

IMS Health response to 'Making Open Data Real: A Public Consultation'

October 2011

Introduction

IMS Health is the gold standard in pharmaceutical and healthcare market intelligence, relied on by virtually all of the world's largest pharmaceutical companies — along with government agencies, policymakers, researchers and financial analysts around the globe.

IMS Health has completed the Cabinet Office's questionnaire on *Quantifying Opportunities for the Growth Review*, has submitted a paper to the Cabinet Office Transparency team on the economic benefits of open data within the life sciences sector and has contributed to the recent Cabinet Office's life sciences workshops on Open Data under the Growth Review. Our response to this consultation is intended to compliment the aforementioned documents.

Our answers to the questions posed in this consultation are based on IMS Health's experience of handling health data. However, many of the principles we discuss will apply to other areas of public sector data. We hope that our consultation response will be useful in highlighting issues to be considered and risks to be averted so as to ensure the success of the Government's Open Data initiative.

Questions

Glossary of Key Terms

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

IMS Health does not have a comment on this question.

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

As stated in the recent Cabinet Office-commissioned review of transparency and privacy, *"The success of the transparency programme will depend on the government not breaking the law and on not losing the trust of the citizen, and to do that it will need, among other things, to avoid releasing information in a form which could lead to widespread compromises of privacy by adversaries even if the releases were legal and the possibility of compromise was initially unanticipated or discounted by the public."*¹ There are therefore two key tests – is the release legal and will it jeopardise or threaten confidentiality?

To ensure that these two questions are answered in a way that satisfies the public, we would support the review of transparency and privacy's recommendation that the Government should establish a health sector transparency panel supported by appropriate technical and domain experts.²

In addition there are other considerations that the Cabinet Office may like to take into account but we recognise that these appear to raise potential conflicts between the two drivers of transparency – public accountability and economic growth. In our other responses to the Cabinet Office, for

example, we have commented that release of some types of data may restrict the ability of private companies to exploit more valuable data in an economic or efficient manner.

As an information services provider we are also reluctant to see information that will inevitably mislead put into the public domain. Based on our long experience we have little confidence that caveats or explanations will be read or used. We understand, however, that to restrict data on the grounds that it may mislead may be viewed as being in conflict with the drive for public accountability. At the very least, however, potential to mislead should perhaps be used to prioritise the order of release of those datasets held by government.

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 public services data, and under what circumstances?

Value for money implies that the benefits to either public accountability or economic growth outweigh the costs of production. We would be interested to know the decision criteria to be used here as this may determine the circumstances under which data requestors should be required to pay, or indeed how much they should pay. As we have stated in our paper on the economic impact of Open Data, there are datasets, which if released, may have a negative impact on economic growth overall.

If, however, the Government decides to charge for new public sector datasets that are not judged to represent value for money it must be careful that its charges do not distort the existing market for information. There must be a level playing field for charges and in charging for public sector data and the Government should put in place policies that ensure that private sector information providers are not put at a disadvantage to the government bodies that hold, release and wish to commercialise 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?

IMS Health does not have a comment on this question.

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

An appropriate mechanism to ensure publication of data by public service providers might be a legal requirement to publish the datasets they collect unless these meet certain criteria which would make them exempt from publication. These exemptions should be set out in detail and include some of the points set out in our answer to question 2 above. A key exemption would be for any data which, if released into the public domain, would enable the identification of individuals, including through a process of jigsaw identification when combined with other datasets already in the public domain. We would recommend that any proposals which follow from this consultation should ensure they contain measures which protect against the risk of jigsaw identification.

An Enhanced Right to Data

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

IMS Health does not have a comment on this question.

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?

IMS Health does not have a comment on this question.

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

The protection of personal data is of particular importance in the health sphere. The information that the health service holds is amongst the most sensitive collected by public bodies. Patients trust that the NHS will protect the confidentiality of their information and it is important to ensure that this trust is not undermined as we move towards a situation where there is a presumption of openness. We note that in the foreword to this consultation the Minister writes that *“It is my intention that no personal data will be shared with any third party as part of this initiative”* and that the consultation commits that *“Transparency will not be extended at the cost of privacy”*. We are encouraged that the consultation document recognises the challenge of ensuring *“that personal data is anonymised, particularly when they are released alongside many other datasets, which have the potential to be merged”* and welcome the contribution that Dr Kieran O’Hara’s transparency and privacy report for the Cabinet Office has made to this debate.³

