

Open Data and the Government Statistical Service

August 2012

The National Statistician

The National Statistician – a statutory office holder – is also the Chief Executive of the UK Statistics Authority and its principal adviser on:

- the quality of official statistics
- good practice in relation to official statistics, and
- the comprehensiveness of official statistics

She is also the Head of the Government Statistical Service (GSS) which is a network of professional statisticians and their staff operating both within the Office for National Statistics and across more than 30 other government departments and agencies.

Enquiries

For enquiries about the content of this report please contact Pam Davies on:

Tel: 01633 455792

Email: nsoffice.enquiries@statistics.gov.uk

© Crown Copyright 2012

The text in this document may be reproduced free of charge in any format or medium providing it is reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright and the title of the document specified.

Where we have identified any third party copyright material you will need to obtain permission from the copyright holders concerned.

For any other use of this material please write to Office of Public Sector Information, Information Policy Team, Kew, Richmond, Surrey TW9 4DU or email: licensing@opsi.gov.uk

Open Data and the Government Statistical Service

Introduction

The GSS is committed to increasing the amount of data released in open and re-usable formats in production of its statistical outputs. This is in line with the UK Government's policy, as set out in the Open Data White Paper¹.

Principle 8 of the Code of Practice for Official Statistics 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.

Principle 5 of the Code of Practice 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.

The Public Data Principles (see Annex A) should underpin the release of Government Statistics in conjunction with the Code of Practice for Official Statistics.

Although this document is aimed primarily at those working on statistics in UK government departments, those in devolved administrations may also find it helpful in bringing additional benefits to users of their statistics.

Context

The UK Government's Open Data Agenda encourages departments to release operational data in electronic open formats that encourage reuse and develop new applications. To support this, it has recently created a right to data, in an amendment to the Freedom of Information Act, and is asking Departments to publish their plans under open data. The GSS must grasp the opportunity of open data, mindful of the need to comply with the Code of Practice for Official Statistics.

Next steps across the GSS

Producers of official statistics will look to:

- release an HTML summary of the key findings from statistical releases to allow re-use
- make increased use of social media to promulgate statistical releases;
- as part of the statistical planning processes, set out a timetable for when underpinning datasets from our key statistical outputs will be released;
- provide a Privacy Impact Assessment with new datasets being released at the time of that release;
- use GSS standard linked data formats where appropriate to ease re-use and release in linked data formats.
- As a short term action, Heads of Profession for Statistics should release by end of October 2012, a list of the datasets planned for release over the next 2 years to end March 2014.

¹ Link to White Paper <http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential>

What is the GSS looking to achieve?

Opening up statistical commentaries for reuse

The GSS is looking for opportunities to open up statistical commentaries for re-use as well as opening up data, but at the same time is aware of the Code of Practice requirements to protect privacy.

At present, the majority of Departments produce their statistical reports in pdf format. This can make re-use difficult. With a few straightforward steps of putting the summary in HTML or XML, possibly posting this out through social media channels, the audience reach can be greatly increased. By doing this, it opens up the statistical releases and allows re-use and direct interaction with many more consumers of information without the filtering of the message by third parties.

Opening up data

The challenge of the “right to data” is to ensure that the underpinning datasets are released in conjunction with the statistics and it gives the right to request these if they are not automatically released. There are significant benefits to the GSS for setting out a forward work-plan for release of datasets as this:

- ensures that we prioritise the most important/valuable datasets for our users; and
- ensures that we can deliver the releases as quickly as possible and within current resource constraints.

As part of setting this plan, each GSS Head of Profession will send into the National Statistician’s Office every year an up-to-date list of the data published as part of their Transparency Plans. This will form the basis of a GSS forward programme.

Protecting Privacy

The Statistics and Registration Service Act 2007, prohibits the disclosure of information about individuals or that enables them to be identified. Other pieces of legislation that affect data collection contain similar provisions.

Even datasets which have personal identifiers removed have the potential to disclose further information if they are not anonymised properly. There are a number of scientific studies into the risk of de-anonymisation. The risks are compounded by the fact that the risk of de-anonymisation potentially grows with every dataset released. Mitigations include:

- the use of Privacy Impact Assessments to show that the risks have been thoroughly considered,²
- testing of anonymisation in order to ensure that as each new dataset is published, it is tested for risks of de-anonymisation against all the data already in the public domain.

There will be some datasets that are too disclosive for wider release but would be of significant benefit to the research community. Departments holding such data assets should consider the set-up or use of data laboratories for approved access by researchers to data.

² Guidance on Privacy Impact Assessments can be found at:
http://www.ico.gov.uk/upload/documents/pia_handbook_html_v2/html/3-app3.html

Linked data formats

There has been much discussion about the advantages and disadvantages of the use of linked data.³ Linked data formats would allow statistical users to quickly get information from across the GSS on all aspects of interest to them.

The key benefits are:

- by use of linked data standards, we can establish the provenance of all National Statistics and their use throughout the web;
- the use of data in this way will allow much freer access to data, including the academic community, and will improve the research capability of the UK;
- it will make bringing compendium publications and overviews together simple and efficient;
- it will increase the ease of analysis across social and economic datasets;
- it allows us to consistently harmonise and analyse across common definitions eg. geography, courts, disability, ethnicity.

The advantages of linked data will only be fully realised by working across the GSS and will be considered further by the GSS Presentation and Dissemination Committee.

The GSS will develop training in producing linked datasets to bring the skills of the staff in the GSS up to date.

Support for GSS Staff

This is a bold and challenging agenda and one that will require support from across the GSS.

Following on from the GSS Seminar on Open Data on 3 April, the GSS will support each other through:

- using exemplars and case studies;
- creating communities of experts in undertaking this work;
- providing training for Government Statisticians, including the e-learning package that has been developed;
- providing a central repository through data.gov.uk of approved linked definitions across the GSS.

3. An overview of linked data can be found at http://data.gov.uk/wiki/Linked_Data.

Annex A - 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 point³;
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.

⁴ www.data.gov.uk