

26<sup>th</sup> October 2011

Dear Sir/Madam

**Making Open Data Real: A Public Consultation - Response from the British Medical Association**

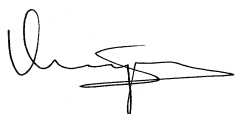
Thank you for the opportunity to respond to the 'Making Open Data Real' consultation. The British Medical Association (BMA) is supportive of widening access to high quality information to ensure that people understand their choices and can make informed decisions.

It is important that the data can be collected as part of the care process and healthcare professionals are not diverted away from providing care due to an additional data collection burden. Data must be meaningful, accurately reflect the services provided and be presented in a way which cannot be misinterpreted to ensure that the public is not misled. Raw health data can fail to measure complexities and subtleties of providing care and it is important that healthcare professional are involved to point out caveats, potential misinterpretations and limits which may not be apparent to others.

We welcome the commitment that privacy will be protected and agree that there needs to be further consideration of how this will be achieved. It is important that confidential data are not inadvertently disclosed by linking datasets.

We hope our comments are helpful and look forward to hearing from you with the outcome of the consultation in due course.

Yours faithfully



Professor Vivienne Nathanson  
Director of Professional Activities

UNCLASSIFIED

## **Making Open Data Real: A Public Consultation – Response from the British Medical Association**

### **1. Do the definitions of the key terms go far enough or too far?**

The definitions are appropriate. It is helpful that the definition of datasets makes it clear that the data will typically be collected as a by product of delivery. We would be concerned about additional burdens being placed on healthcare professionals particularly at a time when the Department of Health is reviewing data returns.

### **2. Where a decision is being taken about whether to make a dataset open, what tests should be applied?**

The first test must be to ensure that the dataset does not identify individuals. This is particularly important in healthcare. Confidentiality is fundamental to healthcare and the relationship between patients and doctors. The requirement for confidentiality allows patients to divulge sensitive information to their doctor without concern that it will be disclosed to others without their consent, except in very limited and exceptional circumstances. Any threat to this principle could have a disastrous impact upon both the individual care of patients and the wider public.

Whilst it may not be possible to identify individuals from a single database, it may be possible to link multiple datasets to identify individuals. The second test therefore should be is it possible to deduce identities or sensitive information by linking datasets? Consideration should not only be given to published datasets but the possibility of linkage between a published dataset and an unpublished dataset held by an organisation or body. An organisation may have been granted access to a dataset of identifiable data for a specific research purpose under section 251 of the Health and Social Care Act. If the government publishes information, it should not be possible for this organisation to infer further information about individuals by linking the identifiable and anonymised published datasets. This is a particular risk with rare conditions, for example, when publishing prescribing data at an organisational level. There needs to be agreed information governance processes to ensure that data are appropriately anonymised or pseudonymised prior to release and measures put in place to ensure that patients cannot be inadvertently identified by opening up access to information.

Data must also be meaningful and accurately reflect the services provided. Publishing raw data can be misleading and damaging. Data need to be presented in a way which cannot be misinterpreted, for example, when comparing performance ensuring that data are truly comparable particularly because misrepresentation could impact upon a clinician's reputation. Data can fail to measure complexities and subtleties of providing care. Anaesthetists, for example, contribute to patients' care rather than a complete episode and whilst a consultant may be named as responsible it is likely that other clinicians will be involved in the patient's care. Another example is when considering the success of smoking cessation services in primary care, the raw data of the number of patients who have stopped smoking needs to be contextualised with information about the practice list size, the percentage of the practice population who smoked to start with and the demographics of the practice. The reporting of mortality figures is a further example of where data can be sensationalised when taken out of context, which can be very damaging. When data are published, clinicians must be involved so that they can point out caveats, potential misinterpretations and limits which may not be apparent to others.

It is important to choose carefully which datasets are published. There is a risk of information overload and the public being unable to navigate the information they need. There should be criteria to establish why it is valuable for the data to be published. There is also a risk that requiring the publication of data may create artificial incentives for certain parameters with the potential that other areas, which are difficult to measure but still important, are neglected.

Finally it is important that the data can be collected as part of the care process and healthcare professionals are not diverted away from providing care due to an additional data collection burden.

