Code of Practice

Secretary of State for Justice's Code of Practice (datasets) on the discharge of public authorities' functions under Part 1 of the Freedom of Information Act

Issued under Section 45 of the Act.

April 2013

Presented in Parliament by the Secretary of State for Justice pursuant to section [xx] of the Freedom of Information Act 2000.

Foreword

1. This Code of Practice (datasets) fulfils the duty of the Secretary of State for Justice set out in section 45 of the Freedom of Information Act 2000 ("the Act") in relation to datasets.
2. Section 102 of the Protection of Freedoms Act 2012 amended section 11 (Means by which communication to be made) and section 19 (Publication schemes) of the Act and inserted new sections 11A and 11B. The practical effect of the amendments is to require a public authority releasing a dataset under that Act to do so in a way that enables the dataset to be used and re-used. This means that the public authority must, firstly, provide the dataset in a re-usable format, where reasonably practicable; and, secondly, grant a licence (in accordance with one of the specified licences in this Code) under which the dataset can be re-used.
3. In addition, Section 45 of the Act has been amended so that there will be a Code of Practice issued under that section to include provisions relating to datasets.

CODE OF PRACTICE

This document provides guidance to public authorities as to the practice which it would be desirable for them to follow in connection with the discharge of their functions under Section 11,11A, 11B and 19 of Part I of the Act in relation to datasets.The Secretary of State for Justice, having consulted with the information Commissioner, issues this Code of Practice under section 45 of the Act on April 2012. (Exact date TBC)

Laid before Parliament on April 2012 pursuant to section 45. (Exact date TBC)

Introduction

This Code of Practice provides guidance to public authorities as to the practice which it would, in the opinion of the Secretary of State for Justice, be desirable for them to follow in connection with the discharge of their functions in relation to datasets under section 11 (Means by which communication to be made), 11A (Release of datasets for re-use), 11B (Power to charge fees in relation to release of datasets for re-use) and section 19 (Publication Schemes) of Part I of the Act. It supplements the Code of Practice on the discharge of a public authority's function under Part I of the Act, which covers all information (not specifically datasets) previously issued separately.

Words and expressions used in this Code have the same meaning as the same words and expressions used in the Act.

Neither the Act nor this Code require the creation of datasets for publication, nor do they require datasets to be updated if they would not otherwise be updated as part of the public authority's function. They are intended to increase publication of updated datasets which are already accessible and to make them available for re-use and, where possible, in a re-usable format.

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. A public authority must disclose any information that has been requested under the Act, including datasets, unless an exemption applies. The

amended provisions in sections 11, 11A, 11B and 19 create additional rights for applicants and additional duties for public authorities in respect of datasets but they do not affect whether or not a dataset should be released under the Act. 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 Commissioner's Code of Practice on Anonymisation.

There is no special provision for incomplete datasets. Even if a dataset is incomplete, or in draft form, it is subject to requests under the Act in the same way as complete datasets or any other recorded information.

The guidance in the Code covers the following areas: a. Scope b. Disclosing datasets in an electronic form which is capable of re-use c. Standards applicable to public authorities in connection with the disclosure of a dataset d. Giving permission for datasets to be reused e. Costs and fees f. Considering publication of datasets as part of a publication scheme g. Advice and assistance to persons making requests for information with requests for datasets h. Complaints

Scope

This Code applies to datasets held by or on behalf of a public authority. The definition of dataset limits the application of the Code to datasets in electronic form that meet the criteria specified at section 11(5) of the Act.

The definition of dataset at section 11(5) is as follows:

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

(iii) 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 first part of the definition (subsection (5)(a)) means that the datasets caught by the Act are those datasets which a public authority has originally obtained or recorded for the purposes of providing services or carrying out its functions, including decision-taking. The purpose of releasing these datasets is to increase transparency and accountability of a public authority's decisions and functions.

