

The Campaign for Freedom of Information

Suite 102, 16 Baldwins Gardens
London EC1N 7RJ
Tel: 020 7831 7477
Web: www.cfoi.org.uk
Email: admin@cfoi.demon.co.uk
Twitter: @CampaignFoI



Draft Code of Practice on Datasets Response to a Consultation by the Cabinet Office

Campaign for Freedom of Information

January 2013

1. The relationship between this code and the existing s.45 code

The proposed code is to be issued under section 45 of the Freedom of Information Act.¹ However, the draft itself does not explain how the proposed code will relate to the existing section 45 code issued in November 2004.² For example, it is not made clear whether the draft code is intended to be a second freestanding code which, in relation to datasets, *replaces* the existing code or whether it is to be seen as an extension of the existing code.

Paragraph 4 of the draft code states:

“Public authorities should handle a request for a dataset in a way that meets their obligations under the Act and that conforms to *the relevant Code of Practice issued under section 45.*” (emphasis added)

This may be taken to imply that requests for datasets should be dealt with only under the proposed code and that the provisions of the November 2004 code do not apply. There seems no reason why this should be the case and we doubt whether the new code, which does not cover all the matters which section 45(2) requires the code to address, is capable of standing on its own. This should be clarified.

2. Good practice in relation to non-datasets

If the draft code is seen as part of, or an extension to, the existing code it raises the question of why certain issues are addressed solely in relation to *datasets* when the same issue arises in relation to *any* FOI disclosures. Some requests may relate to a collection of information which is similar to a dataset but not technically one. We would expect the code - whose statutory purpose is to promote good practice - to encourage such requests to be treated in the same way, unless some specific harm would result.

This is particularly relevant to the use of the Open Government License (OGL). The draft code only deals with the use of the OGL in relation to datasets although it is relevant to *any* disclosure under the Act. However, its use in relation to non-dataset information is not dealt with anywhere in the November 2004 code or the new code. If there is to be section 45 guidance on the use of the OGL it should encourage its use in relation to *any* FOI disclosure not just datasets.

Similarly, the new code should address requests which ask for non-dataset information to be supplied in reusable format and encourage a positive response where this has been

¹ The draft code can be found at: <http://data.gov.uk/library/draft-code-of-practice-datasets>

² Some explanation is provided in the preamble to the online consultation but none appears in the draft code itself. <http://data.gov.uk/consultation>

expressly asked for and is reasonably practicable. Authorities should not be able to respond to such requests by suggesting that the section 45 code does not expect this.

In light of the government's 'Open Data' policy, which is intended to promote a substantial expansion in the release of such information, it would be remarkable if the s.45 code did not address these wider contexts.

We note that the proposed section 11B regulations, on charging for the reuse of information, may apply to *any* disclosed information, not just datasets. This reinforces our view that the draft code's scope should not rigidly limit itself to datasets.

3. Use of the Open Government License

The draft code encourages authorities to release datasets under the Open Government License suggesting that other licenses will need to be used only in "exceptional cases" (paragraph 31). There is little reason to be confident that licenses involving charges will be "exceptional" unless the regulations to be made under section 11B generally exclude them.

The code should place more forceful emphasis on the need to adopt the Open Government License as the default basis for responding to FOI requests. Inappropriate copyright restrictions are still imposed even on non-dataset information which authorities can have no expectation of ever licensing. The code should strongly discourage this practice.

Recent examples of inappropriate copyright warnings include:

- A local authority was asked for the number of prosecutions or cautions for unlawful waste disposal that it had been responsible for under the Environmental Protection Act 1990. It replied "*we have undertaken 16 prosecutions and given 9 cautions*". This brief answer was accompanied by this statement: "*Any information subject to copyright will continue to be protected by the Copyright Designs and Patents Act (1998). This includes information which is copyright of the council. Disclosure of any information by the council to you does not provide you with any rights to use or distribute the information in breach of any copyright.*"³
- Hastings Borough Council were asked how many visitors to the town were attracted by the annual Hastings Chess Congress and what economic benefits the Congress brought. It replied that this information was "not held". It continued: "*Please note: This information is copyrighted to Hastings Borough Council and is supplied for your personal use only. Except as permitted under the Copyright, Designs and Patents Act 1988, the information supplied may not be copied, distributed, published, or*

³ www.whatdotheyknow.com/request/fly_tipping_prosecutions_and_cau

exploited for commercial purposes or financial gain without the explicit written consent of Hastings Borough Council.”⁴

