

Reviewing Aspects of User Orientated Quality Reporting

S. Tucker

Office for National Statistics, Newport, United Kingdom

sarah.tucker@ons.gov.uk

National Statistics Institutes are required by the European Code of Practice to provide quality reports to users alongside official statistics.

To meet this requirement, ONS produces a Quality and Methodology Information (QMI) report for each Statistical Bulletin. QMIs report against the European Statistical System (ESS) quality dimensions and other quality characteristics. They also contain information on the strengths and limitations of data which help users decide upon suitable uses for that data.

QMIs have been published for 5 years, so now is a good time to review how quality information is communicated to users. Are there any gaps in what is provided compared to current user needs? What can be done to extend the use of this information?

The primary purpose of quality reporting has been to help users decide upon suitable uses of the data. It was determined that there could be a step before this, namely to reduce the misuse of the data. Research has therefore been undertaken to look at how ONS can first help users understand how to not misuse the data, and then how the data should be used.

This paper will discuss some work that has been carried out to reduce the risk of users misusing data. The findings from internal focus groups and meetings will be presented, including further work with statistical producers to create webpages for user testing. The paper will explore the results of user testing and the implications for communicating quality information to users in the future.

1. Background

National Statistics Institutes are required by the European Code of Practice (Eurostat, 2011) to provide quality reports to users alongside Official Statistics. For the last 5 years or so, Office for National Statistics (ONS) have met this requirement by publishing Quality and Methodology Information (QMI) reports alongside statistical bulletins. QMIs contain static quality information, descriptions of methodological processes and the strengths and limitations of data. They are designed to enable users to understand and determine suitable uses of the data. QMIs also allow us to report on the European Statistical System (ESS) dimensions of quality.

Since the QMIs were designed, there have been many changes in the digital environment affecting how our users consume our information. ONS has been focussing on the agenda “Digital by Default” with recent work researching user personas and developing a new website. Therefore, it was a good time to review how we communicate quality information to our users with an emphasis on increasing accessibility for different types of users.

As a first step to increasing this accessibility, we considered what a QMI does which, as stated above, is to help users make an informed choice on what the data are suitable to be used for. We then thought about what our users’ first need from quality reporting and determined that before this, the first purpose of quality reporting should be to help users reduce the risk of misusing the data. We then explored the possibility of creating a new quality reporting product which combined the idea of progressive disclosure (presenting the most important information first with less important information following after) with the need to encourage users unfamiliar with quality information to access the product. We investigated the effectiveness of presenting vital information to guard against misuse as a set of 4-5 bullet points combined with an overview, on a separate Quality Summary page, which would then lead into the full QMI.

2. Approach

We then thought about what the content of these 4-5 bullet points should be. Should the content be consistent across all outputs or would bespoke content tailored for each set of data be more beneficial to users? We decided to investigate a bespoke approach as, while consistency across user based products is important, using the same pieces of information for all outputs could risk the information becoming meaningless to our users. A list of potential information was proposed, with guidance presenting a range of options for statistical producers to consider and select from when choosing what quality information to communicate to users.

Statistical producers are experts in their data and what their users need to know about it in order to reduce the risk of misuse. Therefore, it was important to consult them on what should be included in the list to create examples that could then be user tested. This was achieved through a series of focus groups and individual meetings with statistical producers, methodological experts and key stakeholders.

A great deal of information was gathered during this consultation, the information was classified into themes (Guidance, Standards and Definitions, Content, Practicalities and Miscellaneous). The themes that this paper is concerned with are Guidance, Standards and Definitions and Content.

2.1 Theme: Guidance

When thinking about setting guidance for communicating quality information to reduce the risk of misusing data, we first needed to identify a required outcome for the guidance. What do we want this information to do? We want this information to help our users reduce the risk of misusing the data. What does this mean in practise? How can we reduce the risk?

