UK Statistics Authority
1 Drummond Gate
London
SW1V 2QQ

Telephone: 0845 604 1857
E-mail: authority.enquiries@statistics.gsi.gov.uk
Website: **www.statisticsauthority.gov.uk**

Chair of the UK Statistics Authority, Sir Andrew Dilnot CBE

Bernard Jenkin MP
Chair, Public Administration Select Committee
House of Commons
LONDON
SW1A 0AA

20 September 2013

Dear Bernard

STATISTICS AND OPEN DATA

Further to the Committee's call for written evidence on Statistics and Open Data, please find attached a note prepared by the National Statistician's Office and Office for National Statistics officials.

The Statistics Authority remains very supportive of initiatives to further and promote Open Data. The Committee may recall that my predecessor, Sir Michael Scholar, first set this out in correspondence with the Minister for the Cabinet Office two years ago, in October 2011.[1] The Code of Practice for Official Statistics requires that statistics should be made available in 'as much detail as is reliable and practicable, subject to legal and confidentiality constraints', constraints which protect *inter alia* against the disclosure of personal information as proscribed by legislation. Normally the Statistics Authority expects all data underlying official statistics to be published where this is practicable and does not put confidentiality at risk. Judgements about utility and reliability are, in our view, best largely left to users, as long as they are fully informed of the strengths and weaknesses of the data - another requirement of the Code of Practice.

At the same time, we continue to support your view that transparency is not always or necessarily achieved merely by the publication of data. 'Data dumping' can be inimical to transparency and good government. It is the job of official statisticians, and a requirement of the Code of Practice, to produce official statistics which are understandable and readily accessible, with objective and impartial commentary. The publication of data is not enough; the user needs sound professional advice and explanation on how to interpret the data as well.

---

[1] http://www.statisticsauthority.gov.uk/reports---correspondence/correspondence/letter-from-sir-michael-scholar-to-rt--hon--francis-maude---open-data---11102011.pdf

These themes have been reflected in many of the more than 250 Assessment reports that the Authority has published over the past four years, covering over 1,100 statistical outputs. These have contained binding requirements to improve both the accessibility of official statistics and the quality of the written commentary and advice to the user which accompanies their release. While these relate mostly to National Statistics, which we recognise is only a subset of all the data collected and held by government departments and other public bodies, we believe that such binding requirements on producers of official statistics to improve accessibility and understanding for the user has been an important contribution to opening up data and official statistics for the public good.

We hope this assists the Committee in its inquiry and please do not hesitate to contact me should you require anything further.

Yours sincerely

Sir Andrew Dilnot CBE

ANNEX: NOTE PREPARED BY THE NATIONAL STATISTICIAN'S OFFICE AND OFFICE FOR NATIONAL STATISTICS

Introduction

1. The Government Statistical Service (GSS) is committed to increasing the amount of data released in open and re-usable formats in the production of its statistical outputs. This is consistent with the UK Government's policy, as set out in the Open Data White Paper.[2]

2. Principle 8 of the Code of Practice for Official Statistics (see Annex A) states that producers of official statistics should ensure that they are disseminated in forms that enable and encourage analysis and reuse. Producers should release datasets and reference databases, supported by documentation, in formats that are convenient to users.

3. Principle 5 of the Code of Practice (see Annex A) states that official statistics should not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information. Arrangements for confidentiality protection must be sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics.

4. The Public Data Principles (see Annex B) should underpin the release of government statistics in conjunction with the Code of Practice for Official Statistics. Accordingly, these Principles underpin the Office for National Statistics (ONS) releases.

ONS and GSS context

5. ONS is committed to releasing data in a way which will allow maximum reuse of information and which does not compromise its legal obligation to protect the confidentiality of individuals. The ONS approach to Open Data is published on the Publication Hub (www.statistics.gov.uk) website as well as the ONS Publication Scheme at http://www.ons.gov.uk/ons/about-ons/what-we-do/publication-scheme/index.html

6. Each year ONS publishes nearly 500 statistical releases and over 250 analytical reports. These outputs are made available in a transparent way, according to a pre-published schedule and with associated metadata and explanatory text. Outputs are informed by user consultation to help ensure that the published statistics will provide what users need. ONS outputs are published on the ONS websites[3] (which include the Neighbourhood Statistics and the Nomis sites). The majority of these products are free of charge and are published under an open government licence so they are free of restrictions on use or redistribution. ONS also provides links to pre-announced ONS publications on www.data.gov.uk, via Rich Site Summary (RSS) feeds.

