

Administrative Data Quality Assurance Toolkit

Version 1 January 2015

Administrative Data Quality Assurance Toolkit

This toolkit is intended to help statistical assessors review the areas of practice for the quality assurance arrangements of administrative data used to produce official statistics, during the Assessment of official statistics against the *Code of Practice*¹ for possible designation as National Statistics.

The Authority is sharing this toolkit to help explain the nature of the regulatory standard that we are applying when we determine the suitability of practices used by statistics producers for the quality assurance of administrative data.

This guidance can support producers in making their critical judgments that administrative data continue to be suitable to be used in producing statistics, and in providing appropriate documentation for users.

The toolkit comprises:

- a) *Quality Management Actions* (page 3) – a model describing an approach that can be used by statistics producers to satisfy themselves and their users of the ongoing suitability of the administrative data
- b) *QA Matrix* (page 4) – which describes the statistical practices in four main areas for ensuring the suitability of the administrative data quality, with three different levels of assurance as some statistics require a more thorough and detailed approach than others
- c) *Risk/Profile Matrix* (page 9) – which can be used to inform decisions about the level of assurance that is appropriate given the characteristics of the statistics
- d) *Top tips and Helpful Guidance* (pages 10 and 11) – additional guidance material to support judgments about the quality of administrative data

Our assessors will:

- Provisionally determine their view of the appropriate level of assurance required for a set of statistics.
- Request from statistical producers:
 - a) an explanation of their judgments of the required level of assurance for a set of statistics;
 - b) evidence to support the rationale for these decisions;
 - c) evidence of the actions, and the rationale for deciding upon those actions, that the producer has taken to achieve compliance with the chosen level of assurance; and
 - d) evidence which demonstrates that the statistical producer has embedded practices for keeping its quality assurance arrangements under review.
- Identify whether the submitted evidence complies with the *Code of Practice*.
- For instances of non-compliance, require statistical producers to take action to meet the standard of the *Code of Practice*.

¹ <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

Quality Management Actions

Assessors will look for evidence that producers have considered the following quality management actions: *Investigate*: *Manage*: *Communicate*. They represent three types of actions for assuring the suitability of the administrative data and in documenting the findings:

- **Investigate:** Statistics producers should investigate, for example, the types of checks carried out by data collectors and suppliers, as well as the operational circumstances in which the data are produced. They should identify any coverage issues and potential sources of bias in the data collection and supply process.
- **Manage:** Producers should also manage their relationships with suppliers by establishing clear processes for data provision and for managing change. They should also maintain regular quality assurance checks of the data and use other data sources where possible to corroborate their findings.
- **Communicate:** Producers should communicate effectively with their data suppliers and others to ensure users are provided with clear explanations of the strengths and limitations of the data. Producers should work closely with other statistical producers using the administrative data to ensure a common understanding of any quality issues.

As illustrated in the diagram (below), these practices are continuous and iterative, reflecting the ongoing use of the data and the dynamic nature of operational environments.

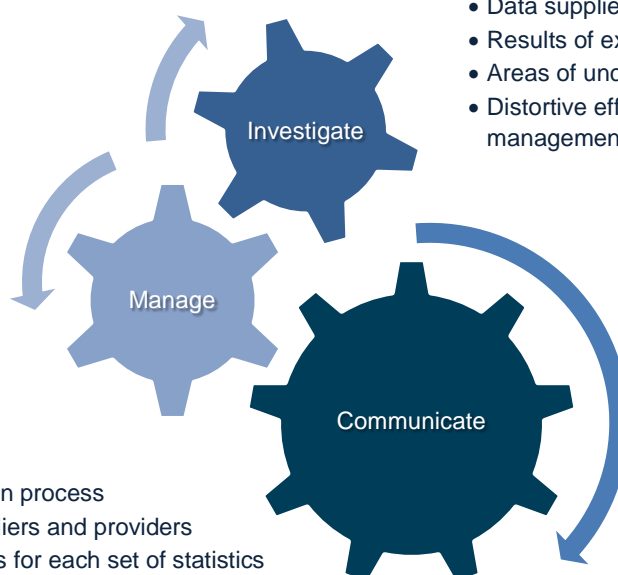
This review of administrative data should not be regarded as a one-off event, but is rather a process that requires repeated evaluation to understand the implications of changes and allow for the ongoing monitoring of the data quality.

