database does not mean that personal data are given away. The information is disclosed and recorded but also remains with the donor.

We would not advocate the use of the concept of property rights as a means of protecting the rights of the data subject, though this might prove both feasible and beneficial in some circumstances. We would rather argue that it is possible to acknowledge a database as property but one that is held in trust by the custodians of the data. Such models have been discussed in the literature (15) and two of us have experience of contributing to the oversight of rare disease registries in which such practices have been incorporated into the oversight and governance of the registry.

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

While some commentary has addressed the effects of general or specific advances in these technologies, here we want to highlight the additional complexity that is introduced by social and cultural responses. As technology changes so do people's expectations of and

approaches to it, and this in turn can have an effect on both privacy and the social expectations around it, especially as technologies are absorbed into mundane social practices and become routinised. For example, against a history of strong bioethical concerns about the leakage of personal genomic information to third parties, some people are now choosing to self-publish their own health data on sites such as PatientsLikeMe and QuantifiedSelf, and in some cases personal health data, MRI scans and an entire genome sequence along with the right to process this data under a Creative Commons licence (16, 17).

Moreover, the increasingly globalised nature of research means that data that have been collected under local governance and national legislation reflecting local social and cultural expectations are also being shared. As a result internationally agreed standards of ethical oversight face the challenge of different cultural perceptions and the possibility of researchers being unfamiliar with what they are/are not permitted to do with data.

A form of governance that keeps in mind both developments in the technologies themselves, and keeps pace with and scrutinises social responses to them, is therefore essential. Several theoretical models have been proposed but a global consensus on an ELSI framework for the sharing of data is still lacking. Simply adding more layers of governance may not be sufficient to ensure that the linking of biomedical and health data yields the promised benefits for those who participate in research and/or the general population, while at the same time upholding and protecting the rights and choices of participants. This is particularly important at a time when the longstanding consensus on the balance between research participant protection and social benefit, enshrined in the Nuremberg Code, original and early revisions of the Declaration of Helsinki and in other instruments, is being challenged more broadly; aspects of protection from violation of privacy and from other harms in the context of linked biomedical data become relevant here.

In line with this we agree with Kaye that technological advances require the current system of 'one researcher, one project, one jurisdiction' to be updated (18). For example, two of us are involved in an international project on rare diseases, RD Connect (EC FP7 funded €12m, until 2018) which is introducing an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research, including harmonised standards for informed consent and global agreements on data and biomaterial exchange (http://rd-connect.eu/platform/ethics/). International collaboration between those concerned with ELSI issues in biomedical and health data linking will be crucial to the design and establishment of a globally agreed framework for the sharing and linking of biomedical and health data. While this is happening now for rare disease as part of RD-Connect and the International Rare Disease Research Consortium (http://www.irdirc.org/), the frameworks being developed will influence developments in common disease and healthcare generally and so should be of wider interest. Involvement by empirical researchers in initiatives such as the Global Alliance for Responsible Sharing of Genomic and Clinical data is therefore to be encouraged (19).

In the interim, we consider that existing mechanisms of ethical oversight should be utilised so that individual research projects are still subject to scrutiny by ethics committees or review boards, but that the global nature of data sharing and the resulting risks and benefits to participants need to be recognised and included in the evaluation of ethical oversight.

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

The basic premise behind linking and sharing biomedical data in research is that linking and sharing will lead to research proceeding more rapidly, effectively and economically, thereby enabling individuals, special groups and the general population to benefit from the promised health outcomes. There are several steps and assumptions in this claim, and it is vital to be clear about which ones are supported by evidence and which are not. In the case of rare disease research, for example, there is evidence: new technologies are considered to be

largely responsible for over 130 new gene identifications in 2012, a more than three-fold increase over the previous three years combined (20). In another example, developments in linking data from different techniques are also promising, for example clinical data mining and genetic analysis are being combined with molecular screening to assess drug effects with a view to repurposing, the aim being to generate new therapies without having to develop drugs from scratch.

Whatever the opportunities and impacts, it should go without saying that researchers, practitioners, policy makers and others involved in the ethical governance of using linked biomedical data in research need to bring a critical appraisal. As we have indicated above, it is important to dissect out the various steps in the process from acquiring and linking data to producing a real outcome (and, ideally, health benefit), and to examine the plausibility of the claims made for and the ethical issues at stake in each. Failure to do this risks misleading the public and funders. As we mentioned, people seem more likely to agree to donate linked samples if they are convinced that linkage increases its usefulness; but if there is no evidence to back this up, their consent was not fully informed.

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

There already exists a substantial body of work examining the basic ethics of balancing the confidentiality of biomedical data and the scope of its future and secondary uses (preventing harm to the donor/participant) against potential benefits of the research. Much of the more recent literature concerns biobanking and specifically addresses the need to link biomedical information to repository samples. To an extent, guidance on legal and governance mechanisms can be drawn from this work.

