NUFFIELD COUNCIL ON BIOETHICS

Response to the Science and Technology Select Committee (Commons) inquiry: The big data dilemma

September 2015

KEY POINTS:

- More research into the harms arising from abuses of data, in particular in the context of health care data, is needed.
- Robust penalties (including imprisonment) for the deliberate misuse of data, whether or not it results in demonstrable harm to individuals, should be introduced.
- Privacy breaches involving individual data should be reported in a timely and appropriate fashion to the individual(s) affected.
- Neither consent nor de-identification is sufficient to protect individual's interests. We provide a framework for the ethical governance of data initiatives.

Introduction

1 This response draws on the conclusions of the Nuffield Council on Bioethics' report *The collection, linking and use of data in biomedical research and health care: ethical issues* which was published in February 2015.This report looks at the ethical issues raised by 'big data' focussing specifically on the context of biomedical research and health care and sets out key ethical principles for the design and governance of data initiatives. The full report is available at http://nuffieldbioethics.org/project/biological-health-data/.

2 Data about individual biology or health is considered by many people to be somewhat more 'sensitive' than other day-to-day information. Partly, this may be to do with social norms, expectations about medical confidentiality, or the fact that some data may reveal stigmatising information. However, from the perspective of data science, whether data are treated as 'biological' or 'health related' data depends on the use to which they are put as much as the source from which they are obtained, or the purpose for which they were originally collected. While our report focuses specifically on the biological sciences and biomedicine, the developments in data use that led to the report are of a general nature, and affect equally fields such as public administration, and the provision of commercial and financial services. Therefore, we hope that our findings may be useful when considering the wider uses of data.

Opportunities and risks for big data

3 Our report starts from a construction of what is novel about the distinctive challenges that big data entails. We are generating more data than ever before, of increasing variety, including about human biology, health, disease and functioning. Meanwhile, advances in information technology and data science provide more ways, and more powerful ways to collect, manage, combine, analyse data. This offers opportunities to generate new insight and extract value from data. Opportunities offered by data in the context of healthcare and biomedical research include:

a. **Making health services more efficient** through better informed decisions about how to allocate resources.
b. **Improving health** by building a stronger evidence base to predict, prevent and treat disease, developing new treatments and using data to personalise treatment and care.
c. **Generating economic growth** by driving innovation in the life sciences. In particular the network of databases within the NHS in combination with genome science is seen as having the potential to generate significant new insights.

4 There is a strong public interest in the responsible use of data to support the development of knowledge, to drive innovation through scientific research, and to improve the wellbeing of all through improved health advice, treatment and care. However, pursuit of these opportunities must take into account the need to manage potential risks of data use, which may include cyber security threats, state intrusion into private life, discrimination, or the misuse of data leading to harm for individuals or institutions.

5 As part of the evidence gathering for its report, the Council commissioned, jointly with the Expert Advisory Group on Data Access (EAGDA), research into evidence of harms resulting from the misuse of data in the context of healthcare and biomedical research. Box 1 summarises potential harms identified as part of this research.

Box 1: Empirical typology of data abuses, their causes and resulting harms

Type of abuse (decreasing intentionality)	Causes of abuse (decreasing intentionality)	Harms caused by abuse (decreasing severity)
■ Fabrication or falsification of data ■ Theft of data ■ Unauthorised disclosure of or access to data ■ Non-secure disposal of data ■ Unauthorised retention of data ■ Technical security failures ■ Loss of data	■ Abuse of data to meet NHS/organisational objectives ■ Abuse of data to protect professional reputation ■ Abuse of data for self-gain (e.g. monetary gain) ■ Abuse attributed to third parties (e.g. hackers) ■ Disclosure by the press or media	■ Receipt of suboptimal care, resulting in detriment to health or death ■ Individual distress e.g. emotional, physical, etc. ■ Damage to individual reputation (e.g. societal, personal or professional) ■ Individual, financial loss ■ Damage to public interest (e.g. loss of faith in

- Non-use of data

- Unauthorised access without clinical or lawful justification (e.g. for curiosity)
- Against the wishes/ objections of the individual
- Abuse as a result of insufficient safeguards
- Abuse arising out of a Freedom of Information request
- Abuse due to maladministration (e.g. failure to follow correct procedures)
- Abuse due to human error (e.g. sending a fax to the wrong recipient)
- Non-use due to misinterpretation of legal obligations

confidential health service, general loss of public trust in medical profession, delayed or stunted scientific progress etc.)
- Damage to organisational reputation (e.g. to NHS)
- Potential for harm to individual, organisation or the public interest in future
- No evidence of harm found due to lack of reported information

Source: Laurie G, Jones KH, Stevens L, and Dobbs C (2014) *A review of evidence relating to harm resulting from uses of health and biomedical data*, available at: www.nuffieldbioethics.org/project/biological-health-data/evidence-gathering/

6 The research also suggests it is likely that the consequences of data misuse are intrinsically difficult to identify and significantly under-reported. There are also a number of obstacles to obtaining redress, including the prohibitive cost of legal action, the fact that victims may not be aware of the harm and the risk of privacy harms being compounded by publicity resulting from the case.

