**Code of Practice FOIA E-mail Responses for FOI 317672**

1. **Campaign for Freedom of Information**

<http://www.cfoi.org.uk/>

**2) Depart for Communities and Local Government (DCLG)**

Attached as pdf.

1. **Department for Environment, Food and Rural Affairs (DEFRA)**

Attached as word doc

**4) East Cambridgeshire District Council**

Attached as word doc

**5) Essex County Council**

Attached as a word doc

**6) Food and Environment Research Agency (FERA)**

Attached as a word doc

**7) Ipsos Mori**

Apologies if this submission is too late, but please find below some feedback to the Consultation - Code of Practice (Datasets).  (Sent by email as my attempts to submit online triggered the spam filter – below I have pasted the relevant part of the code in normal text, followed by my comments in italics).

II, 6: The third criterion, that the dataset “remains presented in a way that (except for the purpose of forming part of the collection) has not been organised, adapted or otherwise materially altered since it was obtained or recorded” again is intended to ensure only ‘raw’ or ‘source’ data is captured within the meaning. Again the purpose here is to exclude from the definition any data which has been manipulated, interrogated or has had any ‘value’ added or ‘expertise’ applied. Accordingly any changes or alterations made to the presentation of a dataset as originally compiled and recorded are relevant in considering whether the dataset is outside the scope of the definition. The key consideration here is whether the reorganisation or adaptation represents a ‘material alteration’ to the original presentation of the dataset. Minor, insignificant or insubstantial changes to a dataset will not take it outside the definition. Decisions will have to be made on a case by case basis by assessing what has been done to the raw data in the dataset.

*Clarification might be needed here to show what level of data manipulation is expected, or permissible, for survey datasets. For instance, I would consider data coding and the calculation of derived variables as meaning that 'expertise' has been applied, but I would think it very unusual if there was an expectation that these things should be excluded. Also, where a report has been published using the dataset, derived variables may have been created in the process, and data users may want to recreate the analyses, or run similar analyses - without these derived variables their chances of successfully doing so will likely be hampered, especially where the raw dataset is highly complex.*

III, 1: When releasing any dataset under the Act public authorities must, as far as reasonably practicable, provide it in an electronic form which is capable of re-use, i.e. a re-usable format. A re-usable format is one that is machine readable, such as Comma-separated Value (CSV) format.

*Some additional guidance could be useful to make clear whether formats such as SPSS or SAS files would count under this definition. It is often inappropriate for survey datasets to be issued in CSV or Excel given the formatting of the variable names and labels, and given that they dataset may contain weighting variables which cannot be applied in these formats. I note that in section VIII, 3 reference is make to specialist software. It might be worth adding that, wherever possible, software which is free to download and use should be included – for instance, PSPP is a free alternative to SPSS and includes sufficient functionality for many analyses.*

VIII, 2: As with FOI requests for any information, applicants should provide a sufficient description of the dataset sought for the public authority to be able to identify and locate it. Public authorities should provide assistance to the applicant that will help him or her describe it more clearly. This assistance could be in the form of published inventories of datasets held by the public authority, with details of their availability, and also provision of ad hoc help.

*It is worth noting that the provision of ad hoc help could become extremely burdensome: the line between providing sufficient factual information pertaining to a dataset to allow for analysis, and providing training and advice on how to run analyses, could easily become blurred. For instance, what should a local authority do if a user has conducted their own analyses but would like the local authority to check that they have not made any errors or misinterpreted anything? Might contractors be expected to carry out this work?*

**8) Land Registry (charged Licence Only response)**

Land Registry currently use the OGL for free data releases and has recently reviewed and simplified its standard licence terms and conditions used for a range of commercial products. These revised terms and conditions were approved in our re-accreditation under IFTS last year.

We have considered whether the terms of the charged licence meet the needs of the licensor or re-user in broad terms given the potential range of products to which it could have application.