**3. If the costs to publish or release data are not judged to represent value for money to what extent should the requestor be required to pay for the public services data, and under what circumstances?**

If the costs of publishing the data are not judged value for money there should be consideration, in the first instance, of whether the data should be published. If the data are still deemed important then we agree that it is important that those providing public services at a local level do not shoulder the financial burden of providing the data. The Health and Social Care Bill includes provisions which permit the NHS Information Centre, as the requestor, to make payment to providers to cover the costs of making the data available and we feel that this is appropriate. In recent years, healthcare providers have stored their data on centralised servers which are supplied by private companies e.g. GP clinical system suppliers. Care needs to be taken to ensure that this remains 'NHS data' and private organisations cannot charge for access to this data.

**4. How do we get the right balance in relation to the range of organisations (providers of public services) our policy proposals apply to? What threshold would be appropriate to determine the range of public services in scope and what key criteria should inform this?**

In relation to healthcare, we agree that it is appropriate that this does not just apply to public services provided by public bodies but also those who have been commissioned to provide services. This will ensure that private providers, offering NHS services, are subject to the same rules of transparency. Policy proposals relating to withholding information due to commercial sensitivity need to apply to all organisations. There needs to be consideration of how to ensure a level playing field so that private providers do not have more scope to plead commercial sensitivity. This could lead to organisations being able to use information to select the most profitable procedures. This might destabilise smaller organisations such as GP practices resulting in a negative impact on patient care. A further example is when working to provide services to children, vulnerable adults and other patients, social and legal services should be subject to the same rules as public providers of healthcare; the data released needs to reflect a team approach to the care of patients.

Bodies within proposed NHS structures such as the NHS Commissioning Board and Monitor should also be subject to these proposals.

**5. What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?**

In healthcare, a wide range of data are already published and collected. For example, on NHS Choices information is available to the public on mortality rates, cleanliness, MRSA and patient comments. To encourage the publication of data there needs to be engagement with those producing the data so that any misunderstandings due to publishing the data can be avoided. There need to be mechanisms in place to ensure the quality of data so that there is confidence in the outputs and there needs to be an agreement of the value of publishing the data so that it can be used by patients to inform their care i.e. to avoid situations where a lot of effort goes into publishing the data and nobody looks at it.

### **An enhanced right to data**

**1. How would we establish a stronger presumption in favour of publication than that which currently exists?**

The advantage of a 'publish by default' principle is that it could help improve the quality of data if those entering data have that mindset. At the same time it is important that those providing public services do not become burdened by producing data and the principles outlined in the response to question two should be applied. This is particularly important in relation to healthcare as some providers will be very small with limited resources.

Alongside the legislation listed in the consultation document, the Health and Social Care Bill also includes provisions for the publication of healthcare data, which create a stronger emphasise on publication of data than currently exists.

**2. Is providing an independent body, such as the Information Commissioner, with enhanced powers and scope the most effective option for safeguarding a right to access and a right to data?**

We believe that the Information Commissioner is the appropriate body to carry out this function. We agree that this power should be infrequently used as there may be entirely valid reasons for not publishing data and it would be inappropriate for time to be diverted away from providing public services as a result of unnecessary bureaucracy.

**3. Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?**

It is not clear whether the question relates to sensitive confidential health information or other forms of personal data. We are concerned, as stated above, that there is a risk of sensitive information being inadvertently disclosed as a result of large datasets being linked. A letter was published in the BMJ earlier this year on the risks of re-identification of patient data when data are linked to other datasets.<sup>1</sup> There needs to be further exploration of these risks, involving experts in the field, to ensure that data remain anonymised as we are concerned that existing privacy measures are inadequate.

## Corporate and Personal Responsibility

**1. How would we ensure that public service providers in their day to day decision-making honour a commitment to Open Data, while respecting privacy and security?**

It is essential that the open data agenda does not impact upon privacy. When the confidentiality of sensitive data are breached it is rarely through malicious intent but due to a lack of understanding of the risks. There must be a commitment to ensure that all staff dealing with confidential data are trained in information governance issues. Those involved in publishing data must understand the risks of inadvertently revealing information due, for example, to small numbers. Privacy and security must take precedence over making data public.

It is important that the publication of data is coordinated to ensure that all organisations are working to similar standards and that the public can access similar comparable datasets for different providers. We agree that there is value in having a panel to provide oversight and set standards.

