

MAKING OPEN DATA REAL: A PUBLIC CONSULTATION

Response by the Wellcome Trust

October 2011

Summary

- We support the Government's commitment to 'open data' and believe that there is immense potential to use more effectively the data held in the public sector to enhance services, improve choice and stimulate research and innovation.
- We urge the Government to take a broad view of 'open data' that embraces all data assets of value within the public sector, including both those that can be made freely available and those that can only be made available in a controlled manner.
- We believe that the complex regulatory landscape governing access to data needs to be streamlined and implemented in a consistent manner. There is a pressing need for the UK to engage actively in discussions on the proposed revisions to the European Data Protection Directive to ensure that research uses of data are facilitated.
- We agree that effective coordination and leadership across government will be vital to realising the opportunities offered by 'open data' and support the proposed objectives for the Public Data Corporation.
- We consider that research access to government data should be provided at zero or marginal cost.

Introduction

1. The Wellcome Trust is pleased to respond to the Government's consultation on its open data strategy. We believe that increasing access to the wealth of data held in the public sector will offer vital new opportunities to enhance public services and to enable new avenues of research and innovation with potential to generate significant economic and societal benefits.
2. As a research charity dedicated to achieving extraordinary improvements in human and animal health, our response focuses primarily on the benefits of open data for advancing research and its application for public benefit, and for enhancing healthcare services.
3. Our response makes a number of broad points in response to the questions set out in the consultation document, "Making open data real". We have also considered the related consultation on the data policy for the Public Data Corporation, and our response incorporates our views on this consultation.

The benefits of 'open data'

4. We strongly support the overall direction set out in the 'open data' consultation, in terms of embedding a 'right to data' and a strong expectation on public sector bodies to make available the valuable data they hold. The case for doing so is strongly articulated in the document – but in summary we believe that the move to 'open data' will:
 - enable the delivery of more efficient and cost-effective services to the public;
 - support joined-up government by facilitating the integration of individual datasets;
 - stimulate innovative research to enhance service delivery in the future;
 - improve choice and decision making by the individual.

5. More specifically, we would highlight three areas where enhancing the availability and use of data could offer enormous benefits for health and research:
 - **Administrative data** – increasing access to administrative data held in public sector and enabling users to link data held in different Government departments could offer exciting new opportunities to understand better and to address complex policy challenges – for example, bringing together health, social, environment, transport and other data might stimulate vital new insights and policy interventions to address social and environmental determinants of ill health and health inequalities;
 - **Health data** – the use of electronic patient records offers enormous potential to enhance patient care across the NHS, and improve planning and service delivery – particularly where records are integrated across primary and secondary care. There is also a tremendous opportunity to utilise electronic patients in medical research - leveraging the unique potential of the NHS as a national health research resource. Uses include: pharmacovigilance; patient recruitment to clinical trials; epidemiological studies; and maximising the value of major cohort studies such as UK Biobank.
 - **Research data** – there is a need to enhance access to the outputs of publicly-funded research, including research papers and data, in a way that enables research users to mine and build on this information to advance knowledge and its application. We believe fundamentally that the published outputs of research should be made available to all in open access form. Research data should be made widely available, with as few restrictions as possible, in a manner which maximises its long-term value.
6. In the healthcare arena, there are several good case studies illustrating the benefits that can be gained from an integrated electronic health system. Scotland provides a key exemplar: sharing of data on the outcomes of diabetic patients in Tayside has resulted in a 40% reduction in amputation and a 43% reduction in the number of people needing laser treatment for eye disease that threatens sight. A further example is Espoo, Finland, which has adopted an eHealth system that supports integrated and self care. This has led to improvements in patient care and savings in healthcare costs. For example, the proportion of diabetes patients who have well-controlled glucose levels has increased from 72 to 82 per cent over four years – generating cost savings due to reductions in surgery and hospital visits.
7. With regard to the possible options for charging for data, as set out in the Public Data Corporation consultation, we would argue that non-commercial research access to data should as a general principle be provided free-of-charge, or on a cost recovery basis where there is a significant administrative burden associated with individual data requests. Because these data were generated through public spending, there is a compelling need to ensure that the barriers to research uses aimed at generating additional public benefit are reduced as far as possible.

Definition and scope

8. We note that the definition of open data provided in the consultation document refers to “data which can be freely used, re-used and redistributed by anyone”. Similarly, we note that the information covered by the Public Data Corporation is defined as “the objective, factual and non-personal data and information which is collected and generated within the PDC or its constituent parts”.

