Writing About Statistics

Guidance for the Government Statistical Service on Preparing First Releases

Contents

Foreword

The Code of Practice for Official Statistics states that we need to provide 'commentary and analysis that aid interpretation', but why is this important? Without interpretation and context, numbers themselves mean very little. Good commentary should leave the reader well informed about the meaning and relevance of the data. This is central to the impact of our statistical outputs, ensuring that they are interesting and relevant to public debate. The Government Statistical Service (GSS) should aim to unlock the stories from data through insightful commentary to encourage wider and better use of statistical releases across the full range of users.

The UK Statistics Authority has identified commentary as one of the main areas in need of improvement, and this is reflected in its assessment reports to date. A number of departments have begun to take action, although there is still much more to be done for our products to reach their full potential.

There is a wealth of guidance already available, and this document aims to consolidate it and bring out the key points of relevance for the GSS when preparing first releases, along with examples. It is intended for use by GSS members as supporting material for their own work in driving forward improvements, and to act as a framework for gathering examples of good practice as efforts continue. As this is a development area for the GSS, the guidance will be updated regularly with further examples.

Three key documents are:

- Making Data Meaningful Part 1 – the concept of a statistical story; key messages; how to structure your publication; what to consider for a web audience.
- Standards for statistical release – criteria for presentation, key messages, language usage and context.
- The National Statistician's message to the GSS on statistical commentary – a clear explanation of the importance of statistical commentary, and priority areas for the GSS.

1. Preparation: understanding the background, uses and users

An understanding of the context in which data are collected, and the factors which could influence the results, is essential in order to produce useful commentary.

Effective user engagement and commentary are closely linked and should be mutually supportive. Good commentary leaves the user more informed, and a strong understanding of users should lead to improved commentary to cater for their needs.

The authority's 'Strengthening User Engagement' report examines ways of identifying users and improving communication, and the Royal Statistical Society is developing an online communications hub, StatsUserNet (to be launched early 2012), to allow direct access to a large user community.

Tips:

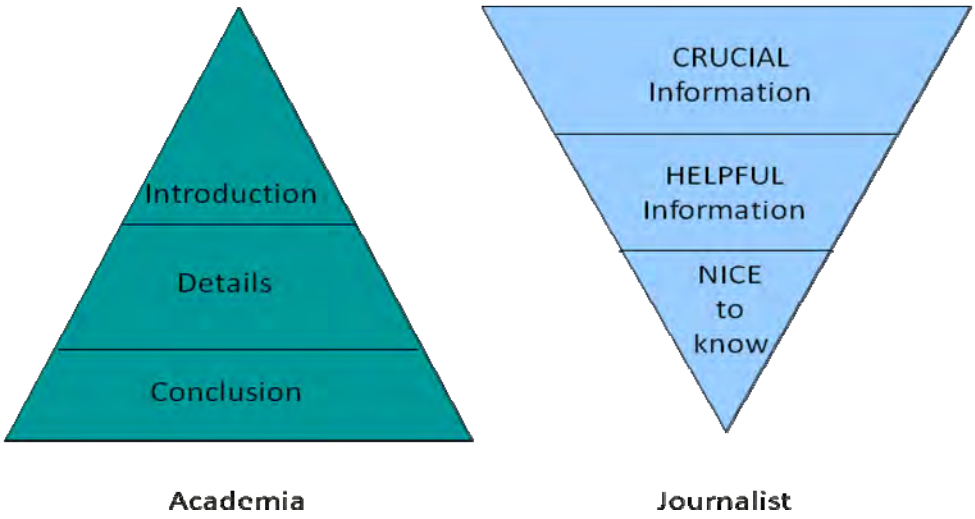
- Get to know the subject area. For example, engaging with policy teams, other analysts and external experts can improve understanding of the background and context of the statistics, and also helps establish whether there are any new initiatives, changes in policies or other factors that could affect the results.

- Consider why the data are collected, who the users are, what they use the statistics for, and why they should be interested in the results. Explain this in the text.

2. Structure

A release should be constructed to allow the reader to scan through the document and pick out the points of relevance to them.

