

**Scotland's 2011 Census Outputs Strategy:
Summary version**

Draft at August 2009

Table of Contents

1.	Background	3
2.	Vision	4
3.	Challenges	5
4.	Addressing the Challenges	6
5.	Other issues to solve	7
6.	Next Steps	9

1. Background

- 1.1 The investment of time and resources in conducting a census can only be justified if the results are made widely accessible and in a format that meets user needs. Output from the 2011 Census will potentially be the most complex and comprehensive set of information about the population of Scotland ever produced.
- 1.2 For the 2001 Census, General Register Office for Scotland (GROS) produced much more analysis of the collected data than was the case for 1991, from summary population counts down to data extracts for specialised users. Compared with 1991, fewer results were published on paper with more disseminated on the web (via SCROL). All these products were, however, essentially still fixed in format and coverage, though with some flexibility provided by the commissioned tables service.
- 1.3 The growth since 2001 in information provision through the Internet has raised user expectations in terms of the content of the 2011 Census results and how these should be disseminated. This has a direct impact on the nature of output provision in terms of access and flexibility, and for 2011 the vision is therefore that the Web will be the primary dissemination route with functionality to allow users to define and carry out their own analyses online, make comparisons between 2001 and 2011 data and access comparable UK-wide statistics from a single source.
- 1.4 User comments about the strengths and weaknesses of the 2001 Census outputs, which have since been reinforced by feedback from consultation about their high level requirements for 2011 helped shape a number of the strategic aims for outputs set out in Annex A of the [Registrars General harmonisation agreement for the 2011 Census](#). These aims include:
 - outputs free at point of delivery;
 - dissemination methods and media to keep pace with technological innovation e.g. on-line flexible table generation, mapping and graphing of standard and user-defined outputs;
 - more on 2001 vs 2011 comparisons;
 - a pre-release output prospectus and timetable that is adhered to;
 - concurrent first release of statistics across the United Kingdom (UK);
 - joined up and comparable UK outputs;
 - common UK statistical disclosure control methods that minimise disclosure risk whilst maximising data utility;
 - fully adjusted database so that all results reflect the coverage adjustment process;
 - meeting the new European Union (EU) regulation requirements; and
 - research access to anonymised microdata.
- 1.5 The Registrars General harmonisation agreement contains an overarching statement that the final census product for 2011 should be "...consistent, coherent and accessible statistics for the UK and for each component country ...". An important dimension of the GROS strategy will therefore be how Scottish census

outputs and metadata can interface most effectively with those for England, Wales and Northern Ireland to give the UK-wide picture sought by many users. As a consequence, it should be seen in the context of the corresponding census output strategies being developed by Office for National Statistics (ONS) and Northern Ireland Statistics and Research Agency (NISRA).

- 1.6 The draft GROS 2011 Census Outputs Strategy does not present final detailed solutions. It reaffirms the strategic aims involved and sets out a range of issues on which further studies and research needs to be carried out to determine the final suite of output systems, services and products that will best meet those strategic aims. The focus of work just now is therefore on finalising plans to address the full range of outputs issues and policies involved, and in particular on investigating in more detail possible tabulation and dissemination software tools.

2. Vision

- 2.1 The overarching 2011 Census business objectives can be summarised as being to:

- provide the most accurate possible census population estimates for different geographic areas across Scotland;
- provide the most accurate possible information about the structures and characteristics of the population in Scotland;
- provide accessible output systems with the right content and functionality;
- ensure the widest possible awareness of census outputs;
- ensure user confidence in the results;
- protect, and be seen to protect, confidential personal census information; and
- provide value for money.

- 2.2 The GROS Census Outputs Project will seek to meet the aims and objectives at 1.4 and 2.1 above by delivering a suite of products and services which include:

- Web as the primary dissemination route, incorporating the flexibility for users to define, tabulate, graph, map and download their own analyses of the data;
- Access to comparable UK-wide data from a single source;
- Formalised comparison of 2001 and 2011 data;
- Tabulation software to provide the mechanism via which tabular representations of the data are produced;
- An increased set of products providing analysis of the census data;
- CD/DVD products to supplement the on-line product set along with a reduced commissioned table service;
- A limited range of printed products, including a report to Parliament and a reference volume of Scotland level statistics;
- Microdata products will be provided via secure mechanisms;
- A feed of census data for datazones into the Scottish Neighbourhood Statistics website; and

- A feed of Scottish census data into a UK database at ONS that will sit under an Application Programming Interface (API) on which developers can build web applications and so provide users with alternative means of accessing UK-wide data.

2.3 A number of high level milestones have so far been identified:

- October 2009 – sign off outputs strategy
- Spring 2010 – conclude user consultation process
- End-2010 – build dissemination systems (populated with 2001 data)
- Mid-2011 to mid-2012 – quality assurance
- Mid-2012 – spring 2014 – production and dissemination of 2011 outputs.