Feedback from the consultation agreed that we can do this by educating our users about the data and how to use it, including context and what they can do with the data. By providing the right information, we can help our users to guard against inadvertent misuse. To do this, we need a set of “rules” to follow to ensure that the right information is provided to users.

In developing these rules the consultation responses underlined the importance of:

- thinking specifically of users on an output case by case basis
- thinking about what users can use the data for and what information is most helpful so that users can make informed judgements on how not to misuse the data
- being transparent
- being helpful
- not being too cautious
- providing examples of good practice to statistical producers

Rules to include in the guidance:

- the most important information comes first. If someone is only going to read one thing about the data, what is the most important thing that they need to know about it? If they are only going to read 2 things, what should they be?

- whilst it is important to include what users want to know, what users *need* to know in order to reduce the risk of misusing the data must always take precedence
- information must always be current, relevant and helpful, and so should be reviewed and updated when necessary
- the most relevant topics will vary from output to output, they must always be the most important and relevant for that data and should never be considered as default topics to be completed on a regular basis
- the information must be consistent with what's in the QMI, there should not be any mixed messages

Therefore, when writing guidance we must create a mind-set of putting the user at the heart of what we are doing, of thinking what is most important for users to know rather than what we want to tell them, or even what just meets our obligations. Information must be tailored for each output and be what is most helpful. All information must be clear and transparent.

2.2 Theme: Standards and definitions

Throughout discussions during the consultation it was appreciated by participants that, with such a diverse range of outputs, a one size fits all approach is not appropriate. Conversely, it was also felt that consistency should be maintained wherever possible without being too constrained. The design and layout of the pages will help to maintain consistency but an agreed set of standards will support this, some examples given were:

- a definitions statement
- a glossary of essential technical terms

When setting up the pages, writing guidance and drafting the text, we will need to use consistent definitions and, where this is not possible, clearly explain the differences between definitions and give clear reasons for these differences.

2.3 Theme: Content

The subject of what content should be on the list to choose from to create a customised selection of information per output generated a great deal of discussion. There was a general

consensus from participants that content needs to be bespoke for each output and not a standard set of questions to be completed. A number of common themes came through strongly (more detail on some of these can be seen below), but it is important to stress that content should not be limited to these themes. Whatever is most useful in helping users to reduce the risk of misusing the data should be included. Statistical producers' experience and knowledge of their data and their users' needs will play a vital role in ensuring content is relevant and useful to users.

Uses - how to use or not use the data was considered to be an important subject to cover for our users. Participants suggested that we could be more proactive in telling our users what the data should or should not be used for. Some examples of what we could specify: how you can use the data; why you should use this data instead of that data and vice versa; how the data is used; what else is there; how outputs relate to each other, where they are similar and where they are different.

Quality commentary - this category has many and varied points to consider while determining what information is most helpful to users regarding reducing misuse. As with all the categories, these points will vary in importance depending on what is going on with the data for each time period. Producers could consider including: the main strengths and limitations of the data; important changes, quality of changes, quality assurance of the data, how the quality of estimates can deteriorate at lower levels; any caveats, warnings or signposts and where particular difficulties in using the data lie.

Uncertainty - we need to clearly communicate where data are estimates and explain what that means. We need to communicate any uncertainty associated with the data and to be clear what this uncertainty means in terms of use.

Accuracy - it was felt that users should be informed of what preliminary and revised means and what to expect. Any important issues that affect the accuracy of the data should be discussed, for example boundary changes that require revisions. Any issues of discontinuity and variance would also need to be considered.

What the data are or are not - clear descriptions need to be given about the data: what it is; what it isn't; what is available and how far back it's available; coverage; what it includes; what it doesn't include.

Coherence and comparability - coherence and comparability were strongly favoured as important points to include by business output areas. Comparisons with other statistics, where they were the same but also where they were different was felt to be of particular interest to users of business statistics. There should also be some consideration given to the harmonised principles used and the resulting comparability between outputs.

