

‘Making Open Data Real: a Public Consultation’

Key Points

- The Nuffield Trust believes that opening more public service data to scrutiny and analysis could help improve the quality, consistency, equity and cost-effectiveness of services.
- We believe that in the analyses of care services data, one of the critical areas is to develop knowledge and tools about the best ways to exploit person level information without compromising an individual’s identity.
- The release of more anonymous or pseudonymous individual-level data could lead to improvements in the accountability, productivity, equity and quality of public services. The analysis of such data can also yield important insights for shaping public policy, and indeed the publication of data may well lead to enhancements in data quality over time, leading to a virtuous circle of improvement.
- We believe in addition to the release of data there needs to be investment in disseminating intelligence about what is possible (and legal) with public data and resources on techniques for exploiting these data
- Innovations in the use of data may emerge from the public sector, from private companies, as well as universities and voluntary groups or individuals. We suggest that it is important to think of ways that reward novel and innovative uses of data and in particular to exploit the ideas and imagination of the public.

The Nuffield Trust believes that opening more public service data to scrutiny and analysis could help improve the quality, consistency, equity and cost-effectiveness of services. Moreover, we believe that research conducted using open data is likely to generate insights that can be used to help shape public policy in the future. We therefore welcome the proposals set out in the Cabinet Office consultation document: *Making Open Data Real*.

The Nuffield Trust is an independent charitable organisation that conducts research and policy analysis aimed at improving the National Health Service. One area in which we have developed expertise in recent years is the analysis of very large linked datasets using information already collected in operational systems.

Throughout the health and social care sector, computerised systems record information about services provided and the health of service users. Through our work, we have demonstrated how it is possible to exploit these operational data systems in order to improve our understanding of how care is delivered. For example, we have been able to construct health and social care timelines that illustrate the care that an anonymised and unknown individual received from both health and social care services over the course of several years (Nuffield Trust 2001a). The ability to use data in this way is a powerful tool to help shape the delivery of care services, the planning of care and for the patients or users of services themselves. For example it can help identify “gaps in care” (where the care delivered does not match their expected needs) and unnecessary duplication of care.

Yet despite the abundance of data there are very few organisations in England able to look across care pathways in the ways that we have. We believe that the principles outlined in ‘Making Open Data Real’ would have helped facilitate the type of research that we conduct, and we therefore welcome the proposals.

Improving public accountability

We believe that the default position should be for all organisations providing public services to make their data openly available. There are three broad types of organisation that will help exploit more open data:

- a) Reviews by statutory regulatory or research bodies. For example, the Care Quality Commission has a duty to regulate both public sector and private sector health and social care, however the data available on from these different sectors differs considerably
- b) Intermediary organisations in public, voluntary and private sectors that undertake analyses to turn raw data into meaningful information. For example through the use of indicators, benchmarking tools, new presentational forms, statistical analyses.
- c) Through information used by individual citizens either to inform their choices or to influence public bodies about issues of concern, such as the accessibility of services

It is therefore important that the move towards greater openness and transparency should not focus solely on one group of secondary data users, but instead it should recognise the potential for multiple uses and multiple users of open data. This is especially important when thinking about the level of detail released and to which type of organisation.

For example we believe it may be helpful to define an intermediate class of data that are not appropriate to be made fully open. There is a case for making such data available to a body of registered users and researchers under strict conditions.

Using data to improve services

At the Nuffield Trust we are particularly interested in applications of existing data that improve services. So for example tools such as risk stratification built on exploiting existing data streams can help target preventive health services (Nuffield Trust, 2011b).

One of the critical steps in yielding the additional value from data is through the ability to link datasets. For example, linking health care data to geographic areas make it possible to look at how differences in social and economic characteristics influence health and health care use at the individual level. Yet linkage of this type relies on standard definitions and terms – in this case the labels of geographic areas being coded consistently in multiple datasets. It is therefore important that there are incentives for the use of agreed terms when collecting and disseminating data.

Yet at the same time we have to be aware of the dangers that an overly restrictive view of what constitutes acceptable data may stifle the collection of information in the first place. A delicate balance therefore needs to be struck to ensure that this does not happen, perhaps by allowing certain datasets to evolve over time.

Encouraging the publication of open data

We agree with the Government that the publication of open data should be encouraged by a combination of 'push' factors (a presumption of publication) and 'pull' factors (a right to data). A useful analogy is the introduction of the Freedom of Information Act 2000, where many public bodies were required to begin pro-active "publication schemes" for the routine release of certain important information.

We support the idea that there should be an expectation that public service datasets should be made open. Clearly there needs to be some exceptions to this eg national security, or individuals privacy but these should not be the norm but subject to separate justification and with some external review.

We believe there may be merits in the proposed approach of rating public sector organisations in terms of how open their data sources are, provided this is undertaken with some sensitivity to the constraints that organisation may face. The publication of such ratings and progress towards greater openness will help reinforce and accelerate change. We would also hope that individual government departments should themselves benefit from the ability of others to link their data and avoid insularity within different sectors of public services.

