wellcome trust sanger institute

Wellcome Trust Genome Campus
Hinxton Cambridge CB10 1SA
T 01223 834244
F 01223 494919
www.sanger.ac.uk



# Nuffield Council of Bioethics consultation:

*The linking and use of biological and health data*

**Response by the Wellcome Trust Sanger Institute**

January 2014

## Key Points

- It is imperative that the linking and use of biological and health data happens. Combining such information would enable better science, make better use of time and resources but, most critically of all, benefit all of society in terms of medical diagnoses, treatment and disease prevention.

- Proposed changes to the EU Data Protection Regulation should in no way be allowed to prohibit the use of anonymised data in research without consent. This would be at the severe detriment of research and, in turn, society as a whole

- It is paramount that everyone is made aware of the significance of the proposed linking and use of biological and health data. We believe that if people are appropriately informed, any fears associated with the linking of such databases could be allayed resulting in resistance from fewer individuals.

## Introduction to Wellcome Trust Sanger Institute

The Wellcome Trust Sanger Institute, one of the world's leading academic centres for genomic and genetic research, is grateful for the opportunity to contribute to the Nuffield Council on Bioethics consultation: The linking and use of biological and health data.

## Q1. Do biomedical data have special significance?

1. It was felt whole heartedly by researchers at the Institute that biomedical data does have special significance and therefore does deserve special protection, however, the level of protection should be consistent with the level of the anonymity of the data, but it should never be prohibitive to research. For example fully anonymised non-identifiable data used in aggregated form could be made openly accessible, whilst those data containing personal identifiable information need to be placed under strict managed access with a robust, yet transparent mechanism for access. Pseudo anonymised (coded) data would need to fall somewhere between these two extremes, and be handled accordingly. One Faculty member stated that researchers are not interested in trying to re-identify individuals and that anonymity was always fully respected.

2. The linking of data is fundamental to advances in both science and medicine. At present there are huge opportunities that are not being utilised yet, particularly given that we have a National Health Service database at our finger tips. By combining these data, studies would have more statistical power and more detailed information, thus improving our understanding of the potential causes of disorders and enabling us to answer previously unanswerable questions, but also allowing us to generate new questions.

3. With regard to genomic research it was felt that an individual has a right to consent to research even though there could be implications for the family and we should be careful not to "re-invent the wheel", i.e., there is a lot of information and experience that has already been gleaned through clinical genetics about consent and we would do well to take counsel from clinical geneticists when reviewing the process of consent. However it should be made apparent to research participants that there are no guarantees of anonymity, and although researchers should be honest in their work and about how participant's data will be handled, they should never make a guarantee about what might or might not happen to the data in the future.

4. The use of genomic data does come with great ethical debate, as these data can have life changing implications for both the participant and their families alike. However, it was felt that with an appropriate level of discussion prior to consent and with appropriate support from genetic counsellors there should be no great challenges posed by the collecting and use of these data.

## Q2. What are the new privacy issues?

5. Whilst it is paramount that there is a need to respect and protect participant confidentiality, it is also in the public interest to pursue research that will share, re-use and link data to accelerate discovery, advance healthcare and treat disease. In addition, such linking of data will allow studies that were not previously possible to happen so, in essence, it will allow us to ask new research questions and answer old ones.

6. Linking of data across large datasets may lead to the identification of individuals (the risk is very small but not non-zero). There was discussion with one investigator who had given considerable thought to the storage of data that could potentially cause the identification of individuals to arise and he suggested that "identifiable personal details" could be held in one arena, whilst clinical/medical/genomic data could be held in another arena. Therefore, any researcher wishing to

work with such data would need, not one but two, key identifiers to link the data and this would therefore possibly restrict any malicious attempts to identify individuals suffering from certain ailments. There was also some discussion around the NHS's ability to deal and cope with the linking of these data, as it was felt that some NHS regions were operating to different security standards than others and this could be very problematic moving forwards. It was felt that realistically this is a task that should be carried out by an outsourced company that has experience in this field.

