Royal Society/British Academy call for evidence on "data governance"

Response by the Wellcome Trust

4 November 2016

Key points

- The health sector offers valuable lessons for data governance that could be applicable across different sectors.

- Governance should facilitate innovation where it serves the public interest and not create unnecessary obstacles, while ensuring data is not shared or used inappropriately.

- We can only realise the benefits of data if there is public support for its use. Trustworthy governance systems and meaningful public dialogue are critical to address low public confidence in use and management of health data.

- The UK has an opportunity to build on its strengths to develop a global leadership role on data governance.

Introduction

1. Wellcome is the UK's largest charitable foundation. Over the next five years, we plan to invest up to £5 billion in biomedical research and the medical humanities in the UK and internationally. We also support the development of new commercial innovations to improve health.

2. Much of the research we fund collects and links health and other types of data to enhance our understanding of health and disease and to develop new health interventions. As a result, data regulation and governance in the UK, EU and globally is a major area of policy work within Wellcome. In the following sections we outline Wellcome's views on four main challenges for good data governance and highlight some potential models drawn from biomedical and health research.

The opportunities for data science

3. There is extraordinary potential for innovation in the analysis of data and data-driven technologies to lead to improvements in human health. This potential mainly falls into two related areas:
   - Depth: Some research fields are now producing vast quantities of detailed data that hold huge potential for advancing biomedical research, if this data is accessible. Genomics and bioinformatics are two key areas of UK strength.

- Breadth: The potential for linkage across different data types can lead to novel insights. In population health, linking data across different sources is enabling researchers to derive important new knowledge about the relationships between different factors affecting health, for example linking housing or income data to health records.

Data siloes

4. While the potential for linkage across different datasets is promising, data is largely managed in siloes. These siloes are defined by sectoral boundaries, boundaries of academic disciplines, or in the case of administrative data, departmental boundaries. As the data is in siloes, so are the governance mechanisms for its access and use. Even within the contained sphere of academic population research, there are a number of data governance mechanisms that use their own nuanced terminology, processes and rules[1].

Concerns

5. These siloes create challenges for governance. When data are linked, it is not always clear who is responsible for the data flows and the enriched, merged data that results. Identity disclosure risks are also higher when datasets are linked.

6. For government-held data, the picture is even more complex as there are numerous legal gateways and barriers operating in different departments and under different rules. Current proposals are insufficient to address this: health data is explicitly excluded from administrative data covered by the Digital Economy Bill, representing a missed opportunity to bring coherence to an otherwise fragmented governance landscape.

7. Established siloes are robust and tend to have developed in a way that works for their specific communities (for example, within one academic discipline, or one government department). These difficulties cannot be overcome solely by technical means as they are often cultural.

What's needed

8. A system of oversight needs to be flexible enough to accommodate the requirements of different statutory rules, data types and risks, yet firm enough to ensure coherence. Governance must be facilitative and support the linkage and use of data that is appropriate and ethically justified.

Legal frameworks

Concerns

9. The UK legal framework for the use of personal data in health research strikes a good balance between permitting research and protecting individuals. However, it is highly complex and confusing due to fragmentation between statute and common

---

[1] See for example the Expert Advisory group on Data Access report on "Governance of Data Access" (June 2015) https://wellcome.ac.uk/sites/default/files/governance-of-data-access-eagda-jun15.pdf

law. This has contributed to a risk-averse culture in sharing and using data for research, delaying and disrupting research in the public interest.

What's needed

10. Assuming the EU General Data Protection Regulation is implemented by May 2018, this provides an opportunity for the UK to clarify and simplify its legal framework. The UK Government must seize the opportunity for holistic reform of the legal framework for the use of personal data in research. For example this could be achieved by;
    - creating a dedicated 'public interest' legal basis for scientific research for private and public organisations;
    - bringing standards of consent and safeguards for health research in the Data Protection Act and the common law duty of confidentiality closer together.