Improvements in data quality

Another likely benefit of publishing open data is that publication could itself lead to improvements in data quality. This phenomenon is observed consistently whenever new data sets are published. For example, quality of the accident & emergency dataset within

Hospital Episode Statistics (HES) has improved considerably since being published a few years ago. The improvements seen may be explained in part by providers being shamed into improving the low-quality data that are now on display, but also by practitioners taking more care over the collection of data because they know that it is likely to be put to a useful purpose.

Developing the infrastructure for handling sensitive data

There are significant amounts of publicly held data that contain information that is sensitive and confidential to an individual and cannot be released in a crude form. However this should not stop the use of approaches that enable the reuse of these data in ways that does not infringe on an individual's privacy. For example in health the use of techniques of anonymising or pseudonymising person level data have enabled a wide range of secondary uses of health information – for those aware of the techniques (Rumbold and others, 2011). In the United States, a legal distinction is made between 'fully' and 'partially' pseudonymous data. Partly pseudonymous data may contain some identifying information and so are only made available to users subject to contractual and technical limitations, and with the agreement that no attempt will be made to identify the individuals concerned (House of Commons Health Select Committee, 2007).

The Patient Information Advisory Group of the Department of Health set out clear guidance on the conditions under which such de-pseudonymisation may be appropriate (Patient Information Advisory Group, 2008). However, its guidance relates solely to the health care sector. The lack of equivalent guidance for social care data means that predictive tools for social care have not been implemented in practice partly for fear of transgressing information governance regulations (Nuffield Trust 2011b).

These tools and techniques need to be supported and developed.

There are moves to refresh the current legislative framework regarding information handling. We welcome these moves, because the current arrangements are currently unclear to many people (Rumbold and others, 2011). In addition to any legislative changes, however, we believe it will also be important to make people aware of what is or is not possible within existing guidelines. There is also scope for improving communication about how datasets may be legitimately used.

Innovation and adapting ways to look at information

We believe that there is almost unlimited scope for people with skills, imagination and ambition to improve how data are used. Allowing them access to the right data can lead to a multitude of insights and applications, which ultimately should help improve public services. These innovations in data manipulation, presentation and analysis arise in all sectors. Indeed many of the most important innovations have their roots in the public sector and in universities. It will therefore be important to ensure that these sectors are not discouraged from accessing open data because of high costs.

We would also suggest that government should consider some investment in fostering innovation. For example an innovation fund for supporting the development of ideas submitted by the public or community groups.

References

Bardsley M, Billings J, Dixon J, Georghiou T, Lewis GH, Steventon A. (2011a). 'Predicting who will use intensive social care: case finding tools based on linked health and social care data.' *Age and Ageing* 40(2):265-70.

Bardsley M, Billings J, Chassin LJ, Dixon J, Eastmure E, Georghiou T, Lewis G, Vaithianathan R, Steventon A (2011b). 'Predicting social care costs: a feasibility study'. London: Nuffield Trust

Healthcare Commission (2007). Independent Sector Treatment Centres: A review of the quality of care. London: Healthcare Commission.

House of Commons Health Select Committee
(2007) The Electronic Patient Record: Sixth report of session 2006–07. HC 422–I.

Nuffield Trust (2011a). 'Individual health and social care event timeline'. London: Nuffield Trust. Available at <http://www.nuffieldtrust.org.uk/data-and-charts/individual-health-and-social-care-event-timeline>

Nuffield Trust (2011b). 'Predictive modelling for social care: Next steps workshop'. London: Nuffield Trust

Patient Information Advisory Group (2008). The Use of Patient Information in the Long-Term Conditions Programme.
www.advisorybodies.doh.gov.uk/piag/piag-ltc-oct2008.pdf

Pollock A (2008). 'Independent sector treatment centres: evidence so far.'

BMJ 336:421

Rumbold B, Lewis G, Bardsley M (2011). 'Access to person-level data in health care: Understanding information governance'. London: Nuffield Trust

For more information about Nuffield Trust,
including details of our latest research and analysis,
please visit www.nuffieldtrust.org.uk



Download further copies of this policy response
from www.nuffieldtrust.org.uk/publications



Subscribe to our newsletter:
www.nuffieldtrust.org.uk/newsletter



Follow us on Twitter: [Twitter.com/NuffieldTrust](https://twitter.com/NuffieldTrust)

Nuffield Trust is an authoritative
and independent source of
evidence-based research and
policy analysis for improving
health care in the UK

59 New Cavendish Street
London W1G 7LP
Telephone: 020 7631 8450
Facsimile: 020 7631 8451
Email: info@nuffieldtrust.org.uk



www.nuffieldtrust.org.uk

Published by the Nuffield Trust.
© Nuffield Trust 2011. Not to be reproduced
without permission.