



Census Data Quality Advisory Working Group

Minutes of Census Data Quality Advisory Working Group: 12th January 2011

Present

Peter Scrimgeour

Laura Murison

David Blue

Kirsty MacLachlan

Allison Craig

Heidi Goodship

Tony Jenkins

Jan Freeke

Alistair Harvey

Andrew Ballinghall

Cameron Thomas

Tom Snowling

Adele Trainer

Hazel McShane

Euan Smith

Organisation

Census Director General Register Office for Scotland (GROS)

Census Data Quality, GROS

Census Data Quality, GROS

Head of Demography Division, GROS

North Ayrshire Council

Scottish Borders Council

Dundee City Council

Glasgow City Council

Edinburgh City Council

Fife Council

Highland Council

Aberdeen City Council

North Lanarkshire Council

Scottish Government (Local Government Finance)

Scottish Government (Scottish Neighbourhood Statistics)

Apologies:

Paul Fensom

Census Paper Data Capture Lead

Paul Davison

Stirling

1. Welcome, minutes and actions from 5th August 2010

1.1 Peter welcomed everyone to the second meeting of the Census Data Quality Advisory Working Group.

1.2 The minutes from the previous meeting were accepted pending a couple of minor alterations. Paul Fensom had been missed from the list of attendees so will be added and Jan asked that the wording in point 22 is changed from 'vacant households' to 'vacant dwellings'.

1.3 Peter then ran through the actions that arose from the last meeting:

Action Point 1: General Register Office for Scotland (GROS) to circulate minutes to Population and Migration Statistics (PAMS) – **Complete**

Action Point 2: Laura to circulate some examples of 2001 anomalies – **Complete**

Action Point 3: GROS to add reviewing a first cut of Census data to the list of responsibilities of the group and to include a statement on confidentiality – **Complete**.

Action Point 4: GROS to circulate a table to be populated with information about local authority specific sources that each local authority will use to compare to Census results – **Complete**

Action Point 5: Add columns for Coverage and Quality in the template to be circulated to local authorities – **Complete**

Action Point 6: Laura to prepare and populate a table with a list of contacts – **Complete**

2. Terms of Reference and note for Population and Migration Statistics (PAMS)

- 2.1 Laura ran through the changes that have been made to the Terms of Reference. Sharing data in a controlled environment to allow discussion and resolution of data anomalies has been added to the list of responsibilities of the group, including a review of the first downstream processing quality checks carried out on Census data. A section on Confidentiality has been added to the document stating that each member of the group will act in an independent advisory capacity and will be required to sign an agreement stating that they will not share data or analysis outside of the working group meetings (except where agreed with the working group chair). Peter stated that the confidentiality agreement is being worked on and will be issued to each member of the working group to be signed when it is complete. It has also been noted that the group will be responsible for representing all local authorities, including those not wishing to actively participate in the working group.
- 2.2 Laura stated that there will be two additional members of the working group, Glen Bramley (Heriot-Watt) and Mhoraig Green (Convention of Scottish Local Authorities - COSLA), who will attend future meetings.
- 2.3 Peter ran through the Note for PAMS. This note will go to the PAMS group and they will be hearing about the progress of this group at the next meeting of PAMS. The Census Quality Assurance (QA) team now have a contact name for someone within every local authority.
- 2.4 Laura explained that Kirsty MacLachlan has requested that local authorities provide GROS with a cut of record level data as close to Census day as possible. Cameron stated that this might lead to duplication of workload and confusion between what Kirsty is asking for and what is being asked for Census QA. Kirsty MacLachlan stated that GROS are seeking legal advice on data confidentiality and will be putting a case together for a joint agreement on data sharing soon. GROS are keen for local authorities to take snapshots of record level data as close to Census day as possible, even if they are not able to provide the data at that time. Jan asked if this had been discussed at COSLA and Kirsty MacLachlan stated that it had and progress had been made with one local authority but it was then decided that it would be better to pursue this for all local authorities so were advised to go through the Society of Local Authority Lawyers and Administrators (SOLAR) group. Jan stated that Glasgow had tried in the past to provide data but their legal services team were apprehensive about sharing data.
- 2.5 Laura stated that the problem Census have is that this joint data sharing agreement is unlikely to be in place in time for Census quality assurance, so the Census QA team is interested in what can be provided in time to be used in the QA of the Census, it is likely that only aggregates can be provided at this time and at the discretion of each data owner. Cameron asked whether GROS contacting

the Information Commissioner would be another way forward, also adding that getting 32 separate agreements would be difficult. Kirsty MacLachlan agreed that contacting the Information Commissioner would be a good idea. Laura asked if the working group members had any thoughts about how to progress with this to avoid confusion between the request Kirsty is making and that of the Census QA team. Heidi stated that they could do the data analysis but whether or not it can be shared is another issue, getting Chief Executive approval would be key. Laura stated that it might be worthwhile the Census QA team putting in their request but stating that GROS are also seeking a joint agreement.