11. We would also like to see the Information Commissioner's Office given greater power and resource to oversee and audit organisations using personal data, with tougher sanctions for misuse of data, including criminal sanctions where appropriate.

The role of consent

12. Some argue for "consent or anonymise" to be the key governance principle for health data. However, it is not always practical to anonymise or seek consent. Technological developments in data science are making 'consent or anonymise' even more challenging, on two fronts:
    - With increasing possibilities for future uses of data that cannot be anticipated at the time of data collection, it is not possible to inform participants of all the potential ways in which their data could be used in the future. It is disingenuous to imply that consent under these circumstances is fully informed.
    - Techniques can be developed for re-identifying individuals from datasets previously thought to be anonymised. This undermines the assertion that individual-level data can be rendered truly anonymised without the possibility of re-identification.

What's needed

13. Good governance can manage the unpredictability about future uses of data. 'Broad consent' is becoming more common and allows participants to delegate future decision-making on how their data is used to a body such as a Data Access Committee (DAC) or Ethics Committee. Independent of the data controllers, a DAC can make decisions about requests for data on a case-by-case basis, based on a strong technical understanding of the datasets, the risks of linkage, and the ethical and legal justification for data use. Further work is needed to establish the limits and conditions of this approach, as well as exploring alternative models such as dynamic consent that give the individual more power over data about them.

14. In some circumstances, it will not be practical to seek consent for the use of identifiable data. It is vital that there are trustworthy governance arrangements and safeguards in place to support the use of data in these cases. For example, section

251 of the NHS Act 2006 enables the disclosure of confidential patient information for medical purposes[2].

15. As a research funder, we seek to facilitate the development and dissemination of best practice in governance as research methods adapt and change over time (see EAGDA summary on p.6).

Public confidence

16. We can only "unlock the power of data" if the systems for managing and using data are able to maintain people's support. For research involving patient information, there are particular sensitivities: records are created under the terms of the doctor-patient relationship, which has respect for confidentiality. Public awareness of how health data is used is low[3], which is a barrier to meaningful discussions about data.

Concerns

17. Following care.data, public confidence in uses of health data, and trust in the government to securely and safely manage this data, are low. There is also a 'data trust deficit'[4] in institutions that manage or control data. Good governance can play an important part in rebuilding this trust. The Caldicott Review into data security and opt-outs[5] is a starting point but it treats health and care data as isolated from other data types, reinforcing issues of data siloes (see 5.-7.).

18. Recent research has shown many people are concerned about commercial organisations using health data[6]. Yet, advances in many areas of research depend on partnerships with industry, for example, drug development, genomic sequencing and analytic software development.

What's needed

19. Governance systems for data need to be comprehensive, proportionate and above all, trustworthy. It is important that people feel confident in the way data is used, and that these uses are in the public interest. This is especially true for new and emerging technologies, which people may be initially suspicious about. Good information security and governance are therefore critical to the success of these technologies. Public scrutiny and engagement also has an important role to play, as demonstrated in the acceptance of other technologies such as mitochondrial donation.

20. Wellcome is supporting a new initiative to improve conversations about what happens to health data, including developing toolkits and resources for researchers, policymakers, patients, media and industry to help them talk about patient health

---

[2] http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/what-is-section-251/
[3] See the Ipsos MORI report commissioned by Wellcome: "The One Way Mirror: public attitudes to commercial access to health data" www.wellcome.ac.uk/publicattitudes
[4] See the Royal Statistical Society's research https://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers
[5] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF
[6] See the Ipsos MORI report commissioned by Wellcome: "The One Way Mirror: public attitudes to commercial access to health data" www.wellcome.ac.uk/publicattitudes

information in a balanced way[7]. It is likely that some of the initiative's outputs will resonate with broader data use and governance discussions. We would be happy to discuss this initiative further.