The second part of the definition limits datasets for the purposes of the Act to factual information subject to the three criteria in subsection (5)(b). The intention behind the first criterion, "not the product of analysis or interpretation other than calculation" is to catch 'raw' or 'source' data. The definition makes clear that mere calculation of information within the dataset does not count as 'analysis' or 'interpretation'. Accordingly, the aggregation of data to form a high-level dataset (such as the creation of regional figures that were collected at district level, or the creation of annual figures from data that were collected weekly) would not by virtue of being aggregated take a dataset outside the definition.

The intention behind the second criterion, to exclude official statistics, is to ensure there is no potential overlap with officials statistics (which may otherwise have been caught within the definition), which are subject to their own regime on disclosure and publication, including under the Statistics and Registration Service Act 2007.

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.

The other key consideration in the definition is how much, if any, of the data in the dataset has been changed or altered. If 'all or most' of the data in the dataset meets the criteria set out in subsection 5, then the dataset will fall within the definition. It should thus not be possible for public authorities to get around the new provisions by making minor or insubstantial changes to the dataset as a whole, or making any changes to a minority of the data within the dataset. Even where a dataset has been significantly changed from the original dataset, such that the new information held no longer meets the definition of a dataset, the original dataset would still subject to the new provisions in the Act (provided of course it is still held by the authority).

Where information requested meets the definition of a dataset, the authority will be under a duty to provide the dataset in a re-usable format (see below), where reasonably practical. However, the 're-use duty' and relevant provisions relating to licensing in this Code do not apply to all datasets. The provisions that relate to licensing and the re-use of datasets apply only to datasets that are, or contain, a 'relevant copyright work'. The definition excludes datasets subject to Crown copyright and Parliamentary copyright (see section 11A (8) of the Act).

Disclosing datasets in an electronic form which is capable of re-use

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.

Datasets are, by their nature, often created in formats that are capable of re-use. Therefore, if a public authority publishes a dataset in a non-re-usable format such as an image file, it should consider also publishing the dataset in its re-usable format before its conversion to an image file, or keeping a re-usable version of the dataset available. Where datasets are only held in non-re-usable formats, and it is impractical or too burdensome to convert the dataset into a re-usable format, the public authority is not obliged to convert the dataset before releasing it.

Public authorities should consider best practice and any guidance on the provisions of datasets issued by the Information Commissioner.

Standards applicable to public authorities in connection with the disclosure of a dataset

When releasing datasets public authorities should as far as possible adhere to the Public Data Principles (http://www.data.gov.uk/library/public-data-principles [49]). These principles are compulsory for central government departments and recommended for the wider public sector. Authorities may wish to publish or signpost to these principles on their website.

When releasing a dataset, authorities should apply the Sir Tim Berners-Lee Five Star ranking system

(http://www.w3.org/DesignIssues/LinkedData.html [51]). Whilst not a mandatory requirement this will help applicants in assessing the suitability for re-use of a dataset they have received.

Published datasets should, so long as there are no good reasons for it not to be provided, be accompanied by a sufficient amount of metadata and contextual information about how and why the dataset was compiled or created in order that users may fully comprehend the dataset they are dealing with and as part of compliance with Section 16 (duty to advise and assist) of the Act.

When procuring new data processing systems, public authorities should reference the Government Principles for Open Standards in new government information technology specifications for software interoperability, data and document formats. The Principles are compulsory for central government departments, their agencies, non-departmental public bodies and any other bodies for which they are responsible for.

Giving permission for datasets to be re-used

Datasets must be released with accompanying details of any licence conditions that apply to re-use of the dataset or otherwise any limitation on re-use by virtue of third party intellectual property rights.

The public authority should ascertain whether copyright and/or database rights in the dataset are owned solely by the authority, that there is no other copyright owner with rights in the dataset. Nothing in the Act removes any rights of third parties in that dataset. If a public authority grants a licence to re-use a dataset where it is subject to intellectual property owned by a third party, the grant may constitute an infringement of the third party's rights. Consideration should also be given to the extent to which it is appropriate to disclose such information, in particular with reference to sections 41 and 43(2) of the Act.