There are a number of terms within our standard terms and conditions which we regard as important to understanding the arrangement between the licensor and re-user. These may be due to the nature of the data licensed or the intended re-use or product. For example, clarity around any warranty provided in terms of accuracy of the data, uninterrupted access to the data/service or that the data is fit for the re-users purpose. We note there is provision for some tailoring of the wording in the schedule, which could be used to cover some of the above points such as liability/warranty clauses. Where appropriate, our own clauses could be used in the schedule, or perhaps the wording from the non commercial charged licence. eg

* *The Information is licensed ‘as is’ and the Information Provider excludes all representations, warranties, obligations and liabilities in relation to the Information to the maximum extent permitted by law.*
* *The Information Provider is not liable for any errors or omissions in the Information and shall not be liable for any loss, injury or damage of any kind caused by its use. The Information Provider does not guarantee the continued supply of the Information.*

The licence is designed for use with a range of data and for datasets in particular, and parties must be mindful of privacy issues with the release and use of data and bulk data. Provision for termination in the event that the ICO or other competent authority advises that the data should not be so released would provide an important safeguard.

The charged licence grants an inclusive range of re-use rights including combining with other data or products. The schedule provides for details of the product or intended re-use, but further clarity of whether re-users should come back to the licensor if they wish to develop another or different product would assist. This would appear to be an appropriate place to set out details of the data format and delivery details.

It may also be helpful to remind them re-users that there are circumstances in which the licensor is obligated to disclose information under FOIA.

**9) London Borough of Bexley**

1. Is the code helpful and does it provide clarity around the circumstances under which data should be published?

The Code is not particularly helpful

* It is not clear how it aligns with the CLG code on transparency for local authorities (which we assume will shortly become mandatory) or the INSPIRE regulations for publishing location data related to the environment (at 4\* level on the Berners-Lee scale).
* The existence of two mandatory codes related to ‘transparency’ and the further statutory requirements under INSPIRE regulations is likely to lead to confusion particularly in relation to how they will be regulated.
* The two codes currently use slightly different terminology and definitions to explain ‘Public data’ or ‘dataset’. Neither definition is completely clear although the ‘FoIA dataset code’ definition is more robust and exempts official statistics (as defined under the Statistics and Registration Service Act 2007) from the code.
* In **Section 40** it refers to an expectation that public authorities should publish inventories of all the data they hold with details of availability; this is also echoed in the CLG code which states that LAs **should** build and maintain such an inventory. Its not clear if this is a requirement or an expectation of good practice, either way it would require a full scale data audit with detailed work undertaken to examine anonymisation and potential mosaic effects. This is significantly resource intensive at a time when LAs are having to implement unprecedented cuts.
* The code assumes a sophisticated level of technical understanding of Open Data that does not exist in most (if not all) local authorities. The Local Public Data Panel’s report into the ‘Open Election Data project’ (see <http://data.gov.uk/blog/publishing-local-open-data-important-lessons-open-election-data-project>) made clear some of the capacity issues local authorities face on this agenda. These skills and capacity issues remain.

1. **Is it clear about:**
2. **The definition of factual information (section 8 (b) ff ) and which data can be or ought to be released?**

* No. The definition of a dataset is not clear to all. It would be helpful if the code came with explanatory notes that provided examples of
  + What is and what is not a dataset?
  + Links to HMG Datasets in all levels of the 5\* ranking system for officers to examine
  + Links to a list of all official statistics that are exempt under the code so that public authorities could easily reference this when responding to FOI requests
* The code is not clear about the implications for data held by contractors/partners, under existing contracts there are likely to be resource implications for providing this data in the formats specified by the code.
* The code is not clear about data the council holds on behalf of partners – for example we hold schools/health data, should a request for data be referred to ‘data owner’ or is the council obliged to respond as it holds the information on its systems

1. **How to publish datasets in reusable formats? (Sections 11-13)**
2. **The use of standards used to release datasets? (Sections 14-17)**

* No. It assumes a level of technical understanding and expertise (i.e. machine readable/CSV formats/Berners-Lee 5\* ranking etc) that currently does not exist across the majority of local authority staff (including system owners and data leads) and creates the need for an additional (unfunded) training requirement.
* Most officers are not familiar with the Berners-Lee ranking scale and would need clear guidance and instruction on how to convert data into this format – the information provided at the website noted in the code (<http://www.w3.org/DesignIssues/LinkedData.html>) is in technical jargon and will not provide assistance to officers in understanding the different formats
* The Code states (Section 16) that data should be published with a sufficient amount of metadata and contextual information on how and why the data was compiled – this is ambiguous, it would be helpful if a best practice example was included in the explanatory notes.

