Submissions to our Call for Evidence

The call for evidence asked for consideration of the following:

A) **Governance for data use: priorities, needs and opportunities**
1) The main opportunities that developments in data use present for your sector
2) The main governance needs and they impact on your work. How these governance needs are currently met.
3) Changes in governance arrangements that would be most beneficial to your area of work, and other, related areas.
4) Your major concerns regarding data governance or the lack thereof. These could be impacts felt within your sector, that affect the wider public or which have potential future impacts/

B) **The data governance landscape**

The governance frameworks, processes and organisations that affect your work.

Examples of organisations and systems for the governance of data use, addressing the following issues as far as possible:

- The sector or sectors it covers (eg, health, research, infrastructure, finance)
- The key aspects of governing data that it concerns (eg, transparency, privacy, security of data, value of data, licensing, lifecycle)
- Lessons that have been, or are being, learned in terms of how use of data is governed through these institutions or processes. Ways does it work, and ways it doesn't
- Ways of reviewing how your organisation functions
- How well it is equipped with changes to how data is used, processed and perceived
- Emerging social, legal and ethical issues should be addressed in any change to the data governance landscape?

Submissions were received from the following organisations and individuals:

- Academy of Medical Sciences (AMS)
- Administrative Data Research Centre England (ADRCE)
- Alan Sturt (individual response)
- Association of Medical Research Charities (AMRC)
- Consumer Data Research Centre, UCL
- Genetic Alliance UK
- Information Commissioner's Office (ICO)
- James Denman (personal response)
- MedConfidential
- National Data Guardian
- Newcastle University
- Nuffield Council on Bioethics
- Population Data Science, Swansea
- Privacy International
- Royal Statistical Society
- TechUK
- UK Statistics Authority
- Wellcome Trust

Summary

- The UK's outstanding medical research base is underpinned by access to data, with significant opportunities presented by increasing capabilities in the collection, linkage and analysis of data.
- The medical research community has been active in exploring data governance, particularly in relation to sharing of health data. The recently published National Data Guardian's 'Review of Data Security, Consent and Opt-outs' outlined steps for establishing a clear governance framework for sharing health data in the UK.
- Secure systems with appropriate safeguards are important to allow reliable and secure data access and linkage whilst protecting individual privacy. There are many examples of good practice in establishing such systems for health data, including the Scottish Health Informatics Programme and the Clinical Practice Research Datalink.
- Ensuring high levels of patient consent will facilitate the collection of high-quality, comprehensive datasets. This is essential as incomplete datasets have the potential to compromise the robustness and validity of research outputs. Extensive engagement with the public, clinicians and wider stakeholder base – using exemplars such as the Million Women Study and UK Biobank – to build transparency and trust around data sharing is key in order to achieve high levels of consent. This engagement will ensure that there is a clear understanding of how, and why, health data might be used for research, the value of such data, and the systems in place to protect data safety. The UK patient data taskforce, initiated by Wellcome, is anticipated to play a key role in facilitating this engagement.

Introduction

The Academy of Medical Sciences promotes advances in medical science and campaigns to ensure that these are translated into healthcare benefits for society. Our elected Fellowship comprises some of the UK's foremost experts in medical science, drawn from a diverse range of research areas, from basic research, through clinical application, to commercialisation and healthcare delivery.

The UK's outstanding research base is underpinned by access to data, which is essential for a large proportion of medical and healthcare research. It is therefore important that any data governance model continues to facilitate access to data for research and support this excellent research base. The medical research community has been active in exploring data governance, particularly in relation to sharing of health data, and the Academy has itself addressed this area through its reports on 'Personal data for public good: using health information in medical research', 'A new pathway for the regulation and governance of health research' and 'Improving the health of the public by 2040', amongst other work.[1,2,3]

Therefore our response focuses on health data, and explores some of the opportunities and challenges in creating an overarching governance framework for use of these data for research.

Opportunities presented by developments in data

As outlined in the Academy's report 'Improving the health of the public by 2040', advances in digital technology are continually opening up ever-increasing volumes of quantitative and qualitative data from a range of health and non-health sources.[3] The nature of 'data' is itself changing, with data increasingly seen as text, image, video and sound, and with many new forms inevitably on the horizon. Changes to data access and ownership are also taking place, with the volume of data held by commercial organisations, for example, dwarfing that held by public bodies; a difference that is set to increase.

The ability for researchers to utilise these rapidly increasing volumes of data is a vital component of any strategy to facilitate research for societal benefit. Enabling this will require a shift in our approach, as a society, to the use and sharing of data for research.

Historically, population data have been the foundation upon which some of the great achievements in medical and health research have been built. Large, representative datasets including those held by the NHS, Government departments, non-Governmental organisations, researchers and the private sector, as well as data generated by individuals, provide extraordinary power to understand the full spectrum and complex interactions of the broad range of factors that drive population and individual health. Combined with increasing computing capability, this will offer unprecedented opportunities to:
- Understand the distribution and determinants of health and disease.
- Explore competing risks and the relative contributions of environmental, behavioural, biological and genetic factors on health and interventions to improve health.
- Develop population-level interventions and diagnostics alongside personalised care, healthcare services and prevention, and evaluate their effectiveness, potentially in real time and at relatively low cost.
- Model future scenarios for non-communicable and infectious disease outcomes.
- Develop early warning and real-time systems for emerging health risks.

Capitalising on this opportunity for medical research will require the development of integrated systems using increasing capabilities in linking various datasets. Organisations such as the Farr Institute are working to find ways to build a better infrastructure for this health data linkage. The development of electronic health records also provide a valuable opportunity to integrate primary and secondary care data with data from across many different care pathways and services in – and indeed outside of – the NHS. With the benefits of having a single national health and social care system, there is potential for the UK to lead the way with health data linkage.

Finally, to avoid vulnerability, such systems must be adaptable to the constantly changing data landscape and to the players within it. It must also operate within a careful ethical and governance framework based on public interest.

Governance requirements

Any data governance model must support and enhance the UK's outstanding research ecosystem, facilitating and strengthening data sharing and linkage with appropriate safeguards. It is essential that any model also facilitates the collection of high-quality, robust datasets, and does not unnecessarily impede access to data for research.

Secure systems for data access

At a meeting on 'Data safe havens' hosted by the Academy, the Medical Research Council and Wellcome, participants considered different models of secure environments for handling data.[4] It was noted that whilst there are many challenges to providing data security, risks can be managed in a number of ways. These include segregation of sensitive data, minimising the movement of data between different locations, effective coding and anonymisation processes of identifiable data, developing agreed criteria for maintaining data safety, and robust recording and archiving of data usage and access.

Some participants noted that access to data should only be allowed for 'approved' researchers.[4] The Confidentiality Advisory Group (CAG) at the Health Research Authority, for instance, currently provides safeguards for access to identifiable patient information by reviewing requests for use of identifiable patient information in medical research where consent cannot be sought, under Section 251 of the NHS Act.

It was also noted at the meeting that appropriate penalties for misuse of data should be put in place to incentivise best practice and accountability in handling of data and to minimise negligent, or even malicious, use. This, however, should go hand in hand with processes and culture that facilitates appropriate data stewardship. The development of a sector-wide training and accreditation programme, directed at individuals and institutions, was suggested.

Greater clarity and harmonisation of guidance and terminology

Consistency across different legislative and policy frameworks, as well as in governance and operational structures, will be essential if we are to capitalise on the full potential of data generated and collated from across different disciplines and sectors. At present, there are numerous sources of guidance on access and standards for health data, with many different bodies involved, such as the Information Commissioner's Office, the National Information Board, the Health Research Authority's CAG and NHS Digital. However, there is no single authoritative voice or source of guidance, which can cause confusion for those trying to navigate this landscape.[2] It would be helpful to consolidate and align guidance where possible, and for further clarity on how the various bodies involved work with one another.

In general there is a key challenge posed by terminology and vocabulary used in communications around data use and governance. There are several terms used for different forms of patient data, such as 'identifiable', 'pseudonymised', 'de-identified' and 'anonymised', amongst others, and it is essential that common definitions are established for these terms to ensure transparency for the system users, clinicians and patients.

Public and wider stakeholder engagement

As identified in 'Personal data for public good: using health information in medical research', health data can be regarded differently, and sometimes as particularly sensitive or private, when compared with other types of data.[1,5] Recent events have led to some erosion of public trust around sharing of health data including NHS England's care.data initiative. Therefore public, clinician and wider stakeholder engagement around use of health data in research is critical. This should help individuals to understand how, and why, health data might be used in research. Clear communication around the value of data sharing and its contribution towards the health and social care system, as well as transparency around data access, will help to build trust around sharing

health data. The recently established UK patient data taskforce initiated by Wellcome will play a valuable role in this engagement.[6]

Consent for data use
There is a risk to medical research if high numbers of patients do not share health data, as this could compromise the robustness and validity of research. A study comparing the care given to affluent and deprived women with breast cancer, for instance, demonstrates how challenges in obtaining consent can introduce bias.[2] At the start of this particular study, patient consent was not required for the review of medical records, but the requirement for consent was introduced later in the study process. Comparing the findings of the original study to a reanalysis of the second smaller dataset of patients who consented showed that the second dataset missed one of the key research findings: that more women from deprived areas, compared with those from affluent areas, presented with locally advanced or metastatic tumours. In addition to averting misleading research findings, access to comprehensive health data is also important to provide the public with equal opportunities to participate in research and to facilitate identification of eligible patients for recruitment into certain studies. Evidence shows that only a small number of patients do not wish to receive direct invitations to participate in research.[1]

There are many nuances to seeking consent for use of health data in line with requirements of relevant regulation and guidance such as the Data Protection Regulation. Seeking and obtaining consent for research can have significant cost implications and be impracticable in some cases. For instance, when seeking re-consent from participants to use data where contact details may be outdated or when seeking consent may cause inconvenience, distress or harm. The Department of Health is currently looking to implement a clear, single framework for governance of health data through a newly proposed consent model. This will be a positive step towards a more transparent, navigable system for data governance, replacing the confused and often opaque systems currently in place where patient opt-outs are interpreted differently across the country. The recent publication of the National Data Guardian's 'Review of Data Security, Consent and Opt-outs' in healthcare outlined next steps for establishing a UK governance framework for sharing health data, proposing an associated 'opt-out' model.[7] The Academy responded to the consultation on this governance model.[8]

Some examples of best practice in data governance and access

Some examples of current systems that successfully allow secure access to linked, anonymised patient data include the **Scottish Health Informatics Programme** and the **Clinical Practice Research Datalink** (CPRD). The **UK Biobank** provides a further example of a successful mechanism for sharing of health data with a long-term follow-up and consent model built into the system.

The **Million Women Study** is a national study of women's health involving more than one million UK women, where disease is monitored through self-reporting, follow-up and record linkage. This study provides a good example of where high-quality, comprehensive data has been used in large research studies, with successful patient engagement and communication. In addition, learnings can be taken from the extensive patient engagement carried out through the **100,000 Genomes Project**, which has also created a secure data governance system for storage and access of sensitive patient data including genomic profiles.

This response was prepared by Liberty Dixon (Policy Officer) and was informed by the Academy's previous policy work and consultation responses. For further information, please contact Liberty Dixon (Liberty.Dixon@acmedsci.ac.uk; +44(0)20 3141 3222).

Academy of Medical Sciences
41 Portland Place
London, W1B 1QH
+44(0)20 3141 3200

info@acmedsci.ac.uk
www.acmedsci.ac.uk

Registered Charity No. 1070618
Registered Company No. 3520281

[1] Academy of Medical Sciences (2006). *Personal data for public good: using health information in medical research. http://www.acmedsci.ac.uk/download.php?f=file&i=13206*
[2] Academy of Medical Sciences (2011). *A new pathway for the regulation and governance of health research.* http://www.acmedsci.ac.uk/download.php?f=file&i=13646
[3] Academy of Medical Sciences (2016). *Improving the health of the public by 2040.* http://www.acmedsci.ac.uk/download.php?f=file&i=37428
[4] Academy of Medical Sciences (2014). *Data in safe havens.* http://www.acmedsci.ac.uk/viewFile/53eb4d247ef80.pdf
[5] Wellcome (2013). *Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data.* https://wellcome.ac.uk/sites/default/files/wtp053205_0.pdf
[6] https://wellcome.ac.uk/news/independent-patient-data-taskforce-announced
[7] National Data Guardian for Health and Care (2016). *Review of Data Security, Consent and Opt-Outs.* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF
[8] Academy of Medical Sciences (2016). *Response to the Department of Health's consultation on the National Data Guardian for Health and Care's Review of Data Security, Consent and Opt-Outs.* http://www.acmedsci.ac.uk/viewFile/57dfa1b898cb9.pdf

Data governance: a Royal Society and British Academy project

Response to the consultation from Professor Ruth Gilbert, Deputy Director of the Administrative Data Research Centre for England (ADRCE) 7th January 2017

Preamble

This response focusses on the use of administrative data for research. Administrative data can be generated by public or private services or by surveys or vital statistics (e.g. birth or death registration) and contain information that is routinely captured about individuals, businesses or places. For example, health data can include information about the conditions affecting a person, education data can include exam results and school absences. In many cases, administrative data can be linked longitudinally between events, places, individuals, families, households or businesses. Administrative data are particularly powerful when these longitudinal linkages are further linked between sectors, for example, when longitudinal information from health records of children is linked with data from their school records to understand whether children with certain chronic conditions are disadvantaged in their education. Key advantages of administrative data include the relatively low cost, large sample size, comprehensive coverage, and their direct relationship to services and hence relevance for policy and practice. Disadvantages include variation in data quality and the fact that data measure events or characteristics captured by the service rather than events or characteristics that actually occurred or were experienced by the individual.

The main uses of administrative data are to run and evaluate the services that generate the data. Research is relatively low on the list of priorities for using the data. However, research is a high priority for generating public benefit to justify the use of administrative data. Hence, concerns about governance need to achieve a balance between enabling use of data for research for public benefit and minimising potential risks to privacy associated with such use. Laurie et al (2015) refer to the need for a social license, underpinned by public benefit, to enable the breadth of use of administrative data.

Problems affecting data governance for administrative data

Failure to take into account the public benefits foregone by not enabling use of administrative data for research.

1. There is no quantification of the public benefits foregone, the delayed discoveries, lives lost or services not improved because researchers were not given timely access to administrative data for research or service evaluation (Jones et al, 2017).
2. The opportunity costs in terms of researcher time lost and research investments not realised because of delays or failures to use administrative data for research are potentially large but unquantified. The waste of research investment is well-recognised by government and by the major research funders (e.g. ESRC, NIHR and MRC).

Prolonged negotiations

3. Examples of delays in the provision of data by data providers abound. These delays reflect limited capacity, changes in personnel and policies, lack of documentation of precedent and the basis for decisions, and institutional, media and political sensitivities. Delays are more likely with new uses of administrative data, for example:

   a. Obtaining data from NHS Digital, the key data custodian for health data in England, can take years. For example, the ADRCE has been working with NHS Digital since November 2013 to develop a new linkage between hospitalisation data for England and school achievement data provided by the Dept. for Education. Despite approval over a year ago by Dept. for Education, NHDS has not yet approved the application.

   b. Farr Institute: A request for a broad use case for hospital data (Hospital Episode Statistics) has been under negotiation with NHS Digital since July 2013.

Data cost too much

4. The Clinical Practice Research Datalink provides de-identified primary care data for research for around 6% of the UK population. The cost for the whole dataset is £150k per year of access, putting this important dataset beyond the reach of most researchers. The dataset is held by the government but much primary care data is held by commercial companies who charge similarly high costs.

Administrative data may not be available at all

5. Some publicly collected data is made available for research only within government departments and only for specific purposes, thereby limiting independent research. Examples include:

   a. Public Health England (PHE) holds a large number of important health datasets, many of which are not available to researchers (e.g.: infection surveillance data). Access even to de-identified record-level data is possible only if researchers are able to establish collaborations with key individuals within PHE and are willing to have their analyses and report writing approved by PHE.

   b. Research for the troubled families report involved novel linkages undertaken within government between datasets from DWP/HMRC, Dept. for Education, Ministry of Justice and local governments. Information is lacking on if and when these linked data will be made available for re-use by researchers.

   a. Research within the Ministry of Justice, Dept. for Education, HMRC, and the Dept. for Business Energy and Industrial Strategy has generated important linked datasets. These datasets could be used to address a variety of questions relevant to policy and to improving methods for data processing and linkage. Eventually, some of these datasets may be made available for research. Access to an independent data safe haven for research (see below) could speed up the process of making de-identified record level data available for research without infringing the original government purposes for linking the data.

Factors underlying problems with governance?

Lack of a research culture to guide innovation in ethics, public engagement, data processing and access

How data are processed and linked is of fundamental importance to the provenance of the datasets used in the analyses and can lead to serious biases in results. Yet data processing and linkage is typically conducted by government data providers (or subcontracted to commercial trusted third parties) without adequate research, evaluation or transparency. Reasons for the lack of R&D and transparency include: a lack of funding to support research into data processing within government; a lack of research analytic capacity; a lack of connection with the research community; the potential to undermine the data product (e.g. by reporting details of data quality) or data outputs (e.g.: evidence on data quality may undermine the credibility of government reports such as hospital performance league tables).

Lack of engagement between data providers and researchers

A further problem is the limited engagement within data providers between decision makers who determine data access and research. They often lack understanding of the way in which administrative data can be used for research, and may not be able to keep abreast of new developments and research in ethics, public engagement and methods for privacy protection. Those making decisions about data access may not be required to appraise the loss of public benefit of not approving access or have the expertise to do so. Data providers may also lack engagement with the public and with advances in understanding of public perceptions or in ethics.

Conflicting interests

Use of data for research may threaten the main business of the data provider or challenge the policy of the government department overseeing the service. For example, improved methods for data linkage could change rankings in league tables for hospitals that are based on mortality rates or readmission rates, which in turn could directly affect remuneration for the hospital. Research that produces different results with the same data can also challenge the credibility of the data provider or their analytics products or may directly challenge government policy. These conflicting interests between government data providers and research are discussed in the report from Australia (see below).