The data governance landscape

21. Here we profile case studies from health research, which may serve as models for different aspects of good data governance. These models have several common features:
   - Independence from the activities they oversee or advise.
   - Membership with a breadth of relevant expertise.
   - Adequate resources to fulfil their remit.
   - Transparent decision making and advice.
   - Facilitate openness or public discussion.

Summary of case studies (see Appendix I for full text)

1	**Expert Advisory Group on Data Access (EAGDA)**	EAGDA is an independent expert group that advises UK organisations who fund a diverse range of research, on data governance. Their advice facilitates a joined up approach between the funders and helps them to anticipate and deal with emerging trends in data use and processing.
2	**UK Biobank Ethics and Governance Council (EGC)**	The EGC is an independent expert group that monitors data governance decisions and related procedures for access to resources from the UK Biobank project. The Council works in a transparent way, reporting its findings publicly and appointing its members through an open process. It is also independently reviewed on a regular basis.
3	**The Global Alliance for Genomics and Health (GA4GH)**	GA4GH brings together organisations from over 70 countries and embeds social and ethical principles into the practical tools and technology it develops to unlock the potential of genomic data.
4	**Managing Ethico-social, Technical and Administrative issues in Data ACcess (METADAC)**	METADAC assesses applications to access data from several population studies, and has harmonised data standards, requirements and language to streamline the process.
6	**National Data Guardian (NDG) Review 2016**	In the recent review of UK health data governance, Dame Caldicott proposed an opt-out model to give people a choice in how their data is used beyond their direct care and set out a model for governance of patient information. Meaningful public consultation is central to the success of these proposals.

---

[7] https://wellcome.ac.uk/news/independent-patient-data-taskforce-announced

5

Appendix I

1) Expert Advisory Group on Data Access (EAGDA)

EAGDA was established in 2012 by Wellcome, Cancer Research UK, the Economic and Social Science Research Council (ESRC) and the Medical Research Council (MRC) to provide an independent mechanism to advise the funders on data governance and to support cohort studies and their Data Access Committees (DACs). EAGDA members are selected on the basis of their world-leading expertise in areas such as genetics, epidemiology, social sciences, statistics, IT, data management and security, law, and ethics.

EAGDA has published research in a number of areas, including the governance of data access[8] and cultural issues where the funders could work together, such as incentives to support data access[9]. When concerns in the research community arose from cases where public data was used to re-identify participants in research studies, EAGDA had the relevant expertise (which is often absent within funding bodies) and independence to credibly recommend how funders should respond. EAGDA's independence and expertise also helps the funders anticipate and deal with emerging trends related to data use, such as the risks of re-identification from jigsaw linkage of genomic datasets.

Through its membership, EAGDA recognises the value of different disciplinary perspectives. While breaking down disciplinary siloes and encouraging harmonisation between funders is good for data governance, EAGDA's work has also highlighted the tensions of common approaches as some communities have different needs.

EAGDA's recommendations to the funders (Wellcome, CRUK, ESRC, MRC) are publicly available. The funders often publish responses to these recommendations, which provides transparency, but they are not accountable to EAGDA. Due to EAGDA's limited powers, at times there is a lack of incentives for funders to make commitments to act on recommendations.

2) UK Biobank Ethics and Governance Council (EGC)

Public trust in UK Biobank is central to its long term success. Wellcome and the Medical Research Council (MRC), the principal funders of UK Biobank, undertook a number of public consultations, which raised the importance of the project's oversight and governance. As a result an Ethics and Governance Framework (EGF) was developed in 2003 and the EGC established in 2004.

The role of the independent EGC is to advise on the development of the EGF, which sets standards for the UK Biobank project to ensure safeguards are in place so that data and samples are only used for scientifically and ethically approved research. The EGC also monitors and reports publicly on how the UK Biobank project follows the EGF, has oversight of information security procedures and advises on the interests of research participants and the public. The Council has no formal power over UK Biobank but does publish annual reviews and can make public statements of concern about the project. Through this reporting, EGC produces transparent assessments of how UK Biobank provides access to data in line with the original terms of consent.

