The collection, linking and use of data in biomedical research and health care:

ethical issues

a guide to the report

Contents   Pages

This guide outlines the main themes and recommendations that are discussed in the Nuffield Council on Bioethics' report **The collection, linking and use of data in biomedical research and health care: ethical issues** (published February 2015).

The report considers the ethical questions raised by advances in information technology and data science in the context of health care and biomedical research.

The report was produced by an expert Working Party. In coming to its conclusions, the Working Party invited contributions from a wide range of people including by holding an open consultation that ran from October 2013 to January 2014.

Introduction

Since the last decade of the 20th century, developments in biotechnologies, health care systems and computing have led to a dramatic growth in the volume and variety of data about people's health and biology.

More data is being generated than ever before, including:

- Electronic medical records
- Genome sequences
- A wide variety of biomarkers
- Body and brain scans
- Data from clinical trials or observational studies
- Lifestyle information collected directly by individuals

Advances in data science mean that there are now also more ways to collect, manage, link and analyse health and biological data in order to generate information for research and other purposes.

A new attitude towards data

Developments in data science that allow researchers to manipulate and 'mine' huge data sets have led to the emergence of a new attitude towards data, where it is seen as a valuable resource that may be re-used, linked, combined and analysed indefinitely, for a variety of purposes.

There are significant opportunities for using data to improve medical practice, produce more efficient services, generate new knowledge and drive innovation.

Data initiatives

The focus of the report is 'data initiatives', which we define as projects involving one or both of the following practices:

- **Where data collected or produced in one context, or for one purpose, are re-used in another context or for another purpose.** This may result in the data taking on a different meaning and significance. For example, biomarker data may be used to inform someone's treatment, but may also be used for the development of therapies, the allocation of resources, or the planning of services, moving between health care, research, financial and administrative contexts.

- **Where data from one source are linked with data from a different source or many different sources**. For example, where data from a disease registry are linked to data about the location of discharges of environmental pollutants to examine or monitor any link between them.

Data initiatives exist at different scales. They may be large – at the scale of a national biobank, health system or international research collaboration – or small – on the scale of an individual research project.

Data opportunities and threats

Decisions and actions informed by the use of data in research and health care may have both beneficial and harmful effects on individuals or on broader groups of people including families, companies, social groups, communities or society in general.

Different people may value these potential benefits and harms very differently – what may be profoundly troubling for one person might be a matter of indifference to another.

Opportunities of data use

There is a strong public interest in the responsible use of data to support the development of knowledge and innovation through scientific research and to improve the wellbeing of all through improved health advice, treatment and care.

The opportunities offered by data use include:

- **Making health services more efficient** through better informed decisions about how to allocate resources.

- **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.

- **Generating economic growth** by driving innovation in the life sciences.

Developments in data science and information technology mean that new opportunities have arisen, and will continue to arise, to extract value from data.

Risks of data use

These opportunities must be balanced against the risks of data use, which may include cyber security threats, state surveillance, discrimination, or the misuse of data leading to harm for individuals or institutions.

Potential harms may include:

• The receipt of suboptimal care through inefficient sharing of data between clinical teams.

• Personal distress through loss of privacy.

• Inhibiting potentially valuable research through loss of public trust in the medical profession.

Many of these potential harms are not recognised or sufficiently dealt with by current legal or regulatory measures.

Independent research commissioned to support this report suggests that 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.

We conclude

• Public and private research funders and the UK Department of Health 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 Independent Information Governance Oversight Panel and the Health Research Authority (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.

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

• 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. [Chapter 2]

Values and interests at stake

There are a number of values and interests at stake when claims are made about whether it is right or wrong to use data in the context of a data initiative.

The significance of data

Medical records are clearly personal, but individual data, such as whether a patient has attended their GP, are not intrinsically any more or less sensitive than any other personal information about people. What matters is the context – for example, records of fertility treatment may be highly sensitive for some people in some contexts.

> The sensitivity of data is highly dependent on the context in which they are used, and how they relate to other information, people, decisions and actions, rather than how the data are categorised.

The value of privacy

Privacy is fundamentally important to individuals and groups in establishing and maintaining their identity and relationships with others. Family, group, community and even national identities may be formed and confirmed by the way information is shared.

Individuals generally see controlling access to personal information as an important aspect of maintaining their privacy. If information is accessed or disclosed against their wishes, it can affect individuals' well-being and infringe their rights. Respecting people's privacy shows respect for them as individuals.

However, there are circumstances in which it may be acceptable to challenge normal expectations of privacy. Sharing the confidences of a family member with strangers may infringe their privacy; on the other hand, concealing instances of domestic abuse could be more unacceptable.

> Determining when it is acceptable to challenge, and potentially breach normal expectations of privacy will depend, to a great extent, on the nature of the relationship between the individuals or institutions concerned (including the state).

Confidentiality and consent

In formalised relationships, such as between a doctor and a patient, the expectation of privacy is enforced through rules of medical confidentiality. But these rules and expectations may be modified in specific cases, for example where people might consent to disclosure of their data for a specific purpose, such as a medical research project.