3. Challenges

- 3.1 In order to achieve our aim of getting the greatest value possible from the 2011 Census data we need to ensure that the outputs are easily accessible and have the relevant content to meet user needs, whilst being mindful of the need to protect confidentiality and achieve value for money.
- 3.2 The extent of functionality that can be provided to users, and the level and number of dimensions of data that can be supported, are dependent upon the technology used to implement the solution. This applies to backend data processing, tabulation software and web delivery mechanisms.
- 3.3 Similarly, determining the type and format of content that can be supported is dependent upon the technology used to implement the solution as well as having a strong dependency on the rules applied to ensure that data are non-disclosive. A review of the available technology suggests that multi-dimensional tables, or 'hyper-cubes', may be the most effective way of transforming back end data for the user to access. This proposed move to hyper-cubes potentially introduces enormous flexibility in the number of dimensions that can be joined and the levels of geography and classifications that can be contained within this. This results in the potential for far more data and outputs to be produced than hitherto, though the balance between data utility and an acceptable level of risk to disclosure is something that needs to be addressed.
- 3.4 The key challenge is therefore to align user requirements for output content with the technical capabilities of the chosen solution alongside the constraints of the disclosure control methods and rules. The decisions that need to be taken to resolve these issues are, in summary, to agree:
- the technology to be used to deliver the backend data processing;
 - the technology to be used for tabulation;
 - the technology to be used for web dissemination (and the extent to which this will be delivered by the UK census offices and/or by partners taking account of the needs of different types of users);
 - metadata collection and delivery integration mechanisms;

- the content of the output products and how to align with technology;
- the number of hyper-cubes and constituent dimensions that can be supported by the technology;
- the degree of perturbation required for each level of the geographic hierarchy taking account of the resulting damage to (and therefore utility of) the data at each geographic level (recognising also the complex geographic hierarchies that exist for Scotland);
- what constitutes a 'safe' hypercube space given issues of disclosure from the full combination of dimensions within each cube.

(There are many other areas of challenge to be addressed which are less critical than those above - these are detailed in section 5 of this document.)

4. Addressing the Challenges

- 4.1 In order to meet the key challenges noted in section 3 above we will run a consultation round with census users, starting some time towards the end of 2009. This will seek to establish views on:
- content of output products;
 - the required functionality for accessing the data;
 - geographic products to be supplied;
 - disclosure control, and the balance between the level of perturbation and the level of detail included in the outputs;
 - licensing, funding and charging models;
 - analytical uses of the data, e.g. linking census with other data and using census data as a sampling frame for follow-up surveys; and
 - microdata samples.
- 4.2 There is an obvious need to align the consultation process with the development of the technical solution for producing and disseminating outputs. The difficulty comes from whether the user requirement is used to dictate the technical solution or whether the constraints of the technical solution are reflected in the final product set and thus impact on how far the user requirement can be met.
- 4.3 There is a similar issue associated with disclosure control in that the products defined as a result of user consultation may reveal technical hurdles that need to be overcome. Our approach to addressing this issue is to base the user consultation on the assumption that a hyper-cube solution will be used whilst also working with traditional table definitions to provide a fall back should the hyper-cube based approach prove unworkable from a technical or disclosure perspective. Throughout the user consultation round we will therefore need to work to confirm the continuing viability of the proposed approach and identify any unexpected complexity and potential blockers to a fully flexible solution.
- 4.4 The final element to address is to ensure that a solution for metadata allows for the metadata products to be integrated with the delivery of data to meet the user requirements for concurrent release of metadata. GROS is contributing to work being led by ONS on this aspect: it is envisaged at this stage that Scottish metadata

will feed into a UK depository but that GROS will also need to maintain its own metadata library.

- 4.5 The ONS Census Outputs team will continue to work with the ONS Web Development Programme (WDP) to build on the initial success of the Application Programming Interface (API) Proof of Concept work by delivering a prototype system with web dissemination functionality in December 2009. They will also engage with partners such as SASPAC and Manchester University to determine the extent of data package formats, API and front end functionality to be delivered by ONS and that to be provided externally. Whilst there is not perceived to be the same need for GROS to deliver a similar API-based solution for its web dissemination of census results, Scottish users will potentially benefit from having another channel to access UK-wide data should the ONS prototype work come to fruition. (If it doesn't, then we will need to consider alternative means of providing users with access to UK-wide census data from a single source.)
- 4.6 Feedback from the forthcoming round of user consultation will be evaluated and reviewed against technical and disclosure constraints, with the aim of developing a fully defined and agreed outputs product set by June 2010. Taking the above approach will allow us to provide answers to the key decisions set out in section 3.4 above, give a clear picture of the products required (and how these will be provided) and enable the timetable for their delivery to be formalised.