- The Metropolitan Police were asked what audiovisual materials were available to people detained in the terrorist holding cells at Paddington Green Police Station. The answer included examples of the material available to detainees (including David Attenborough programmes, a compilation of Premiership Football Goals and Star Wars, Jaws, Indiana Jones and other films). This answer was accompanied by the statement that *“In complying with their statutory duty under sections 1 and 11 of the Freedom of Information Act 2000 to release the enclosed information, the Metropolitan Police Service will not breach the Copyright, Designs and Patents Act 1988. However, the rights of the copyright owner of the enclosed information will continue to be protected by law. Applications for the copyright owner’s written permission to reproduce any part of the attached information should be addressed to MPS Directorate of Legal Services, 1st Floor (Victoria Block), New Scotland Yard, Victoria, London, SW1H 0BG.”⁵*

Paragraph 21 of the draft code suggests that authorities should consider how best to streamline the process of ascertaining whether any third party has copyright in a dataset and obtaining authority to license the dataset. There should be particular emphasis on seeking third party consent to the release of information under the Open Government License.

4. Definition of a dataset

The definition of a “dataset” in section 11(5) of the FOI Act is in our view obscure and the draft code’s attempts to explain it could go considerably further. In particular, the draft does not focus as precisely as it might on the injury which the restrictive definition of dataset is designed to avoid.

We would expect the s.45 code to say that where information does not fall within the definition of a dataset *but it is clear that no harm would be done by responding to the request as if it were a dataset*, that should be done.

5. Analysis or interpretation

Section 11(5)(b)(i) of the Act excludes from the definition of a dataset any information *“which is the product of analysis or interpretation other than calculation”*. Paragraph 10 of page 3 of the draft suggests that the intention *“is to catch ‘raw’ or ‘source’ data”* but beyond that does not explain what harm would be caused by including the product of analysis or interpretation. Paragraph 12 of page 4 suggests that the aim of a separate element of the definition is to exclude data which has had *“‘value’ added or expertise applied”* but again without explaining what harm would result from allowing such data to be released in reusable form subject to the Open Government License.

⁴ www.whatdotheyknow.com/request/hastings_chess_congress

⁵ www.whatdotheyknow.com/request/paddington_green_police_station_2

We think a more precise explanation of the purpose of the provision is needed so authorities can better judge whether or not they are acting in the spirit of the legislation. For example, is the aim to preserve an authority's ability to generate income from data to which it has added value by the use of its own expertise? If so, the code should point out that virtually all of the information produced by an authority by analysis or interpretation will be held for policy or decision making purposes and not to generate income. Such material should therefore routinely be released under the Open Government License, whether or not it constitutes a dataset.

The code should make it clear that the hypothetical possibility of generating income from data at some future date must not prevent its release under the Open Government License at the time it is requested.

It should also make it clear that where data is accompanied by a *commentary* which contains analysis or interpretation, that commentary does not mean that the data itself ceases to be a dataset.

We welcome the development of the Non-Commercial Government License, which should allow voluntary organisations and other non-commercial requesters to obtain data without charge which an authority might otherwise release only on payment of a license fee.

6. Changes to the presentation of data

Paragraphs 12 and 13 on page 4 deal with the circumstances in which changes to the presentation of data may remove it from the definition of a "dataset". These refer to the third element of the definition of a dataset, in section 11(5)(c) of the Act, which states:

(5) In this Act "dataset" means information comprising a collection of information held in electronic form where all or most of the information in the collection...

(c) 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."

The meaning and purpose of this provision is far from clear. Paragraph 12 states that this provision:

"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."

We question this explanation. Any data which has been produced by "analysis or interpretation" is already precluded from being a dataset by section 11(5)(b)(i) of the Act.

Section 11(5)(c) would not be necessary if its purpose was merely to repeat what had already been established by the earlier provision.

Paragraph 13 on page 4 explains that a key consideration in deciding whether a collection of information may have lost its status as a dataset “*is how much, if any, of the data in the dataset has been changed or altered.*” It suggests that a dataset will remain within the definition if the changes “*to the dataset as a whole*” are minor or insubstantial or if any changes affect only “*a minority of the data within the dataset*”. The implication is that only substantial change to the data or the whole dataset could remove it from the statutory definition.

We think this paragraph errs by referring to changes to the *data* rather than *the presentation* of the data. Section 11(5)(c) states that a dataset ceases to be one if the *presentation* of the data has been materially altered since it was obtained. The fact that the *data itself* has changed does not seem capable of removing data from the definition of a dataset.

If a major change is made to the data, for example, by incorporating significant additional data, the only effect, if any, would be to create a *new* dataset. Assuming the remaining elements of the s.11(5) definition are still met both the original and the new collections would be datasets.

We also think the code should explain in what circumstances it would be possible to change the ‘presentation’ of factual raw data held electronically.