The challenge of jigsaw identification which Dr O’Hara identifies in his review is of particular concern in the health sector because of the richness of the datasets that are already or may be released. There are several notable cases of jigsaw identification, the most relevant to the health sphere being from the US where release of certain health insurance records (similar in format to Hospital Episode Statistics data), were used in combination with other data to identify the health records of the then Governor of Massachusetts.⁴

Any increased risk of patient identification from public release needs to be weighed against any public benefit that might accrue from the release of NHS data into the public domain at a more granular level. It is important that safeguards are in place to ensure that the risk to privacy of jigsaw identification is taken into account when considering data release. For example, in terms of current proposals we would argue that coming legislation such as the Health and Social Care Bill should contain more explicit exemptions from release for data that could enable jigsaw identification to reveal the identity of a patient, or at least that there be a formal review process put in place that specifically assesses that risk. We recognise that such an assessment can be difficult and so also support the recommendations made by Dr O’Hara that real world research be conducted into privacy risk and that a health sector transparency panel is put into place. We would also argue that this panel should be put in place prior to any release of any dataset where a review, such as that done by the NHS IC in relation to prescribing data, has indicated the possibility of an increased risk of patient identification.

4. What might the resource implications of an enhanced right to data be for those bodies within its scope? How do we ensure that any additional burden is proportionate to this aim?

IMS Health does not have information that would allow us to estimate the level of resource that will be required by public bodies to process and publish data under the Open Data initiative. However, we would note that for some organisations in the health service the sale of data to information intermediaries in the private sector is a source of revenue. Contracts that are already in place for the sale of data would need to be respected and the loss of a potential revenue stream to health bodies will need to be considered.

It should also be noted that an enhanced right to data will create additional costs for companies in the private sector who handle public sector data. For example, information intermediaries will need to ensure as more public data becomes available that they do not inadvertently enable de-anonymisation of data through jigsaw identification. IMS has experience of implementing systems that protect provider (and thus patient) identity in those cases where it has been established that databases can be easily linked to reveal such an identity. Such systems are expensive – they require a variety of encryption techniques, special processing procedures including processing in memory as well as a variety of access controls.

5. How will we ensure that Open Data standards are embedded in new ICT contracts?

IMS Health does not have a comment on this question.

Setting Open Data standards

1. What is the best way to achieve compliance on high and common standards to allow usability and interoperability?

IMS Health does not have a comment on this question.

2. Is there a role for government to establish consistent standards for collecting user experience across public services?

IMS Health does not have a comment on this question.

3. Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?

There is clearly an important role for information intermediaries in helping to interpret the wealth of data that the public sector generates. This is particularly true in the health sector where there is a need to make the raw data generated by the NHS meaningful for various audiences including commissioners, providers, healthcare professionals, patients and the public. There are currently strict guidelines which cover the management and use of NHS data. These include regulations relating to the Data Protection Act, the Freedom of Information Act, the Law of Confidentiality and the Re-use of Public Sector Information Regulations. Information intermediaries will need to abide by these regulations, and as more NHS data is released into the public domain it will become even more important to have robust information governance standards in place so as to protect confidential patient information.

IMS Health supports the proposal to develop a scheme for accreditation of information intermediaries and notes the NHS White Paper's proposal to introduce *"a voluntary accreditation system, which will allow information intermediaries to apply for a kitemark to demonstrate to the public that they meet quality standards"*.⁵ This would help to give some assurance of reliability to analyses produced by information intermediaries and would help the public in judging which sources of information they can trust, and that the data are being treated in an appropriate manner. IMS Health has considerable experience and expertise in managing sensitive health information and using health data effectively to produce robust and useful insights. We would be happy to contribute this expertise to the development of a voluntary accreditation system for information intermediaries.

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 considerations?

It is vital that public service providers consider privacy and security on a day to day basis when making decisions about whether to publish data. We therefore support Dr Kieron O'Hara's recommendation to create sector specific "transparency panels" that would consider all proposed data releases and provide advice to the relevant public sector body on whether the data released would pose a significant threat to privacy or security⁶. These panels could be made up of independent experts and stakeholders.

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?

IMS Health does not have a comment on this question.

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

IMS Health does not have a comment on this question.