7 In its report, the Council makes a number of policy recommendations. The following relate specifically to potential harms of data misuse, relevant to the Select Committee's inquiry (they can also be found at paragraph **2.50** in the full report):

The UK Department of Health, alongside public and private research funders, should ensure there is continued research into the potential harms arising from abuses of data, and should remain vigilant to any new harms that may emerge.

The UK Government should introduce robust penalties, including imprisonment, for the deliberate misuse of data, whether or not it results in demonstrable harm to individuals.

The UK Government should ensure that privacy breaches involving individual data are reported in a timely and appropriate fashion to the individual(s) affected.

8 In relation to health data specifically, the Council concludes that:

The Independent Information Governance Oversight Panel (IIGOP) and HRA should maintain maps of UK health and research data flows, and monitor and evaluate the hazards and potential benefits of new and existing policies, standards, or laws governing the use of health data.

Public understanding

Limitations of de-identification

9 Techniques to de-identify data include aggregating data into large data sets, removing identifying information such as names or addresses of individuals (anonymisation), or replacing identifying information with a unique code (pseudonymisation). On their own, these techniques reduce the risk of re-identification but they do not reliably eliminate it. Whether or not an individual is identifiable will depend on what other information is or may be available (now or in the future), and on the means and motivation of the person who might wish to re-identify them.

10 The de-identification of individual-level data cannot, on its own, protect privacy as it is simply too difficult to prevent re-identification. This can only be expected to become more difficult as the accumulation of data, and corresponding processing and analytical power, make potentially identifying linkages increasingly possible.

Limitations of consent

11 Consent to data use is usually sought at the time the data is collected. As time goes on, and when it comes to making further use of the data, two obvious problems arise: does the consent still reflect the wishes or views of the individual who gave it; and does the new proposed use still fall within the possible uses that the individual who gave the consent originally intended? While consent acknowledges an individual's right to decide against some uses of data, it does not necessarily prevent harms occurring to them when there may be poorly understood or unforeseen consequences of data use.

12 Where a person providing data about themselves cannot foresee or comprehend the possible consequences of how their data will be available for linkage or re-use, consent at the time of data collection cannot, on its own, protect all of their interests.

Ethical governance of data initiatives

13 The changing context and potential for data re-use means that compliance with the law is not enough to ensure a data initiative is ethically appropriate. Those who manage data initiatives therefore have a continuing duty to promote and protect the legitimate rights and interests of those who have provided data about themselves irrespective of the terms of any consent given.

14 There can, however, be 'no-one-size-fits-all' solution to ensure ethical governance of data initiatives but we propose a set of principles which should be kept in mind when creating a new data initiative:

a. **Respect for persons**: the terms of any data initiative must take into account both private and public interests. Enabling those with relevant interests to have a say in how their data are used and telling them how they are, in fact, used is a way in which data initiatives can demonstrate respect for persons.
b. **Respect for human rights**: the terms of any data initiative should respect people's basic rights, such as the right to protection of private or family life. This includes limitations on the power of states and others to interfere with the privacy of individual citizens in the public interest.
c. **Participation**: decision makers should not merely imagine how people ought to expect their data to be used, but should take steps to discover how people do, in fact, expect their data to be used, and engage with those expectations. Involving people in the design and governance of data initiatives allows their interests and values to be expressed, transformed and reconciled. It can also help to secure their commitment to the outcome and build trust.
d. **Accounting for decisions**: data initiatives should include formal accountability, through regulatory, judicial and political procedures, as well as social accountability through periodic engagement with a broader public, as a way of re-calibrating expectations. Data initiatives must tell affected people what will be done with their data, and must report what actually has been done, including clear reports of any security breaches or other departures from the established policy.

15 In our report, we consider a number of initiatives as examples of good practice, and make recommendations for improving practice in others. The examples of NHS England's care.data scheme, and the Scottish Informatics Programme (SHIP) highlight, in different ways, issues around trust and public engagement (summarised in the box below).

Box 2: Case studies – public engagement & trust

NHS England's **care.data** initiative aimed to upload all GP-held data to a central repository, the Health and Social Care Information Centre (HSCIC), for research and other health-related purposes. Individuals would be able to opt out of having their data uploaded.

