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).