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

The establishment of a Sector Transparency Board to advise the Government on increasing the transparency and usability of NHS data would be of benefit. This is a particularly complex area in which there are vast amounts of information generated and handled by a variety of bodies. The implications of data release will also be wide-ranging, and if conducted in the wrong way could have very serious consequences for the privacy of individuals and trust in the NHS, as well as for the transparency initiative.

Meaningful Open Data

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

We agree that public services should only collect and publish data that are meaningful. We welcomed the Department of Health's recent review of its central data returns⁷. We feel that this was a useful exercise in creating an inventory of datasets collected by the Department of Health and identifying whether these are all of value. Such exercises should be undertaken by all Government departments at regular intervals to assess whether they should continue to collect certain data. It is important that during the review process, stakeholders are consulted about their views on the value of data to ensure that decisions about what to collect/discontinue do not only reflect the internal priorities and needs of public services, but also those of the public.

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

If data are collected by a government department then the principles of transparency would demand that all such data are included within an inventory. We are not convinced that the value of each dataset in the inventory should be assessed by the Government unless such considerations are to affect its decision to release such data. If value is not to be assessed by the Government then it

should perhaps rely on requests from industry or the public to access such data. We would argue that the department should follow the recommendations set out in Dr O'Hara's review, whereby the requestor puts their case as to why the department should release the data and as to how such release will not jeopardise personal privacy. These arguments should then be considered by the health sector transparency panel who would decide whether release should proceed.

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

IMS Health does not have a comment on this question.

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

Data that are not useful to one party may be useful to another. A process to determine "unnecessary data" might follow that recently conducted by the Department of Health, which has stated that it intends to cease the collection of certain datasets and its decision is subject to comments from potential users.

It is clear though that just because a dataset is not made open does not mean that it does not have value.

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 note that the foreword to the consultation suggests that *"given the costs of improving quality, our judgement should be that we will publish data of lower quality in preference to holding it back, while seeking over time to drive up the quality of that data."* As noted above, IMS Health has some concerns about the Government's proposed approach to publishing data although it may not be of high quality. Whilst it should not be the role of the Government to provide detailed analyses of all of the data it publishes, it is important that data that are released are accurate and do not mislead the public.

Questions have been raised about the quality and accuracy of a number of NHS datasets such as Hospital Episode Statistics and programme budgeting data. We acknowledge the argument that if providers of data know the data are going to be put into the public domain, they will take more care to ensure they are of high quality, or at least of a quality that puts their organisation in the best light. However, there is a danger that if the Government allows low quality data to be released into the public domain this inaccurate information could mislead the public, for example about the quality of care being provided by their local health services, and detrimentally influence the behaviour of service users. Public bodies should consider publishing guidelines alongside the datasets that they publish to explain the limitations of the data. Past experience indicates that these are unlikely to be read by many but they will provide a reference document for those who wish to discuss the data in more detail.

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?

IMS Health does not have a comment on this question.

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

In order for the Open Data initiative to have maximum impact we believe that the Government should:

- ***Prioritise those initiatives that will have the greatest public and economic benefit***

Given the current financial climate and strains on NHS resources it will be important for the Open Data initiative in the health service to prioritise those initiatives which could have the greatest impact on improving the quality and patient experience of care, as well as driving improved health outcomes and economic growth.

For example, in terms of one type of data that is being considered for public release – prescribing data – the NHS Information Centre has advised that, unless properly contextualised, prescribing data does not provide a valid indication of the quality of care.⁸ IMS modelling also suggests that the release of practice-level prescribing data by individual drug would make it more difficult to release more valuable data i.e. pseudonymised patient data (real world data) that will enable providers to determine the most effective healthcare interventions. Whilst practice-level prescribing data by individual drug may be technically easy to release, it may not be the dataset that will have the greatest public and economic benefit. In 1999 the market value of prescription data was estimated at £26-32 million and is thought to have declined since, whilst pseudonymised patient data (real world data), which is currently valued at £10 million direct value and £60-70 million indirect value, has the potential to generate £100 million of direct value and an additional £300-350 million indirect value. However, IMS has calculated that the release of practice-level prescribing data at individual drug level would constrain the potential growth in value of pseudonymised patient level data to only £20 million of direct value and £60-70 million of indirect value⁹. In order to maximise the value of prescribing data without damaging the value of pseudonymised patient data IMS would recommend that the Government limit the release of practice-level prescribing data to that by disease/therapy area.