7. In addition to the scheduled products, ONS has handled around 300 Freedom of Information (FOI) requests in the last year which can be accessed via the FOI disclosure

---

[2] http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential

[3] All ONS pre-announced publications can be accessed through www.statistics.gov.uk. ONS manages Neighbourhood Statistics (http://www.neighbourhood.statistics.gov.uk/dissemination/) to meet ongoing demands for good quality small area statistics. Nomis is a service provided by ONS to give users free access to the most detailed and up-to-date UK labour market statistics from official sources and Census statistics http://www.nomisweb.co.uk/

log on the ONS website and a further 1,700 ad hoc requests for data that are listed on the website.[4]

8. ONS collects the data that it uses to produce statistics from the public, from businesses and from other government bodies. Re-use of information held by the public sector is important, but even within central government this can be challenging. However, there has been some progress on re-use of data. For example, improvements to population statistics have been made by using data from many sources, including the Schools Census and DWP's Customer Information System. ONS is a strong supporter of other public bodies following the Open Data principles and publishing their data in a form that would facilitate re-use. There is considerable potential to incorporate such data in the production and quality assurance of official statistics.

9. The GSS has been committed to driving forward the Open Data and transparency agenda. This has included assessing the openness of official statistics, setting out policy and guidance for the GSS, creating e-learning for the Civil Service and identifying a number of exemplary case studies of Open Data good practice. These examples of best practice include www.police.gov.uk, the Department for Work and Pensions' Stat-Xplore, Ministry of Justice reoffending data, and the ONS release of underlying CPI data[5], among others.

10. Since 2012 the GSS has been actively addressing a range of data challenges, including Open Data and transparency, as a whole. This work has led to the identification of a number of key areas which will improve the use and reuse of data, together with making data more widely accessible and available. This work, in the form of the GSS Data Strategy is expected to be published before the end of the year.

Opening up data and statistics for use and reuse – examples and progress
11. The Census is a unique and rich data source that provides high quality information used by government, businesses, academics, the media and the public. The approach used to disseminate the 2001 Census outputs was a major change from previous censuses and was a pioneering early example of opening up data long before that term became recognised. Access to the main results of the 2001 Census was free to all end users and their use of these main results was unrestricted. This contrasted markedly with the 1991 Census which had a complicated charging regime, access limitations and many outputs delivered on paper. The 2001 approach also opened up access to the data through modernised means of delivery and the distribution processes were streamlined to be more efficient and understandable to users. 2011 Census information has been released with innovative data visualisations: www.ons.gov.uk/ons/interactive/index.html. 2011 Census information has informed public debate in the media and the data visualisations have attracted widespread comment and use.

12. ONS is starting to publish its data using open standards that will enable machine to machine exchange of statistical information thereby allowing easier use of the data with

---

[4] http://www.ons.gov.uk/ons/about-ons/what-we-do/FOI/foi-requests/government/index.html and
http://www.ons.gov.uk/ons/about-ons/publication-scheme/published-ad-hoc-data/index.html
[5] http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/responding-to-the-open-data-agenda---an-ons-case-study--consumer-prices-index.pdf

tools such as data visualisation (i.e. the Statistical Data and Metadata eXchange or SDMX, an International Standard ISO 17369:2013 is being used). Accordingly, more detailed outputs from the 2011 Census will be published later in 2013 in an open data format using an enhanced Data Explorer (DE) and an Application Programming Interface (API). These innovations will provide URLs for the datasets that are released using this functionality and that will meet the 4* Open Data standards (see Annex C). A number of datasets will also subsequently be released, for example retail sales, population projections and the Annual Survey of Hours and Earnings (ASHE).

13. The API is an important part of the ONS response to Open Data. It will enable users to take the ONS datasets into their own systems and websites to use in combination with other information. It will also encourage innovative developments such as mobile applications.

14. The next phase of open data releases will focus on economic datasets, serving the user demand in this area. The three initial datasets mentioned above are to assure the approach. ONS prioritised datasets for open data releases based on a number of factors including volume of usage over the last twelve months, whether it is a high profile time series dataset (e.g. GDP), where there is a natural fit with trailblazer data, where common data can be readily combined, and the current system format of the data. ONS plans to release as many of these datasets as possible during 2014 and then update them on a regular basis.

