**Comments on the Website**

**GENERAL COMMENTS**

1. We have three suggestions regarding this draft code of practice:1. The guidance could benefit from a "plain-englishing" exercise to make it easier to understand.2. Currently there is only one reference to "exemptions" - it would be useful if the guidance had a more thorough explanation of exemptions including the approach to datasets which contain personal data.3. The guidance would benefit from illustrative examples.
2. Most data is less than 100pct accurate or completely up to date. It might be useful to have some practices about how the accuracy of data should be communicated.
3. The Association of Chief Police Officers is generally content with the content of the Code of Practice. Although we believe it may be necessary to provide more information as to the definition of a data set. We are encouraged by the fact that at present the Code does not require the creation of data sets nor up dated data sets if they would not otherwise have been produced as part of the Public Authorities function.
4. Summary: Stronger language required for data not covered by the Right To Data Stronger guidance required on the ability for an individual to opt-out of administrative data release Minor amendments needed to meet the new ICO Anonymised Data Code of Practice (released since the consultation began)
   1. This consultation on the Right To Data Code of Practice is welcomed by Privacy International.
   2. The Right to Data should become an important feature of the UK transparency landscape - it may be as important to the social fabric of the country and more important to the economic fabric as Freedom of Information. However it is not without risks.
   3. Some data will relate to "anonymised" individuals and it is this data that we are particularly concerned with in regards to this submission. We have no comment on data not derived directly from records about an individual. The best way to avoid a privacy disaster with data is not to hold it at all. However a complete absence of data is fundamentally useless for policy and decision-making. As such the Right to Data is necessary to support the bridging of this gap.ICO Anonymised Data Code of Practice Privacy must be protected in any data releases. We refer to the ICO guidance on anonymisation here. For ongoing specific data releases a suitable privacy audit should take place as recommended by the ICO. The recommendations made in the ICO Data Anonymisation Code of Practice should be implemented here. We recognise that the Code was not published when this consultation was launched. Not all data can be made available publicly to all. Some data is available for bona fide academic research projects on a request-only basis. Where data is not freely released but is available on a request basis the requests made for such data must proactively and automatically be made public together with information about whether the request was accepted or refused. Data not covered by the Regulations. While the agenda is clear on what data should be included it should also contain a reference to data that is fundamentally excluded as inappropriate for the intent or coverage of such regulations. Opt-out. There should be stronger guidance clarifying the ability and mechanism for an individual to opt out.
   4. The Government has done good work on the Midata agenda led by BIS and with Digital by Default led by the Cabinet Office. Both of these are critical initiatives which allow the Right to Data policy to go further in support of privacy. Midata following Digital by Default principles requires the ability to download your own records and should be accompanied by the ability to opt out. Digital By Default makes this process simple from the consumer point of view.Data released under the Right to Data should have a right of opt-out where it is derived from an individual's records. Missing Data is a known and "solved" problem at the generic level in statistical research literature (e.g. http://www.statisticalhorizons.com/wp-content/uploads/2012/01/Milsap-Allison.pdf). The missing data created by opt-out is therefore not a statistical issue. Where a single case being absent causes a statistical problem we suggest that such work is potentially within the classes of use that the ICO Anonymisation Code of Practice considers as de-anonymised data and as such the organisation doing the work must have acquired informed consent and be registered as a data controller. For anonymised data this is impractical. As a policy issue it is possible for data re-use to prompt significant levels of opt-out. However that reuse can in no way be considered to have informed consent - on being informed consent is removed. The correct policy solution is to limit use that data subjects view as abusive rather than to make opt-out difficult. Digital by Default may make new and subtle solutions possible that were heretofore considered impractical. Pairing the agendas of Midata and Digital by Default allows for individuals to make an informed choice on whether their data is reused.
5. Forest Heath and St Edmundsbury Councils are broadly content with the draft code although note that there will be costs involved in the practicalities of converting data to CSV format. The code appears to cover this through paras 29 and 38 which allow LAs to take account of costs in fees charged and decisions about reasonableness. So if these provisions are included in the final code the system should be manageable and help to increase local transparency.