Focus on purpose not on risks

The current system of data applications focusses on access for a given purpose, after which data should be destroyed. This is very inefficient and wasteful of resources as data applications, validation and cleaning by researchers take months or years. It also increases risks of disclosure as multiple requests for the same linked data require re-linking of identifiers multiple times. Allowing re-use of linked, de-identified data would reduce wasted resources and privacy risks.

The solutions:

1. One solution is to establish data safe havens or authorised data processors that are independent from data providers. There may need to be a number of such safe havens that are able to provide specialist knowledge or roles. Models for such independent safe havens include SAIL (in Wales), ICES (Ontario) and PHRN (Australia; see 'useful documents' below).

   Independent data safe havens could undertake data processing, data linkage, and be the provider of data for research (through remote access or as an anonymised, record level extract, depending on the sensitivity of the data). The safe havens could undertake research to evaluate data processing and linkage methods, to ensure continual innovation and quality control of

methods as data change and new links become feasible. These safe havens could also support data archiving, re-use of extant or dynamic datasets, and could run systems that incorporate new information from researchers into the database (e.g.: by including newly developed codes/flags or algorithms).

In 2017, the Digital Economy Bill is expected to delegate powers to the UK Statistics Authority (UKSA) to accredit data safe havens for linking government data. ONS is expected to be one of these accredited safe havens. There could be substantial benefits to research if the UKSA takes the decision to accredit further safe havens that are entirely independent from government or commercial data providers. Some Farr/ADRN safe havens (e.g. SAIL) may be able to fulfil this role, but in England, there will be a need to further develop this capacity.

2. Advice from a high-level group could helpfully represent the views of researchers and funders to the UKSA in its oversight of accredited safe havens as part of the implementation of the Digital Economy Bill. This proposal is modelled on the new Research Advisory Group, chaired by Dame Sally Davies (Chief Medical Officer for England) and Sir John Savill (Head of the MRC), which was established to improve engagement with NHS Digital in response to serious shortfalls in meeting the needs of the research community.

3. A further solution is to enable re-use of datasets for research. Re-use could be managed by the independent data safe havens who would link, hold, update and archive the datasets for re-use and manage data applications. A plurality of data safe havens could serve this function. UK Biobank is an excellent example of a researcher led, well-governed and managed provider of data (and samples) for re-use, which is keeping pace with an exponential growth in data applications, half of which are from outside the UK. UK Biobank and other cohort investments (e.g. ALSPAC, CLOSER, Scottish Longitudinal Study) have extensive experience in enabling re-use of large, dynamic, linked data resources for research while integrating new findings from research users back into the data. These researcher-led data repositories have also led developments in research and practice relating to ethics, governance, security, and public engagement, as well as some technical aspects of data processing.

4. Funding is needed to enable government departments to widen use of their data for research and to innovate to improve data accuracy and wider linkages. Substantial core funding has enabled healthcare services to support research such as trials and more recently healthcare data. In contrast to the combined annual budget of around £1.8 billion for NIHR and MRC for health research, the research budgets for other government departments that generate important datasets (apart from Defence) are negligible (e.g. £14-40 million per annum).

5. Transparency should be required from all data providers and processors about how data are generated, processed, linked and analysed so that users can understand the provenance of administrative data. Researchers often lack information to quantify the amount of error in administrative data or to take errors into account in analyses. Transparency is also necessary to enable the development of new, more effective methods. Government data providers may not prioritise transparency for various reasons, including lack of resources, commercial contracts, perceptions about security threats, or lack of knowledge or lack of adequate documentation of the processes that they operate.

Useful documents

a. The Law Commission Report on Data Sharing between Public Bodies 2014 found there was often scope for discretion in decisions regarding data access but decision making was often risk averse.

b. The recent Australian Productivity Commission Report on Data Availability and Use recognises the conflicting interests between government data providers and research. The report recommends that government data providers limit their involvement in data enhancement (e.g. linkage or analytics). The report suggests that this should be done by a suite of accredited release authorities (ARAs) who have more expertise in data processing, linkage and other data enhancement and are more independent (e.g. p16, draft recommendation 7.1 and p274).

c. The US Census Bureau has been authorised to mandate linkage of data (government and commercial) to inform evidence-based policy. The research division is considering establishing a separate facility 'sandpit' to enable research into innovative linkages which can be undertaken separately from the main business of the census bureau.

d. The Institute of Clinical Evaluative Sciences (ICES) Ontario - is an example of an independent data processor, linker and provider of data for research and service evaluation. It is independent from the Ministry of Health but receives and links health data from and for the Ministry and there is strong engagement about using data for public benefit. The Ministry of Health also undertakes their own linkages and analyses. Linkage with data from other sectors is developing, for example to examine migration and health.

e. Public Health Research Network Centre for Data Linkage. This service aims to support linkage of health data from public and private data providers across Australia. This group are specifically addressing the complexities of maintaining and updating datasets for re-use for research.

Tel. 01483 566390

e-mail alan.sturt@btinternet.com

www.churingapublishing.com

Twitter @AlanSturt

Mr A.C.Sturt
Churinga
16 Horseshoe Lane West
Guildford

Surrey GU1 2SX

31 October 2016

Professor Ottoline Leyser CBE FRS
Co-Chair, Royal Society/British Academy Data Governance Working Group
The Royal Society
6-9 Carlton House Terrace
London SW1Y 5AG

Dear Professor Leyser,

I am writing to you in your capacity as Co-Chair of the Royal Society/British Academy Data Governance Working Group. I happened to pick up the existence of this group on a 'tweet' from the Royal Society, and had a look through its Terms of Reference. I do not provide 'evidence' as requested (by 4 November), but it is a topic of such fundamental importance that I felt obliged to add another dimension, which is at the heart of the matter, and which those professionally involved routinely eschew or ignore completely. This is homogeneity through time, which really ought to feature in the deliberations. I made an intervention to this effect in a public question after a recent Royal Society lecture.

The assumption of homogeneity through time is fundamental to 'machine learning' and data governance. It underpins the validity of the process, but it is little understood and instinctively rejected by the mathematicians, engineers and physicists involved in this rapidly developing area. Homogeneity through time is quite different from the use of the time-dimension in the normal scalar sense, which features in physics and engineering as the number of time-intervals. It invalidates machine-learning used to predict the response of living entities, because they (we!) are not homogeneous through time: living entities by definition are born, grow and die. They are never the same from one instant of their lives to the next, as they travel along this path. In particular, their responses to the passage of time cannot be predicted. To be more specific, this is true of human beings not only as individuals but collectively in organisations. People in groups often arrive at decisions with which they as individuals disagree and which they know to be wrong, or even unjust. On a much broader scale, the non-homogeneity through time of individuals in response to their environment produces the differentiation which is the mechanism of the evolution of species by natural selection i.e. all life on earth. (It is individuals which evolve, not species. Their differences permeate the following generations through procreation, and so the species as a whole eventually "evolves". Pace Darwin!)

In which case, what does a machine 'learn' which enables it to predict the future? This is not a frivolous question. As John Maynard Keynes, the greatest of twentieth economists and a mathematician, points out in the quotes below, the assumption that the future can be calculated is at the root of many of the difficulties which beset economies, not simply our own but wherever the calculators hold sway. So, from his General Theory:

*Economics is a science of thinking in terms of models joined to the art of choosing models which are relevant to the contemporary world. It is compelled to be this because unlike the typical natural sciences, the material to which it is applied is, in too many respects, **not homogeneous through time**.*

and

It is a great fault of symbolic pseudo-mathematical methods of formalising a system of economic analysis ... that they expressly assume independence between the factors involved ... whereas in ordinary discourse we can keep "at the back of our heads" the necessary reserves and qualification ... Too large a proportion of mathematical economics, is as imprecise as the initial assumptions they rest on, which allow the author to lose sight of the complexities and interdependencies of the real world in a maze of pretentious and unhelpful symbols.

Not only economics! Not only then, but now! Nor is it peripheral; for instance, some very experienced and scientifically informed politicians in Germany fear it may pose a danger to democracy itself, if left unchecked. It is not just a question of convenience.

The fundamental problem on which Keynes puts his finger is that mathematics which is homogeneous-through-time, is being applied to the behaviour of humans, who are most certainly non-homogeneous through time. The result must always be an uneasy fit, and may even eventually become destructive. Some recent economists claim that it is people who are the problem, when the outcome is not what their models predict, because there is 'nothing wrong with their models'. (So what are the models for?) More ominously, some manipulate the data and the models to shape people's information and behaviour to fit the models, and then make supporting policy recommendations. There are echos of this in the reference to "outdated notions of privacy and consent" in the Group's Background note, and the tendency to isolate individual users in their own information bubble by search engines for commercial reasons.

The machine in machine-learning embodies the understanding of the programmer set down in algorithms which are thus fixed at the time and are homogeneous through time, so that they are always applied to the system, however inappropriate to the particular circumstances of the time. The introduction of the word "system" is the key to the problem. Data are in effect lists, however clever the manipulations of programmers. They are essentially static; they have no common sense or creativity. They can only carry out the sequence of calculations as instructed. The system, on the other hand, is the process into which the data are being fed. A system is by definition dynamic; with the passage of time it will no longer be the same system. The individual will be older, perhaps even wiser, the team may learn better to operate/ co-

operate. There are some reports from the US of horrifying sentences being imposed by computers, which makes a mockery of justice. Life is not an algorithm.

The problem for the Working Party is that retention and governance are locked into the process needed to access and use the data. The function of databases is to enable people to make decisions. Mistakes may be particularly dangerous if the databases are interactive, because the interactions themselves open up many more possibilities of misleading outcomes. Decisions by machines are essentially for engineered products, where the rules do not change from minute to minute, so that any mistake is caused by the wrong operation of the machine. The chances of such failure can be minimised by rigorous testing. This, of course, is also true of all the inputs which are homogeneous through time, which are the province of natural sciences, and where the machine successfully lands the aircraft, controls the refrigerator etc. Even so there have been some dangerous situations where people had to struggle with machines and their algorithms to avoid disaster when the sensors feeding data into the machine failed.

All this suggests that we should be very careful about extending the use of 'machine-learning' into systems which are non-homogeneous through time. We need to control what data is stored and how it is used in systems which are non-homogeneous through time; people with experience, insight and learning should be making the decisions. Even in systems which are homogeneous through time provision must be made for human intervention if the machine is taking wrong decisions.

The argument that we must have the whole haystack to be able to find the needle is deeply flawed. Anybody who embarks on such an enterprise does not understand the mathematics of the process. Size itself, especially for interactive databases, becomes a problem rather than a solution, because unwieldy complexity resulting from the interactions causes errors which multiply in greater proportion than the size of the database. The solution is to target the needle, and make databases which are relevant to that. The gung-ho approach of some aficionados who collect data because it can be done may be causing more problems than solutions, a question of the woods and trees.

I have explored the question of homogeneity and non-homogeneity through time in my book on systems *A Degree of Freedom*, a copy of which I enclose for the use of the Working Group as thought fit. Pages 72-82 give a description including Table 6.1 The Components of Systems Classified by Homogeneity Through Time. I have also invented a third category which is semi-homogeneity though time for completeness. These are difficult concepts which I am trying to explain, and I have sharpened up the definitions in the years since I wrote the book. However, the classification in Table 6.1 still stands as a universal summary. Nothing if not ambitious!

I hope this adds something useful for the deliberations of your Working Group, and I look forward to the final report. It is certainly an important area with considerable implications for governance, both now and increasingly in future.

Yours sincerely

A.C. Sturt

Tel. 01483 566390

e-mail alan.sturt@btinternet.com

www.churingapublishing.com

Twitter @AlanSturt

Mr A.C.Sturt
Churinga
16 Horseshoe Lane West
Guildford

Surrey GU1 2SX

21 November 2016

Professor Ottoline Leyser CBE FRS
Co-Chair, Royal Society/British Academy Data Governance Working Group
The Royal Society
6-9 Carlton House Terrace
London SW1Y 5AG

Dear Professor Leyser,

I wrote you a letter on 31 October 2016 in your capacity as Co-Chair of the Royal Society/British Academy Data Governance Working Group, drawing attention to the concept of homogeneity through time, because I thought it was essential for the group to take it into account in any discussions on data retention and governance.

Now, on getting round to filing my paper copy of my letter, I noticed an omission which may need some explanation. For some reason (cutting and pasting?) I described Keynes as the greatest of twentieth economists, which does not make much sense. It should of course read the greatest of twentieth **century** economists, which he really was, though Joan Robinson, his much underestimated fellow economist at Cambridge had similar views.

The quotes are from his seminal work which changed the understanding of the economic process, in spite of the criticisms of fellow economists. They were from his book *The General Theory of Employment, Interest and Money* by John Maynard Keynes Chapter 21 Section III.

I quoted them at the beginning of my own book *The Scale and Scope of Economics (or Economics in Real Time)* Chapter Six Economies of Scope, in which I use the term "scope" to describe the parts of the economic process which are not homogeneous through time.

Keynes was furious at the way some of his fellow economists promptly seized his broad analyses which allow for non-homogeneity through time, and translated them into mathematical equations, which are by definition always homogeneous through time. Hence his strictures in the quotes above about "mathematical concoctions, as imprecise as the assumptions they rest on".

It seems that, provided with lengthy databases and a computer, the temptation to dive into equations and algorithms is irresistible, resulting in what are in effect point

estimates, when their use should be not calculations but an aid to judgment by people with the expertise to make it. This may lead to a growing divorce between economic analysis and the real world. Nor is this confined to economic databases. What about spreadsheets?

Even then the databases need to be continuously updated and re-evaluated. Perhaps the largest database of them all, the GDP, undergoes just such a process for years after the estimates are first produced, because of the complexity of gathering, and allocating data on the national scale, to say nothing of their interpretation.

I hope that this may clarify why I suggest that gathering and governance of data needs to take account of non-homogeneity through time, not to mention any mischievous data which may spread and remain embedded in databases for ever, once they become "pseudomathematically" formalised.

Yours sincerely

A.C. Sturt

AMRC and our response

1. The Association of Medical Research Charities (AMRC) represents 135 of the leading health and medical research charities funding research in the UK.[1] Our members fund research focussed on the needs of patients for better treatments, therapies and interventions designed to improve the quality of life and ultimately prevent or cure their condition. As such, a focus on the patient perspective and patient voice is central to all of our work.

2. In 2015, AMRC member charities:
   - Invested over £1.4 billion of research funding in the UK; more than other public funders of medical research in the UK including the Medical Research Council (MRC) and the National Institute for Health Research (NIHR);
   - Funded around 25% of non-commercial research in the NHS;
   - Funded the salaries of over 15,000 researchers in the UK.

3. AMRC will provide comment on the area of *Governance for data use: priorities, needs and opportunities.*

1. What are the main opportunities that developments in data use present for your sector?

The use of data, also commonly referred to as health information or medical records, is crucial in medical research. Researchers use health information to develop understanding of disease and ill-health, discover new cures and treatments for patients; and improve the care provided by the NHS and provide efficiency and cost savings. We have highlighted case studies[2] showing some of the ways our members, and the researchers they fund, are using health information to save and improve patients' lives.

With the NHS as a single provider, and with a large, socially and ethnically diverse population, the UK has the potential to become a world-leading centre for innovative digital healthcare. If successful, this could improve patient experience, increase efficiency, attract investment and create jobs.

Our members' interests cover all these domains. They are funders and/or users of disease registries; patient/volunteer registers; biobanks; cohorts; medical "apps"; and funders of studies which deploy patient data, including genomic information and tissues; as well as users of data to inform their non-research charitable activities.

The use of mobile health, with access to real time data monitoring, is already empowering patients to manage their health more actively or to live independently. Access to linked genotypic and phenotypic data is having a huge impact in reducing the diagnostic odyssey suffered by patients with rare diseases.

---

[1] For a list of our members see: http://www.amrc.org.uk/our-members/member-directory
[2] AMRC (2016) *How your health information can make a difference*, http://www.amrc.org.uk/our-work/patient-data/how-your-health-information-can-make-a-difference

2. What are the main governance needs and how do they impact on your work?

In the area of the health and social care data, there is a critical need for clearly defined and robust governance processes to give patients confidence that their confidentiality is being respected.

AMRC welcome the publication of the National Data Guardian's review[3]; the recommendations will enhance data security and data handling in NHS and social care organisations. These standards are essential for building trust in the system. If the public do not trust the system, they will be unwilling to share health information for medical research and this will seriously hinder progress on new treatments and cures of diseases such as cancer, dementia, rare conditions and many more.

Implementing these recommendations will be a significant, but important undertaking for health and care organisations. Adequate staff training is essential to support and develop understanding of data security; building confidence and consistency amongst the workforce. The Government must ensure that NHS and social care organisations have adequate support and resource to ensure these improvements take place.

Transparent and proportionate procedures need to be in place to safeguard data and control access. There must be absolute clarity about how data can be accessed, by whom and for what purposes. Clear lines of accountability should be established with ultimate oversight and leadership from a single body which has the necessary authority, competency and expertise. While risks to privacy can never be entirely eliminated, they must be effectively reduced and managed.

Government must ensure that researchers are able to continue to access health information for the purposes of research. Without access to health information, the advancement of medical research will be hampered.

3. What changes in governance arrangements would be most beneficial to your area of work, and other, related areas?

As part of the National Data Guardian for Health and Care's review Dame Fiona Caldicott has proposed a new opt-out model for data sharing in relation to personal confidential information, and a set of data security standards for every organisation handling health and social care information.

Building public trust in the system is a significant but essential undertaking, and will require a comprehensive plan for implementation, which should be communicated transparently with an appropriate timescale. We support the review's recommendation to fully test the opt-out model with the general public before it is implemented and to engage healthcare professionals throughout implementation. It is important that adequate time is given for the model to be implemented successfully and to build public confidence in the system. The experience of care.data demonstrates the need for a comprehensive implementation plan that engages all stakeholders and the public.

We are supportive of the independent taskforce on patient data hosted by the Wellcome Trust, and encourage the Department of Health to engage with the taskforce.