For consent to be valid, it must be freely given and cannot be obtained by coercion or deception. The person giving consent should be aware of the

implications of doing so. This need not mean that they must be aware of every last detail of how the information will be used, but they should be aware of the details that they consider to be relevant to them.

There is considerable debate about how much information and understanding is necessary for consent to be valid.

Public and private interests

Public interest is an important concept, in particular when data initiatives are carried out with the involvement of the public sector or public funding, or are aimed at delivering public good.

As individuals, members of families, groups, communities and nations, we all have both private and public interests. There is a private interest in protecting privacy and in promoting the public good; and there is also a public interest in the protection of privacy and in promoting the public good.

People have different interests, preferences and priorities, which may complement or contradict each other. When we consider which, and whose, interests are relevant in a particular data initiative, it is important to remember that these include not only those of the people to whom data relate, but also of those making use of the data and those who have an interest in the aims or outcomes of the initiative.

Decisions about data use are complicated by the fact that there are often powerful political, economic and scientific interests, which may set out the terms of a data initiative prior to any wider public debate.

The report considers how we should identify and take into account the relevant interests and expectations about how data will be used within the context of a data initiative.

Law, governance & security

A number of legal instruments exist to protect privacy in different ways, but developments in data science and the incentives to link and re-use data have put significant pressure on conventional governance approaches.

In the UK, the Human Rights Act guarantees a right to privacy, except where there is an accepted and overriding public interest.

Data protection law in the UK and Europe controls the processing of certain categories of data and applies enhanced controls to sensitive data such as health data. Specific relationships also generate duties of confidence, such as that between a doctor and a patient.

Where data are to be re-used in other contexts, or for other purposes, procedures to seek the consent of individuals to share data or to de-identify data are typically used in order to ensure their privacy is not breached. However, in the context of modern data initiatives, there can be significant problems with these strategies.

The limitations of de-identification
Examples of how data may be de-identified include:

- Aggregating data into large data sets.

- Removing identifying information such as the names or addresses of individuals (anonymisation).

- Replacing identifiers 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.

We conclude

- 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. [Chapter 4]

The limitations of consent

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.

We conclude

- 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.

- 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. [Chapter 4]

The need for good governance

Whether in health care or biomedical research, the widest access to the richest data is implicitly desirable in order to advance research or improve the efficiency of public services. Those designing data initiatives find themselves in a situation where they are obliged to generate, use and extend access to data, while at the same time protecting privacy.

The limitations of 'consent or anonymise' mean that additional governance arrangements are usually required, including oversight committees authorising access to data; limiting data access through 'safe havens'; or formal agreements on the limitations of data use.

The key issues facing data initiatives are not merely to do with re-identification of individuals. Decisions about how data are used may have consequences for the way different people and groups are treated.

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. Continuing, active participation in governance by those with relevant interests is needed.

Ethical governance of data initiatives

The report does not intend to offer a universal set of instructions to be followed when creating a new data initiative – there can be no 'one-size-fits-all' solution. Rather, it considers the kind of questions that need to be addressed, and the principles that should be kept in mind when doing so.

Data initiatives as human practices

The formation of a data initiative is a complex social practice where tensions and conflicts of interests may exist at many levels: for example, at the level of the individual, of professions, or of the public.

Any data initiative will involve a number of different people, including regulators, commercial firms, doctors, researchers, patients and the wider public, each with their own values, interests and expectations. Nevertheless, there needs to be a legitimate means of reaching decisions about the use of data.

An optimised approach to decision-making

The report argues that an ethically appropriate use of data should respect certain core moral standards, reflecting the basic rights that underpin the legal system, rather than simply aiming to satisfy the requirements of the law. A proposed course of action can be lawful, but still morally questionable.

Involving people in the design and governance of a data initiative 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.

Ethical principles for data initiatives

The report sets out four ethical principles for the use of data in biomedical research and health care:

- **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.

- **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.

- **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.

- **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. [Chapter 5]

The report considers how these principles are reflected in the design of data initiatives with regard to four key elements:

- **Arrangements for data storage**: whether data are retained close to the point of collection, more widely distributed, or gathered together in safe havens.

- **How data are disclosed/accessed**: whether data are published, subject to controlled disclosure or access, or subject only to indirect access.

- **The role of consent or other forms of authorisation**: for example, from explicit individual consent, through implicit consent with opt out, to authorisation by committees and other bodies.

- **The range of approved users**: including academic researchers and commercial users, and how they demonstrate (or not) that their aims are in the public interest.

The report examines a number of data initiatives in health care and biomedical research in relation to our ethical principles and the elements outlined above. Some of these examples are summarised in this guide.

Data initiatives in health systems

There are a number of examples of data initiatives at different levels within the UK National Health Service (NHS). Initially focused on improving business efficiency and the delivery of care, data initiatives within the NHS increasingly aim to support biomedical research and public policy objectives.

The report suggests that without adequate public consultation and involvement, along with trustworthy governance systems, which respect the interests of those involved, initiatives that could have wide public benefits may continue to be challenged and fail to secure public confidence.