5. Other issues to solve

- 5.1 As well as the complex inter-related issues described in section 4, there are a number of other areas which need to be addressed by the project.
- 5.2 Tabulation tools – this will be an integral part of the 2011 Census Outputs solution. There is no intention for GROS to build this software in-house as a number of tried and tested products already exist within the market, likely contenders are SAS and SuperStar. GROS will also need to take into account the outcome of the exercise which ONS will soon be embarking on to procure statistical software for its census and other work.
- 5.3 European requirements – work has been underway for sometime evaluating the impact of European Regulations on the 2011 Census Outputs. This concerns both the content and format of the outputs. The European functionality requirements are consistent with approach being developed by ONS for census outputs (i.e. hyper-cubes, API). The topics covered by the regulations will mostly be delivered via the 2011 Census. Where these topics are not covered by the census, alternative sources will be identified to fulfill the duty laid upon the census offices. ONS leads on this area of work on behalf of all the UK census offices.
- 5.4 Microdata/licensing – This is also an area of work that ONS leads on behalf of all the UK census offices. There has already been some engagement with key users of census microdata, and the Census Microdata Working Group is the primary UK forum for decisions on what is needed in this area and how best to provide this. The user consultation due to start towards the end of 2009 will take forward the

investigation and evaluation of the complex requirements for microdata samples and how these may be affected by the disclosure control policy. In parallel work will progress to evaluate technical solution for the delivery of microdata, primarily focusing on the existing Virtual Microdata Laboratory and the Secure Data Service pilot work. The aim of GROS will be to generate and provide the required extract(s) of microdata for Scotland.

5.5 Charging and funding - The overall goal is to meet the census aims and provide standard census outputs free at the point of delivery with some cost recovery charging for more specialist products. During the coming year ONS will work towards establishing and then disseminating decisions regarding funding of and charging for Census Outputs. Their project faces key dependencies in terms of funding allocation from HM Treasury and Royal Mail/OS licensing policy as well as the extent to which partnerships with other organisations can be achieved. Similar considerations are likely to apply to GROS. The dependencies identified by ONS are summarised as follows:

- engagement with Partners for joint funding bids between July 2009 and December 2010;
- user views sought on charging options from October 2009 to April 2010;
- continued engagement with Royal Mail/OS on licensing issues expected to culminate in September 2009;
- production of provisional charging/funding models to complete in December 2010;
- partnership working proposals to be finalised in December 2009.
- Central Funding Allocation made in Summer 2010.
- funding and charging policy finalised in Autumn 2010.

5.6 Analytical uses - A key aim of the 2011 Census is to ensure as wide a possible use and exploitation of the census results. In 2001 a variety of 'Focus On....' and other reports were produced by the census offices. For 2011, ONS aim to expand on this by using other data sources to provide informative reports about what the census results tell us about trends within the UK. GROS will need to consider the extent to which it can contribute to this work and/or to undertake something similar for Scotland. Consideration will also be required of the extent to which other analytical uses can be made of census data, e.g. for linking census and other data sets and for using the census as a sampling frame for follow-up surveys.

5.7 QA of tables - The 2011 Outputs project has a responsibility to ensure that the products produced are correct and accurate and also adhere to wider quality standards as defined by the UK Statistics Authority (a requirement for National Statistics accreditation). Work is underway at ONS to establish what procedures, processes and resources will be required to achieve this, with the aim of having a full plan of costed activities by December 2009. Similar work will require to be scoped and costed by GROS, including output products that relate to UK-wide data.

5.8 Output Timetable - A key criticism of the 2001 Census was that products were not delivered to timetable. In some cases this meant users sought different sources for the data and did not then return to the census products (particularly in the case of

microdata). It is essential for 2011 that we produce a timetable that is realistic in that it as far as possible caters for all eventualities. Work on preparing this timetable will not commence until the key challenges set out in sections 3 and 5.1 to 5.7 above are addressed so that we can be sure footed when presenting dates. It is expected that the timetable for release will not therefore be available until October 2010.

6. Next Steps

- 6.1 GROS is scheduled to run a formal census consultation round commencing towards the end of 2009. The main focus of this consultation will be on outputs and discussions are underway with ONS to co-ordinate this as far as possible with their programme of consultation (which is scheduled to run from autumn 2009 through to late spring 2010). The topics for the GROS consultation have still to be finalised but will include those mentioned at section 4.1 above.
- 6.2 GROS has also established a time-limited working group of Scottish users to review a number of detailed issues relating to census output geography. The working group, which held its first meeting on 26 August, aims to report back to the Population and Migration Statistics (Scotland) Committee by spring 2010.
- 6.3 Several detailed strategy documents are being developed by ONS in relation to the following elements of census outputs:
 - user consultation
 - microdata
 - technical issues
 - charging and funding policy
 - partnership Working
 - content
 - delivery formats/channels
 - metadata
 - 2001/2011 comparisons.
- 6.4 There is a UK dimension to each of the above strategy elements, and GROS will be participating fully in the working arrangements on outputs which are being established with the other UK census offices to safeguard the Scottish interests and to derive maximum benefit from a co-ordinated approach being taken to the detailed work involved.
- 6.5 Following agreement of the GROS 2011 Census Outputs strategy, a detailed Project Initiation Document (PID) and project plan will be completed (some preliminary work on this has already been carried out). These will seek to specify each strand of work involved, their owners and identify all the key deliverables, milestones, dependencies, estimated resources required, risks and mitigating actions. The PID and project plan will then be reviewed with ONS and NISRA to ensure that each census office has clearly identified the dependencies it has on the others, and that timetables are co-ordinated where they need to be. The aim will be to sign off the PID and project plan by the end of October 2009.