

Lorna Goodey
Information and Knowledge Management
FSA Transparency Group

Open Data Consultation
Transparency Team
Efficiency and Reform Group
Cabinet Office
1 Horse Guards Road
London SW1A 2HQ

25 October 2011

Response to “Making Open Data Real: a public consultation”

1. The Food Standards Agency (FSA) was established in April 2000 as a nonministerial UK Government Department, operating at arm’s length from Ministers and headed by a Chair and Board, who are appointed to act in the public interest. The FSA is an independent national regulator and the central competent authority (CCA) for food and feed legislation. Section 1 of the Food Standards Act 1999 sets out that the main objective of the FSA is ‘to protect public health from risks which may arise in connection with the consumption of food and otherwise to protect the interests of consumers in relation to food’. The FSA is guided by a set of core principles:

- Putting the consumer first;
- Openness and transparency;
- Science and evidence based;
- Acting independently;
- Enforcing food law fairly.

2. Since the establishment of the Food Standards Agency in 2000 the principles of transparency and openness have been at its core, particularly in support of evidence based policy decisions and the publically accessible (and webcast) meetings of the FSA Board. The Food Standards Agency’s Transparency Group, constituted of representatives from across the FSA, have provided this response to the Open Data Consultation.

3. Making Open Data Real – response to the six consultation questions in relation to science- and evidence-gathering work carried out by the FSA:

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

A 1. The definition needs to be clearer in terms of what is within scope, particularly in relation to science- and evidence-gathering work.

The definition of datasets in the glossary of key terms, does not specifically point to data from science- and evidence-gathering research reports.

Factual data, structured or unstructured are defined as:

‘In relation to public services, this data will typically have been collected as a by-product of delivery. This includes, for example, key public datasets about public services; user satisfaction data; and the performance of providers.....’

However, there is an implication that data from science- and evidence-gathering research reports would be included, as Annex 2, Public Data Principle 6 on p.56 states:

‘Public data underlying the Government’s own websites will be published in reusable form for others to use – anything published on government websites should be available as data for others to re-use.....’

We recognise that data may have significant value beyond the original research, for further analyses. In addition by making data available, it will allow for interpretations derived from the data to be tested. We are currently developing a policy on the release of underpinning data from science- and evidence-gathering projects. We support and promote the concept of making underpinning data from science- and evidence-gathering research carried out at public expense, as freely available and in an accessible format as the circumstances of each case allows. Data should be made available as fully and as promptly as possible. However, science- and evidence-gathering research covers many fields of science which give rise to different types of datasets and we are aware that there are a number of practicalities that need to be explored before all underpinning research data could be made freely accessible. Discussions within the wider scientific community on the topic of access to research data are considering some of these practical issues, and this will help the FSA develop its final policy.

Q2. Where a decision is being taken about whether to make a dataset open, what tests should be applied?

A2. Consideration should be given to how to share data at the outset of science-and evidence-gathering work. The following considerations need to be taken into account at the project planning stage:

Data should only be published within the legal constraints of non-disclosure of personal data, or data that would damage commercial interests, which are exempt

under the FoI Act. Some underpinning data would therefore need to be anonymised before it could be published. But it is important to ensure that individuals, organisations or businesses could not be identified, even if data are pieced together.

Data should not be released unless it is feasible to provide supporting information to make it interpretable.

The costs involved in preparing large or specialised data for open access should be taken into consideration. Costs may be significant and may not represent good value for money. It may be unreasonable to put additional public money into making these datasets open, and consideration would need to be given to: who should be responsible for the costs and whether there should be a cost to access the data.

For some research projects, for example co-funded work, or where data used are generated from another project, the Government body may not own the underpinning data and may not have the right to publish it.

There may be some areas of science where it would be difficult to get software that is usable by all, or it may be necessary to develop better software tools so that the data can be more easily managed or accessed. If datasets require specific IT applications to access them, it may only be possible for data to be made available on request.

Experiments in some science disciplines, for example, high throughput metabolomics/proteomics experiments, generate large amounts of data which can only be analysed or processed efficiently using computer software programs. In these cases, it is worth considering if it is necessary for all the raw data to be fully processed to make simple tables that can be put into public archives, or as is often the case if only partially processed data are analysed to find regions of interest in chromatograms, which can be examined manually.

