This response was submitted to the consultation held by the Nuffield Council on Bioethics on *The linking and use of biological and health data* between 17 October 2013 and 10 January 2014. The views expressed are solely those of the respondent(s) and not those of the Council.

The linking and use of biological and health data

A response to the Nuffield Council on Bioethics

January 2014

1. This evidence is submitted by the Royal Academy of Engineering, the UK's national academy for engineering. The views described in this response were compiled following consultation with our Fellows who have expertise in IT systems and security, and biological and health data.

Key messages:

2. This consultation has huge scope. To lead to coherent decisions and action, answers to the proposed ethical questions need to be implemented consistently across all government policies. Piecemeal approaches would lead to inconsistencies and incompatible actions by different groups, committees, departments and institutions.

3. The issues raised rightly focus on where the balance should lie in the conflict between an individual's right to privacy and the public benefits of sharing biological and health data. There are no simple answers and even simple policy interventions may result in unforeseen consequences.

4. Technology is advancing rapidly and any policy that is based on a view of technology is often rapidly overtaken. The benefits of linking and using biological and health data will increase, probably for decades, and the risk of threats to individuals, organisations and society from misuse of the data will increase in the same way. We argue for caution in making data widely available because, with current technology, once data are released they cannot be easily recovered, withdrawn, corrected or deleted.

5. We raise concerns over the issue of consent. Individuals should not lose the right to withhold consent to being involved in research. If their biological and health data have been collected for a specific purpose, the individual may have not given consent for their data to be used in subsequent research or linked data studies. It is important that individuals giving consent for their data to be used in linked research understand the significance and potential implications of their full consent.

6. The potential benefit to society of disclosing someone's biological data could be outweighed by the potential disbenefits to the individual. Therefore as a general principle, personal biological data should be considered private by default, and should not be disclosed unless the individual has given consent.

Do biomedical data have special significance?

7. Yes. From an individual's viewpoint, biomedical data may reveal deeply personal things about your past, lifestyle or members of your family that you would not normally share. In general, the public remain suspicious that their personal data, including biological and health data, will not be kept private.

8. From a collective view point, biomedical data could directly impact the health care system and contribute to the saving of lives for many individuals and populations.

9. We note that there are several data bases already in use for biomedical data. MEDLINE is the most comprehensive. Another is ASSIA, from Strathclyde University, which is an indexing and abstracting tool covering health, social services, economics, politics, race relations and education. Updated monthly, ASSIA provides a comprehensive source of social science and health information for the practical and academic professional. The significance of a specific set of data is best analysed with a systematic review using the Cochrane data base, where such issues of bias are considered.

What are the new privacy issues?

10. In general, there is a potential conflict between the individual's right to privacy of their biological data and society's need to know in order to provide public benefits. It is the role of government to secure the individual's basic rights against excessive control that could hold back potential innovations.

11. Privacy issues around data in the health and clinical sciences are complex. Opening up data and linking of data should be implemented with great care. Concerns are that although data is anonymised, there would be a way of linking it back to certain individuals. Firstly, that this could happen once researchers start integrating datasets from a number of studies, and secondly, if Open Access to genome information were implemented a higher risk of identification would occur.

12. The transfer of data increases the risk of it being accessed by unauthorised persons and misused. Therefore data should only be transferred in circumstances in which it is strictly necessary. Transferred data should be encrypted to prevent it being misused if it reaches the wrong hands. It should always be stored in an encrypted form. Staff from agencies, academic institutions and clinics who will access the data should undergo stringent selections and training.

13. It is often said that "if you have nothing to hide, you have nothing to fear" but we consider this unethical. The statement implies that it is wrong to have something to hide; however, many people have honest reasons to need privacy, such as:

    - medical conditions that could trigger prejudice, such as HIV/AIDS or mental illness
    - deep personal traumas in their past
    - the need to escape an abusive relationship
    - the need to protect adopted children
    - witness protection
    - avoidance of religious prejudices

14. People should not be discriminated against because of their data. If privacy is not the social norm, it will be hard for those who need privacy to lead a normal life. Society should respect these needs and to ignore them is unethical.

15. The statement in paragraph 13 can also be considered unethical because data lasts a lifetime and circumstances change. Things you did, legally and honestly, in the past may be a problem under later circumstances.

16. Personal data would be safer if anonymised; however, it has repeatedly been shown that anonymisation is not reliable. The risk of re-identifying the individual by matching data against other, often public, datasets will become greater in the future as technology and expertise progress.

17. Identity theft has the potential to lead to privacy invasions through the use of personal information for identity fraud and records disclosed through lost and stolen unencrypted mobile devices.

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

18. This is profound. There are no technological fixes for privacy that can be guaranteed to be future-proof. With the growth of funding in big data and analytics, the anticipated issues are likely to become reality.

