

(ONS)

A Framework for Assessing and Reporting the Quality of Information provided by the Neighbourhood Statistics Service

Introduction

The Neighbourhood Statistics Service is designed to provide information to support the National Strategy for Neighbourhood Renewal as well as other requirements for local and regional statistics. The work is being developed progressively using a range of information drawn from a multiplicity of sources. These data acquisition activities will be supported by significant technological innovations to improve the presentation of information, the effectiveness of delivery and overall analytical capabilities.

The first phase of development has been characterised by the acquisition of information from Central Government departments that is designed primarily to meet existing statutory or operational requirements. However, as time passes data will be drawn from a wider range of sources and organisations. In addition, wherever possible, efforts will be made to secure micro-level data so that data matching and linkage techniques may be applied to improve both the range and quality of information.

The Neighbourhood Statistics team will work co-operatively with a range of data suppliers to refine existing systems, increase the amount and type of information available, improve or extend its geographic coverage. This work will be associated with the availability of special facilities to display point data and the development of a limited number of well defined statistical indicators capable of meeting specific policy needs. By bringing together an increasing range of information, more flexible methods of presentation and analysis the Neighbourhood Statistics Service will be able to provide mutually reinforcing evidence to support a range of investigations.

These developments need to be reflected in the plans for assessing and measuring quality in order to ensure that relevant information can be provided in the most appropriate way. In particular, it will be important to take account of the inherent variations in methods of data collection, processing and analysis. The

need for a range of detailed quality measures will need to be balanced by an understanding of the constraints associated with the use of administrative sources, with underlying operational procedures, technical and financial issues. While recognising the need for flexibility, an authoritative framework for assessing and reporting quality can only be developed and implemented successfully if it is based on sound principles, appropriate practices and methods.

At present each supplier contracted to provide information for the Neighbourhood Statistics Service is required to sign a service level agreement. This document includes basic requirements for reporting on the quality of the data delivered and asks for specific information to show that measures have been taken to:

- clean, validate and quality assure the information;
- provide relevant quality measures or indicators; and
- help users to interpret the data.

However, these simple standards are no longer adequate and need to be updated to reflect the increasing range of data that is likely to be delivered as well as the emergence of definitive guidance on quality measurement and reporting. The remainder of this paper outlines the issues associated with the development and implementation of a framework for assessing and reporting the quality of all information disseminated through the Neighbourhood Statistics Service.

Principles Governing the Development of the Neighbourhood Statistics Framework for Reporting Quality

If data suppliers are to provide comprehensive information about the quality of their output, they need to have a clear understanding of overall requirements and how these must be met. The Neighbourhood Statistics Quality Framework will provide the basis for defining and measuring quality by focusing on two separate but interrelated issues namely:

- understanding and reporting key processes; and
- developing quality measures or indicators.

Each of these issues will be considered briefly before considering how they can be applied to the different types of output likely to be published via the Neighbourhood Statistics Service.

Quality Processes

Every data supplier, whether internal or external, will be expected to provide a clear explanation of the methods or systems used to collect, prepare and present their output. By considering such arrangements users should be able to understand, in varying degrees of detail, individual processes or elements in the statistical value chain. In particular the following information will be required:

- details on data collection for both administrative and statistical sources;

- the steps taken to minimise errors or omissions;
- data validation and quality assurance processes, including both voluntary and statutory checks, undertaken by those responsible for data collection, processing or publication;
- geographic referencing procedures; and
- measures adopted to protect confidentiality.

The provision of comprehensive details on each of these activities should facilitate evaluation of the overall effectiveness of individual methodologies, systems or tools. This information, or, in the case of disclosure, relevant extracts, should help users to understand the methods and processes used to produce each dataset.

Quality Measurement and Reporting

Quality is one of the eight principles enshrined in the National Statistics Code of Practice and quality measurement and reporting is an explicit commitment. Although the amount of information that is likely to be available may vary, it will be important to have a consistent basis for measuring and reporting quality. In accordance with proposals outlined in the National Statistics Draft Protocol on Quality Management the Neighbourhood Statistics Service will adopt the quality dimensions of the European Statistical System which is based on six key quality attributes; namely:

- relevance;
- accuracy;
- timeliness and punctuality;
- accessibility and clarity;
- comparability; and
- coherence.

