Using Administrative Data: Good Practice Guidance for Statisticians

Interim Guidance, April 2014

National Statistician's Office

This note sets out some guiding principles around the use of administrative data, with a particular focus on the statistician's role assuring and communicating the quality of administrative data used to produce official statistics. The guidance is based on existing good practice across the GSS. It is not intended to be treated as a set of rules that everyone must follow: the interpretation of the guidance will vary depending on the circumstances in which the administrative data are collected and used. However, it is reasonable to expect that everyone who uses administrative data will have thought about the topics covered in this guidance, for example by considering whether or how the good practice applies within their own specific context and in proportion to the importance of the resulting statistics.

Table of Contents

Introduction

Statisticians across the GSS use administrative data in the production of official statistics. There are good reasons to base statistics on administrative data sources: the data often provide a readily available, rich source of information, without placing any significant burden on data suppliers. However, they are also susceptible to data quality issues.

Common data quality concerns with administrative data include incomplete data, incorrect data format and mis-typed data. There can also be variation in recording practices between suppliers who contribute to an administrative data source. Some of these issues arise because the data have not usually been collected with statistical purposes in mind. There have also been some high profile cases where statistics based on administrative data have been found to be affected by intentional misreporting of the underlying data. Any of these problems not only affects the usefulness of the resulting statistics, they can lead to mistrust in the statistics, adverse publicity and in the high profile cases there have been parliamentary inquiries and a de-designation.

Statisticians who use administrative data to produce official statistics are responsible for the quality of their statistics, but they cannot physically check that each bit of administrative data they use is input correctly at source. The Code of Practice is the rule book for producing all official statistics including those based on administrative data. This note sets out some guiding principles and good practice around the use of administrative data, to mitigate the associated risks. It outlines the reasonable steps that statisticians can take to assure and challenge the quality of administrative data and the information that should be published alongside the statistics.

What is administrative data?

For the purposes of this guidance, a data source is administrative if the data it holds weren't primarily collected with statistical purposes in mind. Any data drawn from such a source is **administrative data**. Administrative data may be drawn from internal systems or supplied by external organisations. Administrative data sources are often used to manage the day-to-day operations within an organisation or to deliver a service. A lot of financial data is administrative.

Who is this guidance for?

This note is aimed at all statisticians who use administrative data. It is particularly relevant for those looking to use a new administrative data source or where administrative data have been used for some time without being reviewed or critically appraised. The guidance will help ensure that the requirements of the Code – and the expectations of our users – are met.

What does this guidance cover?

The guidance in this note is based on good practice that is going on across the GSS, with a focus on the statistician's role assuring and communicating the quality of the data.

The guidance is structured around what good practice looks like at each stage of using administrative data:

- Before using administrative data
- Working with data suppliers
- Validation
- What information to publish
- Managing the ongoing use of administrative sources

The guidance sets out the importance of understanding the data that you are using; seeing how it is collected on the ground; establishing effective relationships with data suppliers; documenting quality assurance processes; knowing about the audit arrangements; publishing clear information about quality; and keeping informed about (and influencing) changes to the data collection systems.

This guidance focuses on what good practice looks like when using administrative data to produce a statistical release. Increasingly there is pressure to publish more data into the public domain (against a backdrop of diminishing resources) through Open Data and Freedom of Information; often this is not through formal statistical releases. Clearly, this guidance will not apply equally across the different release scenarios and a proportionate, risk-based approach is necessary.

This guidance does not cover practical issues around permissions, data transmission and methods, which are often specific to the administrative dataset being used.

Why is this guidance described as "Interim"?

In February 2014 the UK Statistics Authority's Monitoring and Assessment team (M&A) announced a programme of work about administrative data and official statistics. This work will look at how administrative data are audited, the principles that should apply to the use of administrative data and the issues that arise when official statistics are used as the basis of performance targets. M&A has seen a copy of this guidance note and has confirmed that the guidance is consistent with the initial views emerging from its work. The guidance will evolve when M&A completes its work in summer 2014.

How should this guidance be used?

This guidance is intended to be used as a tool in the months before M&A publish their findings. It is envisaged that statisticians and Heads of Profession will want to seek assurance from their teams regarding their use of administrative data, to establish strengths and weaknesses and to put in place plans to address any issues. The guidance is a tool to support this, providing a clear understanding of what current good practice looks like.