---

[3] National Data Guardian for Health and Care (2016) *Review of Data Security, Consent and Opt-Outs*, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF

4. What are your major concerns regarding data governance or the lack thereof? These could be impacts felt within your sector, that affect the wider public or which have potential future impacts

Failure to record, link and share data for care and research is compromising: the safety of today's NHS patients; opportunities for efficiency gains in delivery of NHS services; and research which could transform our ability to predict, diagnose and treat disease.

During the Caldicott review's evidence gathering stage Dame Fiona said she was struck by how little awareness there is generally about the ways in which health and care information is used and safeguarded. A recent IPSOS Mori survey similarly found that detailed awareness of how the NHS uses health data is low[4].

AMRC and our members are extremely concerned that an opt-out system will be introduced without the public really understanding the value of data sharing and the consequences of opting-out. Making the case to the public for sharing personal health information (choosing not to opt-out) is critical for the future of medical research.

Above all, patients, public and healthcare professionals must understand and trust the system. Building that trust is fundamental. We believe information from patient records has huge potential to save and improve lives but privacy concerns must be taken seriously.

For further information please contact:
Katherine Mayes, Policy Officer
(k.mayes@amrc.org.uk; 020 7685 2633)

---

[4] Ipsos MORI (2016) *The One-Way Mirror: Public attitudes to commercial access to health data,* https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

**The British Academy and Royal Society Call for Evidence
Data Governance**

Response from Consumer Data Research Centre, UCL
Sarah Sheppard & Paul Longley

1.0.   Introduction

The Consumer Data Research Centre[1] (CDRC) launched in October 2015, is one of three centres funded by the Economic and Social Research Council, as part of the Big Data Network Phase 2, to *'make data, routinely collected by business and local government organisations, accessible for academics in order to undertake research in the social sciences of mutual benefit to data owners and in ways that safeguard individuals' identities'*[2].

The CDRC brings together world-class researchers at UCL and the Universities of Leeds, Liverpool and Oxford to harness the potential of consumer-related data and seeks to open up consumer data resources to the UK's social science research community through a three-tier data service (open, safeguarded and secure). The centre focuses on ways in which value can be extracted from data to benefit social science researchers, businesses, government and society at large.

2.0.   Data Service

The CDRC have created a three-tier data service giving access to open, safeguarded and controlled data through a data portal[3], and to a mapping visualisation interface[4]. Users of the Open Tier are able to download datasets without restriction, following a simple registration process, while users of the Safeguarded and Secure Tiers are required to apply for access. If the application is successful, safeguarded data is made available to the user by secure download and controlled data at one of the Centre's 3 secure data labs at UCL, University of Liverpool and University of Leeds. The secure lab at UCL is a Police Assured Secure Facility and the labs at Liverpool and Leeds are working towards ISO27001 accreditation.

3.0.   Data Acquisition

The data held in the Centre's Safeguarded and Secure Tiers have been acquired from a range of consumer organisations including retailers, utility providers, transport providers and value added resellers. The Centre promotes itself as a trusted partner that offers secure data services, resources and expertise to our data partners. As the data that we are trying to acquire may include personal and commercially sensitive information, it is essential that we demonstrate that we have robust procedures and secure facilities in place to guarantee the safeguarding of the data that have been entrusted to us. Through a 'ladder of engagement' model, relationships are built and trust developed with prospective partners. Initial collaboration may involve working together on small research projects, or involvement in the Masters Research Dissertation Programme[5]. This may then lead onto larger scale research projects and/or PhD research projects with the top of ladder being reached with the successful acquisition of data to be made available through the service, under the terms of a Data Licence Agreement (DLA).

---

[1] https://www.cdrc.ac.uk/
[2] http://www.esrc.ac.uk/research/our-research/big-data-network/big-data-network-phase-2/
[3] https://data.cdrc.ac.uk/
[4] https://maps.cdrc.ac.uk
[5] https://www.cdrc.ac.uk/retail-masters/

4.0. User Journey[6]

Users wishing to access this data must follow the CDRC Research Approvals Guidelines[7]. The applicant submits an expression of interest to the CDRC Senior Management Team (SMT), whereupon they are assigned a member of the Data Scientist team to support them in submitting a detailed project proposal, including details on their research, methodology, intended outputs and ethical considerations. This proposal is reviewed in the instance by the SMT, followed by the Data Partner(s), who have a right of veto for any application, and finally by the CDRC's independent Research Approvals Group (RAG), comprising of 2 academics from a pool of reviewers and where necessary an expert on Social Science ethics. Once complete the RAG Chair considers all the reviews before making the final decision on whether the application is approved, requires revision or is rejected. Once accepted the User is required to sign the CDRC User Agreement which includes stipulations specific to the particular data partner. The User must have successfully undertaken Safe Use of Research data Environments (SURE) training prior to being granted access to the data, as offered by the UK Data Service (UKDS), Office for National Statistics, Administrative Data Research Network or HMRC. For use of controlled data, once the User has completed their analysis their final outputs are checked by the SMT against the original proposal and then by two CDRC Data Scientists to ensure that there is no risk of disclosure of personal or commercially sensitive data. The processes the CDRC have created around granting access to safeguarded and controlled data have been established following the UKDS 5 safe principals: Safe Projects, Safe People, Safe Environment, Safe Outputs resulting in Safe Data.

5.0. Data Protection and Ethics

The controlled data that the Centre holds may include personal and/or commercially sensitive data. The Centre is conscious of its obligations under the Data Protection Act and to the Data Partner, as well as its responsibilities to ensure that the approved research is ethical. To this end the data partner must agree under the DLA that they are able to share the data with the Centre for the 'purposes of social science research, to harness the potential of consumer-related Big Data for the benefit of society'. Wherever possible the data that is held are either anonymized prior to being shared with the CDRC or aggregated to avoid Data Protection issues. All applications to access controlled data must demonstrate that the ethics of the project have been taken into consideration and that they have been reviewed by the host institutions ethics review committee or evidence provided why this is unnecessary. Where an applicant does not have access to such a committee, the ethics RAG member will review to flag whether it is necessary to have the proposal reviewed by a full ethics panel. In such cases it is intended that they will be directed to the UCL or University of Leeds Ethics Review Committee. This process has still to be tested and it may be necessary to find an alternative review panel to assist.

6.0. Archiving

It is intended, that where possible, the CDRC data products and metadata from all tiers of the service will be archived at the UK Data Archive. Data products created by the CDRC and datasets from those partners who have given permission to do so will have Digital Object Identifiers 'minted' making the data record easily identifiable and accessible now and in the future.

7.0. Conclusions

The service has now been operational for one year and DLAs have been secured with 18 partners and a number of user applications received and successfully completed the research approvals process. We have demonstrated that the governance structure and systems we have put in place both around data acquisition and the User journey are working, but we are constantly reviewing and streamlining the details of the process to improve the service we offer. We have attempted to

provide support to non-academic users to gain ethical approval for their research. However these processes have still to be tested and it is unclear when full review is required whether the support from the host universities will be forthcoming. Understanding of ethical issues around these new forms of data is still relatively unclear.

Working with and learning from the experiences from the other Big Data Network members we are creating a governance framework and processes of best practice for the acquisition, use and archiving of new forms of data, that could be adopted by other organisations. We have the flexibility to be able to adapt these processes under the existing framework to address changes in technology and the data governance landscape.

DATA GOVERNANCE: A ROYAL SOCIETY AND BRITISH ACADEMY PROJECT

Call for Evidence

Response by Genetic Alliance UK, 4 November 2016

Introduction

1. Genetic Alliance UK is the national charity working to improve the lives of patients and families affected by all types of genetic conditions. We are an alliance of over 180 patient organisations. Our aim is to ensure that high quality services, information and support are provided to all who need them. We actively support research and innovation across the field of genetic medicine.

2. Rare Disease UK (RDUK) is a multi-stakeholder campaign run by Genetic Alliance UK, working towards the delivery and implementation of the UK Strategy for Rare Diseases, signed by all four health departments in the UK and published by the Department of Health in November 2013.

3. Data use, sharing and governance are critical issues in the rare and genetic disease arena. We welcome the opportunity to respond to this review. We also work closely with the Association of Medical Research Charities, and support their response, which addresses the uses of data and the data governance landscape in medical research.

Question 1: What are the main opportunities that developments in data use present for your sector?

4. Many rare diseases are severe and life-limiting. For individuals or families affected by most rare diseases, the day-to-day challenges of managing a severe condition are made worse by the absence of an effective treatment or cure. These patients look to research as the source of new therapies to address their unmet health need. In order for progress to be made, patients recognise that the rarity of their conditions means that research relies on the effective sharing and use of their medical data, nationally and internationally (Genetic Alliance UK, 2015).

5. Collating and sharing patient data can help improve our basic understanding of the natural history of rare diseases. This type of information is essential as a basis for making research funding decisions or for enabling evidence-based commissioning. It is important in both cases to be able to estimate disease burden in order to demonstrate the impact that a new treatment or preventative therapy may have. It also provides a research resource itself, either for epidemiological studies, such as recording patient outcomes following a new treatment, or for service provision planning.

6. Patients are generally very willing to share their medical data in order to drive research and ensure that the NHS is able to plan for and provide the necessary treatments and services they

Genetic Alliance UK

contactus@geneticalliance.org.uk
www.geneticalliance.org.uk

Registered charity numbers: 1114195 and SC039299
Registered company number: 05772999

Page 1

need(Genetic Alliance UK, 2015; Genetic Alliance UK, 2016). It is therefore essential that there are clear, functional systems in place to facilitate the sharing of data for these purposes whilst reassuring those that participate that their data will be stored and shared safely and accountably.

Question 2: What are the main governance needs and how do they impact on your work? How are these governance needs currently met?

7. Currently the NHS does not record health information to a sufficient level of detail to assess how many UK citizens are affected by almost any rare or genetic condition. Far less is it possible to tell how old patients with these conditions are and where they live. Without this basic information it is impossible for the Government health departments in the UK and their NHSs to plan effectively and build a health service that is suitable for delivering healthcare for the patients that Genetic Alliance UK represent.

8. The National Congenital Anomaly and Rare Disease Registration Service (NCARDRS) of Public Health England can deliver this information. The visibility and ability to plan alone will be a step change for our community, but the full benefits that such a store of information could deliver are difficult to predict. In short, NCARDRS has the potential to be the biggest influence on improved healthcare for patients with rare diseases in England (and hopefully the rest of the UK) in a generation.

9. In light of the value that data collection and the sharing of this information brings to the rare disease patient community, and given the willingness of the community to share medical data for research, Genetic Alliance UK welcomed the use of the statutory power defined in Section 251 of the NHS Act 2006 for data collection for NCARDRS.

10. The key issue that informs this view is that of how complete a data set NCARDRS can create. It is crucial to the value of the system that it contains as accurate a picture as possible, and as complete a data set as possible is more crucial here than for registers of more common conditions. Small gaps in a data set with a remit to track rare diseases could lead to major inaccuracies in our understanding of rare disease in the UK. For this reason we also support the exclusion of NCARDRS from the general opt-out model.

Question 3: What changes in governance arrangements would be most beneficial to your area of work, and other, related areas?

11. We welcomed the ten data security standards recently proposed in the National Data Guardian for Health and Care's review of data security, consent and opt-outs: these are appropriate and aspirational goals. However, as demonstrated once again by the Care Quality Commission (CQC) report published at the same time, the NHS is still a substantial distance from implementing this across the board in practice. While admirably brief, there is a need for substantial guidance on how the standards are to be implemented, without which it is very difficult to judge their likely effectiveness.

12. As discussed above, patients with rare conditions are both likely to gain a great deal if data sharing is done well (in better coordination and communication in the provision of their direct care and also potentially benefiting from data use in research to develop a new treatment) and also likely to be most affected if it is done poorly. The same applies if their privacy is prioritised above their health, which would be a protection that they do not want (Genetic Alliance UK, 2015; Genetic Alliance UK, 2016).

13. As the community of people living with rare conditions are likely to be disproportionately exposed to any risk of harm due to a chance in data governance arrangements, it is essential

that their voice be heard in any discussion about possible changes. We would be happy to discuss this further with the project team.

Question 4: What are your major concerns regarding data governance or the lack thereof? These could be impacts felt within your sector, that affect the wider public or which have potential future impacts

14. The majority of individuals and families affected by rare conditions will have experienced the difficulties that can result from poor communication of patient data and how this can have a direct negative impact on the quality of care they receive. Patients with rare, genetic or undiagnosed conditions often come into contact with health and social care professionals who know little or nothing about their conditions, and find themselves needing to explain complex details of their medical histories over and over again.

15. We hear many horror stories of patients being treated inappropriately because health and social care personnel lack knowledge of the specifics of their condition, such as for example a different response to a medicine than patients will a similar, more common condition. Appropriate mechanisms for sharing information safely with those who need it have the potential to immensely improve patients' experience of care, however sanctions in the absence of a workable approach may lead to unwillingness to take the risk.

16. If a patient is affected by a rare condition, they are inherently more identifiable than those with common conditions because the number of individuals affected by that condition is small. As a result of this data being at greater risk of allowing patient identification, it is often considered confidential. From our work with the rare disease patient community, we know that they usually do not want the risk of being identified from personal or otherwise confidential patient data to stand in the way of it being shared when it has the potential to improve the quality of the care they receive.

17. In addition, we are concerned that the recent NDG Review did not fully engage with the subject of anonymised data. If we do not address the complexities of anonymisation, we risk further degrading public trust in the NHS' ability to use patient data appropriately.

18. The Information Commissioner's Office (ICO) Anonymisation Code of Practice itself acknowledges that in reality is can be difficult to determine whether data has been anonymised or is still personal data. The suggestion throughout the review that anonymisation of data is a straightforward and widely appropriate solution to concerns about sharing of personal data downplays the challenges of its use. This is particularly concerning as the review does not acknowledge the increased risk of identification of patients with rare conditions.

19. This is an issue on which we feel a direct message to data processors would be valuable. This should acknowledge the trade-off between deeming an individual's data to be identifiable due to the rarity of the condition, and the facility for care and research to be delivered using this data. If a threshold is positioned to restrict sharing data, this would adversely affect people with rare diseases, and would be contrary to their wishes.

20. Additionally, although the code of practice is a useful tool, it is also not directly enforceable unless the incident also in itself constitutes a data protection breach, which limits its usefulness as a safeguard.

21. Studies have repeatedly shown that much of the general public does not understand the differences between personal confidential data, deidentified data and anonymised data, and view individual-level data as personal even if it has been anonymised (for example Genetic

Alliance UK, 2015; Genetic Alliance UK, 2016; Wellcome Trust/Ipsos MORI, 2016). Public concerns about the use of anonymised data were demonstrated in 2014, for example, in the outcry over use of hospital episode statistics by an actuarial body to refine critical illness insurance modelling. The data used was fully anonymised according to the requirements of the code of practice, but many people believed that NHS Digital should not have allowed the report's authors to access the data at all.

22. For this reason it is important not only to discuss what patients are able to opt-out of, but what sort of information sharing they cannot opt-out of. This should also include basic information on how researchers get approval to access anonymised data and how that can be used.

23. We welcomed the NDG's focus on the importance of public trust, and endorse her model of the relationship between public trust, appropriate use of data and strong data security. However this needs to be built on a foundation of a reasonable level of understanding of the issues involved. Both in our recent work on the topic and that of other organisations such as the Wellcome Trust, we have seen that the basic level of understanding among patients and the public about use is health data is very low.

24. In addition to confusions about types of data mentioned above, most people do not have a basic understanding of how their health information is currently used and in what form. There is also a widespread lack of understanding about the role that private and third sector organisations are already playing in our health and care system.

25. The NDG review and the proposed opt-out model assumes a basic level of knowledge and understanding that does not currently exist, and as a consequence, risks the kind of backlash that occurred around the care.data programme. In order to avoid a further erosion of public trust in the NHS's ability to appropriately handle patient data, it is necessary for communication and consultation on the topic to start from where patients and the public are, not where data experts would like them to be. This is likely to require a major communications exercise.

26. We have repeatedly seen that for most people, greater knowledge about the potential uses patient data coupled with greater knowledge of data sharing processes and safeguards, tend to lead to more open and trusting attitudes towards data sharing, but this takes time to develop.

27. We understand that the intention is for NHS Digital to be ready to deliver the new model of data sharing opt outs by March 2017. We are concerned that this is too short a timeframe for the necessary public conversation about health information to have taken place with the consequent risk of a repetition of the public backlash which occurred previously. While it is important for researchers to be able to benefit from access to patient data as soon as possible, it is also essential that this not be rushed if this might set back public trust and willingness to share data, potentially setting the sector back years.

References

Department of Health, UK Strategy for Rare Diseases. Published November 2013, available at: www.gov.uk/government/uploads/system/uploads/attachment_data/file/260562/UK_Strategy_for_Rare_Diseases.pdf

Genetic Alliance UK, Genome sequencing: What do patients think? Published February 2015, available at: http://www.geneticalliance.org.uk/our-work/diagnosis/taking-the-patient-and-family-perspective-to-the-centre-of-genomic-sequencing-policy-making/

Genetic Alliance UK, Genome Sequencing: what do cancer patients think? Published March 2016, available at: http://www.geneticalliance.org.uk/our-work/diagnosis/the-patient-view-on-genome-sequencing-in-cancer-care/

Wellcome Trust/Ipsos MORI, The One-Way Mirror: Public attitudes to commercial access to health data. Published March 2016, available at: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf

ICO response to Data Governance

ico.

About the ICO

The ICO's mission is to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.

The ICO is the UK's independent public authority set up to uphold information rights. We do this by promoting good practice, ruling on complaints providing information to individuals and organisations and taking appropriate action where the law is broken.

The ICO enforces and oversees the Freedom of Information Act, the Environmental Information Regulations, the Data Protection Act and the Privacy and Electronic Communication Regulations.

This consultation seems to be aimed primarily at those using data for research, planning and other purposes – rather than regulators. Therefore the ICO will restrict itself to making a short submission to your consultation pointing out what we see as the main features of an effective data governance system.

The ICO notes that the intention of this consultation seems to be to look at the governance of data across the piece – i.e. collected by any organisation and used for any purpose. This is a very wide and ambitious aim. The ICO recommends that as its work progresses, the Royal Society and British Academy project should focus on some specific issues arising from the use of data and its governance. We make some suggestions below.