1. **Permissions for datasets to be reused and the application of licences for re-use (Sections 18-25)**

This section is not clear. It would be helpful if explanatory notes could include examples of datasets that are covered by all forms of licence in the UK Government licensing Framework (i.e. Open Government/Non Commercial Open Government/Licences where charges apply). In particular examples of where government currently charges for data/information, why and how.

1. **Do you have any other comments about the dataset code of practice?**

* **Pg 3, Section 8 (b) (ii)** – in the definition of a dataset, official statistics are exempted – clarity is needed on what is an official statistic and where requesters can be directed in order to examine a full inventory of official statistics and thus avoid unnecessary requests
* **Pg 5, Section 13** – refers to guidance on the provision of datasets issued by the Information Commissioner – it would be helpful if there was a link here or in the explanatory notes as despite searching the ICO website no such guidance has been found.
* Safeguards from misuse of data – The code or other related information for the potential ‘end user’ of the data contains no guidance on how or what this information should be used for. This remains a concern. Local authorities have significant amounts of sensitive data that if released could be incorrectly manipulated and published to create community tensions. Alongside increasing expert views that genuine anonymisation of data remains an illusion (due to the mosaic effect) we would have concerns about releasing some of our more sensitive data without proper safeguards.
* Alongside a general assumption of a reasonably sophisticated understanding of Open Data formats the code also assumes that those applying it are well versed in the concepts of data anonymisation, data pseudoanonymisation and formal data redaction. This is not the case.

**10) Local Government Association**

<http://www.local.gov.uk/web/guest/local-transparency/-/journal_content/56/10171/3841473/ARTICLE-TEMPLATE>.

**11) Merton Council**

- 1.  Definition  (ii Scope)

I think the definition of datasets is not clear enough to lead to consistent publication of data.  The definition being: "a collection of data in electronic form where all or most of the data has been obtained or recorded ...for any function... and is factual information (not the product of analysis or interpretation, apart from calculation, not an official statistic, and has not been materially altered since it was obtained or recorded)".

 This is a very broad definition and could cover much of the electronic information held by the council.  Many records contain factual information, as well as analysis and interpretation, would these be classified as datasets?

 Because publication of datasets (and updating the published version) will be seen as extra work by councils there could be some resistance and the complexity of the definition provides an easy way for officers to try to avoid classifying their records as datasets.  This could lead to far more cases going to the ICO and far more additional extra work for all involved.

 Consistent publication and the problems associated with it the definition could be addressed if the Code contained a list of the main datasets that authorities are required to publish (when requested).  This would ensure publication matches local priorities but also take the burden of determining what is a dataset away from Local Councils and so ensure consistent publication (when there is interest) across councils.

 The Code of Practice would provide a great deal of assistance if it were to publish a list of datasets.    This clarity would be very welcome.

 -2.  Licensing (v Giving permission for datasets to be re-used)

Copyright/ Intellectual Property Rights can be complex and Local authorities can not afford to obtain legal advice before answering every request.  Further guidance on what licenses should be applied to what datasets (ideally a list of datasets and their corresponding license) would assist authorise in complying with this requirement.

The Code should offer more clarification on the position of major copyright holders (as many of these will be the same for many councils), such as OS and large software suppliers (who may claim copyright over the format of datasets created by their data systems), e.g. Northgate.

Has the extent of copyright restrictions been considered, as this could be considerable?

 -3. Publication (vii Considering publication as part of the publication scheme)

Under Paragraph 37 the code states that authorities must publish the dataset unless it would not be appropriate to do so.  The Code suggests publication might not be appropriate if the data is 'unlikely to be of general interest' - this could be considered to require a subjective judgement and so provides scope for alternative interpretations (leading to a lack of publication of some datasets by some councils).  It could also be difficult for councils to judge correctly what will be of general interest in some cases.  This could lead too little or too much data being publish (leading to council websites becoming clogged with uninteresting data).

 The code also states that councils must update the published data - unless it is not appropriate to do so (but there is no guidance on what constitutes appropriate in this case - so again, given the cost of updating the published version there is scope for some councils not to act on this).

 The code suggests that assistance to requestors can be provided by an inventory of datasets - we do not hold such an inventory.  It would be helpful if the code could include a model inventory for local councils.  This would assist them immensely.

 We are liable to be sued for inaccuracies in some of our data sets (such as for inaccuracies in our local land charges register).  Therefore we don't publish this, because we verify the result of every search (or free inspection) of it by checking the results with other records.  If we are required to publish this - how can we protect ourselves from being sued if this record is found to be inaccurate?  (This is the search carried out before properties are purchased so councils could be sued millions of pounds if they are required to publish these datasets and they are inaccurate).

 Where datasets contain personal or confidential (or free text fields, where confidential data could inserted) - these fields will need to be redacted before publication and this will then necessitate the maintenance of 2 versions of the data set, increasing the chances of error.

 Most current FOI requests are not for datasets they are for answers to particular questions which we answer by interrogation of our records (many of which are data sets). Following such a request would the entire dataset then need to be published?

**12) Ministry of Defence (MOD)**

**1. Para 5.** There is no special provision for incomplete datasets and it would need to be provided if a request is made under FOI. This will result in potentially inaccurate data being presented. We would not publish incomplete datasets on data.gov.uk and don’t believe this is appropriate. A special provision should be considered.

**2. Para 9 & 10**. Makes reference to subsection 5a and 5b when this should read 8a and 8b respectively.

**3. Para 12.** Converting a dataset into an open format (e.g. csv) should not take a huge amount of time. However, this paragraph suggests if it is too burdensome to convert the dataset into a re-useable format, the dataset does not need to be converted before release. This could potentially provide a ‘get out of jail’ card as the definition of burdensome could be subjective.

**4. Section VI - Costs and Fees**. It is not clear, if a fee were to be chargeable for a licence (if issued), how recovery of the fee would be assured and who would assume responsibility for its collection? Also, the amount to be potentially charged by different Government Departments should be considered.

**5. Para 16.** Would we be expected to provide metadata and contextual information about how the data is created? This is not presently undertaken and we currently spend time redacting metadata to avoid any personal information being present in a published document.

**6. Para 20 and 21:** Under the amendment to the FOIA at 11A(1)(c) the requirement is that “the public authority is the only owner of the relevant copyright work,” and therefore if 3rd party rights exist the dataset does not fall within the scope of the Act.  However the Code at para 20 and 21 puts obligations on a public body that the legislation specifically states is excluded from the scope.

**7. Para 33-34 Publication Scheme**. Given the recent changes to the MOD website moving to a Cross Government website, the publication scheme process is being reviewed and updated. The inclusion of datasets as part of the publication scheme (linking to data.gov.uk) can be considered in future as a proactive way of presenting datasets of interest following an FOI request. However, we don’t envisage an increase in the number of datasets published given the low number presently published in line with PM commitments and the MOD ODS.

**13) Open Data Institute (ODI)**

[**http://www.theodi.org/consultation-response/code-practice-datasets**](http://www.theodi.org/consultation-response/code-practice-datasets)

**14) Open Data User Group (ODUG)**

<http://data.gov.uk/blog/odug-response-to-consultation-on-the-code-of-practice>

**15) The Archives and Records Association**

General comment

The Archives and Records Association is the principal professional body for archivists, archive conservators and records managers in the United Kingdom and Ireland. Effective archive and records management have an important role to play in assisting with compliance with freedom of information legislation and a number of our members are responsible for freedom of information compliance within their organisations. The Association welcomes the opportunity to comment on the draft s 45 Code of Practice for datasets.

Sections v to viii

The Association found the guidance on reuse, fees, the publication scheme and the duty to provide advice and assistance clear and easy to use. These sections will provide valuable guidance for information professionals seeking to implement the legislation in their organisations.

Paragraph 16

This paragraph sets out the need to provide sufficient metadata and contextual information to render the dataset intelligible to users. The preservation of context and metadata are core principles of the archive and records management profession. Without this information it is impossible for users to make informed decisions as to the value and reliability of the data. Paragraphs 14 and 15 of the draft highlight external standards with which organisations should seek to comply.