Assessors will identify evidence showing the ongoing review of the administrative data by statistics producers.

Quality Management Actions

Manage – such as:

- Cooperative relationship with suppliers, IT and operational, and policy officials
- Guidance information on data requirements
- QA checks and corroboration against other sources



Investigate – such as:

- Data suppliers' own QA arrangements
- Results of external audit of the admin data
- Areas of uncertainty and bias
- Distortive effects of targets and performance management regimes

Communicate – such as:

- Description of data collection process
- Regular dialogue with suppliers and providers
- Document quality guidelines for each set of statistics
- Description of errors and biases and their effects on the statistics
- Communicate with users

QA Matrix

The QA Matrix helps assessors and producers to determine the types of assurance and documentation required to inform users about the quality assurance arrangements for administrative data. This guidance can support a judgment about the suitability of the data and to identify examples of practices that meet the different levels of assurance.

Producers are responsible for judging the appropriate level of assurance. Producers should be able to explain the basis of their judgments of the chosen levels of assurance.

'No assurance' (A0) is not compliant with the *Code of Practice for Official Statistics*.

The need for investigation and documentation increases at each level of assurance from 'Basic' (A1) to 'Enhanced' (A2) to 'Comprehensive' (A3). It may be appropriate for the levels of assurance to vary among the four practice areas; for example, given specific circumstances it may be appropriate for 'Communication' to be Basic (A1), while 'Data Collection QA' be Enhanced (A2) and for both 'Operational Context' and 'Producer's QA' to be Comprehensive (A3).

Assessors will make an evaluation of what they regard as the appropriate level of assurance for the administrative data during an Assessment of official or National Statistics based on administrative data.

The Authority may decide that given the level of risk of quality issues and the public interest profile of the statistics that a higher level of assurance is appropriate than that judged by the statistics producer. In these cases, assurance levels A1 or A2 may be viewed as not compliant with the *Code*. The Assessment will identify the specific areas of practice that are required for compliance with the *Code*.

Four practice areas associated with data quality

Operational context & admin data collection	Communication with data supply partners	QA principles, standards and checks by data suppliers	Producers' QA investigations & documentation
<ul style="list-style-type: none"> • environment and processes for compiling the administrative data • factors which affect data quality and cause bias • safeguards which minimise the risks • role of performance measurements and targets; potential for distortive effects 	<ul style="list-style-type: none"> • collaborative relationships with data collectors, suppliers, IT specialists, policy and operational officials • formal agreements detailing arrangements • regular engagement with collectors, suppliers and users 	<ul style="list-style-type: none"> • data assurance arrangements in data collection and supply • quality information about the data from suppliers • role of operational inspection and internal/external audit in data assurance process 	<ul style="list-style-type: none"> • QA checks carried out by statistics producer • quality indicators for input data and output statistics • strengths and limitations of the data in relation to use • explanation for users about the data quality and impact on the statistics

QA Matrix

Levels of assurance for four areas of practice related to quality assurance of administrative data regularly provided for producing official statistics (see Annex A to see the QA Matrix on one page):