It is clear that organisations' means of collecting and analysing data – personal data and non-personal data – is developing at a considerable pace. The ICO can certainly appreciate the benefits of this, for example in medical research or evidence-based policy making. This is a golden-age in terms of the availability and use of data. However, to make the most of this, we need to make sure that the public has confidence that its personal information is being used responsibly and for reasonable purposes. The ICO's interest is to make sure that when personal data is being used for a particular purpose, the individuals who are the subject of the information are aware of this and have an appropriate degree of choice and control.

We set out below some specific data governance issues that the ICO believes your project should prioritise:

Transparency: How do we find effective ways of explaining to 'ordinary' members of the public how their information will be used and shared, for example for research purposes? We suspect that many individuals are not unduly concerned about how their personal data is used, provided there are no adverse consequences. However, we believe that there is a general tendency for people to be more interested in the collection and use of their personal information, and to challenge this when they feel that the use of their information is inappropriate or inequitable. We think, for example, of changes made to the operation of social network platforms as a result of subscriber pressure or objections to the 'monetisation' of individuals' online behaviour. The ICO believes very strongly that transparency in the use of personal information is a desirable end in itself, but that it also acts as a catalyst for change when organisations use personal information in a way that individuals find objectionable.

Please see the ICO's recently revised Privacy Notices Code of Practice here: https://ico.org.uk/for-organisations/guide-to-data-protection/privacy-notices-transparency-and-control/ In this document we have tried to present alternatives to long, legalistic privacy notices and to encourage the development of more accessible and engaging ways of explaining to people how their information will be used. We suggest that an effective transparency system is an essential component of effective governance.

Choice? The ICO observes that confusion can arise as to whether individuals have to be given a choice and have to agree to their data being used in a particular way. A good example of this is in the health service. On the one hand there are provisions the NHS Act 2006 that say – in short – that ethically approved research in the public interest can take place without patient consent – even if this would otherwise breach a duty of confidentiality. On the other hand, patients are given the right to opt-out of their personal health information being used for secondary purposes. We are not recommending a particular course of action, merely pointing out that some patients must be confused as to the degree of choice and control they actually have in relation to their personal data. In a strict data protection sense, the law generally provides alternatives to individual consent for data usage. In our view policy makers need to be much clearer as to whether they are giving people a choice, or whether they are going to go ahead without consent – or even in the face of objection – because it is in the public interest to do so. We believe that this is a confusing area for both individuals and data-holding organisations. The role of consent in data governance systems needs to be clarified.

Communication: This is related to the transparency point above. The ICO believes that data-holders and data-providers should do more to explain the form in which individuals' information is made available for research or other purposes. Again, looking at the health service, a good example of this would be the distinction between personally identifiable health information and non-personally identifiable information derived from patients' records. We believe that a better explanation of this distinction would give the public greater confidence that it is possible to use their information for medical research, for example, but in a way that protects their privacy and that will have no adverse personal consequences for them. This is a message data holders must get across more clearly to patients and other data subjects. A coherent set of terminology across the piece would help; 'de-identified', 'anonymised', 'non-personal' etc. The language around privacy and informatics can be very confusing for information professionals, let along the general public. We recommend that this study looks at the possibility of standardising the terminology surrounding personal information use, so that organisations and the public can develop a better understanding of the nature of the information being used for a particular purpose and the privacy risk this poses.

Ethics: The ICO sees data protection – and data privacy more generally – as having a clear ethical dimension; it is about the relationship between individuals and the organisations that keep records about them. This ethical dimension is transposed into data protection law primarily through the concept of fairness. We believe that individuals expect their personal data to be used in a fair and ethical way. We also believe that individuals would be more open to secondary use of their personal information if they knew more about this and had a guarantee that the information organisations hold about them will only be used in a way that is ethical and in the public interest. We believe that there needs to be a clearer articulation of 'the deal' between individuals and the organisations that hold data about them.

Regulation: We realise that the Royal Society / British Academy is at a relatively early stage of its project and that it is not yet clear what its final

recommendations will be. As we said at the outset, this is a very ambitious project in terms of looking at data governance across the piece. The project should consider carefully the consequences of its recommendations for regulators, such as the ICO. Clearly the ICO is responsible for data protection in the UK, including organisations' governance of the personal information they hold. We must make sure that the project's recommendations do lead to any confusion as to the role of the ICO and other regulators with an interest in the data governance space. A possible model might be for the establishment of a set of over-arching ethical principles relating – in particular – to the secondary use of data, personally identifiable and otherwise. (Presumably this would need some form of oversight and promotion by a regulatory body.) Then, 'beneath' those overarching principles would be more specific and detailed rules relating – inter alia – to data protection and overseen by existing regulators. Depending on what those overarching principles are, we should be able to ensure a regulatory system that is cohesive and offers protection to individuals across the piece. It should also help to reassure individuals that information about them – or derived from information about them – is only being used for reasonable purposes and in an ethical way.

We hope these comments are of use, and the ICO looks forward to further participation in the project.

Data Governance

For further information on this submission, please contact Iain Bourne, Group Manager at iain.bourne@ico.org.uk

If you would like to contact us please call 0303 123 1113.

www.ico.org.uk

Information Commissioner's Office
Wycliffe House, Water Lane
Wilmslow, Cheshire, SK9 5AF

ico.
Information Commissioner's Office

DATA GOVERNANCE

Call for Evidence (from Royal Society and British Academy)

Your 'Call for Evidence' is very opportune given:

- the 'Data Sharing' provisions embedded in Chapter 5 and Chapter 7 of the Digital Economy Bill which was recently introduced in the House of Commons:

- widespread public concerns about the potential threat posed by this Bill to the security/privacy of the vast amounts of data held about individual citizens and stored in each of the multitude of datasets held by public bodies, and which will become the raw material for the data sharing arrangements envisaged by this legislation.

Should they be enacted, the permissive data sharing powers proposed within the Digital Economy Bill have the potential to be of enormous benefit not only to the officials, analysts and other stakeholders who have been instrumental in developing them, but also to society as a whole. However, the proposed arrangements for the administration, governance and scrutiny of these permissive powers fall far short of what is required to allay public concerns, and to ensure that these powers are used responsibly and proportionately, and in the public interest.

At the very least, such far-reaching provisions should be underpinned by:

- The establishment of a robust overseer or regulator charged with the statutory responsibility for (a) scrutinising public bodies' plans for utilising these permissive powers; (b) ensuring that such plans strike the correct balance between the needs of data producers and users, and the needs of data subjects; (c) policing public bodies' implementation of such plans; and (d) censuring and imposing penalties on any persons or bodies that transgress the standards of behaviour described below.

- The development and promotion of a robust and consensually agreed body of procedures, principles, and practices to which all data sharing practitioners should be required to adhere.

- The complementary development and promotion of a similarly robust and consensually agreed 'Code of Conduct' or 'Code of Ethics' with which all data sharing practitioners should be required to comply. Such a Code could be modelled on any or all of the extant Codes listed at the foot of this submission.

The governance and scrutiny arrangements that have been proposed within the Bill, to date, fall short of the basic requirements that have been listed above. The most notable weakness is the proposal that the function of regulating the data sharing powers proposed within the Bill should reside with the UK Statistics Authority – a body which supervises the Office for National Statistics and which, therefore, has an

obvious and vested interest in expanding the amount and the degree of data sharing that can be conducted between public bodies. This is akin to suggesting that the Poacher should also be appointed as the Gatekeeper.

There is also no suggestion in the Bill that the proposed arrangements should be supported by a Code of Conduct.

Provision has been made in the Bill, however, for the development and promotion of a set of 'Principles and Procedures' and an accompanying 'Code of Practice'. It remains to be seen, however, whether, and to what extent, these proposed documents will meet the requirements described above.

<u>Links to some existing Codes of Conduct/Ethics:</u>

The International Statistical Institute's 'Declaration on Professional Ethics':
https://www.isi-web.org/index.php/news-from-isi/296-declarationprofessionalethics-2010uk

The UK Government Social Research (GSR) 'Ethical Assurance Guidance'
https://www.gov.uk/government/publications/ethical-assurance-guidance-for-social-research-in-government

The ESRC Framework for Research Ethics:
http://www.esrc.ac.uk/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2015/

The Social Research Association's 'Ethics Guidelines':
http://the-sra.org.uk/research-ethics/ethics-guidelines/

Europe's RESPECT Code of Practice covering socio-economic research
http://www.respectproject.org/code/index.php

James Denman
Statistician
Department for Communities and Local Government
Tel: 030 344 44193
Address: Fry Block, 4/SW, 2 Marsham Street, Westminster, London, SW1P 4DF

coordinator@medconfidential.org

Note to Royal Society Call for Evidence on Data Governance

We could write at very great length,[1] on data governance in the NHS, in the health space more generally, and across the public sector, and we are happy to do so if you wish to aid your report. However, since it's unclear how useful that would be, this is a brief overview of the current view of medConfidential and the reasons for that perspective, mostly as a prelude to a further conversation on areas that are of interest, and a subsequent note on the topics that you feel you need more details upon.

About medConfidential

medConfidential is an independent non-partisan organisation campaigning for confidentiality and consent in health and social care, which seeks to ensure that every flow of data into, across and out of the NHS and care system is *consensual, safe and transparent*.

Founded in January 2013, medConfidential works with patients and medics, service users and care professionals; draws advice from a network of experts in the fields of health informatics, computer security, law/ethics and privacy; and believes there need be no conflict between good research, good ethics and good medical care.

Overview

Within the NHS, we use the phrase "consensual, safe, and transparent" as the goal. That does not prohibit any sort of flow, but it describes the three tests that any use of data should meet. Those terms are context dependent (in issues of contagious diseases and, consent is the consent of parliament, the safety of public health, and the transparency of a the media response to a public health emergency).

In the public sector beyond the NHS, we use the formulation "fair, safe and transparent", for similar reasons.[2]

As to current practices, we expect the Government to publish their response to the Caldicott Review before the end of November 2016, and our submissions to that are available online.[3] We'll look to supplement this submission with details when we know what the Government

---

[1] with copious footnotes omitted from this document.

[2] See enclosure 1

[3] https://medconfidential.org/2016/mid-september-update/, plus this series of higher level and simpler pieces: https://medconfidential.org/2016/caldicott-review-1-the-good/

has accepted, rejected, ignored or outright misunderstood. We especially draw your attention to enclosure 3, which is a very high level overview of the solutions to the problem, rather than the problem itself.

Transparent

Transparency is necessary to ensure that when someone expresses a wish that their data be used, or not, they can **see** that their wish was honoured, and the benefits that came from those uses. Where a patient objected, they can see that their wishes have been honoured, and build the much needed confidence in NHS consent regime from the public.

As much as an organisation uses data in various ways, they must be seen to use it only in the ways they say they will, and seen not to in the ways they say they won't. Whether this is a data usage report[4] or similar, or a data accountability dashboard,[5] there must be a feedback loop through those whose data is used, to find the outcomes of that usage. For the research community that is the fundamental tenet of the "open access" movement in research, as it is only possible in practice due to open access,[6] and was impossible before it. This point is not medical specific.

Safe

This is the simple one in theory - "don't leave the data on the train". Safe data handling is the area that causes the most problems in practice, but those are generally entirely avoidable.

In a medical context, this can also be read as affecting the statutory overrides of consent to keep the population safe (ie in a democracy, it may be the consent of Parliament not the consent of the patient).

Fair (outside the medical world)

While the medical world has "medical ethics" as a guiding principle which is the fundamental basis for most actions, the rest of the data using environment beyond health has only "fair processing" under the Data Protection Act. All data usages must be "fair" - ie a data subject must know what you wish to do with the data, before you do it. (there are a bunch of exceptions and a few grey areas, but in principle, that's what should happen).

We could go into a great deal more detail about the public sector, but can leave that to a follow up submission.

Consensual

For the health world, the principle of consent is vital and underpins almost all activities. It is utterly perverse to suggest that a patient can refuse a life saving blood transfusion for their

---

[4] https://medconfidential.org/2014/what-is-a-data-usage-report/
[5] https://medconfidential.org/wp-content/uploads/2015/12/GDUR_web.jpg
[6] Note the prices / advocacy on the outcomes pages of the data usage report exemplar.

own reason (e.g. religious), but not have any control of how their rich and detailed medical history is used, sold, shared, or exploited.

As such, patients have a consent choice - an opt out - from the uses of their medical information for purposes beyond their direct care (uses for the purpose of direct care are covered by the consent choice for that care).

What this means and how it is implemented, is more patchy than it should be, but the principle is sound, solid, and unshakable.

Other Questions in your call for evidence

Since 2014, when care.data collapsed,[7] the NHS has made significant strides and continues to move towards a system of data flows that are consensual, safe, and transparent.

When we asked HSCIC[8] in March 2014, to whom they had sent data that month, they were unable to answer the question. Following the Partridge Review,[9] HSCIC went on a journey which means they are now capable, as an organisation, of telling each individual patient how their data was used, where their wishes were honoured (or otherwise[10]).

Major concerns are not the direction of travel, but the organisational, bodies, or people who are either so ignorant of the existence of rules that they break them,[11] or who simply think that the rules don't apply to them. This is not a rare or a new problem around data. The only thing that has changed is the scale and speed, which allows someone to accidentally affect millions (or billions) of people.[12]

A brief aside on the Digital Economy Bill (Part 5)[13]

Over the course of medConfidential, it has become clear that the debacle that struck the NHS with the care.data programe was not unique to the NHS. It is not that data handling in the NSH was worse than anywhere else in central Government, it was simply noticed first,

---

[7] The topic of a much longer submission if you wish there to be one.

[8] Now known as NHS Digital.

[9] https://www.gov.uk/government/publications/review-of-data-releases-made-by-the-nhs-information-centre

[10] A whole different topic. https://www.hsj.co.uk/topics/technology-and-innovation/data-confidentiality-row-intensifies-ahead-of-crucial-ruling/7012816.article

[11] https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/ and https://www.newscientist.com/article/2088056-did-googles-nhs-patient-data-deal-need-ethical-approval/ and https://www.theguardian.com/technology/2016/may/06/deepmind-best-privacy-infrastructure-handling-nhs-data-says-co-founder and https://techcrunch.com/2016/08/23/deepmind-nhs-health-data-sharing-deal-faces-further-scrutiny/

[12] https://twitter.com/jamesrbuk/status/798117913869877253

[13] https://medconfidential.org/2016/fertile-breeding-ground-for-fraud-and-misery/

and of interest to more people, and had an increased level of sustained focus.[14] Plus, the NHS had a desire to fix the problem in a manner that is not reflective of the majority of the rest of Government.[15] This is a fundamental problem that the current Digital Economy Bill does nothing to address, as it simple continues the past flawed processes of the past - perpetuating data copying between silos who misunderstand what the data actually means, and use it to make decisions that agree with whatever they would have wanted to do without any data at all.

The Royal Society report is an opportunity to inform such policy debate in the Lords.

There is a place for wholescale, unconsented, copying of a dataset to another "business unit", department, or organisation. However, those are rare than commonly used. Copying is a blunt tool that is easy for the bureaucracy as the pain is paid by others and there is a low coordination cost. APIs to ask questions between data silos are an alternative when private, or specific questions are needed (but require agreement on questions in advance, which is a complex coordination problem for idiosyncratic departments); open publication of statistics required by others is another solution. Similarly, consented sharing of a single record at the data subject's consent is entirely feasible where the data subject (citizen) is involved in the transaction.

Government argues that it wants to use data how the private sector does; but the private sector is generally horrified at the political power plays that go on to influence such copying, and the lack of overarching accountability. No telco would treat data how Government thinks it does.[16]

Long term

The new Information Commissioner was clear in her first speech - it is not privacy *or* innovation, it is privacy *and* innovation.[17] The problem is there is no long term sustainable solution currently encouraging organisations to do the right thing, advocated by those who wish to use data responsibly, as otherwise the secrecy will be utilised by fraudsters and charlatans looking to make a quick dollar.[18]

---

[14] For a non-NHS-but-still-health comparator, the data aspects of this report from the Health Select Committee is very clear: https://www.parliament.uk/business/committees/committees-a-z/commons-select/health-committee/inquiries/parliament-2015/public-health-post-2013-inquiry-15-16/

[15] The other comparator is likely to be DEFRA, and their approaches to data following the 2013 floods, and what they did around opening up data to those who could make use of it to produce actionable information for those who needed that information to make decisions. The leadership on that project came from the Secretary of State, which is also applicable to the progress made by HSCIC (and where that leadership has been lacking, problems have occured).

[16] The possible exception is talk talk, but only as we now have a very good understanding of their internal infrastructure, as it all leaked when they got hacked.

[17] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2016/09/transparency-trust-and-progressive-data-protection/

[18] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2015/10/online-pharmacy-fined-for-selling-customer-details/ - http://www.computing.co.uk/ctg/news/2431162/poisonous-online-pharmacy-pharmacy2u-fined-by-ico-for-illegally-selling-nhs-patient-data - https://medconfidential.org/2015/_pharmacy2u-did-not-keep-a-record-of-whose-data-they-sold/

That is fundamentally the tenet being advocated for in the RS call. It may not be AI, or genomics, it may be whatever comes after those things. To reference a talk I was invited to give to a bunch of AI companies, who basically asked me to come and talk about "data" and "privacy", but what they meant was avoiding ending up on the front page of newspapers. And that requires each reputable company being able to demonstrate what it *did* do, because one of their competitors will get caught doing what the public was afraid of, and when they do that, it's too late to try and claim otherwise. The confidence has to be there in advance, and hoping someone doesn't screw up is an a long term strategy.

When the avalanche has started, it is too late for the pebbles to vote.