Sources - for outputs that are based on administrative data, informing users of the data sources was considered by statistical producers to be of high importance. In discussing data sources, points to consider informing users of could be similarities in sources and differences in outputs, definition of the sources and whether the output is a combination of administrative data and survey data.

Process - whilst describing the process of producing the output to users in order to help them reduce the risk of misusing data there are, of course, many aspects that could be discussed. These include, but are not limited to: sample size and response rate, is this robust and meaningful; periodicity; how is the data calculated (this will be difficult to communicate in the short space intended for this information, but links could be included); link to supporting metadata and methodological documents; non-response rates; whether the data is seasonally adjusted or if it is an index, whether it is based on a basket of goods and when it was it last rebased.

The amount of information available to be discussed is wide and varied, and it should be stressed here that the points given above are not a "checklist" but instead a basis for starting the discussion about what information is most helpful to users of each individual output.

3. Background to user testing

Once the internal consultation was complete, it was important to gain views from our users to ensure that this new quality product would meet their needs.

The development of a new ONS website was in progress, and on the trial website each output had a QMI page that gave a general introduction to the survey. Using the research discussed above, a small team worked together to create an example page for Business Register and Employment Survey (BRES), containing quality information designed to reduce the risk of misusing data. This page included a set of bullet points that gave the most important pieces of

information required to help reduce the risk of misusing data (see Example 1), followed by an overview giving further detail (see example 2).

Example 1: Important points about BRES data included on the example page:

- the Business Register and Employment Survey (BRES) is the official government source of employee statistics by industry
- it provides employee and employment estimates at low levels of geography and industry for Great Britain; however, as it's a sample survey, there is a reduction in the quality of the estimates as the geographies get smaller
- BRES does not include some of the very smallest businesses not registered for VAT or PAYE
- it underestimates the employment measure as it does not include all self-employed
- it was not designed to be used as a time series; BRES represents a snapshot of the GB economy and time series analysis should be treated with caution

Example 2: Overview included on the example page:

Business Register and Employment Survey (BRES) data are used to produce employee and employment statistics and to update the Inter Departmental Business Register (IDBR), the sampling frame used for a lot of our business surveys.

The employment figure is calculated by adding the number of working owners registered for VAT or PAYE to the number of employees employed by a business.

BRES is a point in time survey, based on a certain date in September. Therefore, it is not designed to be used as a time series. BRES is subject to discontinuities over time, such as changes in Standard Industrial Classification (SIC), source data, methodology and reference date. This needs to be taken into account with any time series analysis.

BRES collects employment information from businesses representing the majority of the economy in Great Britain. Northern Ireland data are combined with BRES data to produce high level estimates for the UK. BRES estimates on the ONS website are UK based, those on Nomis[®] are based on Great Britain.

The BRES data and estimates are widely used, both within and outside government, and are a vital source of business employee information. The main users and uses of the output include: [Eurostat](#), the Scottish and Welsh Government, [Department of Business, Innovation and Skills \(BIS\)](#), Workforce Jobs and the [Annual Business Survey \(ABS\)](#). Local Government planning departments are major users of BRES using the estimates published on NOMIS[®] to forecast trends in employment in their specific areas and to claim for Central Government and European funding.

4. Task for user testing

The existing trial website page for BRES and the Example page for BRES were provided to a group of expert users identified by the website user team. Alongside the pages, we provided a task for the user team to test against the pages, the questions we asked are below:

- Does the information in this new format encourage users to read this important quality information, particularly users who would not generally access a QMI?
- Does this kind of quality information give users the ability to make informed judgements on how not to misuse data, or at least an indication that there are things to be taken into consideration before using the data?
- Do the users find this type of quality information more or less useful than the original content, or would a combination of the two be helpful?

The target audience for the test were Expert Analysts and Methodology contacts. The test took place between 17th and 30th November, during which 140 users participated.

As stated above, during the test the users were shown 2 examples of supporting information for BRES, Version 1 the existing trial website page and Version 2 based on the research described in this paper. Users were asked to choose the version that best described what the page was about and informed on how not to misuse data, along with what needs to be taken into consideration before using the data. They were also asked to comment on likes/dislikes on both pages.

Of the 140 participants, 67% chose Version 2 as the example that best described what the page was about and informed how not to misuse data along with what needs to be taken into consideration before using the data.

Some comments on the example version were:

- “It makes it immediately clear to users what it does/doesn’t include.”
- “Liked the “important points”. These immediately explained the limitations of the data”

The team reported that both versions have a number of positive and negative comments with users liking the simplicity of version 1 and the layout and structure of version 2.

User testing recommended implementing the new format example Version 2, but with some additional work to be done on making the information more concise and easier for users to understand.

5. Implementation

With over 100 outputs to create the new Quality Summary pages for, we needed to prioritise the order of implementation. We looked at web metrics and correspondence with users regarding the misuse of data to choose a set of 30 outputs to schedule first.

In order to continue the collaborative process of learning through sharing experience, we decided to set up a series of workshops with 4-5 statistical producers to work together to produce Quality Summary pages for publication. This will help us build up a portfolio of good practice examples to share. To address comments from user testing about understanding the language used, experts from our Editorial team will be attending the workshops to further the aim of easing understanding and expanding the accessibility of quality information to different types of users.

6. Conclusion

Given all the work going on within ONS on developing Users Personas and a new website, it was time to review how we communicated quality information to our users. We identified an

opportunity to increase the accessibility of quality information for users by exploring ways to reduce the risk of misusing our data. We consulted internally with statistical producers to identify what kinds of information our users might need to know about a wide range of outputs. We user tested our approach and are implementing the new Quality Summary pages across the office.

The Quality Summary page discussed in this paper is only one part of a layered approach to our user based quality reporting that also includes QMIs and quality information contained within statistical bulletins.

Fig. 1 A layered approach to quality reporting

	Quality Information in Bulletins (Stand alone in Statistical Bulletin.)	Quality Summary Page (Webpage with QMI PDF attached.)	Quality and Methodology Information PDF PDF attached to Quality Summary Page.
Aim	<p>Aim – to help users understand data and quality implications for that data in specific releases.</p> <p>Information provided should help users understand how to use the data reported on in the Statistical Bulletin.</p>	<p>Aim – to reduce the risk of misuse of data.</p> <p>Information provided should be the most important points in order to reduce the risk of misusing the data, particularly for inexperienced users or users with limited time.</p>	<p>Aim – to allow users to make informed judgement on suitable potential uses of the data.</p> <p>Information provided is designed to help users decide on suitable uses for the data. Template designed to meet requirements of the Code of Practice. The content is used as part of the Regular Quality Reviews and UKSA Assessments for Outputs.</p>
Type of Information	<p>Dynamic information – changes regularly to be specific to the data reported on for each period.</p> <p>Quality warnings/caveats on specific issues relating to the data reported on in that issue of the Bulletin.</p>	<p>Static information – more general across various time periods.</p>	<p>Static information – more general across various time periods.</p>
Length	<p>Concise – fairly high level information.</p>	<p>Concise – high level information – leads into QMI PDF.</p>	<p>More detailed information. Summarised</p>

			descriptions of methods used to create the output (linking to further detail) and reports against the 5 ESS Dimensions of Quality.
Presentation	Frontloaded – critical caveats should be up front in the bulletin content. Should include quality warnings/caveats on specific issues relating to the data reported on in that issue of the bulletin.	Frontloaded – most important points first. Should include quality warnings/caveats on specific issues relating to the most common likely misuses of the data.	Specified template. Should include detailed information on strengths and limitations of data.

The analysis from this research will be reused to inform further improvements to QMIs and quality information contained within statistical bulletins.

The 3 quality reporting products will then work together to provide a range of clear and comprehensive quality reporting products that will be accessible to a wide range of users.

References

Eurostat, (2011), European Statistics Code of Practice, available from:

<http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15>