6. I am generally not content with the Code of Practice (datasets) since I disagree with the underlying approach. I think it is a fundamental error to pursue licensing for re-use of data through the FOIA request process and to conflate access with re-use. I view this approach as a tactic designed to undermine the more robust rights regime articulated in the PSI Regulations. That said I would like to endorse in its entirety the consultation response submitted by the Open Data Institute. In particular I strongly support the ODI's recommendation that the Code should be used to re-enforce open data licensing (and the Open Government Licence) as the default approach to re-use of public sector information. Following are some additional comments and suggestions: Public authorities may receive requests for re-use of information where access is subject to the Environmental Information Regulations (EIR) rather than FOIA. The 'right to data' amendments in the Protection of Freedom Act do not explicitly apply to EIR requests. However in most other respects public authorities are accustomed to treating FOI and EIR requests in a similar manner. It would be sensible to state in the Code that the 'right to data' should be applied to EIR requests also at least as a matter of informal best practice. The Code of Practice should note as a reminder to the public authority that it may also receive requests for re-use of information that are not attached to an access request. The requestor may wish to re-use data that is already accessible just not licensed. The Code of Practice should reinforce the primacy of the access request over any additional considerations relating to re-use of the data. In particular the Code should remind public authorities of their responsibility to meet the statutory timescales for providing access to the data (subject to any exemptions). Even if that means the data cannot immediately be provided in a re-usable format and/or a decision on licensing has not yet been made. Lack of clarity in the Code of Practice on what constitutes a re-usable format is likely to create difficulties particularly as this may affect costs attached to the request. For example although Excel files are not universally considered to be an open format it would be onerous for public authorities to establish charges simply for converting Excel files to another format such as CSV.
7. For example spatial data. At the moment it is deemed acceptable to publish this as PDF maps where it would be much more valuable to GIS community and others to release it as vector shape files or Google kml files that can be reused by others. There is a wealth of this information in Local government planning departments and large projects such as The Olympics.
8. More detailed practical guidance needs to be provided alongside the code of practice to provide a more practical definition of datasets. This could be developed alongside the guidance provided by the Information Commissioners Office.
9. I agree with Tim that the definition of re-usable needs more work. The statement that "a re-usable format is one that is machine readable" is rather dogmatic and not a universally held view. That is certainly the way IT developers tend to think about re-use of data and the statement is more likely to be true for large datasets. However many datasets are re-used directly for analysis (i.e. just read and considered) and does not necessary need to be pre-structured. The Code of Practice should encourage discussion based on the specifics of individual datasets. A single dataset may be suitable for a range of different re-uses and there will not always be an ideal format. When in doubt the public authority should try to strike a balance between making the data re-usable and preserving the data in its original context. (The above comment applies to III 1.)
10. Does this even need to be in the Code? TBL's Five Stars are a useful mechanism for thinking about openness as it applies to data but I'm not convinced they add anything to this particular process. It's important that the public authority makes clear the legal terms of re-use but it's not the public authority's role to coach the applicant on what they can do with the data beyond that. (The above comment applies to IV 2.)
11. Well to the extent that this is reasonably available perhaps. We need to make sure public authorities don't feel they need to write new documentation for every data release particular as some will want to recharge for the time involved.(The above comment applies to IV 3.)
12. The references to third party intellectual property rights are difficult to reconcile with section 11A(1)(c) of the FOIA 2000 which excludes from the ambit of section 11A datasets other than those where the public authority is the only owner of the copyright. The scope of this Code consequently risks exceeding its statutory basis addressing more than just the application of section 11A.
13. In providing an open government license the Council will not be benefit financially from providing the information. It is recommended that the licence provided is one which the Council will benefit from providing the information. Further the additional requirements proposed in the code will place a greater burden on the Councils resources to make information available at a time when the Council is facing further significant cuts to its budget. The recently announced Local Government Finance Settlement will see the Councils budget reduce by approximately £37 million over the next 3 years and this is on top of significant cuts already made to Council funding over the previous 3 financial years.
14. I would suggest removing the phrase "proprietary format". I think it is being used here as equivalent to a format that does not support re-use. However there are some file formats that remain technically proprietary but are generally considered to support re-use because the IP owners have undertaken not to assert rights that would limit their wide use. See for example the Microsoft Open Specification Promise which applies to the .xls file format. (The above comment applies to VI 4.)
15. On the assumption that the Information Commissioner will not be taking over all of the responsibilities of the OPSI. the Code should also require public authorities to make applicants aware of the availability of the OPSI complaints process in respect of PSI re-use -- particularly where the public authority has charged for re-use of data.(The above comment applies to IX 1).