Remember that the guidance won't apply in every situation: for example, how you engage data suppliers will vary greatly depending on how many suppliers are involved and how complex the data supply chain is. And the steps you take will need to be proportionate to the importance of the statistics and reflect the resources you have available.

There is a one page checklist that describes the key issues and the final section of this guidance includes some case studies, which illustrate in more detail how problems can arise and how the good practice in this guidance has been put into practice across the GSS.

1. Before using administrative data

Statisticians who use administrative data to produce official statistics are responsible for ensuring that the statistics they publish are of suitable quality for the main intended uses and they must publish clear, open information about the quality of the statistics, as set out in the Code of Practice. However, they cannot be responsible for the accuracy of the underlying administrative data. The data supplier is responsible for ensuring that the data are of appropriate quality *for their own administrative purposes*. It then becomes the statistician's responsibility to investigate the data to determine whether they are of sufficient quality to be used in the statistics and to determine and communicate the range of appropriate uses of the statistics.

Before using administrative data (or equally, if an administrative data source has been used for some time without a critical review) it is important to fully understand the data collection and to consider the scope for errors to arise in the data. It is better to do this as a conscious step early on in the process rather than rely on quality assurance checks when the data are being used.

Adopt a critical mindset when considering the scope for (intentional or unintentional) misreporting. Be aware of the environment in which the administrative data are collected and used – particularly the existence of any targets. Your internal audit team may be able to help think this through and identify the requirements for data checks at different points in the data supply chain.

a) Investigate the source, asking a set of basic questions, such as:
    i. Why are the data collected? Does this purpose impact on their usefulness for the statistics you want to produce? Will the resulting statistics match user requirements? Are there any data that aren't included, which users might expect to be included?
    ii. Is there any reason / incentive for people to distort the data they supply? Are there any related targets, performance measures or payment by results systems? Are there pressures other than targets, such as public, media or parliamentary attention?
    iii. What is the scope for people to misreport the data, at different points along the data supply chain (unwittingly or otherwise)? To what extent is data collection automated? How many people are involved in data supply? (how many pairs of eyes on the data?)
    iv. Is there any (external / internal) audit of the data? What have audits reported? Audits provide a key means of assuring the quality of the data.
    v. Are there other users of the administrative data? What checks do these users carry out? Do they have any accountability for the accuracy of the data?
    vi. Is there any other source of assurance about the quality of the data? For example, regulations or instructions governing the data standards, definitions, recording practices.
    vii. Are there any alternative related data sources, for comparison purposes?
b) Investigate the practicalities of data supply, including data permissions, security classifications, handling of any personal data, data access agreements.
c) Request a data download and begin to analyse the data to understand it, to discover the issues that may affect the statistics but which are not identified as a concern for the operational use of the data.
d) Only use the administrative data to produce statistics if they are judged to be a good fit to the user requirements, in terms of what the data measure and the integrity of the data. This needs to be an honest appraisal, which can only be taken with a detailed understanding of users' needs.

2. *Working with data suppliers*

Good relations with those who supply the data are critical to ensure that there is a common understanding about the origin of the data, its suitability for use within the statistics and the ongoing supply of data that meets the standards required for the statistics. How you manage these relations, and the extent to which you can implement the good practice below, will vary greatly depending on how complex the data supply chain is.

It is important to remember that the data are being collected for an operational purpose, and the statistical use is secondary, so your use of the data may be seen as an unnecessary burden. When establishing relationships with data suppliers take account of the different values and motivations of those involved in the data supply chain. They may not normally prioritise data accuracy if it is not necessary to achieve the desired business outcome.