Appropriate use of data

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 reaction of GPs, civil society and the media demonstrated that the uses intended by the 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.

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 so that, for example, patients can properly consider what the implications of opting out mean for themselves and the public interest more broadly.

Commitment to public engagement

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 our ethical principles. 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.

Public-private collaboration

A third example is the **100K Genomes** project, delivered by Genomics England Ltd, a company owned by the UK Government. The project was established with strong political backing, to use the NHS as a resource to realise the prospects of genomic medicine.

The project is clear about generating economic returns but, given the private-public relationship, there is a need for a clearer public account, and greater public accountability, regarding how its governance serves the public interest.

We conclude

The report makes a number of recommendations in relation to data initiatives in health systems, including that, for HSCIC:

• An independent, broadly representative group of participants should be convened to develop a public statement about how data held by the HSCIC should be used.

• There should be complete audit trails of everyone who has been given access to the data, and the purposes to which they have been put. These should be made available to all individuals to whom the data relate or relevant authorities in a timely fashion on request. [Chapter 6]

Population research data initiatives

A trend in life sciences research is the increasing use of very large datasets by international teams of researchers studying a wide range of health conditions and diseases. In 2014, it was estimated that one in 30 of the UK population (2.2 million people) were participating in cohort studies.

Adapting to changing circumstances

A common feature of many data initiatives is uncertainty about the specific future uses of the accumulated data. **UK Biobank** relies on a model of broad consent and the project aims to keep participants informed about research carried out, and any results, through newsletters and their website.

UK Biobank has a published Ethics and Governance Framework (EGF), which was developed in consultation with a range of stakeholders, and which is monitored by an independent Ethics and Governance Council. This provides a good example of a participative approach to developing a data intiative, though ongoing participation in governance systems will also be important as the project evolves.

The **UK10K Rare Genetic Variants in Health and Disease** project operates a federated system, where different projects work together under a common EGF. This puts significant emphasis on the local principal investigators to interpret the interests and expectations of participants through consultation with participants themselves.

Both of these data initiatives foreground the role of consent and recognise the challenges of interpreting it in different and changing circumstances.

We conclude

- There should be appropriate mechanisms in place so that governance arrangements can evolve during the life of any data initiative through deliberation with participants, the public, funders and the research community in order to ensure that the interests of participants are respected over the life of the project.

- Partners in research projects should be subject to reasonable surveillance to identify inappropriate data use, and sanctions for data misuse through recognised research institutions. [Chapter 7]

International scientific collaborations

There are many benefits of international collaborations but also significant difficulties. For example, scientists in one country cannot police the activities of those in another country; there may be different laws in different countries; or uncertainty about responsibilities for data protection.

Research suggests that by 2020, 80% of all health care data will pass through a cloud computing provider. The location of data repositories or cloud computing facilities may become a concern (for example, where domestic legislation enables relatively free access by the security services or other agencies of the country in which the repositories are based).

We conclude

- Collaborators on international data research initiatives should agree an explicit ethics and governance framework, and integrate this at their local research site.
- National bodies should publish their policies on the use of cloud computing in health data settings so that data initiatives can communicate this to participants. [Chapter 7]

Participant-led research

Increasing access to digital technologies and the rise of online social networks have facilitated the formation of online communities of people engaged in health research including self experimentation, self surveillance and interpretation of genomic data.

New initiatives, such as **PatientsLikeMe**, have the potential to generate valuable health knowledge. However, while some projects may involve collaboration with conventional academic or commercial research systems, many do not, and consequently may be subject to varying levels of oversight. However, trying to force this research into the conventional mould may well stifle the features that could make it valuable.

We conclude

Biomedical researchers should consider how to maximise the potential of participant-led research to generate health knowledge and secure public benefits, while providing adequate protection of those involved through continuing ethical and scientific assessment. [Chapter 7]

SUMMARY

We are generating more data about people's health and biology, from more sources, than ever before including GP records, hospital notes, laboratory tests, clinical trials, monitoring devices and health apps. Advances in information technology and data science mean that it is becoming easier, cheaper and more valuable to gather, transfer, link, store and analyse these data. This offers significant opportunities to generate new knowledge, improve medical practice, increase service efficiency and drive innovation.

The report looks at the ethics of data use by considering the relationship between privacy and public interest, and how developments in data science and computing have put significant pressure on conventional approaches to information governance, including the approach of seeking consent or anonymising data for use in research.

More needs to be done to ensure that respect for participants and the protection of their data is at the centre of any initiative, through participation and accountability, backed up by good governance, and criminal penalties for the misuse of data. To marginalise individuals who provide data means risking the trust of current and future generations, exposing people to unacceptable risks, and ultimately missing out on the benefits of research.

The report sets out key ethical principles for the design and governance of data initiatives, and identifies examples of good practice relevant to anyone approaching a data initiative, such as a principal investigator in a research project, lead policy official or commissioner of services.

Copies of the short report and this guide are available to download or order from the Council's website: www.nuffieldbioethics.org