Practice area 1: Operational context & administrative data collection

Level of Assurance	Operational context & administrative data collection
A0: No assurance	<ul style="list-style-type: none"> Operational context and administrative data collection by supplier not investigated, managed or documented
A1: Basic assurance Statistical producer has reviewed and published a summary of the administrative data QA arrangements	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided users with an outline of the administrative data collection process, Illustrated the administrative data collection process and main stages, Outlined the operational context, Identified actions taken to minimise risks to quality, Identified and summarised the implications for accuracy and quality of data, including the impact of any changes in the context or collection arrangements
A2: Enhanced assurance Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided users with a fuller description of the operational context and administrative data collection arrangements, eg: <ul style="list-style-type: none"> - a process map detailing the data collection processes, - explanations for classifications, Identified and summarised potential sources of bias and error in administrative system, Identified and described safeguards taken to minimise risks to data quality, Provided a detailed description of the implications for accuracy and quality of data, including the impact of any changes in the context or collection arrangements
A3: Comprehensive assurance Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided users with a detailed description of the administrative system and operational context: <ul style="list-style-type: none"> - explained why the data are collected, who by and how, - identified differences across areas in the collection and recording of the data, - identified issues for individual data items, such as whether objective or based on subjective recording, missing and/or imputed, poorly recorded, Identified issues in design and definition of targets, Identified and described potential sources of bias and error in the administrative system, Identified and explained any safeguards used to minimise the risks to data quality, Provided a detailed and specific description of the implications for accuracy and quality of the data, including the impact of any changes in the context or collection arrangements

Practice area 2: Communication with data supply partners

Level of Assurance	Communication with data supply partners
<p>A0: No assurance</p>	<ul style="list-style-type: none"> • No communication
<p>A1: Basic assurance Statistical producer has reviewed and published a summary of the administrative data QA arrangements</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has outlined the data provision arrangements including: <ul style="list-style-type: none"> - annual statement of needs, - timing and format of data supply, - coordination of data sign-off from data supplier, • Fed back identified errors to data suppliers and recorded their response, • Sought the views of statistics users about the data and resolved any quality issues reported
<p>A2: Enhanced assurance Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has agreed and documented: <ul style="list-style-type: none"> - data requirements for statistical purposes, - legal basis for data supply, - data transfer process, - arrangements for data protection, - sign-off arrangements by data suppliers, • Established an effective mode of communication with contacts (eg with data collector and supplier bodies, IT systems, operational/policy officials) to discuss the ongoing statistical needs in the data collection system and quality of supplied data, • Sought the views/experiences of statistics users and resolved any quality issues reported
<p>A3: Comprehensive assurance Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has established/maintained collaborative relationships, • Has a written agreement specifying: <ul style="list-style-type: none"> - roles and responsibilities, - legal basis for data supply, - data supply and transfer process, - security and confidentiality protection, - schedule for data provision, - content specification, • Used a change management process, • Regularly communicated with the data collector and supplier bodies, IT systems, operational/policy officials eg newsletters, conferences, attending data supplier/IT system group meetings, • Regularly engaged statistics users, resolved any reported quality issues, and held user group conferences

Practice area 3: QA principles, standards and checks applied by data suppliers

Level of Assurance	QA principles, standards and checks applied by data suppliers
<p>A0: No assurance</p>	<ul style="list-style-type: none"> • No description of suppliers' QA procedures and standards
<p>A1: Basic assurance Statistical producer has reviewed and published a summary of the administrative data QA arrangements</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has knowledge of suppliers' QA checks and published a brief description, • Identified whether audits are conducted on the admin data (such as internal or operational audits, external audit such as by regulator), • Described the implications for the statistics
<p>A2: Enhanced assurance Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has provided a fuller description of the main QA principles, quality indicators and checks used by the data suppliers, • Described the role of relevant information management or governance groups in data quality management, • Described the role of audit of the admin data within the collection and operational settings, • Described the implications for the statistics for the quality issues identified by data supply bodies and regulators
<p>A3: Comprehensive assurance Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has described the data suppliers' principles, standards (quality indicators) and quality checks, • Reviewed quality reports for the received data (such as input quality indicators for data accuracy, coverage and completeness), • Identified and documented the findings of investigations and audits conducted on the admin data and associated targets (such as internal and operational audits, and external audits by regulators and professional bodies), • Described the implications for the statistics and determined whether the data continue to be satisfactory for official statistics purposes