19. The effective use of telemedicine, teleheath and personalised healthcare has the potential to reduce some of today's high cost burdens on the healthcare system and

improve access to healthcare professionals for hard to reach populations and individuals.

20. We raise concerns over the occurrence of selection bias that distorts results. For example, online data leaves out approximately 25 per cent of Europe's population who do not use the internet. This subset of the population are more likely to be poor or elderly, so policies based on online data collection will under represent these groups. Similarly, NHS data omits much lifestyle data for healthy people, not the best starting point for developing public health policy. Using 'big data' instead of carrying out controlled longitudinal studies could easily lead to bad policy.

21. It should also be noted that, in the future, algorithms and predictive models will be increasingly used in medical practice and research. It is not just the data but how they are analysed which brings in certain risks and uncertainties.

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

22. The opportunities are significant. Linked biomedical data are creating possible understanding of the causes and treatment of disease. In the case of health research, there are clear benefits of both sharing and opening access to data. Having access to more research could lead to faster progress because results would become more reliable, being drawn from a bigger pool of data.

23. There may be significance for patient groups. By linking data, patients could be presented with fewer requests for tests and participation in research. Health data, especially genome data, are also seen as particularly good for re-use and re-analysis because of their inherent qualities. Genome information has much content, some of which we do not yet understand. In the future we might learn more. This does however raise the issue of 'function creep': using data for purposes other than those for which they was originally intended and given consent to.

24. A person should not lose their right to withhold consent to being involved in research. When an individual gives their biological data, there may need to be a mechanism in place where they can provide consent for the data to be used in linked data studies. If not, then using samples in linked biomedical research in this way constitutes involving them in research without their consent.

25. Consent to the retention of biological data for research should remain revocable. Individuals should be able to request that their profiles be removed from a database. This is because, as technologies develop, the use of their data may change, and an individual may no longer feel comfortable with their data being held. In addition, the constant progress in DNA analysis makes informed consent very difficult. Most individuals will not fully appreciate how their biological data could be used, especially when linked to other biological and health data. For this reason they should be able to revoke consent if further reflection or learning about the issue causes them concern. A possible solution would be to take biological data from individuals or volunteers for the purpose of the investigation, but not to seek consent for those data to be retained and reused at the moment they are collected. Rather, once the investigation has ended, individuals could be contacted and asked whether they consent to their data being retained and linked with other biological and health data. Active consent should be sought at this stage: individuals actively agreeing to reuse, rather than having to actively deny consent. It may also be useful to allow a 'cooling off' period in which people can easily revoke consent so their data are not added to the database straightaway and can be destroyed if the subject so wishes.

26. Big data are a challenge and an opportunity. The use of the web enables access to raw data, and access to mass populations for analysis. Research at a global scale is

increasingly possible, potentially eliminating silo attitudes of researchers but with the risk of misusing accessible raw data.

27. There is currently underway a European Commission Project, RECODE[1] (Policy RECommendations for Open Access to Research Data in Europe). It includes a Work Package on Ethics, but with no outputs at present. However, the project has completed its first Work Package on 'Stakeholder Values and Ecosystems'[2]. Two sections are relevant to this consultation and look at the acquisition of genomic and other clinical data, and large scale modelling in biomedicine. The highlighted report summarises the perception of the value of making data openly available, and the problems of doing so.

28. A number of points emerged that have a less obvious impact on ethical considerations:

    - In order to fully understand and be able to work with linked data the user would need adequate metadata and information on how the separate data sets were generated. The methods and the context that have been used to collect the data are an integral part of the data's validity. This data may not be meaningful to those outside the field, and a potential user may need in-depth knowledge, expertise and specialised software in order to understand and work with the data at hand.

    - Access at different levels may be required. Full access if you are the patient's clinician, access to cohort data (not individual data) if you are a research worker and so on. This would require additional work in presenting the data and a gatekeeping mechanism.

    - Ownership, citation of the data collectors, providing the data with a licence allowing re-use (which may or may not include modification) are important. Ownership is a barrier to making data openly available, particularly when the data have been collected by consortia or consists of an aggregation of data from different sources. When there are a number of different stakeholders involved in various projects, such as pharmaceutical companies, patient organisations, ICT staff, journals and research institutions, each may have a different take on the use and sharing of data.

    - A kitemark may be needed for determining whether data are valid and reliable. Kitemark approval could depend on the context in which the data were collected, on our knowledge of the people or institutions who collected the data, and on the data having been shown to be useful.

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

29. There are many opportunities. With the right expert knowledge, linking data could lead to faster diagnosis and optimisation of prognosis. The inevitable increase in cost of advanced technology and specialisation could be balanced by the potentially significant health gains to patients from linking the managed data in future medical practice.