Given the importance of context and metadata for the interpretation of data, it would be helpful if the Code could adopt a similar approach in paragraph 16, and recommend external standards for the provision of this information.

Paragraph 36

Data is often most valuable when it is available over a time series, allowing the identification of trends and cycles. This paragraph recognises the long term research value of data subject to The National Archives' website archiving programme, and puts in place measures to ensure this information is preserved. However, not all public authorities are subject to this programme. It would be helpful if the Code could set out what other organisations should do to ensure the long term preservation and accessibility of their data

**16) The Chartered Institute of Library & Information Professionals (CILIP)**

CILIP[[1]](#endnote-1)[i] (The Chartered Institute of Library & Information Professionals) thanks you for the invitation to comment on the draft Code of Practice (datasets). In our submission we make the following key points:

* The need for professional information management at all stages of the information life-cycle if the objectives of the Open Data programme are to be achieved
* The importance of data quality and the need to reference within the Code such official guidance as the Lord Chancellor’s Code of Practice on the Management of Records and the Information Principles developed by the Cabinet Office
* The importance of public service providers moving away from proprietary formats to open source software to enable greater accessibility of data for reading and manipulation
* The focus the Code should have on an audience including the newer types of public service providers, many of which will not have access to inhouse information management expertise
* The importance of third party agencies, notably public libraries, in helping the public navigate these new sources of information effectively

**General Points**

We welcome this opportunity and express our continuing support and commitment for the general thrust of the Government’s Open Data Policy. This and associated Government initiatives are vital to promoting transparency in the work of public bodies and it is also clearly right that the wealth of information created and held by such organisations should be fully exploited  for the benefit of the nation.

Much of the Code is focused on the technology and the need for open standards and common approaches based on the Public Data Principles and the 5 Star Open Dataplan as developed by Sir Tim Berners-Lee. We support this move to open standards and systems but note that some so-called open data is in proprietary formats which, whilst commonly encountered, increase software licensing costs for voluntary and other organizations and individuals with limited funds**.**

However our main focus is on data quality, the indexing and retrieval of datasets and the support available for the various user communities. We are also acutely aware of the Open Public Services agenda which is likely to see a large number of smaller service providers enter the market without ready access to the information management expertise available in larger organisations. The Code will be especially important for these smaller providers.

Therefore our specific comments are on section iii (disclosure in re-usable electronic formats), section iv (Standards), section vii (publication of datasets) and section viii (advice and assistance to persons requesting datasets).

**Section iii – Disclosing datasets in an electronic format which is capable of re-use**

The proposals in this section are generally sound and we agree with their inclusion.  But reduced funding in the public and voluntary sectors greatly weakens the business case to purchase commercial software for data manipulation.

To improve this situation so far as possible, Section III.1 should specifically encourage the release of datasets in open rather than in proprietary formats, i.e. the default should be to publish data that can be read and manipulated in open source software.  Further, data should whenever possible be released in a format that readily supports data analysis, e.g. CSV rather than PDF.  We do acknowledge that publishing in formats such as  CSV format may take time for some organizations to achieve, but it must  be the direction of travel.

Experience suggests that the quality of public sector data can be inconsistent, e.g. in the formatting of address data.  Data must be made available that complies with published standards – although this is made more difficult where obvious standards (such as the Postcode Address File) are commercially owned outside the public sector.

 It is important that datasets are not released without the redaction of sensitive information they may contain.  Given that a growing number of small organisations is likely to manage and release datasets in the near future, the guidance should include a specificreminder that datasets must be managed and released in compliance with the requirements of law (such as the Data Protection Act 1998).

**Section iv - Standards applicable to public authorities in connection with the disclosure of a dataset**

The standards specified in this section of the Code are the Public Data Principles, the Five Star ranking system as developed by Sir Tim Berners-Lee, and the Government Principles for Open Standards.  These focus predominantly on technology and format issues. There is little on data quality which was acknowledged in the original Open Data consultation as a particular problem and challenge. It should go without question that the re-use of suspect data will not be to the benefit of any party.  It is notable that datasets regarded as “official statistics” have to be assessed against a quality threshold and it seems strange that other official datasets should not be similarly assessed.