- 2.6 Jan asked if any data had been received from Edinburgh. Kirsty was unsure but stated that there had been good discussions focussing on the data sources held. Cameron stated that it would be a worthwhile checking that the request for aggregated Council Tax data on Census day has been made as he has not heard anything about it. Cameron also suggested that it might be possible to provide this at Output Area level. Laura stated that this might be too low a Geographic level for analysis judging by analysis she has already conducted.
- 2.7 Jan enquired whether this working group will see the first output from Census, adding that the longer this happens before the release of Census outputs, the more pressure it would place on the members in their day to day jobs. Laura stated that GROS will receive data from May onwards but this analysis will be focussed on variable distributions. Population counts will only be available from the beginning of 2012 onwards. Peter stated that each member of the group will need to look at whether they are able to participate or if doing so will make their working environment difficult. Laura stated that it is most likely that the figures will never leave the meetings so this may reduce a bit of pressure.

3. Update on Census Data Quality Assurance Plans

- 3.1 Laura gave an update on the Census Data Quality plans and the progress that has been made since the last meeting of the group. Laura has been pursuing widening the group so that there are members from academia, as well as looking for members from COSLA. The Census QA strategy has been progressing and a document outlining this will be circulated when it is finalised. Also progressing is the list of checks that will be carried out throughout each stage of Downstream Processing, as well as the specifications for the checks that will be carried out in the Data Quality Management System. The list of comparator sources that will be used for Census QA has been finalised, this list only includes sources that GROS are currently aware of so others may come to the attention of GROS. There have been talks around how best to use the Scottish Longitudinal Study in Census QA and these talks are ongoing. Work has begun on building the Data Quality Management System which will run the automatic checks which have been specified. A prototype has been built which met the requirements set out by the Census QA team so now work will proceed toward building the full system. Also work has begun on producing local authority summary profiles which will help with the initial QA of the Census variable distributions. The profiles will contain data from published comparator sources which have comparable variables to that of the Census, allowing analysis of variable distributions.
- 3.2 Laura ran through the an outline of quality assurance showing the different phases of QA and how it all fits together. Laura explained that there are two parts to QA. The first part is to check the quality of the data received from the contractor. Checks will be carried out for, amongst other things, coding errors. The second part is to check the quality of the stages of Downstream Processing (DSP) to

make sure that no errors are being introduced and that the stages are working as expected. Laura explained that there are three phases to Census QA. The first phase is to check how the data looks initially and identify how early DSP processes have changed it. The second phase is to check the population counts and then the final stage is to analyse the quality of the data at low levels of Geography, after the 2011 output areas are assigned.

- 3.3 Laura ran through the Census timeline diagram which shows when the phases of QA are expected to occur. Data will be received from the contractor from May 2011 onwards and then the rest of 2011 will be spent running the first few stages of Downstream Processing and checking variable distributions. The coverage process will then be run from January 2012 until May 2012 and then lower level geographies will be analysed from June 2012 until September 2012. It is only the new lower level geographies (2011 Output Areas) that will be left until the end to check because they are not assigned until late on in the process. Analysis will be carried out on the data using 2001 Output Areas, as well as Datazones, before June 2012. Cameron enquired whether distributions will be provided as counts or percentages. Laura stated that initially the distributions will be provided as percentages at local authority level and Peter agreed but added that in some instances counts may be provided depending on the situation.
- 3.4 Peter asked what the situation is at the moment with regards Census QA at a UK level. Laura stated that the GROS strategy is fairly similar to the Office for National Statistics (ONS) one and ONS are doing a lot of work on tolerances. GROS have also been looking at how to set tolerances and are looking at using the change over time in a source to develop a method to set upper and lower bounds for what is acceptable. Kirsty MacLachlan asked whether the phases ONS are running will be in line with GROS. Laura stated that overall they are very similar but there are some differences, adding that GROS are harmonising with ONS where possible through workshops and working groups running at a UK level. Peter stated that GROS and ONS are planning to publish their figures at the same time.