Council members include experts in ethics, law, biomedical science, social science, public consultation and community and consumer involvement. Members are appointed by an

---

[8] https://wellcome.ac.uk/sites/default/files/governance-of-data-access-eagda-jun15.pdf
[9] https://wellcome.ac.uk/sites/default/files/establishing-incentives-and-changing-cultures-to-support-data-access-eagda-may14.pdf

independent committee through an open process following public advertisement[10]. EGC itself is independently reviewed every five years.

3) The Global Alliance for Genomics and Health (GA4GH)

The aim of GA4GH is to bring together leading organisations in health care, life sciences, information technology and research to unlock the value of genomic data and improve health by helping to establish common approaches to sharing this data[11]. It has a unique set up and strong international focus, involving organisations from over 70 countries.

GA4GH has embedded social and ethical principles into the practical tools and technology it develops and at the heart of its structure with a regulatory and ethics working group. The working group has developed several consent procedures, best-practice codes and guidance documents for data governance and provides advisory support to GA4GH data sharing projects.

4) Managing Ethico-social, Technical and Administrative issues in Data ACcess (METADAC)

The METADAC project aims to understand and deliver best practice in the governance of access to data and biological samples generated by population studies in the UK. It brings together the functions of Access Committee for the Centre for Longitudinal Studies Cohorts, Understanding Society Data Access Committee and the English Longitudinal Study of Ageing, to oversee access to five cohort studies[12]. Wellcome, ESRC and MRC provide the funding for METADAC on a three-year cycle.

The METADAC Access Committee was established in 2015 to assess applications for data access to the studies and includes members with social science, biomedical, legal, data curation, population studies and ethics expertise as well as lay members. The Access Committee's assessment is made against six established criteria[13], including:
- that the application is submitted by qualified researchers;
- is within the ethical permissions of the study in question;
- whether there is a risk of producing information that will allow individual study members to be identified.

A Technical Review Team checks applications for risk of consent and other ethics breaches, incidental findings and other technical issues before passing on comments to the Access Committee.

Bringing together the studies allowed METADAC to harmonise data standards, access requirements and language to make the streamline the process for applicants. This improves the efficiency of data access decisions, and METADAC's combined oversight also allows new linkages of data from different studies. An added benefit of METADAC is that issues that emerge can be shared between the studies and this allows different studies to learn from each other and develop best practice. It also provides a useful mechanism to share issues with the funders.

---

[10] http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/UKBEGC-Annual-Review-2015-small.pdf
[11] http://genomicsandhealth.org/
[12] http://www.metadac.ac.uk/
[13] http://www.metadac.ac.uk/data-access-committee/application-assessment-criteria/

5) The National Data Guardian (NDG) Review 2016

The NDG for Health and Social Care Dame Fiona Caldicott's review of data security and opt-outs is the most recent governance review in a UK health data context. The review proposed an opt-out model to give people a choice in how their data is used beyond their direct care and set out a model for good governance for patient information[14]. Importantly, the review recommended that the Department of Health conduct a meaningful public consultation on the draft standards and the proposed opt-out model, highlighting how central public consultation is to good data governance.

Wellcome broadly supported the NDG's review. However, we highlighted several potential issues with the proposed model including: the lack of realistic timetable for implementation; associated technical challenges; and the considerable policy and communications work needed to make it a success[15]. In the wake of the NDG's review, the Department of Health, Wellcome and several other funders established a new initiative to improve conversations about what happens to health data, including developing toolkits and resources.

There have been two previous reviews about information sharing and the NHS by Dame Fiona Caldicott. This demonstrates the practical difficulty of changing cultures around data use and the length of time it can take to develop a consistent approach to data governance, even in health where the potential benefits are well-recognised.

---

[14] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF
[15] https://wellcome.ac.uk/sites/default/files/NDG-review-data-secrity-consent-Sep16.pdf