To ensure commitment there needs to be an understanding of the value of publishing the data. If organisations see the publication as offering little value then they are unlikely to be committed to the policy. A study by the Kings' Fund found that only 4% of patients said they had accessed NHS Choices information on performance when choosing a hospital<sup>2</sup>. This statistic is unlikely to generate enthusiasm for publishing the information. Further work needs to be carried out with stakeholders, to establish what information it would be useful to publish and how this can be presented in a way that is easily accessible to organisations and the public. There is likely to be a greater commitment to publication once organisations see evidence that their published data is impacting in a positive way.

**2. What could personal responsibility at Board-level do to ensure the right to data is being met include? Should the same person be responsible for ensuring that personal data is properly protected and that privacy issues are met?**

Personal responsibility at Board-level would help to raise the profile of the importance of data within organisations. It would be appropriate for the same person to be responsible for ensuring that personal

<sup>1</sup> <http://www.bmj.com/content/342/bmj.d238/reply>

<sup>2</sup> Robertson, R., Dixon, A. (2009). Choice at the point of referral; early results of a patient survey. The King's Fund.

data are properly protected and privacy issues. There would, however, need to be a commitment to provide appropriate resources and support for this role.

### **3. Would we need to have a sanctions framework to enforce a right to data?**

A sanctions framework would force a commitment to publishing data. It would be preferable for the commitment to be established via other mechanisms including motivation stemming from knowing that the public understand the services on offer and the value of care provided. There may be reasons why the data cannot be published such as a new computer system being installed and the reporting functionality failing. If financial sanctions were enforced it could divert funding away from the provision of care due to reasons which were out of an organisation's control. On certain limited occasions, it may be appropriate to enforce sanctions; this should be a last resort, any extenuating factors should be taken into consideration and an organisation should be given every chance to publish data prior to reaching this stage.

### **4. What other sectors would benefit from having a dedicated Sector Transparency Board?**

We would strongly recommend that there should be a Transparency Board for health due to the nature of personal health information. It is also important to ensure that there is a mechanism in place to escalate concerns or queries relating to the publication of data. Consideration should also be given to establishing a Board for social care or combining this with the health board. If a Transparency Board is set up for health data it would need to include representation from clinicians who understand the complexities of health data.

## **Meaningful Open Data**

### **1 How should public services make use of data inventories? What is the optimal way to develop and operate this?**

The BMA fully supports the principle that public services should only collect and publish data that are meaningful and useful. There needs to be discussion with those providing public services to establish what datasets are meaningful and useful.

We agree that there should be a clear understanding of what information exists and what can be accessed and that information inventories could be a sensible way of delivering this. Organisations across a sector will hold similar data and therefore we agree it is sensible to establish a framework to ensure that data inventories are consistent. This will minimise the burden on organisations if a minimum inventory is agreed, which can be expanded with other relevant organisational information. The BMA is willing to engage with the government to consider the best way that this can be achieved in healthcare.

### **2. How should data be prioritised for inclusion in an inventory? How is value to be established?**

This is a fundamental question and it is likely to be an evolutionary process with different datasets being included and excluded from the core inventory over time. An agreed set of principles should be established. In relation to healthcare, the priority should be the information that will make a difference to a patient's care and inform their decisions. This should include evidence based information about different treatment options and outcomes so that a shared decision can be made on the appropriate pathway of care.

The focus to date has been on providing the public with high level data to choose between different organisations. Whilst the BMA supports meaningful choices we do not believe the patient choice agenda has improved clinical outcomes or offers patients the choices they actually want. There has been much focus on providing patients with high level data. The information patients often request, to help inform their decisions, is not quantitative data but more personalised information about how interventions or treatments will affect their lives. They are often more interested in the clinical effect of surgery rather than high level data on mortality or infection rates. Patients request simple information such as what will

happen, how long it will take, the recovery time, what the side effects will be and whether they will be able to continue to do a certain activity or sport after the operation. Whilst we feel there is value in transparency and making higher level data available the focus should be on providing patients with this kind of information to help support their decisions. This is also reflected by looking at the type of questions patients ask online in various forums; they tend to seek assurances about more personal issues, things they may think are too silly or embarrassing to ask a healthcare professional directly or seek assurances from others who are dealing with similar conditions. Healthcare professionals must be involved in establishing the criteria as they will know the kinds of questions patients ask during consultations.