Is the provision, for example, intended to catch a change in the way in which a set of data is *published* or *viewed*, for example, by switching from the use of a bar chart to a pie chart, moving from a spreadsheet to a word processing document or from an image to an OCR’d document? These would be changes in the “presentation” which do not involve analysis or interpretation. If the intention is to catch such superficial presentational changes when they affect most of the data it would be helpful to explain its purpose - as it is not obvious that it has one. If it is not the intention, the code should make that clear.

Would a change involving the splitting of a column of data in a spreadsheet into two columns be a presentational change capable of depriving a dataset of its status as such if enough of the dataset was affected? This too would not involve the interpretation or analysis of data. If such a change is capable of removing a dataset from the statutory definition, its purpose should be explained. If it is merely a side-effect of a poorly drafted provision, the code should, as a matter of good practice, strongly discourage authorities from relying on it.

We assume that the process of merging two columns into a single column would be regarded as “the product of...calculation” and therefore could not affect the status of an existing dataset.

Is the “presentation” test intended to exclude a database which may periodically be interrogated to elicit different sets of data? This appears to be suggested in paragraph 12 on page 4 of the draft code which states that the intention is to exclude data which has been “interrogated”. If so, in what sense could the “presentation” of the *database itself* be regarded as having been changed by this process?

More fundamentally, we would question whether electronically stored raw data, which is not published, is not held in the form of a recognisable document and is not normally viewed except when interrogated according to criteria which are separately specified on each occasion, can be said to be “presented” at all.

If section 11(5)(c) is in fact redundant, and only serves to exclude data which is already outside the dataset definition because of the product of analysis or interpretation test, it would be better to acknowledge that fact rather than allow authorities to unnecessarily exclude data from the dataset provisions.

Finally, following a meeting with Lord Lucas of Crudwell and Dingwall and the Campaign in January 2012, Lord Henley, minister of state at the Home Office, stated that the proposed guidance would make clear that checks of data while compiling a dataset to ensure their integrity and security could not be regarded as ‘presentational’ changes capable of undermining the status of a dataset.⁶ This statement should be incorporated into the code.

7. Environmental information

Although the new dataset provisions of the FOI Act are not repeated in the Environmental Information Regulations (EIRs), the code should make it clear that they may apply to requests for environmental datasets. This is because the FOI Act applies to environmental information subject to the qualified exemption in section 39. That is, environmental information must be disclosed under FOI unless, in all the circumstances of the case, the public interest in withholding it under the FOI Act outweighs the public interest in its disclosure under the Act.

Where, under the EIRs, an authority is not prepared to release an environmental dataset in reusable form, or to provide it subject to a specified license, the public interest in securing its release in that manner may justify its disclosure under the FOI Act. This should be acknowledged in the code.

⁶ On 1 February 2012, Lord Henley wrote to Lord Lucas stating: “we believe that the concern discussed during our meeting, regarding the possibility of public authorities making 'presentational' changes when compiling datasets to avoid release, is also negated by the words at the beginning of the section "(except for the purpose of forming part of the collection)", which allows for checking of the data to ensure integrity and security of the information when the public authority is compiling a dataset. Accordingly, we believe this definition as currently drafted, and for which further guidance will be set out as above in the statutory Code of Practice, is fit for purpose.”

This conclusion reflects the approach of the Information Tribunal in the case of Rhondda Cynon Taff County Borough Council and the Information Commissioner (EA/2007/0065) which held that the EIRs and FOI Act were not mutually exclusive regimes,⁷ that “there is nothing in EIR or FOIA that says that an applicant must “elect” to use one regime or the other”⁸ and that “the regime in FOIA is providing a potential supplementary right of access to environmental information”.⁹

Several decisions of the Scottish Information Commissioner (SIC) also adopt this approach under the Freedom of Information (Scotland) Act 2002 which, in these respects, is identical to the UK Act. The SIC has stated: “*Where a person requests environmental information, they have dual rights of access under general rights provided by FOISA and under specific rights contained in the EIRs.*”¹⁰

8. Destruction of data

Paragraph 12 page 5 envisages that an authority may create data in a reusable format but publish it in a non-reusable form such as an image file without necessarily retaining the reusable data. In this context it may be worth reminding authorities that if they deliberately dispose of the reusable data after a request for it has been made, in order to avoid having to disclose it in reusable form, they would commit an offence under section 77 of the FOI Act.

9. For the purposes of

A minor point: paragraph 9 of the draft states that the datasets caught by the statutory definition are those produced “for the purposes” of providing services or carrying out functions. In fact, the relevant test refers to datasets produced “in connection with” such services or functions - a slightly wider term.

⁷ Paragraph 24

⁸ Paragraph 28

⁹ Paragraph 32

¹⁰ Scottish Information Commissioner, Decision 120/2008 Mr Rob Edwards of the Sunday Herald and the Scottish Ministers, paragraph 29