These core attributes will be used as a basis for measuring and reporting the quality of all output disseminated via the Neighbourhood Statistics Service. The information outlined below is designed to illustrate the type of information expected:

Relevance - concepts, measurements and products reflect user needs. In the case of Neighbourhood Statistics information should meet either the needs identified in the PAT 18 Report on Better Information or priorities outlined at subsequent reviews and updates such as Blueprint Meetings.

Accuracy - in all cases measures should provide a clear indication of overall coverage and completeness as well as an indication of the types and levels of error associated with data collection systems or statistical processes. These details should be complemented by information about how these assessments have been made (for example, by comparison with other sources, sample checks or follow-up surveys).

Timeliness - details about the currency of the information should be complemented by an explanation of the reasons for any delays and the timetable for updating the results.

Accessibility - data suppliers will be required to provide clear and comprehensive guidance so that users can understand the overall quality of the statistics and the methods employed to produce them. The Office for National Statistics will ensure that information is presented effectively, can be easily accessed and complies with agreed metadata standards.

Comparability - includes comparisons over time and space as well as between domains (for example different industrial sectors or types of households).

Coherence - relates to the reliability with which statistics can be combined in different ways and for various purposes. Wherever possible data suppliers will be required to use common statistical frames, standard concepts, questions, definitions and classifications.

As far as possible individual quality measures should meet recommended standards or be compatible with those adopted for comparable sources. These expectations will need to be discussed in advance with each data supplier and adapted, as necessary, to meet the specific circumstances associated with individual datasets or types of information. The terms and conditions agreed will be set out clearly in the standards section of each Service Level Agreement.

Processes and Procedures for the Quality of Aggregate Data drawn from Administrative Sources

At present most of the data published through the Neighbourhood Statistics Service are provided by Central Government departments in the form of aggregate counts. Whereas some returns are based on data held by a specific department, others are produced from abstracts submitted by individual local authorities or other organisations. Details of the range of non-Census information currently available are outlined in Annex 1.

Administrative sources or systems are not designed to provide statistical information but are set up to meet specific statutory requirements, support operational, financial or management needs. These circumstances will to a large extent dictate what can, and cannot, be provided in the way of both process information and specific quality measures or indicators. Detailed discussion with individual data suppliers, including representatives from the Department for Work and Pensions, the Office of the Deputy Prime Minister, the police and fire services have highlighted the overall importance of specific operational factors including:

- the scale, complexity and lack of stability associated with administrative systems;
- the limitations associated with localised data collection and/or processing arrangements;
- the constraints imposed by external contractors or sub-contractors; and

- the inability to provide comprehensive quality measures for a range of areas.

In view of these issues proposals for reporting the quality of aggregate administrative returns or abstracts derived from one or more administrative sources must be simple, practical and realistic. More specifically, it will be important to combine flexibility with rigour in order to take account of variations in operational or organisational arrangements while safeguarding basic standards. In accordance with the Neighbourhood Statistics Quality Framework previously outlined all data suppliers will be asked to report on key processes and provide relevant quality measures. While the amount and type of information may vary it should be detailed enough to enable users to interpret the data correctly, understand their strengths and weaknesses.

Although the primary responsibility for quality will remain with the data supplier, the Office for National Statistics must be satisfied about overall standards. It is recommended that a simple checklist be used as a guide for assessing the overall acceptability of data based on aggregate administrative returns or abstracts. Each data supplier will be asked to satisfy the following *key criteria for acceptability*:

- data collection - this should cover information on registration or recording practices including statutory requirements and methods of enforcement (for example the use of penalties or incentives);
- statistical frameworks - where possible, data suppliers should use agreed (harmonised) concepts, questions, data definitions and classifications. Any departure from recommended standards should be explained and the implications for the use of the data assessed;
- data processing - methods of data processing, including edit and imputation, quality assurance and validation procedures, must be fully reported;
- quality measurement - appropriate quality measures should be provided. At a minimum there must be a clear indication of the overall accuracy of the data;
- geographic referencing - data suppliers must use approved methods, materials or tools. Where possible, match rates should be reported and any spatial or temporal variations identified;
- measures to protect confidentiality - no information about an identifiable individual, family, household or business will be published; and
- comprehensive metadata must be provided to ensure that users can understand the strengths and weaknesses of the data.

Existing experience suggests that the amount and type of information made available will vary. Individual elements of the checklist may not be applicable or appropriate. For example, edit and imputation may not be used, validation may only be carried out on a sample of data and geographic referencing procedures will reflect different operational arrangements. As a result it may not always be feasible to provide information on overall match rates or more detailed information on the variations between individual areas.