Sam Smith
    medConfidential
    November 2016

Enc:
1. Fair, Safe and transparent briefing[19]
2. Data Usage Report paper (accountability within the NHS)[20]
3. Briefing for the National Data Guardian[21]
4. Mockup of accountability within Government[22]

[19] https://medconfidential.org/wp-content/uploads/2016/11/fair-safe-transparent-october.pdf
[20] https://medconfidential.org/2014/what-is-a-data-usage-report/
[21] https://medconfidential.org/wp-content/uploads/2016/07/2015-09-NDG-presentation-shortenedforweb_.pdf
[22] https://medconfidential.org/wp-content/uploads/2015/12/GDUR_web.jpg

National
Data
Guardian

NDG Response to the Royal Society and British Academy Data Governance project – call for evidence

The National Data Guardian (NDG) advises and challenges the health and care system to help ensure that citizens' confidential information is safeguarded securely and used properly.

Dame Fiona Caldicott was appointed as the first National Data Guardian for Health and Care by the Secretary of State for Health, Jeremy Hunt, in November 2014. The NDG's role is to help to ensure that the public can trust their confidential information is securely safeguarded and that it is used to support citizens' care and to achieve better outcomes from health and care services.

Dame Fiona has chaired three independent, Government-commissioned reviews to examine how the issues related to the use and sharing of confidential health and care information should be addressed and individuals' rights safeguarded. The first, which reported in 1997, established six principles for the protection and sharing of confidential information. The second, which was published in 2013, added a seventh 'Caldicott principle' to the effect that the duty to share information can be as important as the duty to protect patient confidentiality.

In July this year Dame Fiona published a third review, which has recommended stronger safeguards for keeping confidential information secure and a new opt-out for patients and service users to prevent their information being used for reasons other than their care.

The response provided in Section A (below) is based in the main on the work undertaken as part of these reviews.

A)   *Governance for data use: priorities, needs and opportunities*

Information is essential for high quality health and social care – to support the running of the health and social care system; to improve the safety and quality of care, including through research; to protect public health; and to support innovation. Data sharing is essential to identifying poor care. It is clear that more effective data sharing could have enabled some of the recent failures to provide proper care to patients to be identified and tackled earlier. It can also be beneficial to join health data up with other types of information, to provide better services to people.

The use of health and care data is governed by a complex and fairly mature array of overlapping common and statute law, professional principles, mandatory guidance and ethical considerations. Most health and care organisations will have their own governance processes to ensure they are complying with their obligations in regards to the use and sharing of health and care data. In the recent NDG review, the roles of the Senior Information Risk Owner (SIRO) and Caldicott Guardians were highlighted as particularly

important in providing leadership in the governance of information sharing. The report also recommends that the organisation should demonstrate clear ownership and responsibility for data security, just as they do for clinical and financial management and accountability.

When establishing their governance arrangements, it is important that organisations ensure their decision making is transparent to the public, information is readily available regarding the purposes for which data is being used and shared, and the public are given clear information on when they can opt out of such usage.

The need for transparency and greater engagement with the public about the way health and care data is used and the benefits of using such information has been identified as a high priority by the NDG. Evidence from recent years demonstrates that is a need to build trust for the use of health and care data; the ability of the sector to demonstrate robust governance is key to this.

We welcome the Royal Society and British Academy consultation on Data Governance

We have the following general points to make about the consultation.

First, any policy or platform in this area needs to distinguish clearly between data governance and information governance (an issue which has caused a series of problems see for instance Richter and Wilson 2013). In our experience these are rarely the same thing and collapsing the terms or using them interchangeably leads to confusing assumptions regarding the nature of what is being governed, for what purpose and with what intent. Data (plural) is rarely meaningful beyond very specific boundaries and relationships and may be analysed or processed in a number of ways. Information implies the act of informing someone of something, and implies a much more defined interpretative frame if the information is to 'make sense'. We have tried to summarise this point in the title of our ESRC funded Seminar Series in the phrase "what else needs to be shared (or not shared) when we share information?" Because these distinctions are not clearly made in the consultation document, we are concerned that.

Second, the implicit model in the consultation is an individual organisation in a range of transactions with a monolithic state and/or its agents. By its very nature this is a reductive model focusing on just one form of relationship between government and citizen - discrete transactions. Drawing on our empirical work exploring how the various parts of public services use information (or not), we would supplement and extent this view by emphasising the relational nature of information in the interactions between the citizen and the state and the ways in which informational transactions are often experienced by citizens as interdependent, tangled and linked into (often confusing) narratives (see e.g., Dawes, et al. 2009). Recognising, that the governance of data occurs in a complex environment implies that a single platform as such is unlikely to work as it begs question who governs what data which what intents and what recourse? A shared infrastructural approach is required but one which embraces an architectural approach to joining-up the joining-up of data governance at specific localised levels in a dynamic way.

Third, in our empirical research we have identified a much richer set of interactions between the Government, its agents and those who live and work in the UK than is envisaged in much of the consultation document. For instance our work on families (Cornford et. al 2013a) shows seeing the family as a unified (or unifiable) and stable category or specific data item belies the complex nature of family lives. Collecting data and information about families tends to be service specific and the nature of familial relations are represented accordingly, leading to inconsistencies in the nature of what is shared about families.

Households similarly represented the complex nature of these issues. Harmonisation of concepts and identification of social science constructs are problematic: the 'household' obviously exists for people but it does not exist for administrative systems. This is a particular problem being faced by the Beyond 2021 programme of the ONS, as some people can be members of more than one household legitimately for different reasons which would be important to different statistical users. There is a presumption that data exists so as to answer all of our questions, but this does not adequately engage with the needs of data users.

Education data has had a problem recently with the National Pupil Database (NPD) collecting country of birth data. Presently it seems to lack any effective governance at all; where other official data has more or less restrictive compliance (as opposed to governance) models NPD is seen to be headteacher/education system property with no accountability at all.

Linking between health data and education data is essential for instance in improving our understanding child development but require comparable constructs within lives and across countries. Health visitors review development at age 2 and then schools at age 4 but linking these data together has never been attempted. This needs to be done for the sake of understanding the validity of these measures which include data other than cognitive development (i.e. behavioural or socio-emotional development) which is recognised as important but not used in service organisation. Internationally the concept of 'a good level of development' is bandied about but not defined or even aimed at any particular age.

There is a further issue for research when those who own the data have permission from others to use it, rather than having received consent themselves. This is the concern that data is used as a term when what is important to users, commissioners and citizens is information. This becomes obviously more important when we talk about linking data which produces more information or indeed new information. Statisticians helpfully encapsulate the risks as being managed by 'statistical disclosure control'; the government's data science ethics framework considers the purpose a citizen might reasonably have expected their data to be used for. At the end of this all is the realisation that data has been seen as property and entitlement of services, without regard to the expectations of citizens about how information should be used.

We could make similar points about the nature of information sharing in the lives of older adults, especially those with reduced capacity for decision making (whether chronic or acute), or military veterans (who may or not want to be identified as such) (Wilson et al. 2015). These are more complex situations, in which there is an information sharing problem on the citizen sides of the

transaction as much as on the public services side. The ability for local public services to be delivered for those most in needs often relies on the ability to know of, and to know about, such individuals in this wider context (what we have called the 'view from somewhere' – Wilson, et al. 2011, Cornford, et al. 2013b). In these contexts, information is likely to distributed around the various agencies (formal and informal - public/private and voluntary) and across family and other social networks (McLoughlin, Wilson and Martin, 2013).

Data focus implies around areas implies the 'Gateway' model described in the Law Commission review as a significant issue ripe for policy simplification: "The large number of legislative gateways, spread across primary and secondary legislation, is difficult to navigate and creates complexity. There are express and implied gateways, permissive and mandatory gateways, gateways which restrict use or onward disclosure and gateways which do not." (Law Commission, 2014: 73). The consultation seems to have interpreted this statement to mean that the number of gateways should be reduced. We would argue that a more radical interpretation is required, one that questions the model of gateways as the primary way of providing controlled, accountable information sharing in public services and that addresses the mechanisms that have generated a plethora of gateways in the first place (see McLoughlin, Wilson and Martin, 2013). Similar problems for data re-use or research emerge if we consider the problems that the production of data for particular purposes leads to 'gaming' for instance (Lowe and Wilson 2016) If these more fundamental architectural issues are not addressed, we suspect that the success of attempts to reduce the number of gateways and simplify their procedures will be short lived.

Improved – simplified or clarified – data governance of the kind envisioned in the consultation document, while important, is therefore unlikely to result in improved sharing, and therefore more economical, efficient and effective public services, as long as other issues are not addressed at the same time. To make improvements the need for information sharing need to be built into public service professionals' role structure (including training and job descriptions), it needs to be reflected in reward and incentive structures, and it must be reflected in the patterns of relationships and routines – the professional culture of public services (Richardson and Asthana, 2006). Each of these elements needs to be adequately resourced. Without addressing these issues, even well designed changes to data and information governance are likely to fail to make a difference, leading to counterproductive cynicism.

Finally we would observe that previous government policy on co-production in public services has envisaged a more active citizen being more responsible and engaged in the production of services.

This potential can only be realised if the ways in data and information is conceived in public and social policy is represented as an integral part of that vision. It is possible to describe a more sophisticated approach which could lead to better service coordination, practitioner confidence, information sharing behaviour and service delivery: local service communities could work together to improve their information sharing relationships (internally and externally), in parallel with improvements in systems and resources. In the context of the government's devolution agenda for England, the issue of such local 'interpretative communities" (Cornford et al., 2013) and the local governance of information sharing (Wilson et al., 2013) are increasingly important. Central policy is only ever likely to be a part of such an approach.

Rob Wilson (Professor, Newcastle University), Thomas King (Researcher, Newcastle University), James Cornford (Senior Lecturer, University of East Anglia), Sue Richardson (Lecturer, Bradford University), Sue Baines (Professor, Manchester Metropolitan University)*

Cornford, J., Baines, S., and Wilson, R. (2013) **Representing the Family: how does the State "Think Family"?** Policy and Politics 2013, 41(1), 1-18.

Cornford, J., Wilson, R., Baines, S. and Richardson, R. (2013) **Local Governance in the new information ecology: The challenge of building interpretative communities**, Public Money and Management 2013, 33(3), 201-208.

Dawes, S.S., Cresswell, A.M. and Pardo, T.A. (2009) **From "need to know" to "need to share": Tangled problems, information boundaries, and the building of public sector knowledge networks**, Public Administration Review, 69(3): 392-402.

Law Commission (2014) **Data Sharing between Public Bodies A Scoping Report.** London [http://www.lawcom.gov.uk/wp-content/uploads/2015/03/lc351_data-sharing.pdf]

Lowe T, Wilson R**. Playing the game of Outcomes-Based Performance Management. Is gamesmanship inevitable? Evidence from theory and practice**. Social Policy and Administration 2016, (ePub ahead of Print).

McLoughlin, I., Wilson, R., and Martin, M. (2013) **Digital Government at Work: A Social Informatics Perspective.** Oxford: Oxford University Press.

Richardson, S. and Asthana, S. (2006) **Inter-agency Information Sharing in Health and Social Care Services: The Role of Professional Culture**, British Journal of Social Work 36(4): 657-69.

Richter P, Wilson R**. 'It's the tip of the iceberg': the hidden tensions between theory, policy and practice in the management of Freedom of Information in English local government bodies — evidence from a regional study.** Public Money and Management 2013, **33**(3), 177-184.

Wilson, R. and Gray, A. (2015) **Information Sharing: Easy to say harder to do well.** Leicestershire, UK: Centre of Excellence for Information Sharing, 2015.

Wilson, R., Baines, S., Hardill, I., and Ferguson, M. (2103) **Information for Local Governance. Data is the solution… What was the question again?** Public Money and Management 2013, 33(3), 163-166.

Wilson, R., Cornford, J., Baines, S, and Mawson J. (2011) **Information for Localism? Policy Sense-making for Local Governance.** Public Money & Management 2011, 31(4), 295-300.

Response to the Royal Society and British Academy consultation 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 work[1] 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 reidentify, 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.

Data Governance: A Royal Society and British Academy Project

Response from Population Data Science @ Swansea

This document forms the response on behalf of the Secure Anonymised Information Linkage (SAIL) databank and related Big Data initiatives led by Swansea University (referred to collectively here as Population Data Science @ Swansea) to the call for evidence for the British Academy and Royal Society Data Governance project. Population Data Science is an emerging field encompassing the science around the use of data about populations. We welcome the opportunity to contribute to this worthwhile project.

Details of the call for evidence

The call for evidence has two strands:

A) *Understanding priority issues relating to data governance across different sectors and types of organisations in order to better understand the different needs within and across sectors, and to enable better connection between discussions on governance within different sectors.*

B) *Collecting examples of current arrangements for governance of data use in the UK and internationally, and understanding their benefits and limitations, in order to construct an overview of the governance landscape and to develop case studies of governance models.*

With guidance questions as follows:

A) *Governance for data use: priorities, needs and opportunities*
1) What are the main opportunities that developments in data use present for your sector?
2) What are the main governance needs and how do they impact on your work? How are these governance needs currently met?
3) What changes in governance arrangements would be most beneficial to your area of work, and other, related areas?
4) What are your major concerns regarding data governance or the lack thereof? These could be impacts felt within your sector, that affect the wider public or which have potential future impacts

B) *The data governance landscape*
What are the governance frameworks, processes and organisations that affect your work? What reviews of governance are you aware of or involved in previously or currently? Please detail examples of organisations and systems for the governance of data use, addressing the following issues as far as possible:
• What sector or sectors does it cover (e.g. health, research, infrastructure, finance)?
• What are the key aspects of governing data that it concerns (e.g. transparency, privacy, security of data, value of data, licensing, lifecycle)?
• What lessons have been, or are being, learned in terms of how use of data is governed through these institutions or processes? In what ways does it work, and in what ways does it not?
• How long-standing is the organisation, process or activity, and are their ways of reviewing how it functions?
• How well is it equipped with changes to how data is used, processed and perceived?
• What emerging social, legal and ethical issues should be addressed in any change to the data governance landscape?

As B) aims to understand more about implemented formal governance frameworks and seeks views from those that own, administer and report to them, our response focuses on A) in relation to accessibility and use of data for research, development and evaluation.

Categories: • Medicine: clinical, patient records • Medical research • Research and innovation: Machine learning and data science • Research: social sciences. <u>Plus</u> data infrastructure and corresponding governance models.

Brief introduction to Big Data initiatives led by Swansea University

SAIL

SAIL receives core funding from Health & Care Research Wales and is an ISO27001 certified national Data Safe Haven holding a rich array of anonymised, linkable, individual-level datasets about the population of Wales. These include data from General Practice, hospital in-patient and out-patient, ONS births and deaths, cancer registry, screening services, emergency care, maternal and child health, and education. Data can be made available for remote access within the secure infrastructural and analytical environment, subject to technical and procedural controls within a robust, proportionate data governance model.
http://www.saildatabank.com/

Jones KH *et al*. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4139270/

Ford DV *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation
http://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-9-157

Lyons et al. The SAIL databank: linking multiple health and social care datasets
http://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-9-3

UKSeRP

The UK Secure Research Platform (UKSeRP) is an ISO27001 certified independent and customisable technology and analysis platform to allow multiple, complex datasets to be managed, analysed and shared safely, It was developed in response to the growing need for secure data hosting infrastructure. It can be configured to meet data provider and programme owner data governance requirements in accordance with legislative and regulatory requirements.

Jones KH *et al*. The UK Secure eResearch Platform for public health research: a case study.
Lancet (in press, November 2016)

CIPHER

The Centre for the Improvement of Population Health through eRecords Research (CIPHER) is one of the founding nodes of the Farr Institute for Health Informatics Research, established by the MRC and a consortium of other funders. CIPHER is underpinned by the SAIL databank and delivers high-quality, cutting-edge research to advance the health and care of patients and the public.
http://www.farrinstitute.org/

ADRC-Wales

The Administrative Data Research Centre – Wales (ADRC-Wales) is one of four UK centres funded by the ESRC. It builds on existing expertise and an acknowledged reputation for the safe and trustworthy use of data to work with anonymised administrative data to inform social, economic and health related research in the UK.
https://adrn.ac.uk/about/research-centre-wales/

Response

The following response is built around the guidance questions in A):

1) What are the main opportunities that developments in data use present for your sector?

Our sector includes medical, health and social science research, and the provision of infrastructure and data governance solutions to enable such research to take place safely. Many opportunities for research that would otherwise be difficult or not possible are afforded by developments in data use. These include being able to create rich cohorts of multiple linked datasets and augmenting clinical trial data and other study data with routinely-collected data. There are considerable challenges in using Big Data and much of our work is focused on addressing these so that the data can be used safely. This includes developing robust and proportionate governance models with privacy by design so that technical and procedural control measures can be put in place without being unduly burdensome to research. There are many opportunities in methodological research, the development of algorithms and software tools, data visualisation methods, working with emerging data types, working across organisations, capacity building of staff and students for upskilling and enhanced employment options. These opportunities serve to enable high quality data intensive research to take place, to provide a good service to the research community and assurance to regulators and funders that value is being provided for research to benefit patients and the public.

2) What are the main governance needs and how do they impact on your work? How are these governance needs currently met?

Our main governance needs are to strive for good data stewardship in striking the optimum balance between data privacy and data utility without sacrificing either. We have a duty to respect individual privacy, rights and confidentiality, but also a social responsibility to promote the safe use of data to bring benefits and avoid harm. The impact on our work is considerable in terms of the necessity to address the challenges but also rewarding in providing a data infrastructure so that research can flourish. We have developed our data governance models to abide by the relevant regulatory and governance frameworks. We don't hold personal data as defined by the Data Protection Act (1998), since our datasets are anonymised by a Trusted Third Party before they reach SAIL. However, we nonetheless treat the data with great care and require all requests to access data to be approved by an independent Information Governance Review Panel (IGRP), and to have all relevant project-level regulatory and governance approvals in place. Data are provided to researchers who have completed Safe Researcher training and who agree to abide by a Data Access Agreement. There is a suite of technical and procedural controls, some system-wide and some tailored to data use cases, depending on data provider requirements and project-level regulatory and governance approvals. No individual-level data leave SAIL unless there is explicit, informed participant consent (or other lawful mechanism in place, such as s251 approval). Instead, the products of analysis are reviewed by a data guardian to assess disclosure risk before results can be released for dissemination. Although many of the datasets can be accessed remotely, some require the researcher to be on-site. This is the case with some of the administrative datasets connected with the ADRC, if stipulated by the data owner. We are able to accommodate this requirement by receiving researchers in a dedicated safe pod within the Data Science Building. The data governance models of UKSeRP are built on the principles developed for SAIL but, since UKSeRP is not a databank *per se* but is a technology that can hold data, the IG responsibilities of the parties involved are set out in contract, and the control measures are configured and enacted in accordance with requirements.