We therefore recommend that the issue of data quality is given greater emphasis in the section. One way of achieving this would be to specifically reference the Lord Chancellor’s Code of Practice on the Management of Records, issued under the Section 46 of the Freedom of Information Act.

<http://www.nationalarchives.gov.uk/information-management/projects-and-work/records-management-code.htm>

The best way of ensuring quality datasets (and other information records) is to ensure that there is professional information management at all stages of the information lifecycle – as the Foreword to the Code of Practice on Records Management recognises: “Good records and information management benefits those requesting information because it provides some assurance that the information provided will be complete and reliable”.

Similarly we recommend that the Information Principles are also cited as they have been endorsed by the Government’s Knowledge Council and put both the Public Data Principles and the Records Management Code into their proper context. They were developed by the Cabinet Office as part of the ICT Strategy. One of the Information Principles is on “Fitness for Purpose” and so directly addresses quality issues.

<http://www.nationalarchives.gov.uk/information-management/projects-and-work/information-principles.htm>

We are pleased that metadata, one aspect of quality data, is mentioned in the Code (paragraph 16) and that the provision of relevant metadata is one of the Public Data Principles. However given the crucial role of metadata in describing and identifying data we recommend that a minimum set of Metadata for datasets is prescribed and relevant agencies such as the National Archives, Knowledge Council, Statistics Authority and ICO develop a technical note on good practice relating to metadata and datasets. We also note that issues of confidentiality and privacy are covered in paragraph 4 of the code with reference to the ICO’s Code of Practice on Anonymisation.

**Section vii - Publication of Datasets as part of a Publication Scheme**

This is a key part of the Code as it directly impacts on information accessibility and the level and quality of support and assistance possible for the public. It is also another public demonstration of the importance and benefits of a professional approach to information management. A publication scheme requires the identification of information assets and should be an end-product of compiling a comprehensive Information Assets Register. This former government initiative needs to be given a new lease of life and shared more widely in the public sector. We welcome Principle 13 of the Public Data Principles that specifically refers to the publishing of inventories of data holdings and also the similar Information Principle: “Information is published”.

The description of datasets (and other information sources), their indexing in inventories, publication schemes and catalogues and the need to cross-reference entries for similar or complementary datasets provided by other public authorities is not only highly necessary to make Open Data a reality, but also a highly skilled and complex task. The Government needs to take this fully into account when developing its services on Data.Gov.Uk and Direct.Gov.

**Section viii - Advice and assistance to persons making or intending to make requests for datasets**

Our points on data quality, metadata, information assets registers and publication schemes are relevant here. Accessibility and support for persons requesting information is made possible by the professional management of information across its life-cycle.

The obligation to profile the nature of the dataset, its limitations, weaknesses and pitfalls needs to be highlighted more in the Code. We note here that the eighth principle in the Code of Practice for Official Statistics is of “Frankness and Accessibility” – “official statistics, accompanied by full and frank commentary, should be readily accessible to all users”. The first bullet under this Principle is to provide commentary on the quality and reliability of statistics. <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

The Code on datasets should include similar provisions. As well as being an important part of good information provision it also accords with the values of public service as expressed in the Seven Principles of Public Life.

<http://www.public-standards.gov.uk/About/The_7_Principles.html>

Whilst it is important that public authorities are aware and act on their responsibility to support individuals in their requests for information it will require a more positive engagement with the public to promote such use and give them the confidence to understand and use such data. The Government itself has a direct responsibility in ensuring the online help and support provided by Direct.Gov and Data.Gov.UK  are of a higher order. They should also fund and support such third party agencies as public libraries to help the public navigate these new sources of information effectively. The Assisted Digital programme of the Government is aimed at achieving a digital by default delivery of public services by supporting the less confident in the use of online technology. There is a need for joined-up Government policy across all the digital inclusion initiatives currently underway to optimise their impact.

<http://publications.cabinetoffice.gov.uk/digital/assisted/>

**17) The Health and Social Care Information Centre**

The Act conflates Open Data policy with statutory re-use obligations and creates a new set of FOIA obligations that will be challenging to fulfil. The code of practice helps in clarifying some of these boundary issues but none the less the management of data assets under the triumvirate of FOIA/PoF, Open Data and Re-Use Statute will become even more complex.

