

CENSUS OUTPUT METADATA: UPDATE**Background**

1. Advisory Group Paper (00)03 set out the strategy for producing information to support the range of 2001 Census output products – Census Metadata. This update summarises the metadata products and other support material that will accompany the release of 2001 Census output throughout 2003. It refers also to the range of material that is already on the ONS website supporting the release of the first results for local and health authority areas.

Action

2. **Advisory Group members are asked to note the paper and suggest other areas where information is likely to be sought by users to support 2001 Census output.**

Introduction

3. ONS intend that 2001 Census Metadata (now being referred to as Support Information) should be:
 - timely - available before or at the same time as the data;
 - comprehensive - covering all topics and areas that a user might be interested in;
 - coherent – a single source for all Census metadata;
 - consistent in design, format and presentation; and
 - easy to use.
4. Paper AG(00)03 noted that there are likely to be four types of metadata produced for the 2001 Census.
 - *Geographical metadata* – information on geographical hierarchies to be recognised in outputs and the Output Area Production System;
 - *Technical metadata* to enable users to make efficient and effective use of the software products;
 - *Product metadata* - information aimed at assisting users to get what they want from Census outputs in the most efficient way; and
 - *Statistical metadata* - information on statistical aspects of the Census, providing background information on the results, clarifying the meaning and context.

Scope and content of Census Support Information

5. Census Support Information (SI) covers a wide range of information, from background on the selection of topics and the development of questions asked in the Census through to data collection and processing methodologies and assessment of data quality. The range of topics that are proposed for inclusion as SI is provided at Annex A.
6. In deciding what type and detail of information should be included the ONS has considered the customer perspective about how SI might be used, by whom, and for what purpose. General users are likely to require summary-type information, whilst specialised users will seek more specific detail. Furthermore, ONS anticipates that users will want to access SI in a variety of ways. For example, one user may be interested in information about particular sub-groups of the population, such as students or the 'Irish', whilst another may be looking for the full range SI about a particular local area; others, still, may require detailed information about processes for addressing under-enumeration, question non- response or edit and imputation rules.
7. The framework used for storing SI needs to be flexible enough to cope with these different demands. Links will be created between operational processes, themes, topics etc, and information will be available on a *drill down* basis. This means giving summary or general information on a subject to start with, for example an overview of the Census Coverage Survey, and then offering more detailed information if required.

SI product access

8. Users will be able to access SI material via different routes, either by reference to stand-alone products or through packages integrated with the statistical output.
9. *Stand-alone products* are being developed where particular user requirements have been identified. Specific examples are the *Quality Report*, the *Definitions Volume*, a *Geography Information Pack*, and the *General Report*. More detail of these products is set out in Annex B.
10. *Integrated products* will allow SI to be combined with statistical information to provide powerful, seamless output for the user. These products will allow users to view statistical information and SI simultaneously. An example of this would be where metadata is provided with standard tables, so that a user could click on a term appearing in a table, such as 'Employment Status', and be given a definition of that term, perhaps with links to more information on that subject. Or it will be possible for a user to access all relevant information on a particular Census question, covering comparisons with 1991 Census, development and testing of the question, quality of responses, edit and imputation procedures, output classifications etc.
11. The aim is to make all SI available in one form or other, via the ONS website, at least with, or, if possible, before, the release of the relevant statistical output. Consideration will be given to the production of hard copy output, in either printed or electronic form, for stand-alone products, or a print-on-demand service, should there be a sufficient user requirement. **Advisory Group members are asked to indicate whether or not there is likely to be a user demand for such an alternative medium for release.**

What is currently available

12. With the release of the first results from the 2001 Census on 30 September and subsequently, ONS has already made available, via its website, a wide range of interlinked supporting material. Linked directly with the first results, SI includes both commentary and statistical analysis covering:

- Regional summaries;
- Local results and comparisons with national profile;
- Change in the period 1951-2001 at national level;
- 1991-2001 and 1981-2001 change analysis at local authority level;
- Top ten rankings;
- One Number Census Methodology and Quality Assurance;
- The Census Coverage Survey;
- The accuracy of the results;
- Response rates by age and sex for local authorities;
- Implications of the results; and
- A guide to the mid-year population estimates.