3) What changes in governance arrangements would be most beneficial to your area of work, and other, related areas?

4) What are your major concerns regarding data governance or the lack thereof? These could be impacts felt within your sector, that affect the wider public or which have potential future impacts

We've chosen to respond to these questions together since there is a relationship between the two. Data governance arrangements rightly serve to safeguard individual privacy, human rights and the duties of care and confidence. In considering changes to arrangements, there is a need to consider the current arrangements and how they are implemented. Taking into account the main relevant UK legislation and regulatory guidelines, we have the principles of a framework that protects individuals and allows data to be used for *bone fide* purposes. However, it can be argued that there is a degree of disconnect between the elements such that it is not always clear which approvals are required or their sufficiency and relevance to gaining data access when obtained. There is a need for greater awareness of what is actually conferred by the various approvals. For example, researchers are sometimes not aware that although securing s251 approval for access to identifiable data without consent provides a legal basis, it does not guarantee data provision, as this can depend on organisational policies. It would be beneficial to have a more joined up approach to approvals for data intensive research with a recognition that risk to individual identity is not zero, but can be mitigated and made minimal through robust and proportionate control measures.

The UK has (and has had) various programmes for engaging with the public about the ethical, legal and societal implications (ELSI) of Big Data research. It is encouraging to see these grow and provide effective information to people so they can better understand, and input their views into, the uses to which anonymised versions of their data can be used. There is still much work to be done to maintain a proper and consistent message, to incorporate individual choice appropriately, and to convey the concept that the risks of using anonymised data are not zero but are low, provided it is done properly. We often illustrate the benefits of using data for research by showing how the outcomes can affect clinical practice and patient care, or public health benefits. Similarly the harms that can occur when data are misused are publicised, but the other side of that coin, namely harm due to the omission (non-use) of health data use has received less attention. We recently published an international case study exploring the effects of the non-use of health data in clinics, in research domains and connected with governance frameworks. We found that individual instances of data non-use can be associated with harm, but taken together, they can describe a trail of data non-use that may complicate and compound its impacts. There is ample indirect evidence that health data non-use is implicated in the deaths of many thousands of people and potentially £billions in financial burdens to societies. As well as presenting the benefits of data use, also including the negative effects of not using data will provide a more rounded picture that can help create a more balanced perspective in promoting informed choice.

Jones KH et al, The other side of the coin: harm due to the non-use of health-related data
http://www.sciencedirect.com/science/article/pii/S1386505616302039

Data Safe Havens housing large-scale anonymised data have gained greater prominence for the management and accessibility of Big Data, since they can be seen as trusted, secure environments for research, particularly when independent, external accreditation has been attained. However, debate on an effective definition continues.

Lea *et al,* Data Safe Havens and Trust: Toward a Common Understanding of Trusted Research Platforms for Governing Secure and Ethical Health Research http://medinform.jmir.org/2016/2/e22/

One of the challenges is that of effective and reliable processing of the personal data for anonymization purposes without individual consent. While the use of a suitable Trusted Third Party (TTP) to do this is not inconsistent with data protection law, the possibility of having a positive basis such that the processing of personal data for anonymization is put on a proper legal footing could simplify the situation and provide greater reassurance to data providers without introducing

additional risk. It would, of course, carry provisos and require stipulation of acceptable TTPs so that data transfer would be secure and processing would be robust and reliable.

In a piece of commissioned work we carried out for the Nuffield Council on Bioethics and the Wellcome Trust we reviewed instances of harm due to the use of health and biomedical data. It was interesting to observe that the most frequent cause was maladministration of information governance. Similarly, poor governance practice was strongly implicated in the non-use of health data. This raises questions on where the greatest attention is needed in reviewing data governance arrangements – new measures or a better application of existing measures.

Laurie G *et al*, A review of evidence relating to harm resulting from uses of health and biomedical data
http://nuffieldbioethics.org/wp-content/uploads/FINAL-Report-on-Harms-Arising-from-Use-of-Health-and-Biomedical-Data-30-JUNE-2014.pdf

Stevens LA *et al*, Dangers from Within? Looking Inwards at the Role of Maladministration as the Leading Cause of Health Data Breaches in the UK.
Book chapter in press (November 2016)

During its development, concerns were widely expressed that the European Commission's new General Data Protection Regulation (GDPR) would prohibit much medical, epidemiological and social science research. The GDPR was the subject of high profile campaigns and sterling work co-ordinated by the European Data in Health Alliance. The agreed text, passed at the end of 2015, was welcomed because a reasonable compromise has been reached on many of the more worrying issues for research. However some concerns remain, for example, Recital 23 suggests that pseudonymised data should be considered personal data, with pseudonymisation defined in the text (A.4(3b)) as *'the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person'*. Depending upon how this is interpreted, it could have practical implications in relation to participant consent for certain types of data usage that are currently exempt. It is understood that the UK will be implementing the GDPR (equivalency) and it is encouraging that there is scope for derogations at a national level. Work is ongoing involving the Wellcome Trust supported by many research funders, councils and other organisations. We hope that there will be scope for suitable derogations to ensure the GDPR does not introduce new burdens for research.

We very much support the work of this much-needed project and its laudable aims. There is a need to clarify the sufficiency and shortfalls of existing data governance arrangements before proposing new ones. Simplification and not multiplication will be important, in order to avoid introducing more stringent measures that compromise data utility without increasing protection.

Suggestions for change
1. Greater clarity and awareness on which approvals are required for data intensive research and their status when obtained
2. A more joined up approach to approvals for data intensive research with a recognition that risk to individual identity is not zero, but can be mitigated and minimised through robust, proportionate control measures
3. A more balanced view presented to the public of the risks and harms of using and not using anonymised data for research, as well as the benefits of using data

4. A positive legal basis for the processing of personal data for anonymization purposes without individual consent
5. A careful assessment of the causes of problems occurring in relation data governance to inform where to focus efforts for improvement
6. The development of suitable national derogations to ensure the GDPR does not introduce new burdens for research
7. The need to simplify and not multiply data governance arrangements and procedures in all efforts to improve them

For further information please contact:
Associate Professor Kerina Jones
k.h.jones@swansea.ac.uk
01792 602764

Whereas we used to come to technology, technology is now coming to us. Technology is becoming increasingly embedded into the fabric of our society at a time when our devices, networks and services are generating and collecting vast amounts of data on us. From biometrics, smart toys and banking a cascade of data is being generated on all of us. It is currently impossible to live a modern life without generating vast amounts of data on devices and infrastructures that are often designed for *data exploitation*.

Data exploitation begins with the excessive generation and collection of data. Data generation occurs when a sensor turns the physical world into a signal. A good illustration of this are microphones that are embedded in objects such as cars, under the guise of offering hands-free convenience. If the microphone is able to respond to our voice, does that mean is listening at all times? What inferences can be drawn from these data? And can we be sure that these data are not being shared or hacked? Many consumers are unaware about the fact that any microphone is able to constantly generate a signal at alone, let alone whether this is being collected or even analyzed or shared.

For the purpose of this submission, we would like to highlight two cases that illustrate the increasing prevalence of *excessive* data generation.

Excessive Generation by Devices – Barbie and the Internet of Things

In 2015, security researchers discovered that Hello Barbie, Mattel's Internet-connected doll, has a number of security vulnerabilities. The doll functions similar to digital assistants like Apple's Siri and uses Wi-Fi to transmit audio recordings to servers that process speech and prepare responses. Computer security researchers were able to eavesdrop on the communication between Hello Barbie and Mattel's cloud server, allowing attackers to potentially spy on the devices' conversation with the server.

Hello Barbie is just one example of a world that is becoming increasingly connected. This year, Mattel introduced the Barbie Hello Dreamhouse, a smart home that also connects to the internet and accepts voice commands from children. Although the company claims it has fixed the vulnerabilities of Hello Barbie, the case illustrates the substantial security risks associated with internet-connected devices that engage in *excessive data generation*.

Excessive Generation by Services – How a company tracked 16,000 Iowa caucus-goers via their phones

Dstillery is a data intelligence company based in the US that sells targeted advertising information about consumers. During the primaries leading to the 2016 election, Dstillery tracked 16,000 Iowa caucus-goers via

their phones[1]. Of all caucus-goers that had had granted location privileges to the apps or devices that served them ads, the company could identify mobile IDs. Those are anonymized advertising IDs provided by the devices themselves that can be used to pair of caucus-goers with their online footprints. Dstillery uses machine learning techniques to extrapolate behavior. As a result, Dstillery is able to draw intimate conclusions about voter characteristics – such as the political preferences of wrestling fans, new movers, CEOs and even things like DIYers and cigar aficionados.

Excessive generation

The examples above show how data is increasingly generated and collected by the things and infrastructure around us. This generation is often *excessive* in that it exceeds the amount of data that is required for the performance of a device or service.

As a result, the ubiquitous presence of sensors requires a substantial amount of blind trust in the companies that build this new infrastructure. It is vital that the data our devices generate about us is clearly and completely disclosed to the owner or user.

While excessive generation is a serious threat to the right to privacy, it also often a security threat. We have seen many researchers highlighting the technical side of an always-connected car, including the ability to take control of the car[2]. The more devices connected to the Internet the more likely it is that vulnerabilities can arise.

No data should be generated unless absolutely necessary. Even when data is generated, it should only be collected if absolutely necessary. Just because data is collected does not mean it is appropriate to use it in certain forms of decision-making or indeed stored longer than absolutely necessary.

---

[1] http://fusion.net/story/268108/dstillery-clever-tracking-trick/
[2] https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

Evidence to the Royal Society and British Academy on Data Governance

Please give details of the organisation you represent (if any), and which of the following sectors it falls into:

- Other, or cross-cutting sectors

The Royal Statistical Society (RSS) is a learned society and professional body, with more than 8,000 members in the UK and across the world. We are one of the world's leading organisations to promote the importance of statistics and data, and have done so since we were founded in 1834. As a charity, we advocate the key role of statistics and data in society, and we work to ensure that policy formulation and decision making are informed by evidence for the public good. We are pleased to respond to the Royal Society and British Academy's call for evidence on data governance[1], as we would welcome more comprehensive discussion and open public leadership on this topic.

A) Governance for data use: priorities, needs and opportunities.

A major priority for statistics at present is the integrity, and integration, of new data sources. We can highlight three separate drivers of this:

- 'Big data' is driving substantial changes in data availability. Statistics and statistical analysis is needed to make sense of the increasing variety, velocity and variability of data which are produced by the growth of digital and communications technology. As we presented in a paper to the European Commission[2], statistical methods are fundamental to and at the heart of the advances that come from big data.

- In government statistics, and in quantitative social science research, there is a substantial drive not only to integrate some new 'big data' sources, but also to make better use of the administrative data that is already collected by government. Access to administrative data could make data collection more efficient, and also has potential to improve the quality and the timeliness of the research and statistics that are produced.

- Finally, and accompanying these developments, there has been growth in open data publishing, on the part of government but also in other data driven sectors. This is when

---

[1] Royal Society & British Academy (2016) *Data governance: Call for evidence* (PDF). Available from: https://royalsociety.org/~/media/policy/projects/data-governance/data-governance-call-for-evidence-september-2016.pdf (accessed: November 2016)

[2] RSS (2014) Response to the European Commission on Mathematics and Digital Science (PDF). Available from http://www.rss.org.uk/Images/PDF/influencing-change/RSS_response_to_the_European_Commission_on_Mathematics_and_Digital_Science.pdf (accessed: November 2016)

data which is not confidential is openly published as a resource for anyone to analyse or use, as has been the case for example with open mapping data. We have highlighted open data in our *Data manifesto*[3] as an important spur of data-driven innovation and growth.

Accompanying these three drivers, we would like to emphasise three broad sector-wide requirements for the consideration of this review:

A1: We need to develop the capacity to use new data sources appropriately

The growth of potential data sources seems to outstrip the capacity to curate and use them in practice.

- In the commercial sector, there are leaders and laggards in data access and use, and in many cases data is under-used. Recent research by Nesta concluded that 'finding talent with the right domain knowledge, the right mix of skills (e.g. data scientists), experience, and business know-how to apply data in a commercial context is much harder than finding people with the right technical skills (including data manipulation and analysis) [4]'. Domain knowledge in this case includes understanding the theories that explain relationships in the data, and knowing data sources and their limitations[5].
- With regard to government statistics, the Independent Review of Economic Statistics by Sir Charles Bean (Bean Review) considered that new techniques need to be adopted by statistical agencies, such as web scraping, text mining and machine learning, and that this could form an important complement to their existing activities, for example to cross-check data, to fill in gaps, and to explore the significance of new economic phenomena before undertaking more systematic measurement[6]. The Bean Review concluded that statistical services need to be better equipped to use these sources, and should grow the capacity to clean, match and analyse very large data sets.

As adoption of big data and open data increases, there is a pressing need to understand the standards and integrity of the data for onward use. We need access to *metadata* which describes

---

[3] Royal Statistical Society (2016) *The Data Manifesto* (PDF). Available from: http://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS_Data%20Manifesto_2016_Online.pdf (accessed: November 2016)

[4] P. 6 in Bakhshi, H. Mateos-Garcia, J. & Windsor, G. (2015) *Skills of the Datavores: Talent and the data revolution* (PDF). Available from: http://www.nesta.org.uk/sites/default/files/skills_of_the_datavores.pdf (accessed: November 2016)

[5] P. 20 in Bakhshi, H. Mateos-Garcia, J. & Whitby, A. (2014) *Model workers: How leading companies are recruiting and managing their data talent* (PDF). Available from: http://www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf (accessed: November 2016)

[6] 'Strategic Recommendation D: Make the most of existing and new data sources and the technologies for dealing with them', pp. 11-12 in Bean, C. (2016) *The Independent Review of Economic Statistics* (PDF). Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf (accessed: November 2016)

key features of the data that are collected. The Bean Review found that access to the detailed *microdata* is also more necessary than before, as it is crucial to be able to 'dig below the surface' of any data source to establish its limitations.[7]

As we see the rise of algorithms being used in decision making (for example, in predictive policing, and also for job hiring and credit scoring), the Review should also consider whether there are methods that the public can use to hold algorithms to account. We are aware for example that crime data recorded by police is likely to be biased, as crimes that occur in locations that are already frequented by police are more likely to be recorded in a dataset[8]. Different communities can also be more or less willing to report crimes to police, resulting in a different source of bias. A machine learning algorithm trained on such data could guide policing in ways that are more biased about particular areas and particular ethnic groups than would otherwise be the case. Concerns about algorithmic transparency and accountability are growing in public profile – for example a White House report this year focused on the topic of Big Data and Civil Rights in the United States[9], and Germany's chancellor Angela Merkel has called for big internet platforms to make their algorithms more transparent[10]. More consideration is needed of whether and how governance systems can scrutinise algorithms' fairness and accountability.

Governance should encourage sufficient transparency about data, and data standards, with a view to improving the usability of data and ensuring it is used appropriately. Whether there are appropriate mechanisms to do this needs consideration for the future of data governance.

A2: We need to strengthen the legal basis for data sharing

Our *Data manifesto* recognises that data-driven innovation requires well-targeted and proportionate sharing of data for processing and analysis. Privacy safeguards need to be built into any sharing of personal data at the outset.

---

[7] 'Strategic Recommendation E: Become better at understanding and interrogating data', p. 12 in the Bean Review (PDF), Available from:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf (accessed: November 2016)
[8] Lum, K. & Isaac, W (2016) To predict and serve?, *Significance*, 13(5): pp. 14-19, Wiley Online Library. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/full (accessed: November 2016)
[9] Executive Office of the President (2016) *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (PDF), Washington: The White House. Available from:
https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (accessed: November 2016)
[10] Connolly, K. (2016) Angela Merkel: internet search engines are 'distorting perception', *Guardian*, 27 October 2016. Available at: https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception (accessed: November 2016)

In government statistics, the ONS has been assessing the potential for administrative data to produce the next Census[11], and the UK Statistics Authority has been seeking to improve the legal basis for the government and businesses to share such sources[12], most recently in the development of statistical requirements in the UK Government's Digital Economy Bill. Research by academic and third party researchers also benefits from access to the necessary administrative data for public studies, and efforts are also ongoing in the Digital Economy Bill to improve the legal basis for researchers to access data[13].

Legislation is an important reference point, but what is done with data is also affected by the culture of professions and organisations – the law needs to be understood if it is to have proportionate effects. The impact of legislation on governance and leadership needs to be considered.

A3: We need to maintain public participation and public trust

Big data and data sharing poses possible risks to individuals' right to privacy. There are many fields of statistics and research (official statistics, in public health and medical research, in social science, and in market research) that begin by using personal data about or from individuals, based on informed consent. Analysts might integrate a variety of sources of such data to inform conclusions; they typically do so to produce aggregate statistics which are de-identified, and the safeguarding of personal identities is taken very seriously. Nonetheless, statistics are produced by analysis of individual-level or personal data, and members of the public are important participants in this endeavour. If they object to the way in which statistical collections are developed, or think that such collections lack integrity, collections can be placed in jeopardy. An example of this can be seen in the UK government's recent 'care.data' scheme, which would have linked data from hospital records to data from GP records to be used for research purposes. Communication to the public about the scheme is widely regarded to have been inadequate; there were substantial public fears about how widely, and for what purpose, the newly linked data would be shared and used; and follow-up of requests to opt out of the collection was substantially delayed[14]. Care.data was paused and officially discontinued, as a result of public concerns.