7. No-one at the Institute was aware of any harm or even a perception of harm caused by the sharing of biomedical data – it was felt there are many hypothetical scenarios but no real evidence. One investigator felt that if patients are fully aware of what could potentially happen and any risks are accepted, the level of perceived harm would be lowered.

8. We should offer the public the opportunity to benefit society, but not necessarily advocate that this should be the case. It was suggested, however, that moral obligation is never enough, as morals were considered fluid depending on country/ethnicity and time of life. It was felt by one individual that society as a whole would want to share their data as it would be considered a "reasonable thing to do". However, they were concerned that there could be hurdles put in place due to a minority which would prevent the majority from taking part in something which would be for the greater good of society as a whole. In other words, minority concern should not prevent research when it is in the name of public good. Individuals should be entitled to opt out of research; however, these individuals should understand that in extreme cases, such as a global pandemic, they could be overruled.

9. Looking to the future, we should be educating individuals and informing them purposefully that their data will be used for a wide variety of medical research and, in doing so, inform them that ;
    i) Wide variety doesn't include everything such as taking details and offering them to companies, but rather it will be used for everything under the broad spectrum of medical research ranging from, for example, generating a new algorithm to discover genes, through to personalising medicine.
    ii) Decisions to release such data to researchers should be undertaken and moderated by reputable individuals as part of a Data Access Committee.

10. The definition of medical research needs to be defined as broadly as possible. In this way it would allow a data access system to stand the test of time as society changes and would negate the need for continuous reassessment of the term "medical research".

11. The subject of using people's data in ways in which they are unaware, for example as controls, was on the whole deemed to be suitable as long as broad and generic consent was in place alongside Research Ethics Committee approval. However, there was a danger that individuals objecting to this could lead to general loss in confidence which, if handled inappropriately, could escalate to the detriment of all.

12. It was felt by one researcher that, at some point in the future, there could be an incident that could result in the identification of an individual or individuals, or of data being used inappropriately. However, it was felt that it would be paramount at this juncture that people should not overact and, at all costs, "Political Panic" should be avoided as this had led to heightened regulation in the past and regulation should not be written as a reaction to 'scandal', but should be well considered, balanced and proportionate, and never a knee jerk reaction.

13. It was felt that biomedical data should not be treated as legal "property" as it would complicate issues, and inhibit the future of research.

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

14. Genomics is generating vast quantities of data, more so than ever before. Currently however, much of the data generated is held in silos divided by disease type, platform the data was generated on, method of collection and institution the data was generated at. It was felt that advances in data science were now making these data more accessible allowing previous research goals to be pursued. These goals could be realised through linkage between data sets thereby releasing the power of these data to enable greater understanding of human health and disease.

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

15. Without doubt the linking of these data will provide us with new knowledge. Whilst fully anonymised data would be useful in some circumstances it was felt that linked anonymised data would prove of most use as it would allow us to look at all information on an individual such as age, gender, and ethnicity, and in turn allow us to identify susceptibilities to disease based on a variety of factors.

16. We live in a world where people want to keep elements of their lives private, but this varies greatly from individual to individual. Whilst many feel that they have a right to privacy and control over their medical data, others will understand the benefits to be gleaned by sharing such data.

17. There will be a small minority of individuals that will not want to share their medical data and they will be extreme in this regard. Whilst we should respect their views, we should not however be restrained by them and their thinking.

18. Data driven research requires broad consent. Whilst broad consent is becoming more widespread and is conducive to conducting ethically data driven research, the consent process itself comes with significant burden both in time and resource for participant and researcher alike.

19. Whilst consent is considered to be a safeguard and an effective mechanism for the use of personal data, it should be followed with a clarification that "I cannot be sure that your data will not become public." The participant should also be made aware that even if they wish to withdraw from a study it will not be possible to withdraw existing data in its entirety, and that there will be no guarantee that this will stop their data from being used elsewhere.