a) Ensure that those supplying the data understand what the data are used for, so that they are able to provide any relevant information that may be important for the use of the data, and to emphasise the importance of supplying good quality data.
b) Work with the data supplier to ensure you fully understand the data supply chain and the checks and responsibilities along the chain. Produce an end-to-end process map to illustrate this. Please note overlap with the set of questions in section 1 (a) above.
c) Where there is more than one version of the administrative data (for example, unit level and aggregate) use the version for which the supplier organisation has greatest incentive for it to be accurate. Sometimes this may be the raw, untouched unit level data; otherwise it may be the aggregate data, which has been cleaned and checked. As far as possible, align the statistical requirements (in terms of timing, data format) with the relevant source.
d) Visit (one or more of) the data suppliers to see how the systems are being run 'on the ground', to identify any issues which may impact how you can use the data. This visit can be a useful opportunity to promote the importance of the data, to discuss the standards expected and to raise the profile of statistics within the supplier organisation.
e) Establish clear agreements with data suppliers describing the format, content and timing of the data (and any metadata) required and expectations about the quality of the data, including any quality assurance processes that are the supplier's responsibility. Although the data are not collected for statistical purposes, there will still usually be basic quality standards required by the supplier, which can be built into agreements.
f) Agree processes and allocate responsibility for ensuring that the data quality standards have been applied and agree an acceptable route to challenge the quality of any data supplied. It is important to have a good relationship with the data supplier but you must also be able to discuss any data quality concerns with them.
g) Ensure that data suppliers know how to raise questions and issues if any arise and where necessary provide guidance and training to enable them to submit data accurately.

Other good practice, which will apply in many – though not all – situations:

- Provide reports on 'quality' back to data supplier, based on the number of queries raised and the extent of revisions during quality assurance. The supplier may choose to include these reports in their performance management process. In some cases it may be appropriate to publish supplier-level information about the quality of data supplied.

- Spread good practice amongst data suppliers (when there is more than one).
- Where appropriate, make it the data supplier's responsibility to sign off the final data submission – taking responsibility for its integrity. Usually, the more senior the sign-off, the better: this can help raise the profile of the importance of submitting good quality data.
- Where possible, work with software suppliers who provide the administrative systems, to agree the specifications for extracts and the validation rules and integrity checks, so that when the data arrive validation is already complete. The potential gains to be made by investing resource in influencing the design of systems cannot be overstated.
- Identify ways to raise the profile of statistics within the supplier organisation, highlighting the importance of submitting good quality data and explaining how it is used.

3. Validation

Administrative data are particularly susceptible to data quality issues because usually they have not been collected with statistical purposes in mind. Common data quality concerns with administrative data include incomplete data, incorrect data format, mis-typed data and variations in recording practices between those who submit data. Data from administrative systems should never be assumed correct – they should always be validated before further analysis.

It is vital that data quality concerns are followed up with the data supplier, to understand their cause and to mitigate any effect on the published statistics. If you have any questions or concerns about the quality of the data, it is your responsibility to voice them and escalate if necessary. It is also the statistician's responsibility to publish relevant information about the quality of the statistics, so record any concerns or issues with the data to refer to when drafting the statistical release.

a) Understand the possible causes of error in the data from the supplier, including clerical error, processing error and misreporting. Where possible, develop mitigation strategies for these errors, including validation routines that identify and deal with suspicious values.
b) Undertake basic validation of administrative data supplied for each release, looking for outliers, limits, unusual trends or consistency within the data and with external sources where possible. Ensure that any questions raised are followed up with the data supplier. Where necessary, be prepared to question whether the statistics are fit for purpose.
c) Maintain a log of any data concerns or issues, to inform the quality information that will be published alongside the statistics (see next section).
d) Keep a copy of the original data where possible, and maintain an audit trail of the changes and data processing, to aid any future analysis of data errors.
e) Triangulate administrative data with other external sources where possible.

Other good practice, which will apply in many – though not all – situations:

- Establish automatic validation checks at the point of data supply, so the supplier becomes responsible for validating their own data.
- If more than one team / organisation use the data, take advantage of this, but ensure the process is joined up: more pairs of eyes on the data is usually a good thing, but can cause problems if not properly coordinated.
- It may be helpful to revisit the questions in section 1 (a) periodically.

