

Making Open Data Real

A response to the public consultation by Dr Foster Intelligence

Introduction

Dr Foster has demonstrated the potential for private enterprise to turn government data into information products for professionals that improve public services. Dr Foster has demonstrated the potential to use public sector data to create a business generating employment and growth, including export-led growth with sales to the US and EU. While the scale of these activities remains relatively small, we believe that the potential is large and that the industry is still in its infancy. However there are significant barriers to the industry developing. We therefore welcome this consultation and have a number of recommendations around supporting the development of professional services industries using government data.

As well as professional information products, Dr Foster provides information free to the public and is the leading independent source of information about standards in healthcare. There are specific challenges which affect the availability of information to the public. We have a number of recommendations about this area also.

Supporting the development of a data analysis industry

We would draw your attention to the following key barriers to effective use of government data:

- Data is not provided in disaggregated formats that allow value to be added
 - o For example, data on health related benefits is published in aggregated tables of data but access to the underlying data sources is strictly controlled. While some control is necessary at the moment it is a barrier to effective use of the data.
- Legal barriers to re-use of data exist for key data sets
 - o For example, the law on access to ONS deaths data explicitly restricts use for commercial purposes.

- Regulations often limit access to key data elements
 - o For example, access to certain key fields in the Hospital Episode Statistics such as the name of the GP practice that refers a patient to hospital are regarded as “sensitive” and access is restricted. Classification of data as ‘sensitive’ because it relates to professionals or organisation is not appropriate as it limits accountability.
- The processes by which organisations can apply for access to data are often obscure
 - o For example, it is not clear what the process is for application to many clinical audit datasets. Greater transparency about how data can be accessed would reduce barriers to use of data.
- The time taken to process applications for data is often excessive
 - o For example, Dr Foster has had an application for ONS deaths data take two years to be processed. Businesses that wish to use public data are often put off by the apparent lack of interest in helping organisations access data.

Creating value from data

The key to turning data into value is by being able to use data flexibly to inform decision making. Our experience in healthcare has shown that certain features are key to success in this process. They include:

- 1. Being responsive to customer needs:** Being able to cut data to the specific requirements of particular organisations.

For example a local authority may wish to use data to profile particular geographic areas and neighbourhoods. Within neighbourhoods they may wish to profile different age groups.

- 2. Being innovative:** Being able to test different approaches to measurement

For example a hospital may wish to understand the length of stay of patients in hospital in terms of the mean length of stay, the median length of stay, the median length of stay after applying trim points, or the % of patients in the national top decile after adjusting for age.

- 3. Being transparent:** Businesses serving professional users of information created from government data need to help professionals pick apart analyses to confirm for themselves what the information is telling them.

For example, if a hospital has many patients who are readmitted following discharge, they will want to profile those patients in terms of age, social deprivation or co-morbidities.

These are just three examples but the key lesson from each is the same: that creating value from data relies on flexibility. Currently, most government data is published in standard tables of aggregated data. These are of very limited value in helping services understand what the data means. The chances of the precise bit of data that any particular organisation needs being the one that the government has published are vanishingly small.

The government will never be able to respond adequately to the many and varied requirements of different users of data. For this reason, it needs to support the development of an industry of data analysts able to meet end-user requirements.

The focus of government policy should move away from publication of aggregated tables towards supporting the development of an industry in data analysis. Central to this endeavour is enabling access to underlying data sets in anonymised formats.

Creating a legal framework for release of anonymised data

The NHS has been relatively successful at releasing anonymised data in the form of the Hospital Episode Statistics and has created a good process for doing this. This, however, remains the exception. Other health data sets, such as ONS deaths records, are restricted in their use by law. Although ONS deaths records are public documents in paper form, the electronic database of deaths records is only released to approved researchers or under orders from the secretary of state for health. The ONS currently interprets the law around deaths data to mean that it should not be made available to commercial companies to support the NHS or other public services.

Legal restrictions also affect other data sets such as benefits data in connection with disability or carer status that would be of enormous help to planners of healthcare. Reform is needed to make these data sets accessible.

Being responsive to the requirements of the data analysis industry

Currently, organisations that hold large amounts of public services data attempt to meet the needs of external users by publishing data in aggregate format. This keeps large numbers of people inside such organisation employed in these activities. However this approach is not responsive enough to the requirements of public services themselves. Instead organisations should put greater focus on meeting the needs of the data analysis industry and support businesses in meeting the needs of the public and public services.