15. ONS is continually looking to improve its regular outputs based on what users have told us they need. ONS ran a pilot on linked datasets during 2011/12 which enabled datasets to be joined together through common geography codes. This pilot led to European funding of a project that is currently underway to implement linked datasets. Another example is the Open Geography portal available on the ONS website, which gives access to a range of digital geographic data, from postcodes and boundaries to area codes, mapping, and aerial imagery. The portal can help users create statistical maps for their regions or neighbourhoods. Developed by ONS in partnership with Landmark Solutions, the portal uses open source software. Datasets are provided free of charge, mostly under Open Government Licence terms and conditions. The portal allows ONS to comply with the EU's INSPIRE Directive that harmonises the way geographic datasets are supplied across Europe. The tool also meets government aims for delivering 'digital by default' services, and for delivering the higher 5* geography data, directly and through APIs. It also helps ONS reduce the cost of managing and disseminating geographic data.

16. As part of its response to the Open Data Strategy, the Cabinet Office has issued draft guidance to departments on developing the UK National Information Infrastructure. The first step of this guidance is to identify the core national datasets that should be included in an inventory of national datasets held. ONS already publishes many of the underlying datasets for its outputs, where it is appropriate and lawful to do so. The next step is to list other potentially valuable datasets held, for example the datasets that are made available to approved researchers with appropriate metadata and guidance and under controlled access conditions (see paragraph 20).

17. ONS is working with the Economic and Social Research Council (ESRC) to develop proposals for an Administrative Data Research Centre for England (ADRC–E). This Centre will promote research based on linked administrative datasets, thereby reaping

benefits across England as a result of re-use of public data. ONS involvement will include bringing in expertise in managing data standards and providing metadata for researchers in ADRC-E, while contributing to the greater Administrative Data Research Network across the UK.

Protecting Privacy

18. The *Statistics and Registration Service Act 2007* (SRSA) prohibits the disclosure of information about individuals or that which enables them to be identified. Other pieces of legislation that affect data collection contain similar provisions. The SRSA provision makes it a criminal offence for officials to disclose personal information about individuals or organisations.

19. Even datasets which have had the personal identifiers removed can disclose personal information if they are not anonymised properly. There are a number of scientific studies into this[6] and the residual risk can increase as additional datasets are made public. Mitigations include:

   - the use of Privacy Impact Assessments to show that the risks have been thoroughly considered[7]; and,
   - testing the continued effectiveness of the procedures undertaken to anonymise data as new datasets are released into the public domain.

20. There will be some datasets that are too disclosive for wider release but which would be of significant benefit to the research community. ONS holds such data assets and makes them available for approved access by researchers via data laboratories, in particular the virtual micro data laboratory (VML) which has access points at each ONS office in Newport, London and Titchfield as well as in the devolved administrations in Scotland and Northern Ireland. ONS also places research datasets with the UK Data Service at the University of Essex and has done so for many years. ONS's Publication Scheme http://www.ons.gov.uk/ons/about-ons/what-we-do/publication-scheme/index.html sets out the datasets that ONS makes available for research.

Potential Barriers to Wider Transparency and Opening up Data

21. Across the GSS the work on transparency and opening up data is typically being accommodated within existing resources and this may limit the pace of progress. It is necessary to ensure that data acquired from the public for an administrative or statistical purpose is safe to publish, for example it will not disclose personal information about an individual or an organisation, and this work has a cost attached to it.

22. The ONS approach to Open Data sets expectations for standards, for example at a minimum there must be an expectation that metadata will be available for open datasets. Implementation of common standards, such as common classifications or codes, will encourage re-use and help to achieve coherence without which there may be reluctance to access and use data of unknown quality.