9. We agree fully that datasets held by public bodies should be made freely available for re-use in the public domain wherever this is possible and adds value. However, it is vital to recognise that there are some key datasets for which free unrestricted access of this type is not appropriate, but that nonetheless hold considerable value for researchers and other legitimate users. Data contained in electronic healthcare records is one key example of a type of data which offers considerable potential benefits, but for which access needs to be controlled in a manner proportionate to the sensitivity of the data.
10. We would urge that, in progressing the 'open data' strategy, Government departments take a broader view that extends beyond the types of data covered by the definition of 'open data' provided. This should focus on seeking to maximise access to datasets of value in a safe and secure manner. Departments should consider the full range of data assets they hold and, for those which they believe hold potential value, assess whether that can dataset can be shared openly, whether it will require some form of managed access, or whether it is not appropriate to share it at all (for example, if the risk of harm to individuals or population groups is too great).

Regulation – achieving an appropriate balance

11. The consultation document highlights the complexity in the existing legislative and regulatory environment governing the use of public data, and how the lack of consistency in interpretation of this legislation can create a barrier to the effective sharing and use of data. We would emphasise the need for the Government to continue to work to streamline the regulatory environment wherever possible, and to develop clear guidance to ensure its consistent implementation.
12. In the health area, the regulatory landscape governing the use of patient data in research is particularly complex, with requirements under the Data Protection Act overlapping with other requirements – such as Caldicott guardians and the common law of confidentiality, and associated section 251 exemptions. We urge the Government to fulfil its commitment to establish a single Health Research Authority and to empower it to work with the Information Commissioner's Office (ICO) to develop a clearer and more effective governance framework.
13. The recent report to the Cabinet Office by Keiron O'Hara on privacy and transparency describes how enabling the linkage of public datasets will create an increased risk of the identification of individual data subjects. Where such risks exist it is paramount that appropriate safeguards are in place that protect the privacy and confidentiality of individuals and help to ensure public trust. The regulatory framework governing the use of data must therefore aim to be proportionate in balancing the need to protect the privacy of personal data with the need to facilitate legitimate research access in a safe and secure manner.
14. We have particular concerns in this regard in relation to the current revision of the EU Data Protection Directive, on which the UK Data Protection Act is based. It is vital that the UK Government ensures that the development of the Directive and its implementation in UK law does not restrict access to data by researchers and other users in line with the public good. For example, greater clarification is needed on the definition of "personal data", including whether pseudonymised data falls within the scope of data protection legislation. This needs to be clarified in guidance from the ICO and a careful assessment is required of how changes in the revised Directive impact on the use of pseudonymised data in research.

15. The Thomas-Walport review of data sharing examined in depth the issues related to balancing the need to enhance the use of data with the need to safeguard privacy. We would call on the Government to implement its key recommendations, particularly in relation to the establishment of safe havens to allow researchers to access and use sensitive datasets in a secure and controlled manner.

Leadership and coordination across government

16. Achieving the vision set out in the 'open data' strategy will depend crucially on ensuring effective coordination across Government. In particular, the full benefits will only be realised if the data are released in a format that enables them to be used and to be linked appropriately with datasets from other parts of government. We welcome the move to establish the Public Data Corporation (PDC) as a focal point for these activities in government. We assume that the PDC would take on the key roles described in the open data consultation document of aiding discovery of key datasets; ensuring consistent interpretation of regulatory provisions; and driving the development of standards and tools to enable data linkage.
17. With regard to research uses of administrative data, the Trust will take part as a member of the new task force chaired by Sir Alan Langlands that will bring together representatives from government departments, public agencies and research funders to develop recommendations on how future linkages, data access and governance will be managed. We hope that the work of this group will play a major role in shaping an effective cross-departmental approach to maximise the use of data.
18. Looking ahead, there may be scope for the PDC to work with the research community (both in the academic and commercial sectors) to develop demonstration projects on topics of strategic importance that depend on the integration of datasets from a range of government sources. This would help to inform the development of the platforms needed to enable data linkage, and highlight the major barriers to be overcome. In a similar manner, the Trust is currently engaged in discussions to develop pilot programmes in the NHS to demonstrate how e-health systems can be rolled out effectively – building on the successful models adopted in Scotland.
19. There may also be exciting opportunities for the PDC to collaborate with technology companies in developing innovative tools to mine and analyse datasets – and help to stimulate this exciting growth area.
20. In addition to effective coordination across Government, there also needs to be effective leadership within individual departments. To support the move towards a 'right to data', we endorse the proposal to designate departmental board-level responsibility for championing open data and for balancing the need to promote access with the safeguards needed to promote privacy. These individuals should seek to promote a culture throughout their agencies that encourages and incentivises staff to consider the broader value of the data they hold.
21. We believe that effective prioritisation of which data to release will be essential, particularly given current financial pressures on the agencies concerned. This should be based on an appraisal of the likely value of the data to external users and the anticipated demand, together with a cost-benefit analysis on whether the expense needed to share the data justifies the benefit gained from sharing. It is simply not efficient to share data which is not of a sufficient quality or which is not going to be used. In short, the focus must be on quality not quantity, with the aim being to ensure that those data of genuine value are made available in a way that delivers the greatest benefit to society.

The Wellcome Trust is a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health. We are independent of both political and commercial interests.