Where the copyright or database right in a dataset is owned wholly or partly by a third party, a public authority can only give permission for re-use of the dataset if it has been authorised to do so by the relevant third party rights holder. The 're-use duty' only applies where the public authority is the only owner of the copyright or database right. If the authority does not have the legal authority to make it available for re-use it should make it clear to the applicant when releasing the dataset that this is the case. Even when the public authority is not able to license re-use, it must still provide the dataset in a re-usable format, so far as reasonably practicable, if the applicant has asked for an electronic version.

Authorities should consider how best to streamline the process of ascertaining the ownership of copyright and database rights in the dataset, and where relevant obtaining authority to license the dataset, ideally before any requests arrive. Identification and acknowledgement of third party rights is good practice and may prevent unnecessary applications and delay in responding to requests.

Licensing

If the dataset that is being provided, or any part of it, is a relevant copyright work, 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. Authorities are encouraged to use the UK Open Government Licence as most datasets can be re-used without a charge. It is envisaged that the other licences will only be used in exceptional circumstances.

UK Open Government Licence: The Open Government Licence is the main licensing model for the UK Government. Established under the UK Government Licensing Framework, the model encourages the use and re-use of a wide range of public sector information. The Open Government Licence is a non-transactional open licence which enables use and re-use with virtually no restrictions. It permits use and re-use, including for commercial purposes, at no cost to the user/re-user. The Licence can be easily used by public authorities, for example, it only requires public authorities to link to the Open Government Licence which is hosted on The National Archives website [72].

Non- Commercial Government Licence: It is recognised, however, that the Open Government Licence will not be appropriate in all cases, for example, in circumstances where information may only be used for non-commercial purposes. The Non-Commercial Government Licence was developed to meet those situations. As with the Open Government Licence, public authorities can link to the Non-Commercial Government Licence

on The National Archives website [74].

Charged Licence (Beta): If re-use for commercial purposes which involves payment of a fee and/or royalties by the re-user is required, a transactional licence may be used. The licence uses standard licensing terms and forms part of the UK Government Licensing Framework and is available on the National Archives website [76]. [Note: The Charged Licence is currently in 'beta' form and the National Archives is open to receiving feedback about the licence. Specific comments on this beta licence should be made through the National Archives website [76] although this draft Code of Practice consultation welcomes general feedback on use of this licence in the final code.]

Costs and fees

It is important to distinguish between the costs to the public authority of complying with section 1(1) for a request for a dataset, the cost of providing it in a re-usable format under section 11(1A), and the fees that can be charged to the applicant for making a dataset (where it is also a relevant copyright work) available for re-use under section 11A.

Fees Regulations for access made under sections 9, 12 and 13 of the Act apply to the right of access to datasets as to other information. The activities for which a fee may be charged (under Regulation 6) include, if such a preference is expressed by the applicant, provision of a dataset on portable media and posting the dataset on a host facility, which might be the authority's own website or through a third party provider.

Cost of complying with the request

A request for a dataset should initially be treated in the same way as a request for any information under the Act. Under section 12(1) of the Act, a public authority is not obliged to comply with section 1(1), i.e. inform the applicant whether it holds the dataset and, if so, to communicate the dataset, if it estimates that the cost of doing so would exceed the 'appropriate limit' as defined in the Fees Regulations for access. The fees the public authority can charge the requester for providing the information are set out in these Regulations.

