**Code of Practice (datasets)**

**Commentary on the *Code of Practice (datasets)…* for submission to the Defra network consultation – December 2012**

Fera Data Management Unit staff have separately read through the *Code of Practice (datasets) on the discharge of public authorities’ functions under Part 1 of the Freedom of Information Act* and have collated our comments and concerns below.

Foreword: *no comments*

Introduction

3. 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.

Need to be aware of this as it will need constant monitoring. Fera scientific datasets may be collected and augmented over significant periods of time in the normal course of our work – i.e. they ***would*** “*be updated as part of the public authority’s function.”* Does this mean we **would** need to re-publish these datasets? If the dataset is already published then surely this will not come under FOI?

5. 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.

This is a particular concern if a dataset is in draft or incomplete form as this will cause issues in quality, validity, context, unforeseen collection problems and the impact of the work may be damaged in the case of yet to be published papers and reports. If we release data in a draft form, would we need to inform the requester of errors found subsequently in the data once that dataset had been published in its final form?

Scope

9. 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.

We would need to understand how to deal with commercial customers’ information.

12.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.

How do you define whether something is ‘*minor, insignificant or insubstantial*?’ And who would make this decision?

13. 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).

Presumably, the fact that a dataset ‘has been significantly changed’ or ‘no longer meets the definition of a dataset’ would itself have to be made clear somewhere to avoid misleading the requester?

14.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).

Who would decides what is ‘reasonably practical’ and whether a dataset is, or contains, a ‘relevant copyright work’? We would need better to understand how licensing works in Fera?

iii Disclosing datasets....

No comments.

iv. Standards applicable to public authorities....

16. 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.

This makes the assumption that the dataset is complete! How much is ‘sufficient’? What sort of metadata? Is there going to be a template for the metadata?

v. Giving permission for datasets to be re-used

For all elements of this part of the Code, it would require ownership, copyright and IP at contract stage to be fully documented.

20.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.

We **have** to release data in a ‘re-usable format’ even if we do not own it, but **we** cannot licence it for re-use without permission from the owner. Why release in re-usable format it **we** can’t licence it for re-use?

23. **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

This could mean that someone might be able to profit from our work before the project/report has been completed. However care would be needed in choosing the correct OGL to cover this (non commercial Government Licence).

vi Cost and Fees

No comment.

31.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.

What are these exceptional cases?

32. 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]

Guidance has not yet been sorted out!

vii Considering publication of datasets....

33.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

We were unsure what this means. This appears to suggest pre-emptive action on the part of the public authority to publish datasets wherever they WOULD be released if a request for access were to be made. But is this not what we are directed to do anyway?

34.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.

This clearly implies that Fera will need a (formal) retention schedule for datasets. A message to get over to Fera colleagues!

35.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.

When we have a number of datasets how would these be displayed if we cannot have a pick list or drop down menu? In the past we have always worked on putting metadata out with contact number and name. Now we will need to put out data. This seems make FOI redundant as the dataset would already be available.

36.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.

Is our website in scope for archiving? – how do we discover this? If not in scope, what is the procedure?

37.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.

What is ‘general interest?’ How many versions are we required to publish and maintain. Do we have to keep previous versions?

viii Advice and assistance to persons...

40. 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.

This section will require Fera to release a full list of datasets completed and in progress. This may increase the number of FOI requests in itself!

42. 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.

This could lead to a loss of (commercial) confidentiality and Fera might thereby lose customers. However, might there be an exemption under FOI that would apply?