4. What information to publish

The importance of publishing clear and open quality information to accompany the statistics cannot be understated. The information about quality should reflect an honest, critical appraisal of the limitations and caveats associated with using data from the administrative source, and any specific quality concerns identified through data validation. Address any distortion in the data that may be caused by targets, performance measures, payment by results systems or other pressures, including any known 'gaming' (intentional misreporting of the admin data, for example in order to achieve a target). It may be uncomfortable to acknowledge openly, but it is better to be open and (if possible) estimate the likely extent, than ignore the existence of these pressures and be accused of publishing misleading statistics further down the line. Being transparent about the issue may also provide incentive for organisations to drive up the quality of the underlying data.

a) In the published statistics, supply open, clear, messages about the quality of the underlying administrative source. Where there are any concerns about the quality of the data these should be addressed at the start of the publication, not consigned to a note at the back of the publication. It is good practice to include:
    - A short, clear description of the source.
    - Why the data are thought to be fit for use in the statistics.
    - Explain the extent of appropriate use of the statistics, including the potential impact that any quality issues may have on use.
    - How you have assured yourself about the quality of the underlying data.
    - Any remaining limitations and weaknesses that might affect the interpretation of the statistics, including any definitional issues, missing data or other issues identified through data validation.
    - Where there is known or assumed impact from targets and other similar pressures, including 'gaming', publish information about this.
    - Where possible, estimate the likely extent of any data quality issues. If it is not possible to quantify these errors, provide reasoned judgement about likely magnitude and direction.
    - Explain any system changes which have an impact on the continuity or interpretation of the statistics.
    - Provide an end-to-end process map that illustrates the relationship between the data and the resultant statistics and how quality is managed at each stage in the supply chain.
    - If necessary, link to more detailed information about methods and data validation.
b) Publish the detailed data if possible – allowing wider scrutiny.
c) Maintain and consider publishing other documentation that demonstrates you have properly considered and addressed data quality concerns. For example: a quality assurance plan; a review of possible errors or risks, including any pressures that may lead to gaming, and how these risks are mitigated; details about data validation; agreements with data suppliers regarding standards and responsibilities; other sources used for comparison purposes.

5. Managing the ongoing use of administrative sources

There is a risk, when using a data source whose primary purpose is operational rather than statistical, that changes will be made to the source or systems to meet operational demands without considering the needs of, or impact on, the statistics. Changes may affect the format, content or timing of the data. If the statistician is not able to influence the changes they can affect the timing and continuity of the published statistics – or in the worst case, if the statistician is not aware of the changes, they can feed through to the published statistics as unexplained (and misleading) movements in the series. It is also important to know about changes to the environment in which the statistics are collected and any findings from relevant audits.

Engagement with the data supplier overseeing the administrative system and also with relevant networks at a working level will help you be aware of issues affecting the data which may need to feed into the commentary that accompanies the statistics.

a) Monitor the administrative source and the data collection environment, to ensure no changes are introduced that impact on the use of the data in the statistics without statisticians being aware of the changes.
b) Wherever possible, statisticians should sit on the relevant governance boards and or change control groups for the administrative systems, so that statistical needs are considered in the design, management and development of the systems. It is usually cheaper to build statistical requirements in to systems at the outset than amend systems in retrospect.
c) Develop strong links with other organisations that use the same administrative data, to share concerns and work together to ensure quality. Often there is shared responsibility for data quality: for example, if the data are audited and / or are used officially elsewhere (by Local Authorities / Boards etc) then the responsibility for ensuring its integrity should be shared. There is a danger where more than one organisation shares accountability for the integrity of the data, that none assumes responsibility: each assumes the other takes the lead. This makes it important to develop strong links between the organisations. There might also be a case for formalising respective roles in such instances.
d) Know about the audit arrangements for the data that you use, including the timing, content and frequency of audits, and establish processes with relevant audit organisations to ensure that the outcomes of relevant audits are fed through to statistical teams.
e) Take a strong role raising any concerns about the underlying data and take steps to address those concerns. For example; raise concerns with the supplier organisation, your Head of Profession or the National Statistician's Office; identify priority areas for audit at a working level or ask for senior management (or National Statistician or M&A) to influence where audit is targeted. This might include external audit or a supplier organisation's own internal audit.

Other good practice, which will apply in many – though not all – situations:

- Visit (one or more of) the data suppliers periodically to check that your understanding of how the system works 'on the ground' is accurate and that the data collection and processing is in line with expected standards.
- Exchanges, secondments and shadowing arrangements between the statistics producer and the administrative data supplier can be hugely valuable, enabling both organisations to view

the data processes from the other's perspective. This can help move towards more effective working arrangements. There is a risk that the secondee may be 'lost' to the supplier organisation if they choose to find a new job there, but this can still be seen as a positive outcome – they can become a valuable sounding board in future.