Practice area 4: Producer’s QA investigations & documentation

Level of Assurance	Producer's QA investigations & documentation
<p>A0: No assurance</p>	<ul style="list-style-type: none"> • No description of own QA checks
<p>A1: Basic assurance Statistical producer has reviewed and published a summary of the administrative data QA arrangements</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has established regular QA checks on the received admin data, • Published a description of its own QA checks on the admin data, • Outlined the general approach and overall findings, • Identified the strengths and limitations of the admin data, • Explained the likely degree of risk to the quality of the admin data
<p>A2: Enhanced assurance Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has provided a fuller description of its own QA checks on the admin data, • Detailed the general approach and findings for specific quality indicators, • Identified the strengths and limitations of the admin data, • Explained the likely degree of risk to the quality of the admin data
<p>A3: Comprehensive assurance Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit</p>	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> • Producer has provided a detailed description of its own QA checks on the admin data (including validation, sense and consistency checks), • Given quantitative (and where appropriate qualitative) metrics for specific quality indicators (such as input, process and output quality metrics), • Undertaken comparisons with other relevant data sources (such as survey or other admin data), • Identified possible distortive effects of performance measurements and targets, • Identified the strengths and limitations of the admin data and any constraints on use for producing statistics, • Explained the likely degree of risk to the quality of the admin data

Risk/Profile Matrix

The critical judgment about the suitability of the administrative data for use in producing official statistics should be pragmatic and proportionate, made in the light of an evaluation of the likelihood of quality issues arising in the data that may affect the quality of the statistics, and of the nature of the public interest served by the statistics:

Level of risk of quality concerns	Public interest profile		
	Lower	Medium	Higher
Low	Statistics of lower quality concern and lower public interest [A1]	Statistics of low quality concern and medium public interest [A1/A2]	Statistics of low quality concern and higher public interest [A1/A2]
Medium	Statistics of medium quality concern and lower public interest [A1/A2]	Statistics of medium quality concern and medium public interest [A2]	Statistics of medium quality concern and higher public interest [A2/A3]
High	Statistics of higher quality concern and lower public interest [A1/A2/A3]	Statistics of higher quality concern and medium public interest [A3]	Statistics of higher quality concern and higher public interest [A3]

A1 to A3 = Levels of assurance in the QA Matrix

Level of risk of data quality concerns

Use the following descriptions to help determine whether the administrative data should be regarded as low, medium or high risk of having quality concerns:

Low risk – the data may have a low risk of data quality concerns in situations in which there is a clear agreement about what data will be provided, when, how, and by whom; when there is a good appreciation of the context in which the data are collected, and the producer accepts that the quality standards being applied meet the statistical needs.

Medium risk – the data may be regarded as having a medium risk of data quality concerns when high risk factors have been moderated through the use of safeguards, for example, integrated financial audit and operational checks, and effective communication arrangements. It is also appropriate to consider the extent of the contribution of the administrative data to the official statistics, for example, in cases where the statistics are produced in combination with other data types, such as survey or census data.

High risk – the data may have a high risk of data quality issues when there are many different data collection bodies, intermediary data supplier bodies, and complex data collection processes with limited independent verification or oversight.

For example:

- The risk is lower when a supplier team is located within the same organisation as the statistics producer but increased when there are multiple data collection agencies and/or intermediary data suppliers.
- The risk is lower when operational processes include teams whose function it is to quality assure and verify the recorded data and where internal audit is an established

mechanism with the body. The risk is higher when no internal audit occurs and where limited external and independent challenge occurs to the data, such as, provided by regulator bodies.

- The risk is lower when clear agreements are in place and communication is regular and found to be effective by all partners in the data supply process. The risk of data quality issues is increased when roles and responsibilities are not clear and no formal agreement is in place for the supply of data.

Public interest profile of the statistics

Statistical producers and assessors should use the following descriptions to help determine whether the administrative data should be regarded as having low, medium or high levels of public interest (representing the 'public good'):

Low profile – politically neutral subject; interest limited to niche user base, and limited media interest.