[1] EC Research Project RECODE http://cordis.europa.eu/projects/rcn/106728_en.html and http://recodeproject.eu/

[2] D1: Stakeholder Values and Ecosystems, RECODE, 2013, sections 5.2 and 5.3, pages 63-67 http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf

30. We raise caution that every positive impact creates a corresponding liability when a medical practitioner does not take into account some of the data that they had available and a patient suffers (because the data were overlooked).

31. We speculate that there will be impacts on medical practice policy. At the General Practice (GP) level, for example, if software and training were available to analyse linked data it is possible that specific personalised prescriptions could be managed in the setting of the GP surgery.

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

32. The opportunities and impacts are already major and developing rapidly. It is hard to quantify how valuable this data may turn out to be.

33. Biological and health data are increasingly being used to expand our knowledge of fundamental science. This is done with the use of biomimetics in a wide range of industries. By increasing our understanding of fundamental biology and genetics we can learn more about how human biology interacts and changes with environmental pressures and other systems.

34. In addition, biomedical data could be useful in a health and safety context and in consumer goods design, such as through ergonomics.

35. Biological and health data could be viewed as a unique resource to benefit UK plc through new businesses, drug development and lifestyle counselling.

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

36. There are complex issues to resolve and currently there is not a clear pathway.

37. The principle of protecting basic human rights against the misuse by individuals, organisations and states should be safeguarded. The right to privacy will need to be balanced with clearly defined potential benefit to mankind and should be the basis of governance.

38. It is common to see regulation in many industries fall behind rapid technological advances. We suggest exploring the use of a 'safe havens' approach as a model to share complex and sensitive data[3].

39. Legal safeguards need to be employed to prevent misuse of personal, biological and health data. This will be particularly challenging in the international context, where definitions of ethnicity are controversial, public health programmes may not be delivered by a single national health service, and questions arise about who will monitor such studies, or who will be the 'public' whose approval for consent is to be ensured.

Further references

40. The Academy report *Dilemmas of Privacy and Surveillance: challenges of technological change*[4] is a contribution to the public debate on information technology and its

[3] Royal Society, *Science as an open enterprise*, 2012, Box 3.6, page 56 http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

[4] Royal Academy of Engineering, *Dilemmas of Privacy and Surveillance: challenges of technological change*, 2007

possible impact on our privacy. It includes a number of recommendations for improving the benefits and reducing problems likely to stem from advances in information technology that could affect privacy. Recommendations are directed at engineers who design systems, policymakers and commercial organisations that collect data about individuals.

41. The Academy report *Privacy and prejudice: young people's views on the development and use on Electronic Patient Records*[5] presents the findings of a public attitudes research project to explore young people's views on the development of an electronic patient records system in the UK and its use in medical research. It found that young people have significant concerns regarding EPRs and they expressed a strong need to be in control of their own record.

42. It may be relevant to consider human rights laws and charters when reviewing this consultation. We highlight three below.

Universal Declaration of Human Rights[6] (UDHR)

43. The core principles of human rights first set out in the UDHR, such as universality, interdependence and indivisibility, equality and non-discrimination, and that human rights simultaneously entail both rights and obligations from duty bearers and rights owners, have been reiterated in numerous international human rights conventions, declarations, and resolutions. Today, all United Nations member States have ratified at least one of the nine core international human rights treaties, and 80 per cent have ratified four or more, giving concrete expression to the universality of the UDHR and international human rights.

44. Article 12 states, "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

UN Ethics Office[7]

45. The UN established an Ethics Office in 2007. It provides advice and ensures Ethical conduct of UN employees.

European Convention on Human Rights[8]

46. The Convention consists of three parts. The main rights and freedoms are contained in Section I, which consists of Articles 2 to 18. Section II (Articles 19 to 51) sets up the Court and its rules of operation. Section III contains concluding provisions.

47. Article 8, *The right to respect for private and family life, home and correspondence*[9], focuses on information privacy, which concerns the collection, use, tracking, retention and disclosure of personal information.

---

http://www.raeng.org.uk/news/publications/list/reports/dilemmas_of_privacy_and_surveillance_report.pdf

[5] Royal Academy of Engineering, *Privacy and prejudice: young people's views on the development and use on Electronic Patient Records*, 2010 http://www.raeng.org.uk/news/publications/list/reports/Privacy_and_Prejudice_EPR_views.pdf

[6] The Universal Declaration of Human Rights http://www.un.org/en/documents/udhr/index.shtml

[7] UN Ethics Office http://www.un.org/en/ethics/documents.shtml

[8] European Convention on Human Rights http://human-rights-convention.org/

[9] European Convention on Human Rights, Article 8 http://www.equalityhumanrights.com/uploaded_files/humanrights/hrr_article_8.pdf