The 'inverted pyramid' model below highlights the difference between the structures of 'academic' and 'media' publications. For first releases, the GSS should move toward the journalist approach, presenting the most important information at the front of the publication (see also 'key messages' section), followed by more detailed and less critical information.



To ensure that the reader can easily find and digest information, apply the following principles:

- A summary of the key findings at the front of the release;
- A contents list for longer releases;
- Concise and descriptive subheadings to catch the reader's eye;
- Bulleted lists;
- Short sentences and paragraphs, with one 'concept' per paragraph;
- Consistency. E.g. 'per cent' not a mixture of '%' and 'per cent'; consistent rounding.

Six out of ten web users visit Government websites, and the GSS is increasingly looking at ways in which information can be better disseminated on the web. Although many of the general principles apply, there are some **specific considerations for web audiences**:

- Reading is 25 per cent slower. It is even more important that your report can be scanned easily, with the essential information at the top and key words towards the left of the page. See Figure 1 for an illustration of how readers scan contrasting web pages.

- People want quicker results, spending on average two minutes on a site and 47 seconds on a page. Clear identification of key messages is even more important.

More information on web dissemination can be found in the GSS Web Dissemination Strategy.

Figure 1: Scanning patterns on two contrasting web pages. The example on the left is better constructed, with subheadings and shorter sentences.

Source: Nielsen Norman Group

Tips:

- Careful consideration of the content of a publication can help with structure. For example, moving detailed methodology or background information to a 'user guide' can make it easier for the reader to find the essential information in the latest edition.

- Slide packs published on the web can be re-used by visitors, and are as such a helpful way of communicating data and messages.

3. Language

'Making Data Meaningful' outlines the importance and principles of good writing style, the main point being that language throughout should be clear, simple and appropriate to all audiences. Technical language and jargon in particular can be very intimidating, and should be avoided:

Example:

> "gross weekly pay in the bottom income decile was below £276"

could be more clearly expressed as:

> "10 per cent of full-time employees earned less than £276 per week".

or:

> "One in ten full-time employees earned less than £276 per week".

If technical terms are required, explain them on first use in the text. Well known abbreviations and acronyms need no explanation, but err on the side of caution and include explanations of any terms that may not be known to readers.

Tips:

- Seek user feedback on whether they find your publication easy to understand
- Ask a colleague or ideally a non-specialist in your department to peer review the publication. You should also seek feedback from users outside the department. An author can become too close to the writing to be able to judge whether it is understandable to users.
- A glossary hyperlinked from the main text allows you to expand on definitions of technical terms. The NHS IC website has a "jargon buster", defining acronyms and abbreviations.

4. An informative title

A good title allows the reader to identify whether the release contains information of interest to them. It should therefore stand alone as a description of the publication, and contain essential information:

- The set of statistics being published
- The point in time or period to which the latest statistics relate
- The geographical coverage
- The frequency of the release
- If relevant, whether the results are provisional or final

Examples:

"Annual Adoptions in England and Wales, 2009"

"Annual Emissions of Air Pollutants in the UK, 2010 Provisional Results"

5. Key messages

These are critical to ensuring that the main messages of the publication reach the user. Journalists and press offices will use them, often verbatim, and well-crafted key messages help to ensure that they are correctly identified and presented in the media. The messages should summarise objectively and clearly the most important parts of the data. They should be easy to understand, and place the findings in context. They should not seek to summarise all the findings in a publication, instead they should present the most important and relevant messages from the statistics.

Example:

> "Despite mounting financial challenges during the 1990s, young people from moderate and low-income families were just as likely to attend university in 2001 than they were in 1993"

There is much room for improvement in the GSS in this area, especially in identifying the key messages from the results and presenting these in an accessible way. There is in particular a tendency to report on changes since the previous period with no context:

Example:

> "Total formal admissions rose by 2 per cent to 28,700 from 28,100 in 2007-08."

This is not a good example as it does little to inform the reader of the context – it is not possible to tell if the 2 per cent is large or out of the ordinary, because there is no reference to the long-term trend. Reference to long-term trends is very effective at placing the current results into context, and making good use of time series is a priority for the GSS.

Example:

> "The number of adoptions in England and Wales in 2009 was 4,655, a decrease of 5.7 per cent since 2008 and the lowest since 1998.
>
> "Emissions of nitrogen oxides in the UK fell by 13 per cent to 1.08 million tonnes between 2008 and 2009, continuing the decreasing trend"

Tips:

When reviewing the key messages in a publication consider:

- What are the most interesting, relevant and most useful messages? Remember that the most striking change may not be the most important change.

- Could your messages stand alone in a newspaper article without further explanation?
- Are the messages from the last reporting period still the most relevant and newsworthy, or should they be reconsidered? Avoid simply inserting new figures into the text unless appropriate.
- Could the same message be conveyed more concisely, in particular, are there too many figures in the text? Remember that the reader can refer to tables for detailed figures.
- If available, have the results been placed in long-term context? Graph the data to examine trends.
- Get feedback on whether your publication conveys the message clearly.
- Avoid conjecture, and do not comment on the effectiveness of Government policy.
- Social media can be effective in reaching a wide audience and conveying key messages simply and quickly.

6. Context

The publication should seek to explain why the data are collected, what they are used for, and why the reader should be interested in them. If this is difficult, or you can only identify a narrow range of uses/users, it is likely that more research is required into the uses of your statistics. It is acceptable to make assumptions about what the data might be used for and explain this in the text. This is also an opportunity to place the figures in the context of other relevant datasets to allow the reader to develop a fuller picture.

Examples:

Air quality

> "There has been a substantial improvement in air quality in recent years. Despite this, there is still evidence of negative health effects and environmental damage caused by air pollutants. These statistics provide the public with information on current levels of and trends in emissions of air pollutants, and are used to inform air pollution policies and progress towards targets both nationally and internationally."

Environmental accounts

> "Environmental accounts provide data on the environmental impact of UK economic activity, and on the use of resources from the environment in the economy... They are used to inform sustainable development policy, to model impacts of fiscal or monetary measures and to evaluate the environmental performance of different industrial sectors."

Adoption statistics

> "Key users of adoption statistics include the Department for Education who have the policy lead on adoptions... Organisations such as the European Parliament and those in the voluntary sector, for example, the British Association for Adoption and Fostering use ONS's adoption statistics for comparison purposes and also to support campaigns."

It is appropriate to identify policies and targets in factual terms, but not to endorse the policies or comment on their effectiveness. It is also appropriate to list or provide links to current policies and targets.

Example:

> "Although air pollution is an important local issue, pollutants can be carried long distances and the effects may be experienced many miles from their source. The policies in place at the international level set legally binding targets for each country. The UK's targets are presented alongside the results"

Enabling the reader to make comparisons between geographical areas and other countries in the UK or internationally is very effective at placing results in context.

Example:

> "The UK sends approximately 15 per cent more municipal waste to landfill than the EU average. It also has lower recycling and composting rates (34 per cent) than the EU average (39 per cent)"

Depending on the timing of the publication, it may be difficult to do this for the latest data presented in a first release. It is still appropriate to comment on historic data in relation to geographical comparisons, and signpost the reader to sources of further information.

Sometimes it may be misleading to make comparisons without appropriate caveats, for example due to differences in methodology or definitions.

Example:

> "Only broad comparisons can be made between countries because of differences in definitions of types of waste management. The recycling category includes some other recovery options (fuel manufacture, for example), which are negligible in most countries, but account for around 10 per cent of municipal waste in Germany, and 6 per cent in Spain."

Tips:

- Establish where equivalent data and publications are held for other countries in the UK, or more widely internationally. Comment on these where appropriate, and include links to the relevant websites.
- Hold meetings with policy experts and other analysts to discuss and explain the drivers of trends. Providing the analysis is evidence-based, this can legitimately be done in compliance with the Code of Practice, as part of the data validation process.
- Keep in touch with Government Social Research work and academic research on your subject area.

Describing the data - for example in terms of methodology, reliability and coverage – is historically the category that most GSS commentary has fallen into. This is still important, but steps must be taken to ensure that these details do not dilute the messages and story.

It is important to explain issues relating to the quality and reliability of the data, although in general non-technical readers will not need to know the methodology used to ensure the statistics are accurate. It is however appropriate to discuss certain issues around quality when explaining the figures, for example outlining the variability, confidence intervals, statistical significance, and limitations of the data, in relation to how the figures will be used.

