

Re-use of datasets poses potentially onerous obligations

Paul Gibbons analyses new Freedom of Information Act dataset requirements for public authorities.

On 1 September, possibly significant changes to the Freedom of Information Act 2000 (FOIA) came into force. These changes require public authorities to make datasets available to requesters in re-usable form.

The changes, made via section 102 of the Protection of Freedoms Act 2012, amend sections 11 (which is the section of the FOIA that covers the format of disclosure) and 19 (which requires public authorities to have publication schemes setting out what information they make available proactively). It is important to note that these new provisions do not extend the range of information that is available through the FOIA – they do not, for instance, make any change to the coverage of exemptions in the Act. Their effect is purely to change the form in which certain information types are provided.

WHAT IS A DATASET?

Datasets are defined at the new section 11(5) of FOIA. The definition is somewhat tortuous, but certain features can be discerned:

- they must be in electronic form – so a paper document cannot be a dataset;
- they must be a collection – so a single fact is unlikely to be a dataset;
- they must be held in connection with a service or function of the public authority (the Information Commissioner indicates in his guidance that in his view most, though not necessarily all, information held by the authority will fall within this category);
- most of the content must be factual – the Commissioner interprets this to mean “quantitative rather than qualitative” information; he suggests that information in a free text field would not be “factual” as it “cannot be measured or compared in an objective way”, whilst figures or

postcodes for example would be “factual”;

- they must be “raw” data – data that has been produced as a result of analysis or interpretation is not considered a dataset for these purposes;
- they cannot be an “official statistic” as defined by the Statistics and Registration Service Act;
- they must be mostly unchanged since they were collected or created – the Commissioner indicates that this will include datasets that have had to be redacted because data in them is exempt from disclosure.

REQUIREMENTS FOR PUBLIC AUTHORITIES

The main requirement is for public authorities to provide these datasets in a way that allows them to be re-used whenever they receive a request for datasets in electronic form. There are two aspects to this.

Firstly, datasets must be disclosed in a format that allows re-use. Ministers introducing the Protection of Freedoms Bill referred to the practice by some public authorities of disclosing data in portable document format (PDF), preventing the recipients from doing anything useful with the data (unless they manually copied it out). Now authorities will have to release such datasets in a re-usable, machine-readable form such as comma separated variable (CSV) format.

Secondly, if the authority owns the copyright in the requested dataset, and it is not a Crown or Parliamentary work, the requester must be issued with a licence permitting re-use of the data. The new Secretary of State’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 FOIA) sets out three different options for this:

- the default is to make the dataset available for re-use with virtually

no restrictions under the Open Government Licence;

- a non-commercial licence is also available where an authority wants to limit re-use to study or private use;
- finally – and controversially in some circles – there is a charged licence which allows authorities to charge the requester for permission to re-use the dataset (of which more, below).

All three licences can be downloaded from the National Archives’ website.

PUBLICATION SCHEME

The changes to section 19 now require public authorities to “publish...any dataset held by the authority in relation to which a person makes a request for information...”. Note the phrase “makes a request”. The trigger for publication here is the fact that someone has asked for the information, not whether the authority has chosen to disclose it. That raises the obvious question of what should happen where an authority has refused a request for a dataset citing an exemption in the Act. We will return to this question later.

The amendments go further. Authorities are required to publish “any up-dated version...of such a dataset”. So each time a request is received for a dataset it potentially triggers a rolling programme of publication of that dataset.

These requirements – to release datasets in electronic form, to licence re-use, to publish them and to regularly update them – are potentially onerous. Thankfully for authorities, the requirements are ameliorated by some carefully worded limits on their extent.

RESTRICTIONS AND EXCEPTIONS

First of all, authorities only have to provide datasets in electronic form “so far as reasonably practicable”. What does this mean? It isn’t defined within

the legislation, but it is a phrase already used in FOIA to modify the effect of any request for information to be provided in a particular form (section 11(2)). In that section, it explains that in “determining...whether it is reasonably practicable to communicate information by particular means, the public authority may have regard to all the circumstances, including the cost of doing so.” So the cost of making a dataset available in re-usable form might well be a consideration in establishing whether it is reasonably practicable. The Information Commissioner’s guidance on these provisions states that the “time and the cost involved can be relevant factors”. So, for example, the Commissioner suggests that if a dataset is very large or held in a proprietary system, there might be significant cost involved in purchasing additional software or specialist expertise. In making this assessment, the size and resources of the authority will also be relevant. Finally, the Commissioner’s guidance indicates that technical issues might make it impracticable to convert data into an open format. Ultimately, we will have to wait for decisions from the Commissioner and tribunals before we will be sure how “reasonably practicable” should be interpreted.

Secondly, as indicated above, one of the options that public authorities have when licensing re-use is to charge for it. The Freedom of Information (Release of Datasets for Re-use)(Fees) Regulations 2013 (SI No 1977, 2013) allow public authorities to charge a fee not in excess of:

- “the cost of collection, production,

reproduction and dissemination of the relevant copyright work [i.e. the dataset]; and”

- “a reasonable return on investment.”

The latter should be estimated “on the basis of a reasonable estimate of the demand for a relevant copyright work over the appropriate accounting period”. The regulations encourage authorities to develop standard fees for re-use licences. Some have gone as far as to suggest that these regulations and the existence of a charged licence will allow public authorities to derive income from the FOIA. However, given the context of these changes – the government’s desire to promote the open data agenda – it is perhaps unlikely that we will see common use of this provision. This is acknowledged by the Information Commissioner in his guidance which states that it “is expected that in most cases public authorities will use the [Open Government Licence] to make datasets available for re-use...”.

Finally, the obligation to publish requested datasets is restricted by the words at the new section 19(2A) – “unless the authority is satisfied that it is not appropriate for the dataset to be published...”. The question of course is – what can be considered “not appropriate” in this context?

It was noted above that the Act refers to requested datasets, not just those that are disclosed. The Information Commissioner makes clear that where a request for a dataset has been refused, citing an exemption in the

FOIA, this will be a good reason for deciding that further publication is “not appropriate”. Other examples where publication might be “not appropriate” include:

- where there is no wider interest or value in publishing the dataset;
- where it would be expensive to publish the dataset;
- where there are technical issues with publishing the dataset.

Importantly, this condition will also apply to the requirement to keep the dataset up-to-date. For example, where a request has been received for a dataset that is updated monthly, it might only be appropriate to publish updated versions every six months, bearing in mind considerations such as those listed above.

CONCLUSION

This paper started by describing these changes as “possibly significant”. Certainly there is potential for the datasets provisions to have a major impact on public sector organisations that collect large amounts of data. What is not clear at present is firstly how much demand there will be for access and re-use of datasets, or secondly how the various in-built conditions will function to limit the effect of the new requirements. We will have to wait and see on both counts.

AUTHOR

Paul Gibbons is an information rights consultant who blogs about FOI on his FOI Man website at <http://www.foiman.com>