It must be recognised that such difficulties will affect the derivation of quality measures or indicators. While every effort will be made to provide comprehensive details on each of the specified quality attributes, operational or organisational

constraints may preclude this. For example, the collation of aggregate returns by independent organisations, the scale and complexity of administrative systems, issues of cost or burden may restrict the overall availability of information.

Discussion with a number of data suppliers has indicated that it will not be feasible to provide comparable quality measures for a range of areas. For example, detailed checks based on a sample of records may facilitate the provision of indicators at national or regional level but will not support the same level of analysis for smaller geographic units. Nevertheless, it should be possible to make an overall assessment of quality and all data suppliers will be asked to do this. In addition, every effort will be made to ensure that sufficient information is provided to enable users to understand and interpret the data. In particular, special attention will be given to issues affecting comparability. Clear and comprehensive guidance on the nature and extent of any changes and the implications for analysis and interpretation of individual returns will be made available.

The development of mapping facilities and associated release of point data showing the location of public services or facilities will have special implications for quality measurement and reporting requirements. In addition, to the geographic referencing specifications previously outlined it will be important to provide guidance on referencing and matching methodologies as well as explicit standards for match quality and resolution levels. At the same time data suppliers will be asked to ensure that any associated 'attribute data' providing details on the nature of specific services or facilities complies with key elements of the standard checklist. In particular, special attention will be given to the provision of information on data collection and validation, the overall accuracy and timeliness of the data as well as the provision of metadata. There must be clear evidence to show that the data have been obtained from an authoritative and up-to-date source.

These requirements may need to be updated to take account of the inclusion of new types of data or the introduction of more sophisticated analytical capabilities.

Processes and Procedures for the Quality of Information from Statistical Sources and Analyses based on Micro-Level Data

Statistical sources may be defined as those which are designed to meet specific statistical needs and where all relevant operations are controlled or validated by professional statisticians. Statistical information is likely to cover output from a single statistical source, data derived from a combination of sources as well as indicators developed from appropriate statistical or other returns. As work progresses these sources may be complemented by information produced from the aggregation of micro-level data or the matching and linkage of individual records.

Any statistical information should comply with the National Statistics Code of Practice. Although the requirements previously outlined for aggregate data will be relevant, the amount and type of process information as well as the scope and

precision of the quality measures will be significantly different. This may be illustrated with reference to the recently approved series of model-based income estimates.

The availability of a series of estimates of average household income represents the culmination of a major programme of research and development. This has drawn upon the expertise of methodologists from the Office for National Statistics, leading statisticians and academic experts. All stages of the work from the development and specification of the model to the creation and publication of the estimates have been subject to extensive consultation, rigorous quality assurance and validation. Special attention was given to selection of appropriate input data, the evaluation of the model, including detailed assessment of bias and goodness of fit, the overall variability and plausibility of the estimates. The availability of relevant quality measures and comprehensive metadata will enable users to understand and make appropriate use of the estimates.

This case study has underlined the key components which characterise all National Statistics outputs including:

- the application of agreed statistical procedures, approved analytical methods, techniques and tools or the development of new approaches which have been well researched, tested and evaluated;
- the exploitation of the best and most appropriate data sources;
- the use of common statistical frames, definitions and classifications;
- the employment of approved sources, methods and tools for geographic referencing;
- the application of appropriate quality management and assurance processes;
- the publication of key quality measures for all relevant areas;
- the implementation of stringent procedures to protect confidentiality;
- the provision of comprehensive metadata; and
- the availability of further advice and support from relevant experts.

Although all data suppliers will be required to provide process information and report against key quality measures or indicators, the amount and type of information will vary. The nature and extent of these differences between aggregate administrative data and statistical sources is summarised in the table presented in Annex 2.

Quality Standards

The quality of all statistics published on the Neighbourhood Statistics website will need to be assessed and classified so that users are aware of whether they do, or do not, meet National Statistics standards. On the basis of the proposals previously outlined this should be determined in relation to the availability of information on key processes and the provision of explicit quality measures or indicators. More specifically, special account will need to be taken of the inherent differences between statistical and aggregate data.

At present data released through the Neighbourhood Statistics Service are accompanied by one of the following declarations:

'The supplier of this dataset has declared that it does belong within the National Statistics profile' or

'The supplier of this dataset has declared that it does not belong within the National Statistics profile'.

It is recommended that the existing statements should be amended and a new declaration added to cover 'Experimental Statistics'. More specifically, we propose to introduce a series of simple designations including:

- National Statistics;
- Experimental Statistics; and
- Not National Statistics.