- Maintain a log of issues that arise during data supply, which all those involved can contribute to, and use this to draw out and address any recurring problems, for example with processes, systems or software.
- It is important that statisticians are involved in target setting in your organisation, so that targets are developed with full knowledge of the extent and limitations of the statistics that will be used to measure them.

6. *Examples* and case studies

Interpreting the Data – Crime Statistics

Administrative data are not usually collected for statistical purposes, so there can be a mismatch between their original meaning and the final statistical use. It is important to understand where the data are coming from, what the systems were designed to collect and how they do so. Will the resulting statistics mean what you think they mean? Similarly, but separately, will the resulting statistics mean what users will take them to mean?

It is important to consider the impact that performance measures or targets can have, when interpreting the data. Performance measures can distort the data, by encouraging the focus of effort on a particular activity, which may affect not only the data for that activity but also for other activities, which may suffer from a corresponding drop in the level of effort.

For example, crime statistics are produced from police administrative sources. Taken at face value and published as official statistics these may be used as a measure of the amount of crime. However, it is not at all clear that the user's interpretation of what constitutes a crime will match those that are eligible to be recorded on the database. At the simplest level there may be definitional differences to take account of. The data may also reflect police activity rather than the level of crime: police focusing on tackling a particular crime or locality may lead to more offenders being identified and more crime being recorded – but this may not reflect any real change in the level of crime. The statistician needs to understand such issues and then clearly communicate to the users the appropriate uses of the published statistics.

Data Integrity

There are two aspects of data integrity in relation to administrative data: unintentional misreporting and intentional misreporting. Both can be a significant issue with administrative data – more so than is typically the case with survey data, for example. For administrative data, ensuring integrity means adopting a critical, questioning stance rather than accepting the data we are given at face value.

Unintentional misreporting describes a situation where somewhere along the data supply chain there is a genuine and persistent misunderstanding about what data is required (in one or more particular variable). In some cases this will be picked up by data validation routines, but not always: if the data being entered are wrong, but nevertheless within the expected tolerances for the data variable – for example because they are the right data but for the wrong time period. Unintentional misreporting might be quite small in scale, but equally, if it has been caused by some confusion arising from the data requirements / specifications / training (or lack of) then it could be widespread – and probably then harder to spot.

Intentional misreporting is usually a response to a performance measure or payment system of some nature, either formal or informal. The existence of any form of target creates an incentive to misreport, to 'game' the system, particularly if the consequences are perceived as small and the rewards greater. Those inputting administrative data or involved collating the data further along the chain may feel pressurised to meet internal targets, they may be aware that the resulting statistics are used as the basis for financial reward or used to monitor their organisation.

The extent of the assurance statisticians seek about the quality of administrative data needs to be proportionate to the expected use of the statistics and the likely impact of any integrity issues. An element of 'gaming' might be quite widespread, but in fact, even if it exists it may only have a very negligible impact on the published statistics. Conversely, if statistics are used to measure progress against an 'absolute' target (e.g. no patient must wait longer than….) then *any* element of gaming is a concern – it doesn't have to be widespread.

The number of people putting data into the administrative system can be a factor in how difficult it is to misreport data intentionally: the more pairs of eyes involved, the greater the likelihood of someone spotting misreported data. However, if there are more people involved it also makes it more difficult to train everyone to supply good quality data; it is more difficult to establish a relationship with data suppliers; and potentially harder for the statistician to spot misreporting if it does happen.

Often, the nature of administrative systems means that the statistician has no direct control over the data collection environment and little leverage where potential data integrity issues are identified. However, the statistician can (and should) investigate and understand the likely extent of any such issues, by asking the sorts of questions identified in section 1. In most cases, the statistician will also identify that there are suitable assurances, such as audit arrangements, comparison with alternative sources and the statistician's own quality assurance arrangements, which give confidence in the statistics being produced, so that the statistician can in turn give suitable assurance to the users. Where there are any remaining concerns it is the statistician's role to be clear about the limitations in the published statistics and to explain the range of uses that the statistics will support.