Informed consent is required to release human data. Where possible, consent should also take into account any future uses of data, such as the sharing, preservation and long-term use of research data. Otherwise this could lead to potential problems if participants decide after publication that they want data removed. However, participants in a trial, have the right to change the permission, so even where permission was initially granted for data to be made available, a final published dataset may be incomplete.

Q3. 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?

A3. This is an area that will need further exploration for scientific data. It is difficult to anticipate if open access to all science- and evidence-gathering data from FSA-funded work would be useful, other than for interpretations derived from the data to be tested. Where high costs are anticipated to allow open access to the data, the possibility of the requestor being asked to pay to access the data would need to be considered on a case-by-case basis. Even when accessibility may equate to added

value for data, there may be situations where the costs of making datasets available may be so high, that the cost would not represent good use of public money.

It may be reasonable to charge for data where costs are incurred to anonymise large datasets or where specialised computer software programs are necessary for specialist areas of research.

Not all of the FSA's research data are suitable for storage on our open access repository Foodbase, either because of the size of the datasets or due to the specialist nature of the data. Commercial data storage facilities may be required where there may be a fee for data storage or access. In some cases it is preferable for data to be stored within a facility that allows data of a similar nature (such as data to assess foodborne disease burden) to be stored in one place. This has the advantage that those who wish to access the data only have to go to one place rather than needing to use numerous fragmented sources, but there may be a cost involved.

If at the outset of research work it is considered that data may be commercially exploitable, the ownership of the data should be considered when drawing up the standard terms for the contract. Otherwise this may be a situation where the cost could/should be borne by those who wish to access the data.

Q5. What would be appropriate mechanisms to encourage or ensure publication of data by public service providers?

A5. It is acknowledged in the consultation document that there are practicalities that need to be considered before finalising the principles. We support the point made in section 8.7, where it states: It is only once it is easier to publish data on the internet than it is to store it in local files, or on paper, that the public sector can be expected to adopt a more open model.

Although we agree it would help if there was a requirement for all public bodies and providers of public service to proactively publish data that underpins the work carried out with public funding, there are some practicalities that need to be resolved, as outlined in answer to Q2 around the publication of some data for science- and evidence-gathering work.

We therefore agree in the case of some specialist research data, greater access is likely to depend on investment in infrastructure to make access possible at the scale required, e.g. IT platforms which can provide quicker and easier access to data for the user.

Open access to data should be considered at the outset of a project and research contracts should be required to include data management plans. Where there are legitimate reasons why information cannot be released, these exceptions should be agreed at a senior level.

Further questions: The consultation document also lays out a number of other questions with further specific questions (shown in *italics* below).

An enhanced right to data: how do we establish stronger rights for individuals, businesses and other actors to obtain, use and re-use data from public service providers?

- 1. How would we establish a stronger presumption in favour of publication than that which 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?*
- 3. Are existing safeguards to protect personal data and privacy measures adequate to regulate the Open Data agenda?*
- 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?*
- 5. How will we ensure that Open Data standards are embedded in new ICT contracts?*

We support the idea that underpinning data should be made as fully and promptly as the circumstances of each case permits. But until the practicalities around publication of science- and evidence-gathering data are resolved, it would be difficult to embed the principle that data should be open by default in existing legislation.

Setting transparency standards: what would standards that enforce this right to data among public service providers look like?

- 1. What is the best way to achieve compliance on high and common standards to allow usability and interoperability?*
- 2. Is there a role for government to establish consistent standards for collecting user experience across public services?*
- 3. Should we consider a scheme for accreditation of information intermediaries, and if so how might that best work?*

It would be helpful to have a Code of Practice for science- and evidence-gathering work that formalises what is required and which makes clear the minimum that can be expected on publication and quality of data, which will include compliance with the Public Data Principles.

Corporate and personal responsibility: how would public service providers be held to account for delivering open data through a clear governance and leadership framework at political, organisational and individual level?

- 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.*

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?*
3. *Would we need to have a sanctions framework to enforce a right to data?*
4. *What other sectors would benefit from having a dedicated Sector Transparency Board?*

Strengthening the role and broadening the membership of the Public Sector Transparency Board may help to address some of the specific issues that arise for publication of science- and evidence-gathering data.

Meaningful Open Data: how should we ensure collection and publication of the most useful data, through an approach that enables public service providers to understand the value of the data they hold and helps the public at large know what data is collected?

