

Response to the Royal Society and British Academy call for evidence on data governance

November 2016

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 considers current arrangements for governing the use of data 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 status are often considered to be more 'sensitive' than other data. This may be related to 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' 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.

Data opportunities and risks

- 3 In health care and biomedical research settings, digitisation has allowed an escalating accumulation of data including:
 - clinical care data (e.g. primary care and hospital records)
 - data from clinical trials and observational studies
 - patient-generated data (e.g. from 'life logging' or consumer genetic testing)
 - laboratory data (e.g. from imaging, genome sequencing and other 'omics')
 - administrative data or metadata
- 4 Given the UK's strong research base in the biomedical sciences and the unique resource and infrastructure of the UK's national health services, the use of health data has become a strategic focus. Opportunities offered by data in these contexts include
 - Increasing efficiency and transforming service delivery
 - Generating improvements in medical treatment
 - Generating economic growth from the life sciences.

- 5 To achieve these outcomes a number of policy orientations have been set in the UK and elsewhere, such as:
- increasing IT intensity and introducing new infrastructure in health systems
 - establishing partnerships between the public and private sectors to promote resource exploitation and innovation
 - centralising data resources to facilitate analysis of linked data
 - promoting 'open data' and 'data sharing' to encourage the widest possible use of resources
 - Investing in 'big data' and in the knowledge economy to foster development of new tools, methodologies and infrastructures.

For a more detailed account, please see Chapter 2 of [our report](#).

Concerns about data use and governance

- 6 There is a clear public interest in the responsible use of data to improve well-being through improved health advice, treatment and care, as well as through increasing economic prosperity more generally. However, the pursuit of opportunities must take account of the need to manage a number of threats to welfare. These threats take a number of forms, for example:
- **Misuse of data leading to harms to individuals and institutions** (ranging from detriment to health, loss of privacy, financial loss, reputational damage, stigmatisation and psychological distress).
 - **Discriminatory treatment**, ranging from targeted advertising to differential pricing that compounds social disadvantage, to discrimination in insurance and employment.
 - **State surveillance of citizens**, particularly in the light of revelations about the US National Security Agency, which is greatly facilitated by large databases and linked information systems.
- 7 Independent research commissioned to inform our work¹ found that the negative impacts of data misuse are potentially much wider than are those recognised by legal and regulatory systems. Furthermore, the nature of privacy harms and of the judicial and regulatory systems means that they are likely to be under-reported by the victims and obtaining redress is difficult.
- 8 The Council's report includes a number of policy and governance recommendations to address these concerns, including continued research into the potential harms arising from abuses of data, and robust penalties, including imprisonment, for the deliberate misuse of data, whether or not it results in demonstrable harm to individuals. Since the publication of our report, we have been in dialogue with other national medical and research organisations about how these recommendations might be implemented.

¹ Laurie G, Jones KH, Stevens L, Dobbs C, *A Review of Evidence Relating to Harm Resulting from Uses of Health and Biomedical Data*, available at <http://nuffieldbioethics.org/wp-content/uploads/A-Review-of-Evidence-Relating-to-Harms-Resulting-from-Uses-of-Health-and-Biomedical-Data-FINAL.pdf>

Limits of current governance arrangements

9 A number of overlapping legal measures exist to protect individuals' privacy, principally: formal privacy rights, which guarantee freedom from interference, albeit that they may be qualified by certain public interest considerations; rules of data protection, which control the 'processing' of various kinds of 'personal data'; and duties of confidentiality, which protect against unauthorised or unreasonable breaches of confidence.

De-identification

10 Technical measures may also be applied to prevent the identification of individual subjects and reduce the risk of privacy infringements:

- **aggregation** of data makes it harder to distinguish individual cases, although it is not wholly secure in the face of modern statistical techniques; it also makes further linking of data difficult
- **anonymisation** by the removal of identifiers also makes individuals difficult to re-identify, although re-identification may still be possible depending on what other data or information are available
- **pseudonymisation**, the replacement of identifiers with a code, enables linking of data where the correspondence between the code and the case is known, although data may still be vulnerable to inferential re-identification

11 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.

Consent

12 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.

13 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

14 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.

15 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:

Ethical principles for data initiatives

The use of data in biomedical research should be in accordance with a publicly statable set of morally reasonable expectations and subject to appropriate governance.

- **The set of expectations about how data will be used in a data initiative should be grounded in the principle of respect for persons.** This includes recognition of a person’s profound moral interest in controlling others’ access to and disclosure of information relating to them held in circumstances they regard as confidential. This does not mean that individuals’ interests may never be overridden, but that they may only be overridden where there is a legitimate reason to do so.
- **The set of expectations about how data will be used in a data initiative should be determined with regard to established human rights.** This will include limitations on the power of states and others to interfere with the privacy of individual citizens in the public interest (including to protect the interests of others). This principle seeks to avoid potential rights conflicts and violations rather than leaving them to be dealt with retrospectively through judicial processes.
- **The set of expectations about how data will be used (or re-used) in a data initiative, and the appropriate measures and procedures for ensuring that those expectations are met, should be determined with the participation of people with morally relevant interests.** This participation should involve giving and receiving public account of the reasons for establishing, conducting and participating in the initiative in a form that is accepted as reasonable by all. Where it is not feasible to engage all those with relevant interests – which will often be the case in practice – the full range of values and interests should be fairly represented. This allows the identification of relevant privacy norms and the development of governance measures (such as design of consent and authorisation procedures) in relation to these norms; it allows preferences and interests to be expressed and transformed through practical reasoning, and account to be given of how these interests are respected in decision making, helping to foster trust and cooperation.
- **A data initiative should be subject to effective systems of governance and accountability that are themselves morally justified.** This should include both structures of accountability that invoke legitimate judicial and political authority, and social accountability arising from engagement of people in a society. Maintaining effective accountability must include effective measures for communicating expectations and failures of governance, execution and control to people affected and to the society more widely. This ensures that data initiatives remain in touch with changing social norms.

Practical precepts for data initiatives

- 16 A key aim of data governance in the context of biological research and health care should be to ensure sustainable public understanding, trust and participation in data initiatives. For this to be possible, an essential element 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.

Governance in practice

18 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).

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 (and eventually dismantled) 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.