Linked documents or a methodology section/annex within a document are often a simpler way to provide further background information. Make sure that these notes are as easy as possible to understand and not over-detailed. They may cover:

- Data sources
- Coverage - time periods, geography etc.
- Discontinuities
- Methods of analysis
- Definitions and classifications
- Reliability
- Information about how the methods and definitions used relate to European Union or international concepts and classifications
- Where data definitions or methodology have recently changed, adequate description of these changes, including numerical comparisons that enable users to see the extent of differences with the previous data series
- Where data are normally subject to later revision, a clear explanation that these are initial estimates and when they are likely to be revised

The GSS Quality Measurement and Reporting Guidance provides a template for reporting on the quality of outputs

7. Interpretation

Interpretation is a key area for improvement, but is also the most challenging in terms providing insightful commentary without straying into opinion and conjecture. Your text should explore relationships, causes and effects, to the extent that they can be supported by evidence. It should be impartial and objective, and present a balanced picture.

Examples:

Prison populations

> "The prison population grew rapidly between 1993 to 2008 – an average of 4 per cent a year. This rapid rise was driven by:
>
> - Increased numbers of people sentenced to immediate custody from 1999 to 2002;
> - Increases in the average custodial sentence length and increased use of indeterminate sentences;
> - Increase in numbers recalled to prison following breaches of the conditions of licence and these offenders spending longer in prison once recalled;
>
> However, the rise in the prison population has slowed considerably since the summer of 2008 with an average annual increase of 1 per cent (Figure 1.1).
>
> The flatter trend seen since 2008 partly reflects the introduction of the Criminal Justice and Immigration Act (CJIA) 2008, which changed sentencing and offender management in ways which helped to reduce growth in the prison population. For more information see CJIA 2008 http://www.justice.gov.uk/publications/docs/offender-management-stats-qtly-bullet-july-sept10.pdf"

Broadband speeds

> "There is a large difference in average broadband speeds between urban and rural areas, and rural areas have a slower average broadband speed. One reason why there is such a difference is that it is harder for network operators to recoup the fixed costs necessary for upgrading exchanges and cabinets in rural areas, where there are lower population densities, and therefore fewer end customers.

Hospital admissions

> "Whereas last year the increase in formal admissions was in such admissions to independent sector hospitals, this year's increase was solely in formal admissions to NHS facilities."

Air pollutant emissions

> "There have however been substantial long-term reductions in emissions due to measures such as introduction of catalytic converters, unleaded petrol, and reduced benzene content in fuel."

In some cases there may be plausible but less certain explanations. It is important to apply sound professional judgment here, but with careful wording, less certain explanations can also be included. Choose your language carefully:

- 'affect', 'cause', 'consequence', 'effect', 'impact' suggest causality
- 'association', 'correlation', 'corresponding', 'equivalent', parallel' suggest a relationship but not a causal association
- 'expect', 'believe', 'think', 'predict', 'envisage', 'forecast' suggest a more provisional explanation

Commentary should avoid simply describing the data and excessively quoting figures which are generally better presented in a table.

Example:

> "In 2008-09 78 per cent of section 136 Place of Safety Orders in NHS facilities (6,200) ended in an informal status and 22 per cent ended with the patient being formally admitted to hospital. In 2007-08 the comparable figures were 71 per cent (4,900) and 29 per cent."

In addition to being confusing for the reader, each of the highlighted figures needs to be cross-checked against the data tables. This is both resource intensive and prone to error should the underlying data change.

Tips:

- A strong knowledge of your topic and its theoretical context helps with interpretation, for example whether one variable could have had an effect on another, without straying into opinion and conjecture.
- Consulting in advance with policy teams or other specialists can establish if there have been any policy, societal, or economic changes, or new initiatives that could cause a change in results.
- If you find it possible to simply insert new numbers without updating the text each time you issue the release, is it just recording rises and falls without explaining what is driving the changes? Commentary must add value to the statistics.
- Publications should be reviewed to ensure that interpretation has kept pace with change.
- Consider how your interpretation can feed into the key messages.