---

[11] ONS (n.d.) 'Administrative Data Census Project' (webpage), available at: https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject (accessed: November 2016)

[12] UK Statistics Authority (2016) *Delivering better statistics for better decisions: Why we need new legislation for better access to data* (PDF), available from: https://www.statisticsauthority.gov.uk/wp-content/uploads/2016/03/Delivering-better-statistics-for-better-decisions-data-access-legislation-March-2016.pdf (accessed: November 2016)

[13] Cuthill, V. 'Dreaming about legislative change: an opportunity for data-driven research that must be seized' (webpage), *ESRC Blog*, 26 September 2016. Available at: https://blog.esrc.ac.uk/2016/09/26/dreaming-about-legislative-change-an-opportunity-for-data-driven-research-that-must-be-seized/#more-1392

[14] Flanagan, O. (2014) 'Care.data: a year of bungling and confusion', *StatsLife*, 25 November 2014, Available at: https://www.statslife.org.uk/features/1921-care-data-a-year-of-bungling-and-confusion (Accessed: November 2016)

However, when the context, safeguards, and purpose of collection are clear and well-explained, public fears diminish. Engagement with members of the public through focus groups and deliberative research has shown qualified support for data to be used for scientific, social or statistical research, where this has an intended public benefit, and where there are sufficient privacy safeguards in place[15]. The willingness of the public to support research can be seen in well-established data-linkage studies that are supported by participants. To provide examples we abridge here two case studies of success factors for ongoing studies, drawn from a report published by the Wellcome Trust[16]:

- The Avon Longitudinal Study of Parents and Children (ALSPAC) is a large birth cohort study established in 1991 based in Bristol, England. ALSPAC have followed-up the health, well-being and development of multiple generations of study family members. ALSPAC has also linked data from its participants with a wide range of administrative data sources about them, including hospital records, care records, education records, and demographic data. This has taken place on a transparent basis whereby participants are informed of how the study intends to use their administrative records, and are offered a means to object and withdraw from this process if they so wish. The study has demonstrated high levels of support from participants, and arrangements for consent are coupled with 'safe haven' governance arrangements to show that ALSPAC is a reliable custodian of data.

- The SAIL (Secure Anonymised Information Linkage) databank established in Wales in 2006 has uploaded and linked over 2 billion records from multiple health and social care providers. The databank saw an increase in participation from 40% to 70% of primary care providers in Wales from 2014-2015. This appeared to be the result of a successful drive by a dedicated GP engagement team. In response to the 'data trust' issues that were emerging from the care.data scheme in England, SAIL also reviewed its communication policies to ensure sufficient communication with the public, and transparency in the form of more internal and external audits. The author of this case study says that increased participation from GPs resulted from communication about the benefits of linked data research.

---

[15] Economic and Social Research Council (2014) 'Public dialogues on using administrative data' (webpage), Available at: www.esrc.ac.uk/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/ (accessed: November 2016)
RSS (2014) 'RSS research finds data trust deficit …' (webpage), *StatsLife*, 22 July 2014. Available at: https://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers (accessed: November 2016)
RSS (2016) 'Public supports sharing health data for research…' (webpage) *StatsLife*, 14 March 2016. Available at: https://www.statslife.org.uk/news/2721-public-supports-sharing-health-data-for-research-poll-says (accessed: November 2016)
[16] Public Health Research Data Forum (2015) *Enabling Linkage to Maximise the Value of Public Health Research Data* (PDF). Available from: https://wellcome.ac.uk/sites/default/files/enabling-data-linkage-to-maximise-value-of-public-health-research-data-phrdf-mar15.pdf (accessed: November 2016)

Governance arrangements are key to act upon the intentions of the legislative framework for both data protection and data access, and to check privacy safeguards at the outset of new data sharing developments. In particular, we believe any governance framework must ensure attention to privacy concerns is balanced with public engagement on the benefits of data use.

B) The data governance landscape

What are the governance frameworks, processes and organisations that affect your work? What reviews of governance are you aware of or involved in previously or currently? Please detail examples of organisations and systems for the governance of data use.

Data protection law, which has recently been reformed in Europe, forms a key reference point for governance. The content of European reforms has been a matter of interest to us, and alongside many other organisations we co-signed a series of statements by the Wellcome Trust, and welcomed that the reformed law maintains substantial provision for data to be shared for scientific and statistical research purposes[17]. The government Minister for Culture Media and Sport has made clear that the reformed regulations in Europe will take effect in the UK, as the General Data Protection Regulation (GDPR) is scheduled for adoption in 2018 [18], so adoption will not be affected by the outcome of the recent referendum.

Domestic law and arrangements are also important. This Review will be aware that there are specific statutory gateways for data sharing, and beyond this, non-statutory policies and codes of practice. The Law Commission completed a scoping review of data sharing between public bodies in 2014[19], which found that approaches to data sharing were in some cases overly cautious, that statutory provisions for data sharing were in need of further review, and that the work of the Information Commissioner's Office and other bodies which provide advice and guidance should be further explored.

Regarding the law for official statistics, we have welcomed the Digital Economy Bill's intention to amend the Statistics and Registration Services Act 2007, to strengthen the basis for government and business to share data for statistics. This amendment was called for by the Bean Review and has the support of the UK Statistics Authority, whose briefing addresses why they need new

---

[17] Thompson, B. (2016) Analysis: Research and the General Data Protection Regulation (PDF), Wellcome Trust. Available from: https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf (accessed: November 2016)

[18] Denham, E. (2016) How the ICO will be supporting the implementation of the GDPR, *Information Commissioner's Office blog*, 31 October 2016. Available at: https://iconewsblog.wordpress.com/2016/10/31/how-the-ico-will-be-supporting-the-implementation-of-the-gdpr/ (accessed: November 2016).

[19] Law Commission (2014) *Data Sharing between Public Bodies: A Scoping Report* (PDF), available from: http://www.lawcom.gov.uk/wp-content/uploads/2015/03/lc351_data-sharing.pdf (accessed: November 2016)

legislation for better access to data. Although we are in favour of the principle, <u>our evidence to the Bill's scrutiny committee</u> has questioned whether the clauses proposed are strong enough to have the intended effect[20]. Changes to the regulation of UK official statistics may also be needed; the <u>Bean Review</u> recommended that the UK Statistics Authority's regulatory function should be strengthened[21].

Broadly, in the area of softer law and governance arrangements, we would like to highlight the conclusions of our workshop report *The Opportunities and Ethics of Big Data*[22], which recognised a need for more proactive leadership on data ethics. A National Council exists in the United States, and we have called for a new national Council for Data Ethics to proactively consider the ethical challenges of new data science developments in the UK.

Finally, drawing on our experience of government data and statistics, we believe that the proliferation of data across government needs clearer governance, leadership and capacity across all work areas, both for analytical/statistical and service delivery purposes. It appears that the current mechanisms for data governance in government are too fractured, which risks a lack of strategic oversight for this key area of work. The way in which governance on key issues has been helped or hindered by the complexity of the UK's data landscape might be considered in this review.

RESPONSE ENDS

Submitted by RSS' Policy and Research Manager, 4 November 2016

---

[20] Royal Statistical Society (2016) *Briefing on the Digital Economy Bill 2016* (PDF), available from: http://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS-evidence-on-research-and-statistics-in-Digital-Economy-bill-Sept-2016.pdf (accessed: November 2016)
[21] 'Strategic Recommendation F: Strengthen the governance framework so as to help support the production of high-quality economic statistics', pp.13-15 in Bean, C. (2016) *Independent Review of UK Economic Statistics* (PDF), available from:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf (accessed: November 2016)
[22] Royal Statistical Society (2016) *The Opportunities and Ethics of Big Data* (PDF), available from: http://www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-opps-and-ethics-of-big-data-feb-2016.pdf (accessed: November 2016)

techUK response to the British Academy and Royal Society
Data Governance call for input

techUK welcomes the opportunity to provide written input to the British Academy and Royal Society Data Governance consultation. techUK is the industry voice of the UK tech sector, representing more than 900 companies who collectively employ over 700,000 people, about half of all tech jobs in the UK. These companies range from innovative start-ups to leading FTSE 100 companies. The majority of our members are small and medium sized businesses.

In today's current digital economy an increasing amount of data is being created, processed, managed, shared, stored and deleted by organisations all at the click of a button. Information is becoming faster, more agile, more mobile and more valuable than ever before. Data analytic technologies are enabling multiple data sets, of structured and unstructured data, to be analysed in real time. As a result, organisations, across both the public and private sectors, are gaining greater insights and understanding from data that would have been unachievable before.

The UK has a fantastic opportunity to be a world-leader in this new era of data driven growth and development. However, this will only be fully realised if citizens data privacy and security concerns are addressed and citizens continue to share or allow their information to be used in this way. Data governance approaches, frameworks and strategies, along with effective communications and engagement, have a vital role to play in building a culture of data trust and confidence that will underpin the future development of the UK's data driven economy and society. It is therefore important that questions related to whether traditional data governance approaches and strategies continue to remain appropriate are discussed and addressed.

Before responding to the questions raised in the call for input, techUK believes it is important to first clarify the definition of data governance that will be used as the basis for our response.

Data governance focuses on ensuring the ongoing management of data throughout its lifecycle including data's availability, integrity, confidentiality and security. This is achieved by businesses developing and deploying appropriate data governance strategies. This is achieved via a number of tools including industry standards, privacy impact assessments and privacy policies. The following response has therefore been developed based on this definition of data governance.

Governance for data use: priorities, needs and opportunities

1. **What are the main opportunities that developments in data use present for your sector?**

Big Data is already making a significant contribution to the UK economy but we are at just the beginning of that opportunity. The Centre for Economic and Business Research estimated that by 2017 Big Data could contribute £216 billion and generate 58,000 new jobs in the UK

and Ireland[1]. It is also predicted that the market for Big Data technologies, software and solutions will grow significantly in the next five years at a much faster rate than the overall IT market. However, the importance and value of data goes beyond the tech sector.

Data is a vital enabler of both the UK and EU economy and the delivery of public and private sector goods and services. Sectors including finance, retail, transport, agriculture, sport, energy, education, media and healthcare increasingly depend on effective use of data. Half of all global trade in services already depends on cross-border data flows.

Big Data analytics provide UK businesses with the opportunity to combine, consolidate and then analyse large datasets to find previously hidden insights and knowledge. For example by identifying patterns in customer behaviour, common customer likes and dislikes and previously unknown customer requirements, businesses can develop personalised goods and services based on consumers' needs and wants. The opportunities for private sector adoption extend right across sectors and industries from retailers such as Argos[2], John Lewis[3] and Tesco[4] to media companies such as Netflix[5]. Organisations are using personal and non-personal data to provide users with products and services that are personalised and responsive, in real time, to customer needs. It is important to recognise that not all Big Data will be personal data.

Data is also playing an important role in the digital transformation of public service delivery. In particular, intelligent big data analytics can help public service organisations to increase the operational efficiency of public service delivery, reduce expenditure and costs whilst delivering increasingly personalised services to citizens. This is particularly important at a time when further budget cuts to public sector organisations are expected whilst the demands on public services continue to rise. Speaking at a techUK event on realising the potential of Big Data in the Health and Social Care, Dr Geraint Lewis, Chief Data Officer for NHS England, outlined the opportunity of using large data sets to address the NHS' "Triple Aims" of improving the experience of care, advancing health and care, and reducing the cost of care.

2. **What are the main governance needs and how do they impact on your work? How are these governance needs currently met?**

As outlined above the role of data governance is to ensure the availability, integrity, confidentiality and security of data. These data governance needs are achieved by businesses developing and deploying appropriate data governance strategies. This is achieved via a number of tools including industry standards, privacy impact assessments and privacy policies.

As the amount of data being created is increasing at an unprecedented rate, organisations are navigating an environment where their roles, requirements and regulatory responsibilities in relation to data protection, data sharing and data security are changing. Data governance policies, procedures and frameworks already play a key role in helping organisations to bring order and standardisation particularly to how large data sets are classified, managed, protected and used.

[1] Centre for Economic and Business Research (2012). *Data Equity: unlocking the value of big data*. Retrieved from http://www.sas.com/en_us/insights/articles/big-data/big-data-big-value-huge-opportunity.html
[2] ComputerworldUK (May 2014). *Argos invests in big data to personalise customer experience*. Retrieved from http://www.computerworlduk.com/news/data/argos-invests-in-big-data-personalise-customer-experience-3514755/
[3] Splunk (accessed August 2015). *John Lewis: Operational Intelligence Supporting Online Growth for a Billion Pound Website*. Retrieved from http://www.splunk.com/view/splunk-at-john-lewis/SP-CAAAM7C
[4] Splunk (accessed August 2015) *Tesco.com Accelerates Development Through Deep Understanding of Customer Behavior*. Retrieved from http://www.splunk.com/view/splunk-at-tesco/SP-CAAAHVG
[5] Wired.com (accessed August 2015). *Big Data Lessons from Netflix*. Retrieved from http://www.wired.com/insights/2014/03/big-data-lessons-netflix/

For example applying data governance principles to big data can increase the quality and accuracy of large data sets and as a result increase the effectiveness of data analytics tools when applied to vast data sets. Data governance therefore could play a role in enabling organisations to gain valuable insights from big data as a result of using data analytics. By doing so data governance can become a powerful tool in building confidence that investment in big data and analytics providers and the tools and services being offered can deliver real business results.

Through data governance technology providers can also demonstrate how their solutions and services can ensure data is protected and secured throughout its lifecycle. Data governance therefore has a role to play in not only acting as an enabler of market development but also as a competitive advantage for technology providers that can demonstrate (through the use of industry standards) their data governance credentials.

Data is a key enabler of the UK's economic growth, long term productivity and job creation. Data-sharing and international data transfers underpin trade right across the economy, with over half of all global trade in services depending on cross-border data flows. In a post Brexit world it is vital that data continues to be able to flow within the UK and between the UK, Europe and beyond.

Now is the time to ensure the UK is in the best possible shape to enable, encourage and support data driven innovation. Having clear and consistent data governance policies, procedures and frameworks that are aligned with recognised European and international industry standards is key to ensuring companies continue to store and access data in the UK, and view the UK as their digital gateway to Europe and beyond. At this time what is needed is clarity, certainty and consistency that current data governance policies and procedures will continue, as well as assurances that the UK Government will implement the new European General Data Protection Regulation (GDPR) and gain adequacy which will be vital to ensuring data continues to flow.

3. **What changes in governance arrangements would be most beneficial to your area of work, and other, related areas?**

In answering this question it is important to highlight that the introduction of additional legislation or regulations specifically related to data governance is unlikely to keep pace with the speed of technological change and therefore should be avoided. Instead of changes to current governance arrangements what would be supported is an approach to data governance arrangements that ensures data is protected and also encourages, and supports, data driven innovation.

4. **What are your major concerns regarding data governance or the lack thereof? These could be impacts felt within your sector, that affect the wider public or which have potential future impacts**

UK technology companies working across all sectors take their data privacy and security responsibilities very seriously. They place data governance at the heart of their organisational structures and put in policies and procedures to ensure consumers feel confident and informed about how their data is used. For example data will be frequently anonymised and pseudonumised to ensure it is secure and encryption technologies are employed to keep data secure when it is on the move or at rest.

Unfortunately there will always be organisations who have no intention of following data protection rules, will conduct ethically questionable business affairs and not put in place data governance policies, procedures and frameworks. While this is a key concern it is suggested that those organisations will simply be left behind as data driven technologies

evolve and market competition see those demonstrating data governance best practice will push ahead.

Given the increasing importance of data it is right to stop and consider any concerns relating to whether existing data governance frameworks in place are appropriate to protect the availability, integrity, confidentiality and security of data throughout its lifecycle. However, techUK to date is not aware of any evidence of concerns relating to the effectiveness of current data governance frameworks in place.

It is important to remember that one of the most exciting aspects of the data driven economy is that it is an ever evolving market, with new innovations appearing all the time. For example the move from data informing decisions to data predicting decisions, and eventually automated decisions based on data. It will be important that any concerns relating to the transparency of how data driven decisions are identified and addressed as we move forward. But it is also important that steps are not taken at this stage of our data development which could stifle the development of a competitive and innovative data driven economy and society.

The UK is not at the start of its data driven journey, and many of the data governance issues being explored by this call for input are already being addressed by organisations operating today and by existing legal frameworks and technical and contractual measures. For example as explained below the starting point on any discussion around data governance should be the current UK data protection legal framework and the new General Data Protection Regulation (GDPR) that will undoubtedly require changes to current data governance frameworks.

The data governance landscape

1. **What are the governance frameworks, processes and organisations that affect your work? What reviews of governance are you aware of or involved in previously or currently?**

It is suggested that the starting point on any discussion around current data governance frameworks, policies and processes should be the current data protection legal framework that is enforced by the UK's Information Commissioners Officer (ICO). As an independent and trusted regulator the ICO plays an important role. It provides organisations sustainable guidance and advice on how to meet data protection and data governance requirements and ensures appropriate and proportionate data privacy and security measures are put in place and communicated clearly to citizens.

Having a workable data protection and privacy legal framework is key to ensuring public trust and confidence that data is being handled in an appropriate way. techUK therefore welcomed the review of the current EU Data Protection Directive (95/46) which has been in place since 1995. Following this review the new General Data Protection Regulation (GDPR) will come into effect in May 2018.

The GDPR is a wide-ranging regulation that aims to strengthen consumer protection and enhance trust and confidence in how personal data is used and managed. These new legal requirements will be the starting point for all organisations in developing data governance policies, procedures and best practices to demonstrate their accountability and compliance with the new rules. It is therefore suggested that this consultation considers carefully the changes that will be introduced by GDPR to future data governance frameworks and procedures.

However, when it comes to building trust and confidence on data, the new data protection law will provide a starting point, and baseline requirement, and many data driven

organisations will go far above and beyond their legal requirements, by the development to advanced policies and procedures, in order to build trust and confidence with partners, customers and consumers.

Organisations with good data governance procedures will actively promote data security and privacy and how they go above and beyond the baseline legal requirements. There are already examples of this practice today. For example currently there are no legal requirements to inform either the ICO or consumers in the event of a data breach (apart from telecoms and internet service providers). However good organisations will inform their customers in the event of any data breach to ensure consumers can be confident that something is being done to correct the incident. For example, in 2016, Oracle's MICROS Point-of-Sale Division, suffered a data breach from hackers.[6] Despite being under no legal obligation to, Oracle confirmed this breach and outlined the steps it was taking to address the situation[7].