Cost of providing the dataset in a reusable format

If the cost of complying with the request would not exceed the appropriate limit and the information is not otherwise exempt, the public authority must provide the dataset (subject to any right to charge a fee). If the requestor expresses a preference for the dataset in electronic form, the public authority must provide it in a reusable format, so far as reasonably practicable. In deciding whether it would be reasonably practicable to provide the dataset in a reusable format, the public authority can take account of all the circumstances of the case, including the costs of doing so. These costs may include, for example, the cost of converting the dataset from a proprietary format to a format such as CSV. If the cost of using an external contractor is taken into account such that it is considered not reasonably practicable to provide the dataset in a re-usable format, the need to use that contractor and the cost of doing so must be justifiable. The Act does not prescribe any cost limit for determining what is reasonably practicable. The size and the resources of the public authority may be relevant here. If the public authority concludes that, in all the circumstances of the case, it would not be reasonably practicable to provide the dataset in a reusable format, then the authority must still provide the dataset in another format (again subject to the appropriate limit and/or any right to charge a fee).

Fees for providing the information

Fees for providing a dataset, as for any other information, to an applicant may only be charged in accordance with sections 9 and 13 of the Act and the Fees Regulations for access.

Fees for allowing re-use of a relevant copyright work

In exceptional cases a public authority that is the owner of a copyright work in a dataset being released may exercise the power to charge a fee for re-use of that dataset.

Section 11B of the Act provides that the Secretary of State may make regulations about the fees for making relevant copyright works available for re-use; public authorities can charge a fee, in accordance with any regulations under section 11B, for making that copyright work available for re-use. Alternatively, if a public authority is entitled to charge a fee under any other legislation for the re-use of a relevant copyright work, then it may do so on that basis instead. [Note: Guidance will be provided here once the regulations have been drafted and approved]

Considering publication of datasets as part of a publication scheme

Public authorities are encouraged to subject existing and newly created datasets to the same tests that would apply when considering a request for release of the dataset. If the dataset would be released on request, the public authority should consider publishing it through the authority's publication scheme. The Information Commissioner's Office produces guidance on publication schemes for public authorities.

Public authorities should consider their long term plans and processes for the collection and storage of datasets, keeping in mind that they should be made easily accessible and in a re-usable format for requests or publication as part of their publication scheme as well as for normal business purposes. Maintaining inventories of datasets can contribute to better management of datasets as well as to compliance with the Act.

When publishing a dataset on their website, public authorities, should, where possible, publish it in a machine reachable format, so that the data can be directly downloaded from a given URL without further human intervention. Data should not be hidden behind a login, drop down menu or pick list.

Authorities whose websites are in scope for archiving by The National Archives should first check that the data has been captured and can be downloaded from the UK Government Web Archive before removing the data from their own websites. Following this practice will avoid relevant authorities' websites from becoming cluttered with out of date data, whist still keeping such data available.

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.

Furthermore, when the authority publishes the dataset under its publication scheme, it must (as for responding to a request) provide it in an electronic form that is capable of re-use, where it is reasonably practicable to do so. What is reasonably practicable will again depend on the circumstances in each case and include the same considerations when dealing with a request, e.g. the cost of doing so, the work involved in doing any conversion and whether any specialist equipment or software is required.

Advice and assistance to persons making or intending to make requests for datasets

Public authorities are required to provide advice and assistance to people who have made, or who propose to make, requests for datasets just as they would for any other information. This is under section 16 of the Act (a public authority's duty to provide advice and assistance to applicants and potential applicants).

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.

Where specialist software is required to access and/or re-use certain datasets, authorities should advise applicants where they can obtain such software. However, care should be taken not to unnecessarily endorse any particular manufacturer’s product.

Public authorities should make it clear if any copyright or database rights in a dataset are owned by a third party and, if so, the identity of that third party. This will alert potential re-users of the limitation on re-use and the need to contact the rights owner to seek permission to re-use the dataset.

Complaints

The redress mechanisms of the Freedom of Information Act apply for datasets just as for any other information; and the Information Commissioner’s functions under Part IV apply in the same way. The Information Commissioner will have the jurisdiction to consider any complaints in respect of datasets, including any complaint in respect of the re-use duty.