Some information can be prioritised on the basis that it is information that the public regularly wants to know. A dataset may be accessed by small numbers because a condition is rare but be incredibly valuable to those who are accessing it; the criteria for inclusion cannot be judged on the percentage of the public accessing it alone.

Consideration should also be given to ensuring that there is transparency around the use of data for secondary uses. There is limited public awareness around the extent to which data are used for research and the benefits this can bring to the health service. The open public data agenda could be an opportunity for not only providing transparency around the data that are held but also how data are used.

### **3. In what areas would you expect government to collect and publish data routinely?**

It may be helpful to report on activity rates on a yearly basis and other data that are required for contract monitoring and payment. Hospital and GP outcome data should also be routine providing that they are published appropriately i.e. they accurately reflect the care provided.

### **4. What data is collected 'unnecessarily'? How should these datasets be identified? Should collection be stopped?**

The BMA supports efforts to streamline data collections. The Department of Health is currently consulting on a fundamental review of data returns and a clear methodology has been established for the review. Stakeholders have been engaged in the process and this model could be applied for other sectors. We agree that if datasets do not meet the agreed criteria then collection of data should stop.

### **5. Should the data that government releases always be of high quality? How do we define quality? To what extent should public service providers 'polish' the data they publish, if at all?**

We believe that the data should always be high quality. One of the aims of the open data agenda is to make data available so that the public can make informed decisions. Data must be high quality to ensure that the public are not misled. Poor quality data results in a lack of public confidence and will also affect commitment to the open data agenda. If the quality of published data across the sector is poor, organisations are less likely to invest time improving their data because there will be fewer benefits for example, it will not be possible to compare and benchmark services. It is important that both the public and providers value the data that are published.

We recognise that there needs to be a balance between allowing organisations to quality-assure data prior to publication and publishing timely data. We have concerns about the 'publish and improve' policy<sup>3</sup>; the media report on data when they are first available and this could have a detrimental affect on organisations and individuals. Real-time data can be very valuable and at times inaccuracies can be identified more readily. Dashboards are a good example of how real-time data can be presented to patients in A&E departments on waiting times, infections etc without the need for data cleansing. It is likely that a 'one size fits all' approach will be difficult to implement as some data will require longer to produce whereas other data can be made available immediately. This may depend on the IT systems in place at different organisations.

---

<sup>3</sup> Liberating the NHS: An Information Revolution:  
[http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH\\_120080](http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_120080)

The quality of data should be judged in terms of accuracy and in terms of how meaningful the data are. Data which have been contextualised can be more valuable than raw data alone which can be misleading.

### Government sets the example

**1. How should government approach the release of existing data for policy and research purposes: should this be held in a central portal or held on departmental portals?**

A central 'health portal' for signposting the public to available information could be very helpful as data are currently dispersed widely and can be difficult for the public to navigate. Ensuring that these data are easy to find could ensure that available data can be applied to research rather than repeating requests for the

same data. Whilst the central portal would serve to signpost information, local portals could tailor the view of the data to make it more relevant to the local community.

**2. What factors should inform prioritisation of datasets for publication, at national, local or sector level?**

We suggest that the focus should be on publishing data that are already being collected and used for health service planning as this will be more accurate and locally useful. This would also be a better use of resources rather than deciding what should be published and then asking for it to be collected.

The demand for the data should be another factor; the restructuring of the NHS will create new intelligence needs which need to be addressed as a priority to ensure decisions are based on evidence. The resources required to publish the data will also be a factor at both national and local levels.

**3. Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?**

This is difficult to respond to as it will vary depending on the data. We feel that the priority should be to ensure that existing data are presented in ways that are easy to use, navigate and view in different formats to encourage a more intelligent use of data for both providers and the public. The NHS is often referred to as data rich and information poor. The priority should be to ensure that existing data are high quality, can be trusted and can be used by both health organisations and the public to drive standards.

### Innovation with Open Data

**1. Is there a role for government to stimulate innovation in the use of Open Data? If so, what is the best way to achieve this?**

The presentation of data is extremely important and we agree that innovation should be encouraged. We agree that commercial companies will play a role in making data available to the public. Commercial involvement must not compromise security or the privacy of data. We would also be concerned if this resulted in data being published in a biased way or healthcare organisations having no opportunity to challenge misrepresentation. This must be a collaborative arrangement with providers and innovators working in partnership to meet the data demands of the health service at both national and local levels.



UNCLASSIFIED