Key aspects of this would be have open and transparent mechanisms for requesting data; having mechanisms in place to allow access to underlying event-level data in anonymous formats; and responding to requests for data in a timely fashion.

Meeting consumer needs for information

Dr Foster was established in 2000 to provide more information to the public about standards in healthcare. Dr Foster has explored a number of potential business models to support this activity including working with patient organisations, seeking sponsorship from the private sector and working with the media.

While this is a growing area and one that will see development in the coming years it should be recognised that commercial models to support provision of data to the public remain relatively weak compared to professional demand. There is therefore a strong argument in favour of government activity to support the development of such a market.

If such an approach is taken, we would recommend, as far as possible, trying to amplify existing consumer demand rather than substitute for it. So, for example, subsidising public facing information services based on the extent to which they successfully engage the public (measure e.g. in phone calls or web visits) would be preferable to block grants.

We would also recommend that the government put in place a system to do this which enables them to engage equally with the private sector and the voluntary sector. The voluntary sector is often given preference because it is seen to have greater legitimacy in speaking on behalf of users of public services. While this approach is reasonable, it should be recognised at the same time that many private sector organisations are more skilled in communicating with certain groups of people.

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

It would be helpful to explicitly recognise that businesses providing services to the public sector or directly to the public have a right to request data from central government which they need to support these activities.

We would like to see a legal framework which ensured that:

- a) All organisations holding data were empowered to release it to commercial organisations under appropriate conditions
- b) That commercial organisations have a right to request data in formats not published and that the presumption is that such request will be met
- c) That any limits or conditions to meeting such requests are public and explicit
- d) That requests are met in a timely fashion
- e) That where the provision of data in formats not published results in additional costs, a charge may be applied but that this charge will not exceed the marginal cost of meeting that request

1.2 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?

Yes.

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

Further work is required to establish the framework for release of data in anonymised event-level formats where there is a potential risk of re-identification .

1.4 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?

If the public sector organisations are themselves tasked with meeting the myriad different requirements of external data users, the resource implications will be very large.

If instead, holders of data support the development of an analytical industry, it would very likely result in a diminution of central data processing activity and results in a more demand-led public data infrastructure.

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

No answer

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

In the main, data standards need to be set for specific data sets and data requirements. This is a key function for central government bodies. However, such data standards setting should be a transparent process and one which engages the data analytical industry.

Q2.2 Is there a role for government to establish consistent standards for collecting user experience across public services:

Yes

2.3 Should we consider a scheme for accreditation of information info-mediaries and how might this best work?

No. Prior accreditation of info-mediaries will limit the development of the industry. Stating standards to which organisations using public data are required to comply would be useful. This should include a system of recourse for those complaining that public data is being mis-used. But prior accreditation will create unnecessary burdens on industry.

3.1 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?

By placing legal obligations on the directors of such organisations.

3.2 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?

Yes. There should also be responsibility at board-level for ensuring that data is recorded accurately.

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

Yes

3.4 What sectors would benefit from having a dedicated Sector Transparency Board?

The key areas relevant to public service delivery including health, education, criminal justice and benefits.

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

This is a complex and long-term endeavour but one that is well worth supporting. It requires development of a systematic approach to cataloguing data assets by type of data, type of service, type of organisation and type of respondent. The catalogue would include frequency and dates of data collection and who is custodian of the data.

In time the catalogue would expand to include details of the data elements within each data set. Finally, each data element itself would be 'inventoried' with metadata tags.

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

The priority should be large centrally held data sets that relate to public services.

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

As described above, we are less concerned with the government publishing data than with the government allowing access to data to enable others to publish and/or support public service delivery organisation.

4.4 What data is collected ‘unnecessarily’? How should these datasets be identified? Should collection be stopped?

By being more responsive to the needs of the data analytical industry, demand for data will help identify which data sets are of most value.

4.5 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

Greater efforts should be made to enable access to ‘unpolished’ data as this will make the process of improving data quality faster.

5.1 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?

No comment.

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

The objectives set out in the consultation paper give a good framework for prioritisation. Those that will best support growth, public service improvement and productivity, choice, accountability and social growth. These datasets are the key large routine data sets held about education, health, policing and benefits provision.

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

Existing data at a more detailed level.

6.1 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?

As described above there is a legitimate role for the government to stimulate the development of services that support public use of data. Commercial models to support provision of information to the public about public services are weak compared to professional demand for data. Stimulating this market would be of benefit.