Public Sector Finances, HM Treasury and Office for National Statistics

ONS and HM Treasury produce public sector finance statistics based largely on administrative data from departmental accounting systems. This is a more controlled environment than would be the case for many other administrative systems, with departmental accounts audited annually by the National Audit Office, but it is useful to consider as an example of good practice in some respects.

- Two departments are involved in the production of the statistics (ONS and HMT) so there are two sets of eyes on the data. Both departments can access the underlying data system to carry out checks.
- The monthly statistics are produced from data recorded on an accruals accounting basis. Data are also collected on a cash accounting basis, which provides a useful comparison source. Any significant discrepancies can be identified and investigated.
- Statisticians are involved on the boards that oversee the accounting system, for example the Financial Reporting Advisory Board.
- HMT publishes the raw data to open it up to scrutiny. The data are so detailed that it is questionable how much effective scrutiny this introduces, but it might provide greater incentive for accurate reporting.
- HMT provides guidance and training for users of the administrative system. As part of a project to update the data provision system, HMT worked with each department to quality assure the mapping from their departmental accounts system to the HMT system to minimise transcription errors.
- HMT provides each department with a quality report each month, describing the accuracy of the forecast data, the timeliness of the submission and the extent to which the data need to be adjusted. This is supposed to be incorporated into the performance reviews of the department's Finance Director and Permanent Secretary.
- HMT is aligning the data requirements and timing with the relevant departmental Board's requirements to ensure it receives the most accurate data available.

7. Other Useful Information

Conflict between official statistics and management information is covered in detail in the National Statistician's Guidance: Use of Administrative or Management Information.

There are some other examples of good practice in *Examples of Practices Consistent with the Code of Practice*, published by the UK Statistics Authority's Monitoring and Assessment team.

ONS's Quality Methods and Harmonisation Tool can help you think through some of the issues when using administrative data, particularly how closely the data will match user requirements.

Eurostat's ESSNET administrative data project has produced useful information including:

- Common sources of error in administrative data;
- List of basic quality indicators for administrative data;
- International good practice examples for communicating with data suppliers;

8. Coherence With Other Guidance

This guidance note is intended to build on, rather than replace guidance that already exists. It is intended to be entirely consistent with the Code of Practice and to supplement the existing National Statistician's Guidance.

Protocol 3 of the Code of Practice specifically addresses the use of administrative sources for statistical purposes. This guidance takes full account of the requirements of Protocol 3, whilst also drawing on other Principles and Practices from the Code, particularly in relation to understanding user's needs, assuring quality and frankness and accessibility.

The National Statistician's Guidance: Use of Administrative or Management Information focuses on tensions between Management Information and Official Statistics. It also emphasises the need for statisticians to understand the underlying administrative systems; to publish transparent information about quality; and to create and maintain a process map that illustrates the relationship between the data and the resulting statistics. It doesn't explicitly talk about the potential for 'gaming' or suggest ways to investigate this further. There is an annex that provides a template for good metadata, which could helpfully be updated as it doesn't currently include any specific mention of data integrity.

The National Statistician's Guidance on Quality, Methods and Harmonisation says the following:

- "Where statistical releases are dependent on administrative data sources, the statistical requirements of the administrative systems should be promoted and risks to the integrity of the statistics minimised."
- "Statistics should be derived either from data collected objectively and independently or from predefined administrative data that follow established procedures and are auditable."
- "When dealing with statistics produced from administrative sources, process quality measures may be the number of queries from the statistical producer to the administrative data supplier, and the percentage of data items changed during quality assurance."
- "..for statistics produced from administrative sources, output quality measures such as the proportion of administrative records with missing data, documentation of the main uses of administrative data and the known sources of error."

The UK Statistics Authority's Monitoring and Assessment team (M&A) has a programme of work underway on administrative data and official statistics, details of which are published on its website. This guidance note will evolve further as M&A publishes the outcomes of its work. At the point of publication, as far as possible this guidance note is intended to be consistent with M&A's initial views, as set out in three sources:

- the documents published alongside its programme of work on this topic;
- "Principles to guide the Statistics Authority's assessment of quality assurance practices relating to statistics produced from administrative data" circulated to Heads of Profession on 15 April 2014 and published on the GSS website; and
- an email to Heads of Profession on 7 Feb 2014 the text of which is repeated below.