1. *How should public services make use of data inventories? What is the optimal way to develop and operate this?*
2. *How should data be prioritised for inclusion in an inventory? How is value to be established?*
3. *In what areas would you expect government to collect and publish data routinely?*
4. *What data is collected ‘unnecessarily’ ? How should these datasets be identified? Should collection be stopped?*
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?*

Externally commissioned science-and-evidence-gathering work should be subject to external peer review before publication of final reports. This provides an independent opinion and assessment of the technical work undertaken, the key findings made, the validity of findings, their significance and utility. Peer-review may be effectively achieved through a variety of means, for example, the peer review can be by experts in the field of study, and/or expert or scientific advisory committees or working parties.

When considering the long-term usability of data, attention should be given to useful and appropriate data format and software.

In order to ensure that data are meaningful, there should always be comprehensive data documentation to ensure that users understand their origin, content and purpose and any manipulations that may have taken place.

We support the development of data.gov.uk and other digital channels to support users in finding and accessing relevant high quality data and easy to use tools and applications.

Government sets the example: in what ways could we make the internal workings of government and the public sector as open as possible?

- 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?*
- 2. What factors should inform prioritisation of datasets for publication, at national, local or sector level?*
- 3. Which is more important: for government to prioritise publishing a broader set of data, or existing data at a more detailed level?*

Openness and transparency are core values of the FSA, and we are fully committed to the full publication of the results of all science- and evidence-gathering research projects as soon as possible after the final report has been accepted. A summary of the findings is published on the Agency's website and a copy of the full report on our open access repository, Foodbase, and we support the publication of the underpinning data in a suitable repository or data archive at the same time, or as soon as possible after that time. We also encourage the publication of articles in peer reviewed journals. However, sometimes researchers express a reluctance for us to publish the full project report promptly before they have been able to publish articles in peer-reviewed journals and/or there is a reluctance to make all underpinning data associated with a project freely available at the same time as the report is published on Foodbase before they have had the opportunity to get a paper(s) accepted for publication in a journal. Researchers want to be able to fully analyse the data first and to get their paper(s) accepted in peer reviewed journals, before others have the opportunity to, as their research reputation is judged by the number of papers they publish and their chances of getting their own work published may be reduced if papers have already been published by others.

Innovation with Open Data: to what extent is there a role for government to stimulate enterprise and market making in the use of 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?*

It is unclear how much data generated from the FSA's science- and evidence-gathering work could/would be exploited for other purposes, but there may be advantages in public service providers developing new collaborative ways of working with data users, including commercial users in some areas of research.

4. Broader comments on "Making Open Data Real" consultation

Further questions: The consultation document also lays out a number of other questions, which require a more general view in addition to the science- and evidence-based perspective set out above.

Meaningful Open Data: how should we ensure collection and publication of the most useful data, through an approach that enables public service providers

to understand the value of the data they hold and helps the public at large know what data is collected?

An identified corporate responsibility at board level for data in an organisation, especially where large amounts of personal and social data are being managed, would indicate that an organisation takes its data management seriously. Recognition should be given to the established roles in Government, such as the Senior Information Risk Owner (who should sit on the board already), the Departmental Records Officer, Data Protection Officer, Chief Scientist, etc. All of whom have varying responsibilities in relation to the managing, sharing and reuse of data and information. Some organisations will also have Chief Knowledge and Chief Information Officers. Clarity and streamlining of these responsibilities in organisations is becoming essential and the existing regulatory and governance approach needs to be clear and transparent. This aspect of how central Government manages its approach to data, information and knowledge needs to be reviewed and a clear framework communicated through a single authoritative voice.

The third point of the draft Public Data Principles set out in Annex 2:

Public data will be released under the same open licence which enables free re-use, including commercial re-use – all data should be under the same easy to understand licence. Data released under the Freedom of Information Act or the new Right to Data should be automatically released under that licence.

Clarification of this principle needs to be made as public bodies do collect/hold, in the course of public service delivery, a considerable amount of information in which copyright is vested in third parties e.g. journal or newspaper articles, which might form part of an FOI response. A public body can provide one copy of an article under the personal study exemption, however if a requester wants to make commercial re-use of any non-Crown copyright information, they will have to contact the copyright owner for permission. In other words, where ownership of copyright is vested in third parties, it is not in a Government Department's gift to release such information for commercial re-use.

5. I hope that you find these comments helpful. If there is anything you would like to explore further with officials feel free to contact me and I will ensure that your team can make contact with the appropriate respondent.

Lorna Goodey
Information and Knowledge Management
Chair: Transparency Group