Additionally good organisations will ensure their terms and conditions set out their data governance procedures in an intelligible way. The UK Information Commissioner's Office (ICO) has recently released a code of practice on communicating privacy information to individuals entitled "Privacy notices, transparency and control". The ICO makes it clear that being transparent and providing accessible information to individuals as to how their personal data will be used is a key element of the Data Protection Act 1998 and the new General Data Protection Regulation (GDPR).

2. What emerging social, legal and ethical issues should be addressed in any change to the data governance landscape?

The adoption of big data, data analytics and the development of machine learning and Artificial Intelligence driven technologies is leading to academics, philosophers, lawyers and technology experts increasingly asking questions about what it means to be human and distinction between humans and machines. For example to what extent we will be content as a society to transfer responsibility and control in certain situations from a human to a machine. These discussions are happening with the development of driverless cars which is leading to legal questions on responsibility and liability as well as a much broader debate about how automated decisions will be made based on data.

techUK has supported the proposed creation of a UK Council of Data Ethics as recommended by the House of Commons Science and Technology Select Committee's Big Data Dilemma report. The Committee's report into Robotics and Artificial Intelligence has also recommended the creation of a Commission to consider the social, legal and ethical issues that our AI driven future may raise.

At a time when the creation of multiple bodies is a possibility it is vital that careful thought is given to whether a single, or multiple bodies, are needed and clarity on the remit, membership, focus, role and responsibility of any body created.

While we see a role for Government in establishing a body it is important that once created it should be independent of government. This will ensure the longevity of its activities and its ability to take a long term, future looking approach to emerging social, legal and ethical data related issues.

---

[6] http://krebsonsecurity.com/2016/08/data-breach-at-oracles-micros-point-of-sale-division/

[7] http://fortune.com/2016/08/08/oracle-data-breach-retail-hacks/

It is important that there is clarity and common understanding of the remit and focus of any body created. Being unclear could cause a lack of clarity and confusion by businesses on what is required which could result in data become less, rather than more, protected.

As highlighted above for techUK members data governance means the ongoing management of data throughout its lifecycle including data's availability, integrity, confidentiality and security. This is achieved through the development and deployment of appropriate data governance strategies based on requirements within legal frameworks (such as the Data Protection Act and in the future GDPR) that are regulated by the UK's Information Commissioners Office (ICO). As an independent and trusted regulator the ICO plays an important role in supporting organisations to meet their data protection and data governance requirements.

It is suggested that any new body created should not seek to duplicate the work of the ICO on data governance and attempt to create additional legal requirements. Instead additional resources should be given to the ICO to continue its important work. Particularly given the growing importance of data to the UK.

However, as we move forward it is vital that we have a way for organisations, of all size and sector, to ask, discuss and consider the long term legal and ethical questions about how data could be, and should be, used in the future. There are big issues and profound questions that will need to be addressed in the near future, questions that neither industry nor academia can tackle alone.

The membership of any new body created must be multidisciplinary and involve representatives from government, academia, citizen groups and direct industry engagement. Industry input will be vital to ensure a balanced long term focused discussion that reflects market developments and is future looking in considering long term implications of data related issues. Membership should be broadly based and involve people who think about these questions in different ways.

techUK stands ready to work with Government, academia and other key stakeholders to consider the best way to bring people together to discuss and address the longer term social, legal and ethical questions and data issues.

UK Statistics Authority
1 Drummond Gate
London
SW1V 1QQ

Tel: 020 7592 8613
E-mail: heather.savory@ons.gov.uk
www.statisticsauthority.gov.uk/nationalstatistician

Heather Savory | Deputy National Statistician for Data Capability

Louise Pakseresht
By Email only
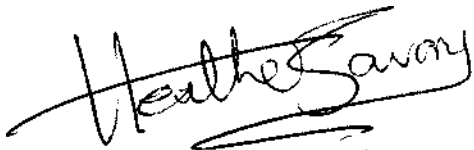
2 November 2016

Dear Ms Pakseresht

I write in response to the British Academy and the Royal Society's call for evidence on the governance arrangements that are currently in place as part of data collection, curation and use.

The UK Statistics Authority is an independent body at arm's length from government, which reports directly to Parliament. The *Statistics and Registration Service Act 2007* established the UK Statistics Authority with the statutory objective of promoting and safeguarding the production and publication of official statistics that serve the public good. Within this context, the UK Statistics Authority and its executive office, the Office for National Statistics, have established Governance arrangements to ensure data is used legally, ethically and safely.

We recognise the importance of building public trust in the use of all types of data and support all efforts to do that. We welcome all ventures which look to build trust with the public in relation to data use and are happy to work and share best practice with parties involved in this venture.

The attached notes set out some of our arrangements and comments relating to this very important area in further detail. The first describes "governance for data use: priorities, needs and opportunities" (Annex A), the second outlines the data governance landscape within the UK Statistics Authority: the National Statistician's Data Ethics Advisory Committee; the Approved Researcher Scheme; and the Administrative Data Research Network Board (Annex B).

Yours sincerely

Heather Savory

Annex A - Governance for data use: priorities, needs and opportunities

What are the main opportunities that developments in data use present for your sector?

1. The strategy for UK statistics sets out a vision for better statistics informing better decisions. This centres on producers of statistics being able to access data held by government departments and other organisations. New data developments present opportunities for UK statistics to realise its vision of 'Better Statistics Better Decisions'[1]. This means the statistical system using more pre-existing data sources which will improve the timeliness, frequency and costs associated with statistical production. It will also help manage falling survey response rates and reduce the burden on the public and businesses to respond to surveys. Making new data sources available for statistical production and research will facilitate the production of better statistics, which have an essential function at the heart of decision making across the UK.

2. The governance for statistics is set out in the Statistics and Registration Service Act 2007. This established the UK Statistics Authority as an independent body reporting to Parliament.

What are the main governance needs and how do they impact your work? How these needs are currently met?

3. Our main data governance needs relate to the access, use and sharing of data for statistical and research purposes for the public good.

4. Existing legal and ethical governance: [2]

   i. **Limited to statistics and research purposes:** The *Statistics and Registration Service Act 2007* (SRSA) limits the functions of ONS (as the executive arm of the UK Statistics Authority) to the production and publication of official statistics that serve the public good. The Authority cannot exercise any functions beyond the scope of the SRSA.

   ii. **Criminal penalties for misuse:** The SRSA provides for a strong criminal penalty on the unlawful disclosure of data. The Statistics Authority is also subject to the Data Protection Act, the law of confidence, and the Human Rights Act.

   iii. **Statutory independence:** The UK Statistics Authority has statutory independence from ministers, operates at arm's length from government, and is directly accountable to Parliament. The UK Statistics Authority Board has a majority of non-executive members, and the Chair of the Authority is appointed after a pre-appointment hearing before a parliamentary committee and a formal motion debated on the floor of the House of Commons. The Chair and senior executives are held publicly to account and routinely provide evidence to parliamentary committees.

   iv. **Transparency and standards:** ONS operates transparently and publishes guidance about what data it uses and when, and the public value that is derived from the data and information supplied to it for the purposes of producing official statistics and statistical research. ONS's Information Charter sets out how ONS carries out its responsibilities for handling personal information. ONS's Respondent Charters for business surveys, and household and individual surveys set out the standards that respondents can expect.

   v. **Strict security controls:** ONS has a strong record in protecting and safeguarding the security of data and information supplied to it, not least in its rigorous protection of personal Census information collected over the past 200 years. ONS imposes strict controls around physical security, personnel security and procedural security of the identifiable data it holds. All ONS staff must sign the ONS Confidentiality Declaration to confirm they understand strict obligations to keep information safe and secure, and the penalties for any infringement. ONS also adheres to the Government's Security Policy Framework.

---

[1] https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-betterstatisticsbetterdecisionsstrategyfor2015to202_tcm97-44175-5.pdf

[2] https://www.statisticsauthority.gov.uk/wp-content/uploads/2016/03/Delivering-better-statistics-for-better-decisions-data-access-legislation-March-2016.pdf

vi. **Statistical disclosure control:** All outputs from ONS research are subject to Statistical Disclosure Control which prevents the identification of individuals, households and businesses (and their attributes).

vii. **Codes of practice[3]:** The Code of Practice for Official Statistics has statutory underpinning in the SRSA and statisticians are under an obligation to adhere to its ethical requirements, including its principles of integrity, confidentiality, and the use of administrative sources for statistical purposes. As part of the current legislative process for the Digital Economy Bill to improve the UK Statistics Authority's access to data the Authority has prepared an initial draft of a Statement of Principles and Procedures[4] underpinning the new statutory framework for ONS to access data held across government and beyond. The Authority will early next year undertake a public consultation on a final draft ahead of finalising and laying the Statement before Parliament and the devolved legislatures.

viii. **Ethical scrutiny:** The National Statistician established the National Statistician's Data Ethics Advisory Committee in November 2014. The committee provides ethical consideration of proposals to access, share and use data. The committee has a majority of independent and lay members from outside Government, and operates transparently with all papers and minutes published. This committee provides independent scrutiny of data shares across government and with other data providers and advises the National Statistician, who reports to the UK Statistics Authority Board.

What changes in governance arrangements would be most beneficial to your area of work and other, related areas?

5. **Data Sharing:** Official statistics produced by the Office for National Statistics (ONS) play a vital role in supporting the development of economic and public policy and informing public and commercial decision-makers. The current legal arrangements governing the production of statistics are however increasingly costly, cumbersome, inhibitive of methodological innovation and lag far behind those of many of the UK's international partners. Information Sharing Orders[5] take a minimum of six months and frequently much longer to complete. The operation of these arrangements in practice is incompatible with meeting the needs of users of statistics across Government and beyond for timely and responsive data from the statistical service to inform better decision-making. Legislation covering access to data from businesses dates from the 1947 *Statistics of Trade Act*. It does not have the flexibility to support the public good need for access to statistics with the necessary efficient, effective and proportionate mechanisms.

6. New legislation is necessary to modernise the way statistical data is collected and to ensure ONS has access to the data it needs to produce fit-for-purpose official statistics that meet the challenges of a modern administration and the evolving needs of statistical users and policy makers. The Digital Economy Bill will give ONS a right of access to new sources of data. These will improve the quality of existing statistics and support the development of new statistical outputs that will give greater insight into the UK's society and economy. They will reduce existing compliance burdens on respondents including businesses, households and individuals. Without this new legislation we will struggle to leverage new sources of data to modernise our approach to delivering statistics and our strategy.

---

[3] https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-codeofpracticeforofficialstatisticsjanuary2009_tcm97-25306.pdf

[4] http://www.publications.parliament.uk/pa/cm201617/cmpublic/digitaleconomy/memo/DEB55.pdf

[5] Secondary legislation made under the Statistics and Registration Service Act that allows the Board to receive information from, or disclose information to, other public authorities where it would have otherwise been unlawful to do so.

What are your major concerns with regarding data governance or the lack thereof? These could be impacts felt within your sector, that affect the wider public or which have potential future impacts.

7. The primary concern for official statistics relates to the new powers in the Digital Economy Bill not being enacted. Delivering better statistics would be more challenging and would mean policymakers, businesses and citizens are not as well-informed as they could be in making decisions.

8. We recognise the importance of building public trust in the use of all types of data and support all efforts to do that. We are determined to build and maintain trust and confidence in our work, which as defined in statute is for the public good. We welcome all ventures which look to build trust with the public in relation to data use and are happy to work and share our best practice with parties involved in this venture.

9. We would like governance mechanisms which can help clarify what are and are not the acceptable uses of bulk personal data by both public and private enterprises.

10. We would welcome governance mechanisms which are simplified and transparent to address the uncertainty, for citizens, about the ways in which certain businesses are currently able to leverage the data they collect as part of 'Business-As–Usual' and deliver general analyses as a 'value-add' revenue stream.

Annex B - The data governance landscape

National Statistician's Data Ethics Advisory Committee[6]

What Sector or sectors does it cover?

11. The National Statistician's Data Ethics Advisory Committee (NSDEC) provides advice to all organisations and bodies which fall within the National Statistician's remit. This includes the UK Statistics Authority and its executive office, ONS, the Government Statistical Service and Devolved Administrations. At the request of the Administrative Data Research Network Board the committee also consider proposals originating from government and the third sector that require use of ADRN infrastructure for statistical or research purposes.

What are the key aspects of governing data that it concerns (e.g. transparency, privacy, security of data, value of data, licensing, lifecycle).

12. NSDEC provides advice to the National Statistician on proposals which access, use or share data for statistical or research purposes. Its overarching aim is ensure proposals have a clear public benefit and that the use and sharing of data is communicated accessibly and transparently to the public. NSDEC has a defined application process and has an agreed set of principles which the committee uses when evaluating proposals. These principles were developed by the committee through consideration of existing ethical frameworks from academia, government and health. They also instil key values from the Statistics and Registration Service Act 2007 and the Code of Practice for Official Statistics.

NSDEC's ethical principles

   i. The use of data has clear public benefits for users and serves the public good.
   ii. The data subject's identity (whether person or organisation) is protected, information is kept confidential and secure, and the issue of consent is considered appropriately.
   iii. The risks and limits of new technologies are considered and there is sufficient human oversight so that methods employed are consistent with recognised standards of integrity and quality.
   iv. Data used and methods employed are consistent with legal requirements such as the Data Protection Act, the Human Rights Act, the Statistics and Registration Service Act and the common law duty of confidence.
   v. The views of the public are considered in light of the data used and the perceived benefits of the research.
   vi. The access, use and sharing of data is transparent, and is communicated clearly and accessibly to the public.

How long-standing is the organisation, process or activity and are their ways of reviewing how it functions?

15. NSDEC was established following agreement at a meeting of the UK Statistics Authority Board in late 2014. The Committee has met quarterly since its inception. Following a year of operation the committee undertook a self assessment which included review its functions.

How well is it equipped with changes to how data is used, processed and perceived?

16. Members are kept informed of developments in the use of data for statistical and research purposes. This has been through training provided by those employing the latest methods for interrogating data (linkage, scraping etc) and through presentations from analysts, legal experts and data processors. The Committee includes members with varied backgrounds including expertise in analytics, data processing, governance and public acceptability research.

---

[6] https://www.statisticsauthority.gov.uk/national-statistician/national-statisticians-data-ethics-advisory-committee/

17. The UK Statistics Authority has also contributed to better understanding public perceptions of use of novel methods and data. Earlier this year the Authority part funded public acceptability research in to the use of data science by government. This work was led by the Government Digital Service, which has been developing ethical frameworks for government data scientists. NSDEC was kept abreast of this work and its findings were presented to members by the Government Digital Service. This join up ensures the committee is aware of current public perceptions for using novel data sources and data science methods when evaluating relevant proposals.

18. Between meetings, the secretariat provides updates to members. These updates include changes within the data ethics landscape, updates on proposals previously considered by the committee and media coverage relating to data such as data loss or hacks.

19. In the context of statistics and research, legal issues relating to data sharing should be addressed in the data governance landscape. Removing barriers to data sharing for statistical and research purposes, through the Digital Economy Bill, will provide professional statisticians access to the sources of data needed to produce the statistics that decision makers and citizens expect for them to make better more well-informed decisions.

Approved Researcher Scheme[7]
What Sector or sectors does it cover?

20. The Approved Researcher Scheme is used by ONS to grant secure access to de-identified data for statistical and research purposes, as permitted by the Statistics and Registration Service Act 2007. Access can be granted to individuals from a range of sectors (academia, government and commercial) subject to the researcher gaining accreditation. It is governed by the Microdata Release Panel which was established by the National Statistician to ensure that unpublished data are only made available by ONS for statistical research in line with legislation, the Code of Practice for Official Statistics and ONS policies.

What are the key aspects of governing data that it concerns (e.g. transparency, privacy, security of data, value of data, licensing, lifecycle).

21. Data made available under the Approved Researcher Scheme can be accessed through the ONS secure research environment known as the Virtual Microdata Laboratory (VML).

22. Approval of projects within the VML is governed by the Microdata Release Panel (MRP).

23. The Microdata Release Panel ensures that:
    i.   the confidentiality of data made available for statistical research is protected;
    ii.   a consistent decision making approach is taken when considering all requests for access to published information is kept confidential and secure, and the issue of consent is considered appropriately; and
    iii. access is only granted for research that serves the public good.

22. Approved Researchers access approved projects within the VML through safe setting locations with government across the Public Services Network (PSN).

23. Before a researcher can be approved to use the VML they are required to successfully complete training which has been jointly developed by ONS, Her Majesty's Revenue and Customs (HMRC), the UK Data Service and the Administrative Data Research Network (ADRN).

24. All research analysis and outputs are checked within the VML by specialised ONS statisticians before leaving the research environment. Whilst delivering public benefit from research using the Approved Researcher Scheme, these checks ensure that data subjects' confidentiality is maintained at all times.

---

[7] https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

How long-standing is the organisation, process or activity and are there ways of reviewing how it functions?

25. The Approved Researcher Scheme was established following enactment of the Statistics and Registration Service Act 2007. At the beginning of 2015, ONS carried out a public consultation on the criteria and process used to grant Approved Researcher status and the Safeguards used to ensure the confidentiality of all personal data held.

The Administrative Data Research Network Board[8]

26. The UK Statistics Authority provides the governance for the Administrative Data Research Network (ADRN), through the ADRN Board. The Board reports to the UK Statistics Authority Board, which in turn reports to Parliament.

27. The ADRN is a UK-wide partnership between universities, government departments and agencies, funders and the wider research community.

28. The Network provides accredited researchers from academia, government and the third sector access to linked de-identified administrative data within secure environments.

29. Researchers wishing to access the Network are required to satisfactorily complete accreditation training, have their proposal ethically reviewed and subsequently approved by the ADRN approvals panel, which assess projects based on feasibility and scientific merit.

30. Ethical review of proposals originating from academia is undertaken by the host institution's ethics committee whilst proposals originating from government and the third sector are considered by the National Statistician's Data Ethics Advisory Committee.

---

[8] https://www.statisticsauthority.gov.uk/national-statistician/administrative-data-research-network-adrn/

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

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