These simple designations should be accompanied by short statements to underline the implications for the quality of each dataset or output. It is suggested that the following descriptions, based on compliance with the requirements set out in the National Statistics Code of Practice, should be applied:

- **National Statistics** - this information complies fully with the National Statistics Code of Practice and related protocols.

Users can be confident that all information classified as 'National Statistics' whether it be statistics derived from a single source, output derived from a combination of sources, or data based on the matching or linkage of micro-level data, will meet the highest standards. Data suppliers will be required to provide comprehensive information on quality using agreed measures or indicators based on the quality dimensions defined for the European Statistical System. In all cases the information must be sufficient to enable users to make appropriate use of the data and understand their strengths and weaknesses.

- **Experimental Statistics** - this information has been developed in accordance with the principles set out in the National Statistics Code of Practice but has yet to be fully accredited as a National Statistic.

This designation will apply to output based on the development and application of new statistical methods or techniques such as the use of small area estimation to produce model-based estimates of income for wards in England, Wales and Scotland. The status of such statistics may reflect the fact that new methods require further testing, validation or quality assurance, or that output may be modified in response to user feedback. Nevertheless, users will know that all aspects of the work have been conducted in accordance with the requirements set out in the National Statistics Code of Practice and its associated protocols. Once it can be shown that the new output or product is fit for purpose and acceptable to users then the National Statistician or the relevant Head of Profession will consider formal accreditation as a 'National Statistic'.

- **Not National Statistics** - this information based on administrative data does not comply fully with the National Statistics Code of Practice.

In this case operational constraints associated with the collection, processing or validation of administrative data will mean that output, whether derived from a single source, or a combination of sources, cannot comply fully with the stringent requirements of the National Statistics Code of Practice. For example, it may not have been possible to make use of recommended definitions or classifications, to provide a full set of key quality measures or offer comparable information for all specified areas.

Whereas all types of statistical outputs will be covered by existing arrangements and will have been classified as 'National Statistics' or 'Experimental Statistics' by the National Statistician or the relevant Head of Profession, all other data sets will need to be reviewed by a Quality Assurance Panel. The Panel, consisting of staff from the Office for National Statistics, the data supplying organisation and the Neighbourhood Renewal Unit will work co-operatively to ensure that the data meet an acceptable standard. More specifically, all types of aggregate data will be required to meet the following *key criteria for acceptability*:

- the provision of explicit information on data collection and processing methods;
- the use of standard concepts, questions, data definitions and classifications. Where this is not possible, the differences together with the implications for the use and interpretation of the data must be fully explained;
- comprehensive details on validation and quality assurance including methods for checking and correcting inaccurate, incomplete and missing returns;
- a clear statement on the overall quality of the data, supplemented, where possible, by relevant quality measures;
- details on geographic referencing arrangements including, where possible, details on the overall number of matched and unmatched records;
- the application of appropriate methods to safeguard the confidentiality of information on identifiable individuals, families, households, or businesses; and
- the provision of comprehensive metadata.

The Neighbourhood Statistics Service reserves the right to reject any dataset where one or more of the key criteria are not met. In particular, special attention will be given to the overall effectiveness of data collection, methods of validation and quality assurance, the standards of geographic referencing and to disclosure control. By considering these issues, members of the Panel will be able to assess the overall acceptability of the data and their suitability for publication.

Recommendations

In accordance with information outlined in previous sections of the paper it is recommended that the following proposals be approved by members of the Statistical Policy Committee and adopted as the framework for evaluating and

managing the quality of aggregate data, point data and all types of statistical output. These recommendations take account of the fundamental differences between aggregate data based on administrative information and all types of statistical output.

- aggregate statistics and point data based on administrative sources will be assessed in relation to the *key criteria for acceptability*;
- all types of statistical output including statistical indicators and statistical data derived from the matching and linkage of micro-level data will comply with the standards enshrined in the National Statistics Code of Practice and its associated protocols;
- every data supplier will be expected to provide a clear explanation of the methods or systems used to collect, prepare, validate and present their information;
- where possible, data suppliers should use harmonised questions, concepts, data definitions and classifications;
- the quality of individual outputs will be reported using the six key quality attributes of the European Statistical System;
- procedures for reporting quality must be agreed with each data supplier and included in their Service Level Agreement. At a minimum there must be a clear indication of the overall quality of the data;
- geographic referencing must be undertaken using approved materials, methods or tools;
- no information about an identifiable individual, family, household or business will be published;
- comprehensive guidance must be provided so that users can understand the strengths and weaknesses of the information and make appropriate use of it;
- the expected standards for aggregate and statistical data will be different. Whereas aggregate output or point data derived from one or more administrative sources will usually be classified as 'Not National Statistics', all statistical outputs will be either 'National Statistics' or 'Experimental Statistics'.