4. Coverage Adjustment Methodology

- 4.1 Neil ran through a presentation explaining the Coverage Adjustment Methodology.
- 4.2 Kirsty MacLean asked why some babies were left off the questionnaire in 2001. Neil stated that in some cases about it is likely that the parents forgot to include babies on the questionnaire. It is thought that those babies who were still in hospital at the time of the Census were most likely to have been missed off the questionnaire. In other cases respondents may have needed to request a continuation form for a baby, as they couldn't fit them onto the household questionnaire, but did not do so.
- 4.3 Tom stated that it looked strange that in 2001 Aberdeen was particularly well enumerated when compared to other cities. Neil agreed that this does look strange but there is no evidence to suggest that it is not accurate and the reason for it being well enumerated compared to other cities is unknown. Another strange figure shows Argyle and Bute being the second worst enumerated local authority in Scotland. Neil explained that GROS think this is purely down to Faslane Naval Base being poorly enumerated.
- 4.4 Peter asked who is doing the programming of the Coverage Adjustment process. Neil advised that ONS are doing the IT work and GROS will adapt the algorithm where necessary to meet Scottish needs. Peter also asked how the Coverage

Adjustment process fits in with Statistical Disclosure Control. Neil stated that the method for Statistical Disclosure Control is 'Record Swapping'. This method will swap records between different geographical areas to create doubt so that it is not possible to definitely identify someone. The Coverage Adjustment process and edit and imputation create uncertainty before Statistical Disclosure Control is run so this will be taken into consideration and will reduce the amount of Record Swapping required.

- 4.5 Heidi asked if imputation will be used for Statistical Disclosure Control and Neil confirmed that, whilst it was looked at as a potential method, it will not be used. Heidi then asked if Coverage Adjustment will be discussed at this group and Laura confirmed that it will when that stage in DSP is reached.

5. Local Authority Templates

- 5.1 David ran through the templates which have been prepared to be sent to local authorities for completion. The three templates are 1. Comments on Comparator Data Sources to be used in Census QA, 2. Local Authority Datasets and 3. Local Authority Knowledge. The Local Authority Datasets template has been designed to allow each local authority to provide details about any dataset which they are able to provide for use by GROS. GROS would be looking for information about the source, such as the breakdown of the data, the level at which it is available at, the coverage and quality of the data and the time period it relates to as well as when it would be made available to GROS. It was suggested that it would be good to have an example in the template to show what information is desired.
- 5.2 The Local Authority Knowledge template allows each local authority to provide additional information about their local authority. GROS are interested in any issues or potentially useful information as well as details of the geographic level at which they occur.
- 5.3 The Comments on Comparator Data Sources to be used in Census QA template lists all the data sources known and available to GROS with space for each local authority to comment on the quality of the data from each source for their local authority, highlighting issues and providing evidence to back up comments. It was suggested that this document would require a lot of work and could be off putting to someone looking to complete it. Laura suggested that it would be a good idea to list two or three core sources that GROS would be looking for detailed comments on and then have an 'Others' section where each local authority can comment on other sources as they deem necessary. Euan suggested that it would be beneficial to speak to someone involved in the Scottish Household Survey to see how they weight their figures because if they use the Census then this could become circular. It was suggested that it would be a good idea to take these documents to the PAMS group for feedback.

Action Point 1: Take templates to the next meeting of PAMS for feedback.

- 5.4 David then ran through the Local Authority Summary Profile template which will be used as a comparator when checking the variable distributions from the Census to make sure that they are looking sensible. All of the information on the profiles will come from published sources. The Summary Sheet will contain all the variable distributions from the 2001 Census (sourced from Scotland's Census Results On-Line (SCROL)), apart from those which are no longer relevant or comparable. As well as this the Summary Sheet will also provide data on other key topics (such as ethnicity, sex ratios and marital status) from comparator data sources. So far a lot

of the information has come from the Scottish Household Survey and the Scottish Neighbourhood Statistics website. Kirsty MacLachlan suggested that it would be beneficial to have sheets in the spreadsheet for Deaths and Migration as key topics. Jan asked if GROS were producing this document for all local authorities and if it would be made available to them. Laura confirmed that GROS will be responsible for producing this document and it will be made available to each local authority. Jan then asked if there would be a summary sheet for Scotland and it was confirmed that there will be.

6. Example Data Anomalies

- 6.1 Laura ran through some examples of the different types of data anomalies that may be encountered, some of which were evident in 2001. The three main types of errors are 1. Respondent Errors, 2. Capture and Coding Errors and 3. Downstream Processing Errors. Respondent errors include not mentioning babies on questionnaires and mistakenly ticking the civil partnership option for the marital status question when the person is actually part of a cohabiting couple. Peter mentioned that during rehearsal there was a problem with some people deeming a civil partnership to be a marriage which was not in a church. Capture and coding errors include individual forms which have been crossed out being captured as person records. There is a stage in Downstream Processing which should delete these records if there are any that are captured initially as real people. Downstream Processing could introduce error such as edit and imputation introducing unexpected variable distributions. Peter stated that it will be a case of keeping an eye on this and being aware that issues could crop up. Euan asked if any work will be done on data errors and characteristics of the respondent. Laura stated that this has not been mentioned so far but could be identified through Edit and Imputation. Alistair asked if any work has been done on poorly answered questions from 2001. Laura stated that the Census Quality Survey will look at the quality of responses to questions.