13. This material to date, in total, is estimated to amount to the equivalent of a report of over 800 pages. In addition, a range of other, more general, regularly updated support material continues to be accessible via the website. This covers:

- Census News;
- Output Prospectus;
- Census background;
- Final specifications of Standard Area Tables;
- Review and Evaluations;
- Strategic development;
- Access to previous census statistics;
- Census education;
- Census methodology in depth; and
- Useful links

14. Clicking-on to each of these enables the user to drill-down to a range of periodically updated material at varying levels of detail and user interest, and inter-linked to other related web pages. For example, *Census News* currently contains material covering:

- One Number Census: further information
- 2001 Census output schedule
- Consultation on SARs;
- The Census Data System
- First and final releases of Census results;
- More on the first results
- Final specifications of Standard Area Tables;
- Postcode Sector Profiles;
- Census Advisory Group meetings;

- Forthcoming events
- MRS Geodemographics Knowledge Base;
- Distribution License and Approved Suppliers Scheme;
- Revision of the Standard Industrial Classification;
- Multi-source topic reports; and
- Technical aspects of bulk delivery service.

while *Census methodology in depth* currently links to:

- Consultation on SARs;
- Copies of the 2001 Census forms;
- Guide to the One Number Census;
- Census exchange;
- Parliament's role; and
- Advisory Group papers.

15. Drilling further down, by clicking on *Census News* and then on *One Number Census: further information* users can currently access details on:

- An evaluation of the One Number Census;
- Census Coverage Survey sample sizes and response rates for each local authority;
- Paper on dependence within the One Number Census;
- Commentary of QA themes:
 - Students
 - Babies in areas with large ethnic populations
 - Borrowing strength from babies
 - Post-stratification
 - Prisons
 - 1991 Census under enumeration adjustments;
- ONC illustrative Quality Assurance Pack

16. Using the navigation menu also enable users to explore other linked websites, including those of:

- GROS and NISRA;
- Royal Statistical Society;
- a range of international census agencies;
- Public Records Office;
- Family Records Centre;
- 1901 Census website;
- HM Stationery Office; and
- Eurostat.

Advisory Group members are asked to note the paper and suggest other areas where information is likely to be sought by users to support 2001 Census output.

ONS, January 2003

ANNEX A CENSUS SUPPORT INFORMATION CONTENT

The list below attempts to identify the scope and coverage of Census Support Information categorised by the several types of metadata. Support information will generally be presented from a UK perspective with variations that apply in different parts of the UK being highlighted and explained. Information that is planned to be available through stand-alone products is indicated by * and is described more full in Annex B.

Geographical metadata

- Output Area design
- Information Pack*

Technical metadata

- Bulk delivery service
- How to run SuperTABLE
- CSV file specification
- Technical requirements for SuperTABLE

Product metadata

- Prospectus and timetables
- Table specifications
- Indexes
- Catalogue and prices
- Cell layout
- Value added supply
- Census Access
- Terms and conditions of supply
- Lists and tables

Statistical metadata

Census questions

- The Census forms
- Questionnaire design (question and small-scale testing, business cases etc.)
- Classifications (including standard derived variables)*
- Coding indexes
- Analysis of 1997 Census Test
- Analysis of 1999 Census Rehearsal
- Comparability with 1991 and earlier Censuses
- Harmonisation with other government social surveys

Edit and Imputation

- Edit and Imputation rules
- Question imputation rates

Disclosure and confidentiality control

- Legal aspects
- Methodology (different levels of detail for different users)
- Record swapping
- Thresholding
- Evaluation
- 1991/2001 comparability

One Number Census (ONC)

- Methodology aimed at various levels (e.g. easy, technical)
- Response rates
- Evaluation
- Comparability with 1991/1981

Census Coverage Survey

- Methodology
- Evaluation
- CCS forms

Data Quality

- Census Quality Report*
- Census Quality Survey (CQS)
- Validation processes
- Validation rules and tolerance checks
- Coding quality

Data Processing

- Clerical vs. automatic recognition (OCR and OMR)
- Scanning quality
- Methodologies
- 1991-2001 differences
- 100 per cent processing
- Steps to address systematic errors

Data Collection

- ED design
- Field procedures
- Postback rates
- Effectiveness of difficult to enumerate procedures
- Field quality checks
- Non-compliance

Output

- Lists and descriptions of products
- Definitions
- Population bases

Legal background

- White Paper
- Census Acts
- Census Orders and Census Regulations

Community Liaison

- Background
- Impact on the Census

Links to sources of external evaluation of the Census

- Papers on 2001 Census from academics

Uses of Census Information by Users

- Census Offices could analyse this information to provide justification of undertaking a Census.