Medium profile – wider user and media interest, with moderate economic and/or political sensitivity.

High profile – economically important, reflected in market sensitivity; high political sensitivity, reflected by Select Committee hearings; substantial media coverage of policies and statistics; important public health issues; collection required by legislation.

Top Tips

These five tips summarise the main pointers for statistical producers to develop a good understanding of the quality issues of administrative data:

Don't trust the safeguards	<ul style="list-style-type: none"> • check if safeguards are functioning effectively
Get involved	<ul style="list-style-type: none"> • work and share with suppliers, such as through secondments and webinars, to develop a common understanding
Raise a red flag	<ul style="list-style-type: none"> • identify potential data quality concerns using input and output quality indicators and investigate anomalies
See the big picture	<ul style="list-style-type: none"> • identify what investigations and audits have been conducted and what they found
Corroborate the evidence	<ul style="list-style-type: none"> • confirm the levels and the trends shown by the stats derived from the admin data

Helpful Guidance

1. UKSA admin data webpage with case examples, talks from RSS event, Exposure Draft report and tool for non-statisticians:
<http://www.statisticsauthority.gov.uk/assessment/monitoring/administrative-data-and-official-statistics/index.html>
2. NSO Guidance, Use of Administrative or Management Information:
<https://gss.civilservice.gov.uk/blog/2014/05/interim-administrative-data-guidance/>
3. Using Administrative Data: Good Practice Guidance for Statisticians:
<https://gss.civilservice.gov.uk/wp-content/uploads/2012/12/Interim-Admin-Data-guidance.pdf>
4. ESS network project on administrative data for business statistics – including Deliverable 2011/6.5, *Final list of quality indicators and associated guidance*, and Deliverable 2011/6.1, *Basic list of quality indicators* <http://www.cros-portal.eu/content/admindata-sga-3>
5. BLUE ETS Deliverable 4.2, Report on methods preferred for the quality indicators of administrative data sources: <http://www.blue-ets.istat.it/fileadmin/deliverables/Deliverable4.2.pdf>
6. ESS network Work Package 2a, Deliverables 2.2 and 2.4 – to check the suitability of an administrative data source before first use: <http://www.cros-portal.eu/content/admindata-sga-3>
7. US Federal Committee on Statistical Methodology, US Census Bureau data tool for checking a new administrative data source: www.bls.gov/osmr/datatool.pdf
8. ONS Quality Centre, Guidelines for Measuring Statistical Output Quality:
<http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/index.html>

Contacts

If you have any questions about this toolkit, please contact:

- assessment@statistics.gsi.gov.uk

Further advice about statistical practice can be sought from:

- the National Statistician's Good Practice Team (GPT):
goodpracticeteam@statistics.gsi.gov.uk
- ONS Quality Centre: ons.qualitycentre@ons.gsi.gov.uk