**COMMENTS UNDER INTRODUCTION SECTION**

1. INTRO: 1: Wouldn't it be better to have a code of practice based on a consensus of relevant stakeholders rather than the opinion of a Secretary of State.
2. INTRO: 3:"Neither the Act nor this Code require the creation of datasets for publication “As set out in the ICO's new code of practice on anonymisation certain datasets may require anonymisation before publication to ensure that personal data is protected appropriately. Would this be regarded as creating a new dataset?
3. INTRO: 4:"In deciding whether to release a dataset a public authority should consider any exemptions which may apply and in particular consider the exemption in section 40 of the Act relating to personal data and the Information Commissioners Code of Practice on Anonymisation."It may also be worth noting that there could be a risk of anonymising the 'wrong' field. A dataset may be anonymised and then proactively released. If subsequently a request is made for the same dataset with the previously anonymised field disclosed then this may not be possible due to the risk of re-identification based on the comparison with the original dataset. Hence the benefits of proactively disclosing some datasets may be reduced unless there is detailed consideration of the purposes to which they may be put.

**COMMENTS UNDER SCOPE SECTION**

1. SCOPE: 2: Point (b) (iii) - It is unclear (to me anyway) what this is getting at. Data is necessarily organised. Datasets often consist of many data items that are collected over a period of time and to publish them may require them to be anonymised filtered categorised or checked for accuracy.
2. SCOPE: 2: I agree with John Crab. The intent of ensuring raw data is good but the definition appears to exclude valid alterations like anonymisation and accuracy-checking.
3. SCOPE: 2: The definition of a dataset should have an 'and' or 'or' after the word authority. Paragraphs 9 and 10 suggest it should be 'and'.
4. SCOPE:6: There are many cases where a requester might have a valid reason to request a non-raw dataset but to want access to this as machine readable data. For example if an authority has taken a raw dataset of primary data on parking then applied a process to aggregate this to generate spreadsheets on traffic or parking hot-spots and this has then formed the basis of policy a requester may be interested in the analysis spreadsheet as well as or instead of the raw data. These requesters appear to have no right to data under the current wording. Is there a valid justification for this? The desire to tightly define raw data should not lead to 'cooked' or prepared data only being available in PDF or print-out formats! I realise the Act is worded with this restrictive definition that centres on raw data - and that unless my reading is wrong it does not give a right to the processed data in machine readable form (even though this processed data could still be subject to the Freedom of Information Act). However could the guidance highlight that whilst raw data should be provided when requested and the tight definition is necessary to ensure this the definition should not exclude authorities from providing processed data products in machine readable forms if this is what the requestor wants.
5. SCOPE: 6: An example of what would be considered a minor or insubstantial change would be useful here. Also it may be worth clarifying that a request or remains entitled to make a request under the FOIA for the data or information that underlies a dataset (they may be interested for research purposes for example) even if that dataset has been republished in useable format and public authorities should deal with those requests in the usual way under the Act.

**COMMENTS UNDER DISCLOSING SECTION**

1. DISCLOSING: 1: Perhaps it would be more useful to pure text-based format such as CSV because it's probably not unknown for someone to convert a CSV to a PDF and send that out!
2. DISCLOSING: 1: The definition of re-usable needs to be expanded and clarified here. Points might include that a re-usable data is one for which A requester can open the data file in freely available and widely accessible software packages- this means that formats such as CSV should be preferred over or offered in addition to formats like Excel and proprietary formats which can only be opened with commercial or specialist software should be avoided. It is possible to process the data directly carrying out any appropriate operations on it such as sorting columns filtering rows running aggregates of values - this requires well structured data. Where possible the meaning of the data should not be contained in the layout. Common elements in the dataset are expressed in uniform ways- for example dates are always in the same format codes or names are always in the same case and numbers are expressed consistently (e.g. 1 000 or 1000 but not a mixture of the two). The meaning of fields and values is clearly documented - either through clear naming of fields or through accompanying descriptions provided along with the data. In addition where possible machine readability is enhanced if: The dataset uses common standards where they exist - including standard identifiers and standard field names. These might be standards like the public spending vocabulary developed for government or third-party standards such as KML for indicating 'points of interest'. It is important to also note that it may be possible to provide data in a variety of machine readable formats and where possible the authority should correspond with the requester to identify the best format. For example points of interest could be provided in a CSV spreadsheet form or KML. Ideally both would be provided: though depending on the context and the re-user one may be more appropriate than the other.
3. DISCLOSING: 1: To back up these points the following blog posts might be useful: Open Data takes more than CSV When Machine Readable Data Still Causes Issues“Wrangling Dates¦ I'm sure there are many others on the same theme...
4. DISCLOSING: 1: I agree with Tim Davies that this needs a fuller definition. It would also benefit from a wider range of examples. Not all datasets can sensibly be produced in CSV - collections of documents or photographs for instance could meet the definition of a dataset. And re-usability like openness is a range rather than a dichotomy - for a table PDF is more re-usable than JPEG but less re-usable than Excel or CSV.
5. DISCLOSING: 1: For example Spatial data. At the moment it is deemed acceptable to publish this as PDF maps where it would be much more valuable to GIS community and others to release it as vector shape files or Google kml files that can be reused by others. There is a wealth of this information in Local government planning departments and large projects such as The Olympics.
6. DISCLOSING 2 Worth a point that authorities are encouraged to respond to requests by proactively engaging with the requester and understanding how they might move towards providing more structured and regularly released open data in future?
7. DISCLSOSING 3 It may be useful to explicitly draw attention here to data protection guidance, and [the recently released Anonymisation Code of Practice](http://www.ico.gov.uk/news/latest_news/2012/new-anonymisation-code-sets-out-how-to-manage-privacy-risks-and-maintain-transparency-20112012.aspx).