Specific comments:

* There needs to be more precision in the language around ‘publication’ to distinguish between data which are on a website or otherwise publicly  available and data which would be released only with an accompanying licence.
* Page 1, Code of Practice, 2nd and 3rd paragraphs – references to April 2012.  Should this be April 2013?
* Page 2, Introduction, 1st paragraph – last sentence makes reference to ‘the Code of Practice’.  Consider adding FOIA section 45 to aid understanding. Consider re-wording/clarifying last sentence.
* Page 4, Para 12 it is not clear whether work undertaken to improve data quality would mean that the resultant data would be captured by the act or not and if not this would mean that the incorrect / invalidated data would be captured by the act which would not appear to be a sensible approach. Further clarification would be helpful.
* Page 5, Standards applicable to Public Authorities’ in connection with the disclosure of a dataset, Para 17 – relates to standards when procuring new systems rather than the disclosure of a dataset.  Consider moving to a more relevant section.
* Page 6, Para 20 –Final sentence states that even if the Public Authority is not able to licence re-use i.e. because it does not own copyright, it must still provide the dataset in a re-usable format if the applicant has asked for an electronic version. This would not be the case if the data is held by the Public Authority on behalf of another which may well be the case where the other body retains copyright. This point would benefit from clarification.
* Page 9, Para 37 – states that it may not be appropriate to publish a dataset ‘if there is unlikely to be a general interest in the dataset’.  I’m not sure how helpful this statement is. Will this be a subjective judgement by the Public Authority? How should this be established?

**18) The Health and Social Care Information Centre (charged Licence Response)**

As the Health and Social Care Information Centre (HSCIC)  does not currently impose re-use  charges / fees there is no immediate impact for the HSCIC. Should the HSCIC move to re-impose re-use fees then the draft agreement could be quickly adopted

However we note that whilst the Freedom of Information Act, via s102 of the Protection of Freedoms Act 2012, retains the presumption that the public body  may make a charge for the commercial re-use of data, the basis on which such charges may be established remains unclear.

Specific comments on the draft licence are:

* Page 2, Term – Why does the licence continue beyond the specified term (unless terminated by either party)
* The period of the licence is not defined.  In some circumstances this will be for an instance of a dataset at a point in time for indefinite use (see also comment below).
* Re section 8.  It would be reasonable for a user to request and obtain a copy of a dataset at that point in time, use it to produce a product, and not then update the product each time the dataset was updated provided the product made clear that that was the case [i.e. that the underlying information was correct at [date] but had not been updated since].  To threaten withdrawal of information (which may wreck a product) or enforce the payment of a fee for update would be unreasonable in many circumstances.
* Section 10.  Should there be a right to require the return or destruction of the information at termination of the licence (see also however comment relating to indefinite periods)?

**·19) Transport for London (TFL)**

Firstly, we welcome the code. TfL have been making datasets available for some time and we have seen the public benefit in getting the data out there to developers. We have set up a Developers Area on the TfL website <http://www.tfl.gov.uk/businessandpartners/syndication/default.aspx> and this gives us some experience of publishing live feed data.

It is unclear from the Code whether live API data would be covered, but if it is we would urge that user registration be allowed. Our experience of releasing live tube information via the Trackernet feed was that customer demand of 10 million hits in 2 weeks compromised our internal systems. User registration is necessary for the following reasons:

1. If we do not have registration we it will limit our ability to consult with our data consumers. Our release of Bus Countdown feeds has been largely informed by the developers who use the data
2. Likewise, if there are changes in licensing, or if there is likely to be disruption to data supply it is important that we are able to communicate with those who link to the datafeeds
3. User registration allows us to manage and respond to demand. Without this control we would be unable to maintain a stable service
4. User registration is a barrier to denial of service attacks
5. User control allows us to ensure that there is fair usage of data, and to restrict access to users whose use of data impacts on other users.

I have just two other points. The Open Data Licence allows users not to include attribution statements where data is combined from multiple sources. This could be replaced with a requirement that the statements are provided on a linked page.