Level of Assurance	Areas of practice related to quality assurance of administrative data regularly provided for producing official statistics			
	Operational context & administrative data collection	Communication with data supply partners	QA principles, standards and checks applied by data suppliers	Producer's QA investigations & documentation
A0: No assurance	<ul style="list-style-type: none"> Context and data collection by supplier not investigated, managed or documented 	<ul style="list-style-type: none"> No communication 	<ul style="list-style-type: none"> No description of suppliers' QA procedures and standards 	<ul style="list-style-type: none"> No description of own QA checks
A1: Basic assurance Statistical producer has reviewed and published a summary of the administrative data QA arrangements	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided users with an outline of the administrative data collection process, Illustrated the administrative data collection process and main stages, Outlined the operational context, Identified actions taken to minimise risks to quality, Identified and summarised the implications for accuracy and quality of data, including the impact of any changes in the context or data collection 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has outlined the data provision arrangements including: <ul style="list-style-type: none"> annual statement of needs, timing and format of data supply, coordination of data sign-off from supplier, Fed back identified errors to data suppliers and recorded their response, Sought the views of statistics users about the data and resolved any quality issues reported 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has knowledge of suppliers' QA checks and published a brief description, Identified whether audits are conducted on the admin data (such as internal or operational audits, external audit such as by regulator), Described the implications for the statistics 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has established regular QA checks on the received admin data, Published a description of its own QA checks on the admin data, Outlined general approach and findings, Identified the strengths and limitations of the admin data, Explained the likely degree of risk to the quality of the admin data
A2: Enhanced assurance Statistical producer has evaluated the administrative data QA arrangements and published a fuller description of the assurance	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided users with a fuller description of the operational context and administrative data collection arrangements, eg: <ul style="list-style-type: none"> a process map detailing data collection processes, explanations for classifications, Identified and summarised potential sources of bias and error in administrative system, Identified and described safeguards taken to minimise risks to data quality, Provided a detailed description of the implications for accuracy and quality of data, including the impact of any changes in the context or collection arrangements 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has agreed and documented: <ul style="list-style-type: none"> data requirements for statistical purposes, legal basis for data supply, data transfer process, arrangements for data protection, sign-off arrangements by data suppliers, Established an effective mode of communication with contacts (eg with data collector and supplier bodies, IT systems, operational/policy officials) to discuss the ongoing statistical needs in the data collection system and quality of supplied data, Sought the views/experiences of statistics users and resolved any quality issues reported 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided a fuller description of the main QA principles, quality indicators and checks used by the data suppliers, Described the role of relevant information management or governance groups in data quality management, Described the role of audit of the admin data within the collection and operational settings, Described the implications for the statistics for the quality issues identified by data supply bodies and regulators 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided a fuller description of its own QA checks on the admin data, Detailed the general approach and findings for specific quality indicators, Identified the strengths and limitations of the admin data, Explained the likely degree of risk to the quality of the admin data
A3: Comprehensive assurance Statistical producer has investigated the administrative data QA arrangements, identified the results of independent audit, and published detailed documentation about the assurance and audit	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided users with a detailed description of the administrative system and operational context: <ul style="list-style-type: none"> explained why the data are collected, who by and how, identified differences across areas in the collection and recording of the data, identified issues for individual data items, such as whether objective or based on subjective recording, missing and/or imputed, poorly recorded, Identified issues in design and definition of performance measurements and targets, Identified and described potential sources of bias and error in the administrative system, Identified and explained any safeguards used to minimise the risks to data quality, Provided detailed description of the implications for a quality of the data, including the impact of any changes in the context or data collection 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has established/maintained collaborative relationships, Has a written agreement specifying: <ul style="list-style-type: none"> roles and responsibilities, legal basis for data supply, data supply and transfer process, security and confidentiality protection, schedule for data provision, content specification, Used a change management process, Regularly communicated with the data collector and supplier bodies, IT systems, operational/policy officials eg newsletters, conferences, attending data supplier/IT system group meetings, Regularly engaged statistics users, resolved any reported quality issues, and held user group conferences 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has described the data suppliers' principles, standards (quality indicators) and quality checks, Reviewed quality reports for the received data (such as input quality indicators for data accuracy, coverage and completeness), Identified and documented the findings of investigations and audits conducted on the admin data and associated targets (such as internal and operational audits, and external audits by regulators and professional bodies), Described the implications for the statistics and determined whether the data continue to be satisfactory for official statistics purposes 	<p><i>Consider the following types of activities:</i></p> <ul style="list-style-type: none"> Producer has provided a detailed description of its own QA checks on the admin data (including validation, sense and consistency checks), Given quantitative (and where appropriate qualitative) metrics for specific quality indicators (such as input, process and output quality metrics), Undertaken comparisons with other relevant data sources (such as survey or other admin data), Identified possible distortive effects of performance measurements and targets, Identified the strengths and limitations of the admin data and any constraints on use for producing statistics, Explained the likely degree of risk to the quality of the admin data