In addition, there is a specific risk of personal data leaking out the tracked changes or hidden fields of machine-readable files that authorities previously published only in PDF or print forms - and so in either this guidance, or the ICO guidance, it may be neccessary to address this privacy issue explicitly.

It could be addressed here, or in the section on the costs of preparing a dataset for machine readable distribution, to note that this should include an audit of any potential privacy implications of releasing the data in the format of choice.

**COMMENTS UNDER GIVING PERMISSION FOR DATASETS TO BE REUSED SECTION**

1. GIVING: 3: "if the applicant has asked for an electronic version." should be removed IMO
2. GIVING: 3: Notwithstanding our general comments below relating to extent that this Code should be addressing datasets that fall outside the scope of section 11A it is worth mentioning that where the dataset comprises plans drawings or maps the public authority is mandated to use a particular formula of wording in order to indicate that a third party enjoys reserved rights in the data. Details can be found in the Copyright (Material Open to Public Inspection) (Marking of Copies of Plans and Drawings) Order 1990 and The Copyright (Material Open to Public Inspection) (Marking of Copies of Maps) Order 1989.
3. GIVING: 3: We would recommend adding some explanatory text to the following sentence let it be misinterpreted as meaning that the public authority must positively facilitate the unauthorised use of third party intellectual property! (The purpose of this provision is of course to facilitate the re-use of the information once the required third party permission has been obtained).Even when the public authority is not able to license re-use it must still provide the dataset in a re-usable format.; This leads us on to a more serious point: the Code might well put a duty on public authorities to supply datasets in electronic format: however the user will not necessarily be able to lawfully view the data in that electronic file (let alone put the data to use) if it contains third party material. This is because by viewing data that has been recorded in electronic format the information is temporarily stored in the memory of the user's computer which is an act restricted by copyright.
4. GIVING:4: Beyond our general comments (see below) relating to extent that this Code should be addressing datasets that fall outside the scope of section 11A we are concerned about the wording of paragraph 4. The provisions relating to "obtaining authority" could be read as meaning that public authorities should have a process in place for obtaining authority to license a dataset if so requested. To do so goes far beyond any obligation hitherto placed on public authorities and risks being extremely burdensome. Ordinarily the extent to which an authority has the power to permit the re-use of third party information will have been agreed contractually when the dataset was first commissioned: re-users cannot expect public authorities to go back to the data owners in order to renegotiate such agreements
5. GIVING: 5: We believe that this section of the Code should be clearer in explaining that the provisions relating to reuse and charging do not apply to Crown-owned copyright works or Crown-owned database rights per section 11A(8) FOIA 2000.
6. GIVING: 5: This wording "the public authority must make that work available for re-use in accordance with the terms of one of the specified licences in this Code" could be unduely onerous on public authorities who hitherto have enjoyed a wider freedom to draft terms and conditions tailored to the nature of the data that they are licensing. For example when licensing a dataset the public authority would generally want a custom set of indemnity and limitation clauses included in the licence contract in order to precisely define the extent of its responsibilities and liabilities should there be a problem with the data.