The public debate ahead of the initiative's launch and reactions of GPs, civil society and the media demonstrated that the uses intended by the Health and Social Care Information Centre (HSCIC), while provided for in law, were not consistent with people's expectations about how their data would be used, including by companies outside the NHS. As a result, the programme was postponed in order to create the opportunity to establish more appropriate governance measures. In addition to the involvement of the HRA Confidentiality Advisory Group and the appointment of a National Data Guardian, broader public engagement could help to address questions about what uses of data are

ethically appropriate.

An alternative approach was taken by the **Scottish Informatics Programme (SHIP)**. A key feature of SHIP was its commitment to public engagement – both in determining the acceptability of the initiative, and as an integral part of its continuing governance.

SHIP demonstrates a number of elements of good practice according to the Council's ethical principles for data initiatives. Risks and benefits are assessed on a case-by-case basis, focusing on context rather than simply the type of data used. The initiative aims to respect public and private interests, partly through public engagement; and it takes seriously the need for public trust and concerns about the involvement of commercial interests. Through its system of research authorisation it also acknowledges the importance of responsible behaviour on the part of professionals over and above the duty to respect the consent of patients, even where data with a low risk of re-identification are used.

Practical precepts for data initiatives

16 The key to ensuring sustainable public understanding, trust and participation in 'Big Data' initiatives will be to maintain the engagement of, and oversight by, patients and other affected people not just as a new initiative is being developed, but as it evolves over time. It is important that the promoters and operators of data initiatives using health and biomedical data give careful thought not just to how they secure moral acceptability and provide adequate transparency at the beginning, but also how this is to be maintained as the system evolves. Failure to maintain a workable reconciliation of moral, legal, social and professional norms, as much as a failure to produce it in the first place, can lead to a loss of public trust and compromise both the respect for private interests and the attainment of public benefits.

17 The Council's ethical approach gives rise to a series of precepts for someone approaching a data initiative, such as a lead policy official or a commissioner of services.

- **Identify prospectively the relevant values and interests in any data initiative.** Some process of stakeholder mapping and reflection on this will be essential as an initial step to understand where these interests are located and what informs them. These will include private interests but may also include economic and political interests, for example. Explicating their moral content may allow them to be set in the same light as other moral interests. This critical reflection may very often reveal that what appear to be 'hard constraints' or 'strategic imperatives' rest on moral assumptions or prior value commitments that ought themselves to be brought into question.

- **Take special care to identify those interests that may be especially at risk or that arise from diverse values.** Identifying situational vulnerabilities (i.e. why the consequences of a particular data initiative might disproportionately affect certain individuals or groups) and understanding how

different people value the potential benefits and hazards of data initiatives is essential to explore what forms of respect for individual freedoms (e.g. consent) and forms of governance may be required.

- **Do not rely simply on compliance with the law to secure that data use is morally appropriate, particularly where it does not fully reflect moral norms.** The norms enshrined in legal instruments, while they determine how data may be used (and, in certain cases, how it must be used) are insufficient to determine how they should be used. It should never be assumed that compliance with the requirements of law will be sufficient to ensure that a particular use of data is morally reasonable.

- **Establish what existing privacy norms are engaged by the contemplated uses of data.** These will have a number of different sources, including social conventions, value and belief systems, and needs of individuals, groups and communities. This might include, for example, norms of professional confidentiality, of data sharing within families or social groups, or of wider acceptance of data use. Findings from consultation or public opinion research will be informative at this stage (but caution should be exercised when relying on existing research as the circumstances, values and interests may differ from one data initiative to another). Resistance among the public to the involvement of profit-seeking commercial actors may be an important phenomenon in this context. If private sector organisations are going to play a role in the delivery of public services and public goods, this must be engaged with in formulating reasonable expectations. Attempts to shift norms or impose new norms without engagement risks undermining trust and therefore the objectives of the initiative.

- **Involve a range of those with morally relevant interests in the design of data initiatives in order to arrive at a publicly statable set of expectations about how data will be used.** Participation helps to ensure both that different values and interests may be represented and that expectations are statable in a way that is intelligible from different perspectives. It also helps ensure that an account is given of how morally relevant values and interests are respected. Structured public dialogue or other forms of deliberative engagement, including direct participation of representatives in the initiative, will often be valuable.

- **State explicitly the set of morally reasonable expectations about the use of data in the initiative.** These are likely to include who will have access to data and for what purposes, the way in which disclosures will be authorised (including the form of any relevant consent procedures) and how the conduct of those with access to data will be regulated or accounted for.

- **Involve a range of those with morally relevant interests in the continuing governance and review of data initiatives.** What constitutes morally reasonable expectations may alter over time as new opportunities and threats emerge and as norms shift. Measures such as monitoring relevant social research, periodic consultation or a standing reference panel of participants are desirable.