M&A's initial view on what it expects of producers (7 Feb 2014 email to Heads of Profession):

"In cases where official statistics are produced from administrative sources, and particularly where responsibility for the application of rules, classifications or definitions rests with a number of organisations, we will look for evidence both that appropriate quality assurance processes, including statistical quality audits, are conducted; and that the statistical implications of this quality audit work are made clear in order to help the user interpret the statistics.

…the main things we would expect a producer to do include:

- Review the arrangements for auditing the quality of administrative data to ensure their suitability for use in producing official statistics.
- Engage with external experts to review and improve the management of administrative data, and the production of official statistics from them.
- Publish a process map of the statistical production process, and describe the way that quality is managed at each stage.
- Document the risks associated with data recording in the administrative process, and identify a mitigation plan. Include within this the range of errors that can arise including clerical errors, processing errors and those arising through 'gaming'. Document the pressures that could lead to the manipulation of data, and the steps that are taken to minimise those.
- Describe, simply and prominently, the strengths and weaknesses of the statistics (including non-sampling and sampling errors), in relation to their use and potential use. Quantify (estimate) the size of those errors where possible and present them clearly. Where the size of errors can't be estimated, provide reasoned judgment about the likely magnitude and direction of the various errors, and the potential impact on the use of the statistics. Consider an inherent bias in the statistics. Describe the impact on the statistics of any changes to the administrative processes.
- Document the validation procedures applied to the administrative data, including additional analyses undertaken, triangulation against other data sources etc.

But this is not definitive, and is intended to be an invitation to discussion."

Using Administrative Data: Good Practice Summary Checklist

There is more detail on each of these issues in the relevant section of the guidance.

1. ***Before using administrative data*** *(page 4)*

- ☐ Investigate the source, asking a set of "challenge" questions (page 4) and analyse a sample of data to ensure that you understand the data supply chain and the scope for misreporting.
- ☐ Establish the extent of audit and other quality assurance procedures.
- ☐ Only use the administrative data if they are judged to be a good fit to the user requirements.

2. ***Working with data suppliers*** *(Page 5)*

- ☐ Ensure that those supplying the data understand what the data are used for, the importance of supplying good quality data and how to raise questions and issues if any arise.
- ☐ Provide relevant guidance and training to enable data suppliers to submit data accurately.
- ☐ Visit (one or more of) the data suppliers to see the data collection 'in real life'.
- ☐ Where there is more than one version of the administrative data (for example, unit level and aggregate) use the version for which the supplier has greatest incentive for it to be accurate.
- ☐ Establish clear agreements with data suppliers describing the format, content and timing of the data required and setting out roles and expectations about the quality of the data.
- ☐ Agree any checks that you will use to ensure that the data quality standards have been applied.

3. ***Validation*** *(Page 6)*

- ☐ Understand the possible causes of error in the data from the supplier and establish mitigation strategies, including validation routines. Compare data with external sources where possible.
- ☐ Ensure that any questions raised are followed up with the data supplier.
- ☐ Keep a copy of the original data. Maintain an audit trail of the changes and data processing.

4. ***What information to publish*** *(Page 7)*

- ☐ In the published statistics, supply open, clear, messages about quality, including how you have assured yourself about the quality of the underlying data, any limitations (including missing data and known or assumed 'gaming') and the extent of appropriate use of the statistics. Address any significant concerns about quality at the start of the publication. See complete list on page 7.
- ☐ Produce and publish an end-to-end process map illustrating the data supply chain and the checks and responsibilities along the chain.
- ☐ Maintain and consider publishing other documentation that demonstrates you have properly considered and addressed data quality concerns.

5. ***Managing the ongoing use of administrative sources*** *(Page 8)*

- ☐ Monitor the administrative source and the data collection environment, to ensure no changes are introduced that impact on the use of the data in the statistics without statisticians being aware.
- ☐ Wherever possible, statisticians should sit on relevant governance boards or change control groups.
- ☐ Develop strong links with other organisations that use the same administrative data, to share concerns and work together to ensure quality.
- ☐ Know about the audit arrangements for the data that you use and ensure that statistical teams are aware of the outcomes of relevant audits.
- ☐ Take a strong role raising and addressing any concerns about the underlying data – with the supplier team / organisation, HOP, National Statistician's Office or audit organisations.