Finally, a lot of data is likely to include a geographic element which will cause further issues for those wrestling with derived data from Ordnance Survey. Freedom to syndicate this data, or the development of a fourth licence covering OS derived data would be a great help!

**20) University of Oxford**

Attached as word doc

**21 )Universities UK (UUK)**

This response takes into consideration the collective views of the higher education sector, although we have also encouraged individual institutions to provide input and outline any specific issues they might have. We would like to highlight the following points:

1. Although not directly related to this guidance, Universities UK is extremely pleased that the government has committed to introduce an exemption to the Freedom Information Act for on-going research. This is in response to concerns that the publication of incomplete data and research results before the peer review process has taken place, and before the research in question can be rigorously assessed by experts within the scientific community, could undermine the reputation of individual research groups, institutions and the UK as a whole. Although the exemption has not yet been introduced it will be important that there is clear guidance covering the interface between the new regulation controlling access to data sets and the new exemption for pre-publication research.  It will also be important to understand what interim or temporary protections may be put in place to cover pre-publication research data during the gap between April 2013 and whenever the new research exemption is to be enacted.
2. The assumption in the guidance is clearly in favour of providing open, free licences. We would therefore welcome further clarification on what constitutes the ‘exceptional circumstances’ under which a Charged Licence can be granted.
3. We would welcome further clarification on the following areas regarding costs:

* what is likely to be deemed ‘reasonably practical’ in terms of the cost limit for making data available in a reusable format?
* how available resource will be calculated. Higher education institutions may be at an unfair disadvantage if ‘the size and the resources of the public authority’ have to be taken into consideration when deciding on reasonableness?

1. We would welcome clarification on how generally accessible the electronic re-use requirement has to be. The legislation and guidance only seems to imply a duty to make it available in machine readable format. It is quite possible that research data will be created by specialist research equipment and held in electronic format that is perfectly reusable, but only for those with access to the specialist equipment and software that created it. We would welcome clarification on the requirement on institution here. We believe that further research sector specific guidance may be required. Although the development of specific guidance may not fit with the timeframe for this consultation, it could be something that is considered by and developed through the new Research Sector Transparency Board.
2. We would welcome further clarification on the following, regarding fees:

* if and when the Secretary of State will make regulations about re-use fees
* in their absence, what guidance there will be for calculating what is deemed to constitute as ‘reasonable return on investment’ and in particular whether this include potential earnings, as well as expenditure to date

1. It would be helpful to provide information on how this guidance will interact with research funders’ requirements with regard to the availability and use of data.  Many funders now require data to be made available on open access terms in any case, so it would be helpful if funders’ requirements and guidance in this area could be aligned. Again, this may be something that could be included in specific guidance for the research sector.
2. We welcome the positive dialogue we have experienced with government and the ICO on issues relating to transparency and freedom of information and would be happy to discuss any of these points further.

**22)Wakefield Council**

Would you please clarify something for me? When reading section (V) Giving permission for datasets to be re-used, and Section (V1) Costs and Fees, the above code seems to be moving from purely access to information to the addition of re-use as it covers licensing and charging. I have no problem understanding the wording, but my problem is trying to put it all in context with the proposed changes to the Re-use of Public Sector information Regulations (Amending Directive 2003/98/EC on re-use of public sector information) which also stipulate the application of licensing and charging. The way I read the two documents, there appears to be some conflict?

**· 23) West Yorkshire Police**

 The draft code of practice is on the whole very informative, easy to understand and references relevant and important further information.

 Practitioners may benefit from further clarification/detail, either in the body of the Code or via links to trusted, credible information sources, in relation to:

 (ii) 14 - 'Relevant Copyright work'

(iii) 11 - Machine Readable Format - i.e. further examples of open source formats that would be acceptable

(iv) 16 - The requirements around 'sufficient amount of metadata and contextual information'

(vi) 31 - 'Exceptional cases' in which a fee can be charged when a public authority has copyright of the information.

(vii) 37 - The lack of requirement to publish 'if there is unlikely to be general interest on the dataset'

 It would also be useful to be of any anticipated review of the first version of the code post commencement order.

1. [↑](#endnote-ref-1)