

The Campaign for Freedom of Information

Suite 102, 16 Baldwins Gardens
London EC1N 7RJ
Tel: 020 7831 7477
Web: www.cfoi.org.uk



**Submission to the
Protection of Freedoms Bill Committee
on clause 92 of the Bill**

*Campaign for Freedom of Information
4.4.2011*

Disclosure of datasets under the FOI Act

Further to the Campaign's oral evidence to the Protection of Freedoms Bill committee on March 24, we would like to make a number of additional point about the provisions on "datasets" in clause 92 of the Bill.

Clause 92 amends section 11 of the Freedom of Information Act 2000, by providing, amongst other things, that where an FOI request is made for a dataset in electronic form:

- it must be supplied in a form which is capable of being reused; and
- if the public authority is the copyright holder, no copyright restrictions may be imposed on its reuse other than those set out in a 'specified licence'. We understand the intention is to automatically permit reuse, subject to the modest conditions set out in the Open Government Licence or a variation of it.

The clause also amends section 19 of the FOI Act, requiring authorities to publish any requested dataset as part of their 'publication schemes' and keep it up to date, unless the authority is satisfied that this is not appropriate.

These are welcome provisions. However, we have concerns about the definition of the term 'dataset', which underpins the measures:

"(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—

- (a) has been obtained or recorded for the purpose of providing a public authority with information in connection with the provision of a service by the authority or the carrying out of any other function of the authority,
- (b) is factual information which—
 - (i) is not the product of analysis or interpretation other than calculation, and
 - (ii) is not an official statistic (within the meaning given by section 6(1) of the Statistics and Registration Service Act 2007), and
- (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."

Paragraph (c) of the definition provides that a dataset ceases to be a dataset if any change is made to the way in which the information in it is presented. On the face of it this means that even a modest change in presentation, such as the merging of two columns of data into one, or the separation of one column into two, would mean that the information ceased to be a dataset.

Many datasets will be adapted over time as circumstances change, for example, where there are changes to the physical boundaries of a monitored area, to legal definitions affecting the monitored population (eg a dataset showing benefit claims by particular groups may be adapted when the entitlement to the benefit changes) or to reflect changing policy objectives or public expectations. For example, public concern about the injuries to children in road traffic accidents may lead to an existing dataset being modified to distinguish between adults or children or between children of different ages.

Under paragraph (c) of the current definition, any such change would mean that the information ceased to be a dataset. The authority would no longer have to release it in reusable form or publish it under its publication scheme and could then impose copyright restrictions preventing FOI requesters from publishing the data without permission.

This would lead to the new dataset provisions being circumvented by relatively modest changes to the way in which data is presented. It would also permit authorities to deliberately modify a dataset in order to reduce the disclosure requirements. We assume this is not the intention, but it appears to be a consequence of the current drafting.

The Bill's Explanatory Memorandum refers to the purpose of paragraph (c) of the definition of dataset as follows:

339. New subsection (5)(c) requires that the information within datasets has not been materially altered since it was obtained or recorded. Datasets which have had 'value' added to them or which have been materially altered, for example in the form of analysis, representation or application of other expertise, would not fall within the definition for the purposes of new subsection (5).

There is an obvious mismatch between this explanation and the actual text of subsection 5(c) which is clearly not limited to changes which add value to the data (for example, by combining two separate datasets of different types of data) or changes involving the application of special expertise.

If the purpose of 5(c) is to exclude data which is the product of sophisticated analysis or other expertise, it appears to be redundant, since this is already achieved by clause 5(b)(i). This excludes from the definition of dataset, information which is "the product of analysis or interpretation other than calculation". Given this provision, we have some difficulty understanding the purpose of the further constraint set out in 5(c).

In any event, we question the purpose of section 5(b)(i) itself. Why should information which is the product of analysis or interpretation automatically become subject to restrictions on its reuse?

Suppose an authority holds a dataset on road traffic accidents, and suppose that it later adapts that dataset to show whether road calming measures have helped reduce accidents. The improved dataset might then show which accidents occurred on roads with particular types of traffic calming measures, and whether those measures had been properly implemented.

- The decision on whether a calming measure had been properly implemented would involve “analysis or interpretation” which, under paragraph 5(b)(i), would exclude the information from the definition of dataset.
- The newly added data would mean that the presentation of the information would have been “organised, adapted or otherwise materially altered” since it was originally recorded, thus excluding it under paragraph 5(c)
- The dataset would have had “value’ added...in the form of...expertise” thus falling foul of the policy objective set out in paragraph 339 of the Explanatory Memorandum.

In our view, none of those arguments should be sufficient to justify the reimposition of copyright restrictions. The dataset has been adapted to help assess the success of an existing policy. If an FOI requester obtains that data and wishes to reproduce it on their own web site, together with their own analysis of the data or critique of the policy they should be entitled to – without having to seek copyright permission from the authority or face the prospect of having to pay for a copyright license.

We suggest that the only circumstances in which the authority should be entitled to impose copyright restrictions are where the dataset is being commercially exploited by the authority or where ‘value added’ changes have been made in order to make it suitable for such exploitation. Where a dataset has been modified simply to reflect changing circumstances or to contribute to the authority’s decision-making, as in this example, these restrictions should not apply.

This example illustrates why, in our view, it is not just datasets which should be excluded from copyright and reuse restrictions but *all information* released under the FOI Act where (a) the authority is the copyright holder and (b) the information is not being commercially exploited by the authority or being processed with a reasonable prospect of being so exploited.

When publication of a dataset is “not appropriate”

Clause 92(4) of the bill inserts a new section 19(2A) into the FOI Act, requiring authorities to publish any dataset which has been requested under the Act under its publication scheme unless “the authority is satisfied that it is not appropriate for the dataset to be published”.

We are not clear why this “satisfied that it is not appropriate” test has been adopted, particularly as it involves a subjective element (that the authority is not satisfied) which will be difficult for the Information Commissioner to oversee.

The obvious circumstances in which authorities should be entitled to refuse to publish a requested dataset are where:

- There is no obligation under the Act to release the dataset because the information is exempt from disclosure (and the public interest test, where it applies, is not satisfied)
- The cost of locating, retrieving and extracting the necessary information exceeds the cost limit which applies under section 12 of the Act

If any further restriction on proactive publication is necessary, a more suitable test would be that publication “is not reasonably practicable”. This would avoid the subjective element of the proposed text and corresponds to the reasonably practicable test used elsewhere in the Act, eg in section 11(1) and in the other amendments proposed in clause 92.

This formula will still allow relevant guidance to be included in the code of practice under section 45 of the Act as envisaged by the amendments in clause 92(5) of the bill.

*Maurice Frankel
4 April 2011*