The majority of people who allow their biomedical data to be used for the improvement of healthcare do so for complex but primarily altruistic reasons – to support future research, to benefit society, and sometimes out of a sense of solidarity with future others who may benefit from improvements in biomedicine. Legal and governance mechanisms need to bolster this voluntariness by ensuring that those who allow their data to be used are not disadvantaged – either individually, or as a group – when doing so.

We agree with Lemke's consideration that any ethical oversight should take account of "ongoing community involvement in oversight committees, policy development, and educational initiatives" (5 p376). We would add that there is also a major need for improved cultural awareness to ensure that sections of the population who are traditionally harder to reach – certain ethnicities, socioeconomic groups, and (in some cases) people affected by illness or disability – are adequately represented, informed, and consulted. In saying this, we are not blasé about the difficulty of engaging some communities in dialogue about research endeavours where both risks and benefits may be hard to convey.

In this vein we want to highlight the need for careful thinking around the most responsible way of making appropriate use of people's willingness to allow use of their biomedical data in research. Sometimes the context in which these data are obtained, or consent is requested, may make it hard for people to think clearly about the potential risks and disadvantages. For example, one of us has been involved in research involving people who give tissue samples to help identify human remains in the aftermath of a disaster; given the context, it is unsurprising that donors give little consideration to any potential longer term risks through linkage to their biomedical records.

Conclusion

In this response we have concentrated on areas which are particularly pertinent to our ongoing work and interests and we have attempted to present analysis of issues which appear to be contested or unclear in our empirical contexts. We recognise that this

represents but a portion of this debate, which we deem timely and necessary. We would hope that the outputs of this consultation will not only shape deliberations around bioethical debate but will be used to inform policy, to contribute to public understanding, and to the ongoing negotiation around the balance of rights, autonomy and technological development.

References

1. Mascalzoni D, Knoppers BM, Aymé S, Macilotti M, Dawkins H, Woods S, et al. Rare diseases and now rare data? Nat Rev Genet. 2013.
2. Academy of Medical Sciences. Personal data for public good: using health information in medical research. London: 2006.
3. Hoeyer K. Donors Perceptions of Consent to and Feedback from Biobank Research: Time to Acknowledge Diversity? Public Health Genomics. 2010;13(6):345-52.
4. Kaphingst KA, Facio FM, Cheng MR, Brooks S, Eidem H, Linn A, et al. Effects of informed consent for individual genome sequencing on relevant knowledge. Clinical Genetics. 2012;82(5):408-15.
5. Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. Public and Biobank Participant Attitudes toward Genetic Research Participation and Data Sharing. Public Health Genomics. 2010;13(6):368-77.
6. Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Genomic research and wide data sharing: Views of prospective participants. Genetics in Medicine. 2010;12(8):486-95.
7. Levy D, Splansky GL, Strand NK, Atwood LD, Benjamin EJ, Blease S, et al. Consent for Genetic Research in the Framingham Heart Study. American Journal of Medical Genetics Part A. 2010;152A(5):1250-6.
8. Brady RO. Enzyme replacement for lysosomal diseases*. Annu Rev Med. 2006;57:283-96.
9. Aartsma-Rus A, Van Ommen G-JB. Antisense-mediated exon skipping: a versatile tool with therapeutic and research applications. Rna. 2007;13(10):1609-24.
10. Tabor HK, Stock J, Brazg T, McMillin MJ, Dent KM, Yu J-H, et al. Informed consent for whole genome sequencing: A qualitative analysis of participant expectations and perceptions of risks, benefits, and harms. American Journal of Medical Genetics Part A. 2012;158A(6):1310-9.
11. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013;339(6117):321-4.
12. Malin B, Sweeney L, editors. Re-identification of DNA through an automated linkage process. Proceedings of the AMIA Symposium; 2001: American Medical Informatics Association.
13. Roberts D. NSA mass collection of phone data is legal, federal judge rules. Guardian. 2013 27 December 2013.
14. Johnsson L, Helgesson G, Hansson MG, Eriksson S. ADEQUATE TRUST AVAILS, MISTAKEN TRUST MATTERS: ON THE MORAL RESPONSIBILITY OF DOCTORS AS PROXIES FOR PATIENTS' TRUST IN BIOBANK RESEARCH. Bioethics. 2012:no-no.
15. Godard B, Schmidtke J, Cassiman J-J, Aymé S. Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. European Journal of Human Genetics. 2003;11(2):S88.
16. Sporny M. The Beautiful, Tormented Machine. 2011; Available from: http://manu.sporny.org/2011/public-domain-genome/.
17. Iaconesi S. la cura. 2011; Available from: http://opensourcecureforcancer.com/.

18. Kaye J. From single biobanks to international networks: developing e-governance. Human Genetics. 2011;130(3):377-82.
19. Global Alliance . Creating a global alliance to encourage responsible sharing of genomic and health data: white paper. 2013.
20. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013;14(10):681-91.