**COMMENTS UNDER COSTS AND FEES SECTION**

1. COSTS: 2: Doesn't this suggest a fee would be chargeable for asking the authority to host it on their own website? If so that seems odd. Perhaps it's an acknowledgement that some authorities will be charged by their IT provider? If that's the case the cost should be absorbed by the authority - the requester should not be punished because the authority failed to choose an IT provider that doesn't have an archaic business model that charges for simple website modifications. There is a greater benefit in showing up such ridiculous charging regimes by forcing the authority to suffer for agreeing to it. Tough love you might say.
2. COSTS: 2: Some authorities already have a disclosure log up and running. Is this section suggesting that those authorities are now going to have to start charging to put information up on their websites? This seems strange seeing as we have been doing it for several years now and not taking any payment.
3. COSTS: 2: "a fee may be charged""may be" means it's an option not a requirement
4. COSTS: 4: Even when an authority judges that a dataset cannot cost effectively be provided they should still be asked to apply an appropriate license to the non-machine readable release that they provide. This will ensure third parties are not restricted in then turning the non-machine readable product into a machine readable product following its publication
5. COSTS: 4: The guidance may also want to encourage authorities to actively engage with re-users in working out the appropriate format for the data and managing the costs associated with providing machine readable data: for example consultants may charge a much higher price for something a civic hacktivist would be willing to do for no cost; or parties requesting data may be willing to provide advice and support to the authority on extracting data in an appropriate way. The guidance may not be the only place to do this - but encouraging authorities to explore collaborative rather than antagonistic relationships with data requesters where possible would be good.
6. COSTS: 4: The guidance could explicitly ask agencies to where refusing to provide a machine readable dataset on grounds of cost to provide details of the hourly or daily rate of the contractor and anticipated time that has led to the cost exceeding the threshold. This is information that if the request was under the threshold may be published through the public spending open data requirements or that may be FOIable independently and so there should be no reason for this not to be explicitly included in any refusal response to a request for machine readable data.
7. COSTS: 4: When refusing on the grounds of cost the costs should be stated and there could be the option for the requester to fund either the excess above the threshold or the full cost.
8. COSTS: 4: What happens to the data set when it is released to person A?; I haven't seen anything that tells you what to do next. I would have thought that organisations should be then publicising the data on their own website data.gov.uk etc so that other people are aware the data set exists and is in the public domain to be used - This in my opinion would be good practice but we all know good practice isn't always implemented. I'm also assuming some sort of retent and disposal schedules would need to be applied to the data sets. This can probably be ignored now as the next part of the consultation covers this comment.

**COMMENTS UNDER CONSIDERING PUBLICATION OF DATASETS AS PART OF A PUBLICATION SCHEME SECTION**

1. CONSIDERATION: 1: I think this should go further than encouragement
2. CONSIDERATION: 3: "If a dataset has been requested from a public authority under the Act then the authority must publish that dataset in accordance with its publication scheme unless the authority is satisfied that it would not be appropriate to publish it. It may not be appropriate to publish it if for example the dataset is exempt from disclosure under one of the exemptions in the Act or if there is unlikely to be general interest in the dataset. If the authority holds an updated version of the dataset it must also publish that updated version unless it is satisfied that it is not appropriate to do so."Should they be encouraged to publish to data.gov.uk as well/instead of their own site?
3. CONSIDERATION: 4: Should this apply to county records offices too?
4. CONSIDERATION: 6: when the authority... Is there a need for clarity that when a dataset is published in non-machine readable format through a publication scheme a requester is within their rights to ask for this in a machine readable format and the prior presence of this document in non-machine format in the publication scheme is not a reason to refuse or close the request? Should authorities be encouraged to provide means for users to discuss the formats they would like to see currently non-machine readable data in to help the authority set priorities for reasonably practicable future publication of datasets?

**COMMENTS UNDER ADVICE AND ASSISTANCE SECTION**

1. ADVICE: 2: If an applicant requests a dataset which doesn't exist as such but could be derived from datasets which are available how much help is the authority required to give? Should they deny the request because the dataset requested doesn't exist say that dataset doesn't exist but suggest other existing datasets which might be relevant identify specifically which existing datasets could be used to derive the one requested explain how to derive the one requested from the existing datasets derive the requested dataset and provide it to the applicant?
2. ADVICE:3: A need for specialist software may be unavoidable in a few cases but shouldn't the guidance recommend avoiding it wherever possible? Usually it will be possible to extract the data into a non-proprietary format and this is more open. Also advising people where to find software that can read a particular specialist format without unnecessarily endorsing a particular product might require knowledge of the software market that authorities won't always have.