ANNEX B OUTLINE OF PROPOSED STAND-ALONE PRODUCTS

2001 Census Quality Report

1. As previously announced, the Census Offices plan to produce a *Census Quality Report*. Quality will be considered in its widest sense and the report will draw together information from a range of sources to assist users to make informed decisions about the appropriateness of 2001 Census data for particular applications.
2. The aim is to publish this report, in electronic and hard copy formats, around June 2003. The proposed content covers:
 - *Population base, definitions and field methodology* – information on the 2001 population base, as well as issues associated with comparing 2001 Census output with that from the 1991 Census or other sources. The report will also cover field methodology as well as other definitions, such as households and communal establishments.
 - *Information from the testing programme* – drawing on the wealth of information already available on the extensive programme of question testing.
 - *Coverage* - information on the Census Coverage Survey and the One Number Census process, and about the coverage achieved.
 - *Comparability with data from other sources* – information which will assist users to assess the suitability of the 2001 Census data for comparison with data from other sources, including the degree to which Census questions are harmonised with questions in other Government surveys.
 - *Information from processing* - information about processing methodologies, including the comprehensive quality assurance strategy being implemented.
 - *Edit and imputation* - a full account of edit and imputation, describing the edit rules applied, the level of non-response and the imputation methodology.
 - *Validation of results* – description of the validation process and any unexpected results.
 - *Census Quality Survey* – information on the methodology and outcomes of the Census Quality Survey.

2001 Census Definitions and Classifications

3. The Census Offices plan once again to produce a *Definitions Volume* (around mid March 2003) that will provide customers with definitions of concepts and terms used in the 2001 Census outputs, in a dictionary format for ease of reference. The report will also attempt to guide the user through the pitfalls of comparisons with 1991 Census by highlighting differences in questions, definitions and population bases.
4. Details of the full range of basic data classifications, including derived variables, will also be included as material that will, in effect, be the finalised version of the Consultation Document *Census 2001 Classifications* issued in September 2001. This information will, in particular, assist users who wish to define and commission their own customised output.

Census Geography Information Pack

5. Planned for release with, or prior to, the Output Area statistics, this product will include a look-up file showing the constitution of each geographic area level in terms of postcode and Output Areas, as well as giving a reference date and indicating whether census data is assigned to each geography on an exact fit or best-fit basis. A list of those administrative areas with boundary changes between the 1991 and 2001 Censuses will also be noted.
6. The Pack will contain details of OA centroids and will also include headcounts (all residents, males, females and households) at the unit postcode level.
7. A geography/product matrix will show which output tables are available for which geography.

General Report

8. As reported in Annex A of Paper AG (01) 07) the *General Report* will provide details on all aspects of the Census operation. More specifically, the preparation of the *General Report* will enable ONS to provide a strategic overview, and an integrated review, of specific issues such as foot and mouth disease and the 'Welsh' tick-box. The Report is currently scheduled for publication in hard copy form in Autumn 2003, though it is intended that particular chapters, drawn from individual Census project reports, should be available via the ONS website as and when they are completed.
9. The scope of the proposed content of the *General Report* will cover:

Using the Census results

Geography
Output Policy and Dissemination
Census Access

The quality of the Census results

Census Coverage Survey
Processing
Edit and Imputation
One Number Census
Downstream Processing (including IS strategy)
Development of Census Questions (results from testing programme)
Data Quality (including disclosure control)
Output Production

The Census operation

Planning and Consultation
Publicity
Data Collection Development
Data Collection Support
Legislation
Programme Management