The standards and procedures set out in this paper may need to be updated to take account of changes to the National Statistics Code of Practice. At the same time it may be necessary to review these arrangements to ensure they comply with other relevant policies or procedures on quality promulgated by the ONS Quality Centre or the Policy and Quality Division of the Office for National Statistics.

Conclusion

This paper presents a framework for assessing and reporting the quality of all information released through the Neighbourhood Statistics Service. If the proposals are accepted by members of the Statistical Policy Committee they will provide a consistent basis for specifying an agreed set of quality measures which will be included in the Service Level Agreement with each data supplier.

Annex 1: Table to Show Non-Census Information Currently Available on the Neighbourhood Statistics Website

Domain and Dataset Title	Period Covered	Coverage
<i>HOUSING</i>		
Changes of Ownership by Dwelling Price	2001	England & Wales
Dwelling Stock by Council Tax Band	2001	England & Wales
Social Rented Housing: Demand and Supply	2000 - 2002	England
Dwelling Stock by Tenure and Condition	2000 - 2002	England
Social Housing Rents	2001 - 2002	England
<i>ECONOMIC DEPRIVATION</i>		
Family Credit Claimants	1999	Great Britain
Income Support Claimants	1998 - 2000	Great Britain
Child Benefit Claimants	1998 - 2000	Great Britain
Working Families Tax Credit Claimants	2000	Great Britain
<i>CRIME & SAFETY</i>		
Notifiable Offences	1999 - 2001	England & Wales
<i>EDUCATION, SKILLS & TRAINING</i>		
Pupils at Key Stage 2	1998	England
University Admissions	1997 - 1998	England
<i>INDICES OF DEPRIVATION</i>		
Indices of Deprivation	2000	England & Wales
Countryside Agency's Definition of Rural Areas	2000	England
<i>HEALTH AND CARE</i>		
Attendance Allowance Claimants	1998 - 2000	Great Britain
Disability Living Allowance Claimants	1998 - 2000	Great Britain
Hospital Episode Statistics	1999 - 2000	England
Welsh Patient Episode Statistics	1999 - 2000	Wales
<i>PEOPLE & SOCIETY: POPULATION & MIGRATION</i>		
Population Estimates	1998	England
Vital Statistics	1998	England & Wales
Parliamentary Electorate	1997 - 1998	England
Parliamentary Electorate	1994 - 1999	Wales
<i>WORK DEPRIVATION</i>		
Jobseekers' Allowance Claimants	1998 - 2000	Great Britain
Incapacity Benefit Claimants	1998 - 2000	Great Britain
Severe Disablement Claimants	1998 - 2000	Great Britain
VAT Registered Enterprises by Industry	2000	United Kingdom
VAT Registered Enterprises by Sizeband	2000	United Kingdom
Employee Jobs	1998	Great Britain
Claimant Count	1998 - 1999	England
Claimant Count and Claimant Count Rate	1998 - 2001	United Kingdom
<i>PHYSICAL ENVIRONMENT</i>		
Floorspace Rateable Value - with Accommodation	2000	England & Wales
Floorspace Rateable Value - without Accommodation	2000	England & Wales

Annex 2:

Table Illustrating the Availability of Information on Quality for Administrative and Statistical Sources

STATISTICAL SOURCES									
Area Level	Key Processes				Key Quality Measures				
	Data Collection	Processing/ Validation	Geographic Referencing	Disclosure Control	Relevance	Accuracy	Timeliness	Comparability	Coherence
Output Area	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ward	✓	✓	✓	✓	✓	✓	✓	✓	✓
Local Authority	✓	✓	✓	✓	✓	✓	✓	✓	✓
Region	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country	✓	✓	✓	✓	✓	✓	✓	✓	✓
ADMINISTRATIVE SOURCES									
Output Area	✓		?✓	✓					
Ward	✓		?✓	✓					
Local Authority	✓	?✓	?✓	✓					
Region	✓	?✓	?✓	✓		?✓			
Country	✓	✓	✓	✓	✓	✓	✓	✓	✓