7. Final List of Comparator Data Sources

- 7.1 Laura ran through the comparator sources in the final list, making reference to some key points. This list contains the data sources that GROS are aware of and will be available for use during Census QA. Other sources may arise as communication with each local authority continues. GROS will look at using NHS Central Register (NHSCR) data at Datazone Level but are aware that there are some issues, such as list inflation. Births data will be used to identify missing babies and it is thought this will be quite an accurate source for comparison. The Demography Armed Forces Survey is in the final list of sources but as yet GROS are unsure if this will be going ahead. There will be access to Communal Establishment data but analysis already done shows the Census will probably be more up to date. Analysis has also been carried out on the Assessors Portal data and this is thought to be a very strong source. Some sources will be used to focus on particular sub groups and some will be used for overall population figures. Laura also added that discussions are ongoing with data owners to see what can be discussed at this working group.
- 7.2 Allison wondered if the Scottish Government departments who have provided these sources would be able to provide contact details for the local authority representative who provided it in the first place. Hazel stated that the International Passenger Survey is not very reliable so Local Government Finance take a three year average when using it. Kirsty MacLean stated that the hope is the

International Passenger Survey is going to become more reliable. Peter added that it is likely to be OK at Scotland level but not be of much use below.

- 7.3 Tony enquired as to whether it is necessary to provide details of a source on the Local Authority Datasets template if it is already mentioned on the Final List of Comparator Data Sources. Laura stated that it would not be necessary to mention it on this document but it would be good to have comments on it in the Comments on Comparator Data Sources to be used in Census QA template.

8. Population Sub-Groups and Prioritisation of Checks

- 8.1 David ran through the spreadsheet showing the checks that the Census QA team are looking to carry out at each stage of Downstream. The spreadsheet lists all the variable distributions that should be checked and cross tabulations after each stage of Downstream Processing has been completed. Most of the checks will be carried out after the data has been loaded initially so that the quality of the data received from the contractor can be checked and then after the imputation process to check the quality and accuracy of the data. Variable distributions will be checked after every process which could effect them to make sure that the distributions are continuing to look sensible and Downstream Processing is not introducing unexpected changes. The Local Authority Summary Profile will be used to help with this analysis.
- 8.2 David then ran through some of the checks which the Data Quality Management System will perform. There are two main types of checks that the system will carry out and these are 1. Population Count Checks and 2. Household Count Checks. The figures from all these checks will be compared to a comparator source and a tolerance level will be set to determine whether the Census figure requires further investigation or not. The method for calculating tolerance levels is still being decided upon. All the checks will initially be viewed at local authority level, with the option to drill down to Datazone level if deemed appropriate and the comparator data is available at this level. Census figures which fall outwith tolerance will be flagged for further analysis. The population count checks will be broken down by five year age groups and where the comparator data allows this could then be broken down to looking at figures by single year of age.
- 8.3 Cameron asked what the Census QA Team hoped to gain from the Fertility check. Laura stated that over-counting or undercounting of females could be identified by this check. Also measure is quite a high profile one so it would be necessary to check that the figures aren't vastly different from what would be expected. If they were then this again would need to be analysed.
- 8.4 Peter asked if the tolerances can be altered in the system. David stated that this will be the case and the system is being built in a flexible way to allow this.

9. Communication (e.g. contacts within each local authority, information for PAMS)

- 9.1 Laura enquired as to what the best method would be for keeping local authorities updated. Alistair suggested that this would be best done by sending a regular update to PAMS, as well as keeping the Census Liaison Officer up to date since they will be a regular contact and are not always members of PAMS. Cameron stated that sometimes, within an local authority, the Census Liaison Officer and PAMS member will not be in contact so it would be important to make sure that things don't go missing and that work is not duplicated.

10. Data Sharing for Quality Assurance Purposes

10.1 Laura stated that pre-release access has been investigated and sharing Census data within this group for the purposes of quality assurance should not be a problem. More work is currently taking place with data owners to clarify what can be shared regarding other data sources.

11. Any Other Business and Date of Next Meeting

11.1 No other business.

11.2 Date of Next Meeting: Before end June 2011