20. Big pharmaceutical companies should be actively encouraged to use data and tap into all resources, as it is the only way that research will be accelerated. This is a mutually beneficial relationship between all stakeholders, and ultimately by assisting big pharmaceutical companies we are ultimately helping the patient.

21. There was surprise expressed by researchers that commercial companies have taken so long to respond to the patient calls to release clinical trial data, and that only now funding bodies are beginning to request specific data sharing plans. There needs to be a much stronger push on this as, by the release of such data, there would without doubt be faster progress in improving health, a better use of money and higher quality science.

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

22. With appropriate governance and management the benefits of linking these data will give us even more data *per se* and greater statistical power. The linking of data will give us a stronger understanding of epidemiological research, as it will allow us to understand more about the causes of disease, how to detect outbreaks and to see the effectiveness of interventions.

23. The main hope would be the ability to personalise medicine – leading to better use of limited resources and quicker selection of appropriate care for individuals.

24. Whilst it was felt that the public would want to share their data, it was also suggested that the public do not have a deep enough understanding of the value of this data and that more should be done to inform and engage them. The public need to have confidence in how their medical records and information will be used in research – again these fears could be allayed by appropriate education.

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

25. The main causes for fear are the use of biological and health data beyond research and health care for example by insurance companies or marketers.

26. The ethical implications of using predictive analytical research tools with biomedical data are fraught with danger and their use and implications of their use should be very carefully considered.

27. With regard to individuals profiting from the use of their biomedical data, this was considered a "tricky" question and the response of the Institute was very much split. One individual commented that if you can donate your body material for clinical trials and be paid for it, why you shouldn't be paid for donating DNA?

28. It was felt that data from direct to consumer health tests should be made available to the research community in an anonymised form, so as much data as possible could be used within the constraints of privacy considerations. However, it was understood that there should be an awareness of any risk sharing data may pose to the businesses involved.

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

29. There is grave concern that the draft EU Data Protection Regulation in its current form could have massive implications for genomic research. The proposed amendments will prevent the use of identifiable data in research without consent and make the use of pseudo-anonymised data in research without specific consent very difficult, if not impossible. Societal benefits of research, including improvements to healthcare and public health, and corresponding economic benefits are likely to suffer heavily under these proposals.

## Further comments

30. Currently there is a great sense of data paternalism throughout the UK, but we should not forget that patients have a right to share their data as much as they have the right to protect it. Whilst much could be gained by the linking of biological and health data through the NHS on a national level, broader thought and consideration must be given to linking of data on an international level, as there is good evidence through international collaborations, such as the International Cancer Genome Consortium (ICGC), that much more can be delivered by research using an international scale approach. For example, currently there are many rare diseases throughout the world that are yet to be diagnosed, however, if this data were to be linked internationally much more could be done in terms of understanding these diseases, understanding how to manage them, but more critically identifying ways of possibly treating or preventing them.

31. Another necessity critical to the linking and use of biological and health data is consent. For example currently the data from the DECIPHER project is held privately by each individual collaborating centre, with restricted access available to other centres. Whilst this is considered an appropriate level of security for such developmental disorders it also has its drawbacks, as although the DECIPHER project holds a wealth of information, its usefulness is occasionally prohibited by inadequate consenting of patients due to lack of clinician time, which results in a lack of ability to share data. Careful consideration therefore must be given to the consent process if the linking and use of biological data is to succeed.

32. There was also concern expressed about the inconsistency of Caldicott Guardians' decisions regarding sharing of diagnostic data. This needs to be reviewed and a level of consistency achieved across all NHS organisations.

The Wellcome Trust Sanger Institute would like to thank the Nuffield Council of Bioethics for the opportunity to contribute to this consultation.

David Davison
Director of Corporate Services
Wellcome Trust Sanger Institute