- ***Explore where value and insight can be gained from linking existing datasets***

The NHS collects a wealth of information that is useful to commissioners, providers, patients and the public. Often, however, such databases, like the prescribing data collected by the NHS Business Services Authority, are unable to be used effectively to drive up the quality of care because such data cannot be linked to diagnosis, procedures, outcomes or care processes at a patient level. The NHS Information Centre has established a linkage service where such links across care settings and types of care can be made. This will be an essential service going forward and it should be supported. IMS Health has already made use of it to link hospital prescribing to patient characteristics, diagnoses and procedures in the Hospital Episodes Statistics database and to General Practice care in the GPRD database to create a Secondary and Primary Care Longitudinal Database (SPLD). This longitudinal database could contribute greatly to the Government's aims of the 'Information Revolution' in the NHS by enabling meaningful health outcomes data to be collected in a consistent and effective manner.

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

Broader sets of data often raise questions that can only be answered by data at a more detailed level. The question should not thus be whether or not to publish a new set of data or more detailed data but rather whether or not sufficient data can be released such that key questions can be answered. In our previous responses we have commented that provider level data (for example data about GP practices or Trusts), whilst having some value, is not as valuable as pseudonymised patient level data which in turn is not as valuable as identifiable patient level data.

In addition, there is a danger that releasing certain data at a more detailed level may damage the potential value of other datasets. For example, the release of practice-level prescribing data by individual drug (currently only released at PCT level by disease/therapy area) may limit the type or volume of pseudonymised patient level data that can be released. Pseudonymised data has the potential to offer valuable real world evidence on the impact of healthcare interventions on patient outcomes.

We would therefore recommend that the Government focuses on how to release pseudonymised patient level data rather than releasing provider level datasets. Provider level datasets are not only of less value (from both an economic and quality of care perspective) but may also restrict organisations' ability to store and process pseudonymised patient level data effectively or efficiently due to the risk of jigsaw identification.

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?

Whilst the Government may be able to play a role in stimulating innovation in the use of Open Data, it is also important that it does not stifle innovation that already exists in the private sector or unwittingly harm innovative work.

For example, the life sciences industry is already actively engaged in unlocking the potential economic benefits to the UK that may accrue from Open Data such as the provision of 'Real World' data to prove the value of medicines and/or from data to aid recruitment in clinical trials. However, the success of these initiatives is dependent on continued access to pseudonymised patient level data but only where the confidentiality of patients is assured. Innovation in this context is thus a balance between data access and privacy.

Likewise the Government should ensure that the powers given to the bodies responsible for the collection and release of data are exercised in such a way that private sector information providers may compete on equal terms with those bodies. Access to the raw data collected by those bodies should be on the same terms and timings as the body itself.

REFERENCES

¹ *Transparency Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*, Dr Kieran O'Hara, September 2011, pg.11 available at

<http://www.cabinetoffice.gov.uk/sites/default/files/resources/transparency-and-privacy-review-annex-b.pdf>

² *Transparency Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*, Dr Kieran O'Hara, September 2011, pg. 65 available at

<http://www.cabinetoffice.gov.uk/sites/default/files/resources/transparency-and-privacy-review-annex-b.pdf>

³ *Transparency Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*, Dr Kieran O'Hara, September 2011, available at

<http://www.cabinetoffice.gov.uk/sites/default/files/resources/transparency-and-privacy-review-annex-b.pdf>

⁴ Described in *Broken promises of privacy: Responding to the surprising failure of anonymization*, Ohm P., Available at http://www.law.depaul.edu/centers_institutes/ciplit/niro_symposium_09/pdf/paper_ohm.pdf

⁵ *Equity and excellence: Liberating the NHS*, Department of Health, July 2010

⁶ *Transparency Government, Not Transparent Citizens: A Report on Privacy and Transparency for the Cabinet Office*, Dr Kieran O'Hara, September 2011, pg.65 available at

<http://www.cabinetoffice.gov.uk/sites/default/files/resources/transparency-and-privacy-review-annex-b.pdf>

⁷ *Fundamental review of data returns: A Consultation on the Recommendations of the Review*, Department of Health, 30 August 2011, available at http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_129725

⁸ *Wider release of NHS prescribing data: Consultation document*, The NHS Health and Social Care Information Centre, 2008

⁹ IMS, data on file