---

[6] See Elliot, M., Kalra, D., Singleton, P., and Smith, D. (2012) Statistical Disclosure Control and Protecting Privacy in a Clinical Data Warehouse, The Cathie Marsh Centre for Census and Survey Research; and Willenborg, L. and de Waal, T. (1996) Statistical Disclosure Control in Practice, Springer

[7] Guidance on Privacy Impact Assessments can be found at: http://www.ico.gov.uk/upload/documents/pia_handbook_html_v2/html/3-app3.html

23. It is important to acknowledge that there is a risk that data will be misinterpreted with poor consequences for policy making and public understanding. Open datasets should be quality assured to facilitate use and reduce misunderstanding and the potential for inefficiency or for wrong decisions to be made on the basis of misleading data. This risk is best managed by the data holders opening up their public datasets with clear descriptions of the data and its quality.

Annex A: Principles 5 and 8 of the Code of Practice for Official Statistics[8]

Principle 5: Confidentiality - Private information about individual persons (including bodies corporate) compiled in the production of official statistics is confidential, and should be used for statistical purposes only.
Practices
1. Ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information.
2. Keep confidential information secure. Only permit its use by trained staff who have signed a declaration covering their obligations under this Code.
3. Inform respondents to statistical surveys and censuses how confidentiality will be protected.
4. Ensure that arrangements for confidentiality protection are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics. Publish details of such arrangements.
5. Seek prior authorisation from the National Statistician or Chief Statistician in a Devolved Administration for any exceptions, required by law or thought to be in the public interest, to the principle of confidentiality protection. Publish details of such authorisations.
6. In every case where confidential statistical records are exchanged for statistical purposes with a third party, prepare written confidentiality protection agreements covering the requirements under this Code. Keep an operational record to detail the manner and purpose of the processing.

Principle 8: Frankness and accessibility - Official statistics, accompanied by full and frank commentary, should be readily accessible to all users.
Practices
1. Provide information on the quality and reliability of statistics in relation to the range of potential uses, and on methods, procedures, and classifications.
2. Prepare and disseminate commentary and analysis that aid interpretation, and provide factual information about the policy or operational context of official statistics. Adopt formats for the presentation of statistics in graphs, tables and maps that enhance clarity, interpretability and consistency.
3. Make statistics available in as much detail as is reliable and practicable, subject to legal and confidentiality constraints, offering choice and flexibility in the format according to the level of detail required by the user.
4. Publicise official statistics in ways that enable users to identify and access information relevant to their needs. Make access to official statistics as straightforward as possible by providing easy-to-use entry points.
5. Ensure that official statistics are disseminated in forms that, as far as possible, are accessible to a range of different audiences, including those with disabilities.
6. Ensure that official statistics are disseminated in forms that enable and encourage analysis and re-use. Release datasets and reference databases, supported by documentation, in formats that are convenient to users.
7. Manage official statistics in accordance with relevant public records legislation and codes of practice on records management. Deposit official statistics (accompanied by information about their purposes, design and methods) with the relevant national archive as required in legislation.

---

[8] http://www.statisticsauthority.gov.uk/assessment/code-of-practice/code-of-practice-for-official-statistics.pdf

Annex B: Public Data Principles

The Cabinet Office released draft Public Data Principles in 2010 which should underpin the release of data. The Principles were re-launched in the Government's Open Data White Paper, published in June 2012 with the announcement that they have now been adopted as UK Government policy. The principles are:

1. Public data policy and practice will be clearly driven by the public and businesses who want and use the data, including what data is released when and in what form;
2. Public data will be published in reusable, machine-readable form;
3. Public data will be released under the same open licence which enables free reuse, including commercial reuse;
4. Public data will be available and easy to find through a single easy to use online access point;
5. Public data will be published using open standards and following the recommendations of the World Wide Web Consortium;
6. Public data underlying the Government's own websites will be published in reusable form for others to use;
7. Public data will be timely and fine grained;
8. Release data quickly, and then republish it in linked data form;
9. Public data will be freely available to use in any lawful way;
10. Public bodies should actively encourage the re-use of their public data; and
11. Public bodies should maintain and publish inventories of their data holdings.

Annex C: The Open Data Five Star Scheme

The Open Data Five Star Scheme as published in the White Paper [9]:

* Available on the web (in any format) with an open licence that makes it free of restriction on use or redistribution

** Available as structured data (e.g. Excel instead of an image scan of a table)

*** Available in an open non-proprietary format (e.g. csv or xml)

**** All of the above plus Uniform Resource Locators (URLs) so that people can 'point' to them directly, rather than just to an area on the website

***** All of the above, plus a link from the data to other people's data to provide context.

---

[9] http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential