Statistical Quality Management Strategy

Contents

Introduction

The quality of Welsh Government statistics is important to ensure that they are fit for purpose for our users needs. People who use our statistics should have confidence in the quality of our statistical services and products and as such we endeavour to ensure that we are positively recognised for:

- the quality of our advice

- the quality of our data

- the relevance and impact of our analysis

- our role in supporting and co-ordinating wider statistical partnerships by promoting good practice.

The quality of our statistics will impact on the trust our users have in them and us. When making decisions, users need to be clear about the quality of the data they are using and we must strive to provide high quality statistics. We ensure our statistics are fit for purpose; we use appropriate processes and are transparent about our methods. We also ensure the factual and presentational quality of our statistics meet the requirements of our users. When using data from administrative sources we will work with data providers to improve quality and to provide users with appropriate levels of information on processes.

It is our responsibility to adhere to the Code of Practice for Statistics. Our adherence to this is assessed independently by the UK Statistics Authority.

Quality is a key pillar within the Code and Principle Q3: *Assured quality*, in particular, deals with statistical quality practices. The Code requires us to:

- be clear on our approach to quality management (T4.5)

- undertake systematic and periodic quality reviews (Q3.5)

- ensure our staff are suitably trained in statistical quality management (T5.4)

- assess, minimise and explain any limitations of the data and statistics (Q1 and Q3.1)

- ensure the quality of data and statistics is monitored and reported on regularly (Q3.3).

This strategy is reviewed annually by the Statistical Quality Committee. The purpose of this committee is to provide the Welsh Government's Statistical Services Management Team (and the National Statistician) with greater assurance over the quality of Welsh Government official statistics, in addition to supporting the strategic management of quality across Statistical Services Division.

What is quality?

The Government Statistical Service defines quality to mean that statistics fit their intended uses, are based on appropriate methods, and are not materially misleading. Quality requires skilled professional judgement about collecting, preparing, analysing and publishing statistics and data in ways that meet the needs of people who want to use the statistics.

Quality and our users

In order to be a high performing statistical producer we must provide information to enable our users to know and understand the quality of our statistics. This will allow them to decide if our outputs meet their needs and also to understand the limitations of our data when they are using that data to make decisions; measure performance; or allocate resources. In turn, this will allow improved dialogue with users about their ongoing requirements. Identifying and responding to users' needs is a key component of the Code of Practice.

The Statistical Quality Objectives

In line with the Code of Practice for Statistics, Welsh Government's Statistical Services Division has four strategic objectives for quality:

1. to ensure staff are trained in effective quality assurance and quality management

2. to encourage staff to apply intellectual curiosity, question their data and take ownership of the quality assurance for which they are responsible

3. to publish appropriate quality reports or indicators for our statistics

4. to conduct regular reviews of our processes and the outputs we produce, ensuring we work with data providers to understand the end-to-end processes.

Implementing the Statistical Quality Objectives

Ensure staff are trained in effective quality assurance and quality management.

As stated above, all staff have a part to play in ensuring the quality of the statistics they produce. To maximise the culture of continual improvement and self-assessment staff should be trained in effective quality assurance and data checking approaches according to their roles and responsibilities. See Annex 1 and 2[1] for guidance on data checking and data validation. Statistical quality is fundamental to the induction process for new staff, with all new staff receiving training in quality management and data checking.

Encourage staff to apply intellectual curiosity, question their data and take ownership of the quality assurance for which they are responsible

All staff have a (differing) responsibility for ensuring the quality of the statistics we produce. By establishing a culture of continual improvement the quality of our statistical outputs will improve as all those involved in producing statistics implement ideas of best practice. By adopting an inquisitive mindset and investigating curiosities within the data as part of quality assurance, real value can be added to our outputs and potential errors discovered. For example, the number of road casualties in one local authority was highest at 3am on Sundays. This does not pass a common sense test. But when the team looked into it they identified this was due to one incident where a car drove into people departing a nightclub. Knowing this is important context for the data.

Publish appropriate quality reports or indicators for our statistics.

To increase transparency and trust in our statistics, we publish quality reports or indicators for all our statistics, in the most appropriate manner. Quality information is either included in individual outputs or published as a separate Background Quality Report (available on our website under the "Background information" tab for the relevant output), such as the quality report (PDF) for Homelessness in Wales. These include details of how our statistics meet and respond to users' needs and details of appropriate use of our statistics, including any limitations to their use. One element of quality management is ensuring appropriate and transparent handling of revisions and errors. We use our Revisions policy for this.

Conduct regular reviews of the processes and outputs we produce, ensuring we work with data providers to understand the end-to-end processes.

To continually improve our processes and outputs we conduct peer reviews and critical self-assessments, for example using Government Statistical Service (GSS) quality management tools and guidance.

With regards to statistics that use administrative sources we will follow the UK Statistics Authority guidance for the quality assurance of administrative data. We will ensure that we are working with data providers to understand the quality of the underlying data and related administrative issues that might impact on the data and to continuously seek ways to improve the end-to-end process.

---

[1] Annex 1 and 2 are taken from existing internal documentation.

Welsh Government's Statistical Quality Committee

The Welsh Government's Statistical Quality Committee contains representation from all statistical branches within Statistical Services Division. The Committee's purpose is to support the strategic management of quality across Statistical Services. The Committee's role is:

- to take forward the Welsh Government's Statistical Quality Management Strategy

- to promote and advise on the use of the Government Statistical Service (GSS) quality guidance and tools

- to promote and manage the rollout of quality training

- to regularly review quality incident reports on issues that have occurred (e.g. errors) and ensure an appropriate response is issued, disseminating experiences across Statistical Services as appropriate

- to undertake a regular programme of peer reviews of official statistics outputs

- to review and challenge quality assurance processes in the production of both official statistics outputs and internal work such as resource allocation and analytical modelling (in line with guidance contained in the Aqua Book: guidance on producing quality analysis for government)

- to review and challenge disclosure control risk assessments for statistical outputs

- to review and challenge arrangements in place for the quality assurance and auditing of administrative data sources used in the production of official statistics

- to identify potential statistical special events impacting on Wales

- to provide support in statistical quality matters to other official statistics producers outside of Statistical Services in Welsh Government and its arm's-length bodies

- to review emerging assessment findings (within and outside Welsh Government) and consider implications.

In addition to its internal Statistical Quality Committee, Welsh Government also plays an active role in the Government Statistical Service Quality Champion Network

Statistical services 'Special events' policy

Statistical special events are events which are identifiable, do not recur on a regular cycle and which have at least the potential to have an impact on statistics (for example an event that leads to a brief discontinuity in a time series or impacts on data quality/responses).

In line with the Office for National Statistics (ONS) policy on special events and Government Statistical Service guidance (PDF) on special events, we will determine whether to treat an event as 'special' with regard to:

- whether it has a general effect across a number of outputs

- whether it is restricted to one (or at most a few) periods

- whether the effect is, or is likely to be, noticeable (specifically, any effects which are difficult to distinguish from the normal variation in a series will not be deemed special)

- the views of users

- its impact on Wales.

It is expected that special events will be uncommon and it is possible that special events for Wales may not be deemed 'special' for the UK and vice versa (for example the London 2012 Olympics and Paralympics were special events for the UK but not for Wales as the effects were not noticeable in statistical outputs about Wales).

Welsh Government's Statistical Services Division will routinely identify events which may affect the series it publishes via its Statistical Quality Committee. In these circumstances, commentary on them will be provided alongside published figures as appropriate. Those special events that are identified as being in addition to the ones listed by the ONS will be shown below.

There will be co-ordination across the affected outputs to gather and summarise the available information. This may include co-ordination with other official statistics producers (such as notifying them of a special event or obtaining additional data from them).

Statistical Services, Welsh Government

Email: stats.info.desk@gov.wales

Version: 4.0

Last reviewed: March 2018

Next review date: March 2019

Annex 1 - Checking data

What does this section cover?

The aim is to minimise the number of errors which can occur at any stage of the processes in getting data from providers to the end user. This section draws together a number of general points covering a wide range of circumstances of data checking. The focus is on statistical data, but many of the principles apply to factual information of any sort that might be provided. Other types of errors in our outputs, such as poor spelling or grammar, can undermine customers' confidence in us and our work.

Why check data?

Because it is fundamental to the service we provide. If users of statistics cannot rely on the data we give them it is no use to them and our reputation suffers. In certain areas, such as resource allocation, the consequences of providing wrong data can be expensive and impact on the delivery of services to the people of Wales. It can also have a severe impact on trust in official statistics as well as government in general.

All Welsh Government official statistics should be subject to a checking process.

The nature of the checking process will vary depending on the circumstances. Ideally, it will be documented, with key stages signed off by individuals where appropriate (e.g. for work relating to statistical releases, bulletins and reports). The most basic is the reasonableness check by the person providing the data. This may be the only one available in some circumstances (e.g. an urgent telephone request), and is often heavily dependent on an individual's knowledge and experience, but is a critical part of quality assurance.

Make checks as independent as possible

Where possible data or calculations will be checked independently by another member of staff, and will be cross-checked with other data sources. Asking a colleague from another statistics branch to provide an independent check is also recommended where proportionate to do so.

Make the level of checking commensurate with the use of data.

There will never be enough resources to make all checks exhaustive. We take into account the use of the data when determining the number and extent of checks required for a particular piece of work, to ensure the quality is fit for the purpose intended.

Comparability and Coherence

We check against other published sources, either from the Welsh Government or other organisations such as the Office for National Statistics or the Department for Work and Pensions. This may either be exact checks (e.g. against the same data source) or sense checks (against similar data sources, but perhaps with different coverage/definitions/time periods).

Final checks should be applied as closely as possible to the source data.

The final data should be cross-checked against the original source where appropriate.

Users should be given advice on the level of checks applied to the data. This is particularly important if checks have not been as comprehensive as they could be, perhaps through lack of time or other resource constraints. But this does not absolve statisticians of the responsibility of providing the highest quality data available under the circumstances.

Assume the data being checked has an error in it.

There is the risk that if we assume that data we are checking is right, because we trust the person providing it, we won't be as alert to mistakes as we should be. For this reason, checking should be carried out assuming errors are present as default.

Build time for checking into all projects.

There is absolutely nothing to be gained in rushing a process through to meet a deadline only to find that the errors that slip through cause embarrassment and major problems to the user, and a subsequent cost to resolve the consequences.

Use IT to make checks as efficient as possible.

For example, using a spreadsheet to check row and column totals independently can often save the need for manual checks of individual numbers. But…

Beware of spreadsheets

Errors can be caused by an incorrect formula in a cell, a badly executed 'copy and paste' or any number of other pitfalls, such as ensuring links between documents are maintained. Databases are usually a more secure way of holding data and less vulnerable to many of the errors that can occur with spreadsheets.

Beware of visual checks

We often see what we expect to see. An error can slip past two or more people, even though they think they have checked the data carefully.

Show an intellectual curiosity for what we are publishing (or treat all data with suspicion)

We should be questioning our data and the reasons for changes, by considering other sources, sense checking or talking to data suppliers.

As well as being a safer vantage point from which to undertake a checking process (see above), it can save major embarrassment when simply providing data from apparently authoritative published sources. This also supports us in understanding the data and what it is telling us, if we understand the reasons for unusual movements in the statistics.

Promote a culture of checking

Managers should challenge staff by seeking confirmation that outputs have been checked and to encourage a culture of quality assurance and checking.

Share good practice and ideas

If you have found an efficient way of automating checks, discovered a new pitfall in a process and a way around it, or have tips for minimising errors, share them with others. Look out for ways to anticipate and prevent error and share them.

Recognise the value of checking

Quality assurance is a critical part of the statistical production process. It is a 'proper job' in its own right and should never be undervalued. Time spent in careful, efficient checking is never wasted, even if no error is uncovered. The value added by the effort is the degree of confidence that everyone can put in the data once it has passed all the necessary checks.

Where errors occur, look for process failures rather than individual culpability

Most errors are a result of system failure rather than individual neglect.

Annex 2 - Data validation

What does this section cover?

This section covers various forms of data validation. Data validation is a crucial element of data handling, done to varying levels of detail dependent upon the information being collected. It can be broken down into three key areas, as follows:

- For data being keyed in by hand, it is essential to 'punch error check' the data. This is to ensure that there have been no errors in the keying-in process.

- Data must be arithmetically consistent. For example if one part of a form requests a figure which is then also used in a later part of the form, these figures should match.

- The data should make sense. For example the data should be consistent over time, after allowing for any functional changes. We should be questioning the data and wanting to understand the reasons for any changes over time, or between categories.

The aspects above are covered by this annex although much of the associated work can be significantly reduced by means of a well designed automated electronic data collection process.

Electronic data collection

Well designed electronic data collection and management systems can remove the need for punch error checking, and often some (if not all) arithmetic validation and even some sense checking.

Spreadsheets

Data collection spreadsheets can have automatic data validation set up on specific cells. There can also be a data validation worksheet which brings together relevant checks within the spreadsheet that would otherwise be performed after the form was submitted.

Spreadsheets can also be pre-loaded with data from other data collections to compare with current data and providers can be asked to provide an explanation when a set validation check fails.

Spreadsheets should be loaded into databases using automated processes where possible (as opposed to manual processes) to limit the potential for errors.

SQL databases

Structured data management following collection can improve the quality and validation of data. Where appropriate, data should be stored in a 'star schema' format to allow total flexibility in querying and validation.

Data should be stored in SQL database tables with strict permissions on who is allowed to access or amend data. The database tables should be well designed with primary keys to ensure referential integrity and there is no duplication of data.

Data should be imported into SQL databases from electronic data collection sources automatically with no need for any manual cut-and-paste operations.

When changes to data are required, the old data records should be deleted and the amended data collection form re-imported rather than manually updating data.

When a data collection has ended, the data can be moved and stored in a read-only SQL table to ensure that the finalised data does not change.

Data exchange

In some areas there are automated systems which follow the same principles of validation at source. For example, the Data Wales Exchange Initiative (DEWi) provides real time, on line validation of statutory returns from schools and local authorities, shifting the burden of data cleansing from the Government onto the data suppliers. As soon as a file is uploaded the supplier can view a detailed report listing all the errors in the return, together with some information on possible reasons for the error and what action they need to take to correct the error.

Further, the validation routines within DEWi are integrated into the management information systems of all schools in Wales, facilitating consistent data validation across the whole sector and allowing returns to be validated, if necessary, independently of DEWi. Some of the validation rules are also used at the point of data entry, to prevent invalid or incorrect data ever being recorded.

We also use the AFON service which extends the ideas above to datasets other than education and also for the secure transfer of electronic data files. The greater the amount of data that can be collected using data exchange methods the better as they can reduce the need for additional validation after receipt of the data.

Punch error checking

All data that is manually captured (hand punched as opposed to electronically captured) should be checked against the data source wherever possible.

Data captured should be 'double keyed' (two separate operators) to minimise the error rate – keying methods such as 'single keying' (one operator) and 'double verification' (one person keys and a second verifies the data) do not provide sufficient confidence in data quality.

However, one cannot assume that the data returned back in its raw format is correct until tertiary data quality control checks are done. Examples might be consistency checks on known relationships between data in a particular row of column or row of numbers, or comparisons with previous year's data and these checks

can be built into holding spreadsheets (or into databases, SAS or other software packages).

Arithmetic validation (sometimes referred to as internal consistency checking)

This should only take place once the user is happy that the data has been input correctly.

There are clear reporting implications if an identically defined quantity appears twice or more on the same form but the numbers reported against it differ. At first, the idea of the same quantity appearing twice or more on the same form seems unnecessary, but the following examples illustrate the possibility.

There is often more than one way to break a total down, for example total pupils may be broken down by year group and ability to speak Welsh. It is unnecessary and burdensome to ask for the two-dimensional detail of the breakdown, instead two separate breakdowns suffice, in which case the totals of both should match.

A figure in one part of a form may feed through into a calculation later in the form, for example the total expenditure on education is part of total local authority expenditure. The total figure taken from the education part of the form must agree with the figure recorded in the summary of total expenditure later in the form.

Arithmetic validation can be used in cases such as these to avoid the potential pitfalls of reporting the same quantity as different amounts. For data held in databases or well designed spreadsheets, arithmetic validation is best handled by making simple comparisons via queries or basic spreadsheet tables.

Arithmetic validation can be performed within the electronic data collection spreadsheet. This allows users to spot potential errors and correct them or provide an explanation before submitting the spreadsheet.

Sense checking (sometimes referred to as external consistency checking)

Data should always be subject to some kind of sense checking. This should take place after punch error checking (if relevant) and arithmetic validation has been concluded. Essentially the process is identical to arithmetic validation in that queries or spreadsheet tables should be set up to compare the data against whichever item is relevant. The following are examples of possible sense checks:

- check of consistency over time

- check of consistency with other data collections

- check of consistency between the data of different providers

- check of completeness, that is data is not blank where figures are expected

- checks on the data itself, such as data being of the correct sign, within certain control figures or non-zero.

We cannot rely on automated sense checking. Staff producing the data, and the responsible statistician should be questioning the data as if they were an external user wanting to interpret the information. We should be interrogating the data to ensure that we are happy that the information represents reality as far as possible.

- Why has local authority x seen its number of something halved?

- Why has the overall total gone up/down by that amount?

- Why has the age/mode/sector profile of this category changed significantly since last year?

- What is the story here? Is that a real change or a function of the data?

Part of sense checking will be asking questions of data suppliers to see if they can explain the changes. There have been times when we have asked a data supplier about a change and they have suddenly revealed a change in process, or that the data was simply provided incorrectly.

If we can satisfy ourselves with the answers to such questions then we are also in a better position to interpret the information and improve our commentary and metadata.

Validation sign-off with data suppliers

Where time permits we should ask data providers to sign-off data by sending a certification statement. This is especially important where the contents of the form have large financial implications.

How to handle errors

Where differences are discovered, it is necessary to contact data providers to correct the errors but it is important to note that changing one figure, particularly a summary total, will often require other figures to change to maintain the internal consistency of the data. It is therefore useful to have a good understanding of the data and thus an idea of what might be done to correct the error before contacting the data provider. It is imperative that inexperienced members of staff are offered suitable instruction before being asked to carry out this element of the exercise.

It is best to carry out all correspondence in respect of arithmetic validation via e-mail. If correspondence is entered into via the telephone, it is essential to ensure that accurate records are made of conversations and paper records are kept up-to-date. All changes should be well documented with names and dates.