# 2011 Census

# Outputs Strategy

**Introduction**

Output from the 2011 Census will be the most complex and comprehensive set of information about the population ever produced.

For 2001, ONS produced much more analysis of the collected data than for the previous 1991 Census, from summary population counts down to data extracts for specialised users. In the intervening years, users have become even more demanding, encouraged by the growth in information provision especially through the internet. This has a direct impact on the nature of output provision in terms of access and flexibility.

As stated in the 2011 Census Business Case *"The ultimate benefits of the census are realised when the users of census data make use of the published outputs."* Therefore the investment of time and resources in a census can only be justified if the results are made accessible and the outputs produced meet user needs.

In 1991, the majority of outputs were paper based. As a result, the products were fixed in coverage and format. Flexibility was provided, to a limited degree, by the Commissioned Tables Team.

In 2001, less paper was produced following the rise of web dissemination. The products were, however, still fixed in format and coverage. Flexibility continued to be provided by the Commissioned Tables Team.

Projecting this trend forward, the vision for 2011 is to provide only the set of paper products that may be sold via publishers. Web will be the primary dissemination route and will be enhanced to include flexibility for end users to create their own products, including formalised comparison of 2001 and 2011 data.

This document sets out the ONS Strategy for dissemination of outputs from the 2011 Census.  A separate consultation currently underway will define the content of the output dissemination systems.

## Context

**Statement of Agreement between the National Statistician and Registrars General**

The National Statistician and the Registrars General for Scotland and Northern Ireland have made a Statement of Agreement about the conduct of the 2011 Censuses across the UK. This includes an annex relating to the agreed aims for 2011 Census Outputs, based on feedback from users of the 2001 Census outputs.

The essential elements of these aims are:

- Provision of 2011 Census data via systems and services which exploit modern technologies enabling users to interact flexibly with the data at various levels
- Outputs that are comparable across the UK

The strategy set out in this document specifically refers to England and Wales, but we are working closely with the General Register Office for Scotland (GROS) and the Northern Ireland Statistics and Research Agency (NISRA) to ensure UK comparability, meet the aims in the Registrar Generals' agreement and ensure a single point of access to UK-wide outputs.

**2011 Census business objectives**

The following are the overarching 2011 Census business objectives

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

All have a relevance and need consideration in the delivery of 2011 outputs, but it is the functionality and content which have the greatest relevance and which the Outputs Project has sole responsibility for delivering.

**User consultation**

From the first release of 2001 Census results feedback has been actively sought regarding all aspects of census outputs. The continuation of user forums such as Census Advisory Groups, and specific interest groups, such as the Microdata Working Group, have provided mechanisms for gaining an understanding of the census users continuing and developing needs.

The 2011 Census Project has also, in more recent years, undertaken focused consultation work utilising mechanisms such as web surveys, blogs and wikis. The outcome from all these consultation mechanisms has been collated and used to influence the decisions taken in defining the 2011 Census Output Policy and the resulting vision and strategy for 2011.

Further detail on user requirements will result from a user consultation round undertaken between October 2009 and April 2010, building on the outcomes from earlier consultation work. It will include all topics which affect the production and dissemination of census data including: the mechanisms of delivery; disclosure control; geographical information; and funding and licensing.

While none of the topics for consultation is overly problematic in itself, there are a range of issues and dependencies in relation to each that will make engagement with users more complex. This is particularly true in the early stages of consultation where many dependencies are unresolved.

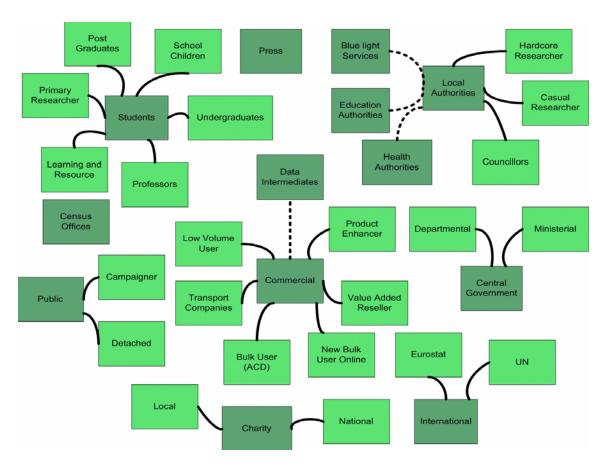**Vision for 2011**

**Overview**

Building on the 2001 experience, the consultation to date, and the development of available technologies, for 2011 the Census Outputs Project will meet these aims by delivering a suite of products and services which include:

- Web as the primary dissemination route, meeting the common needs of users, incorporating the flexibility for users to create their own products

- Formalised comparison of 2001 and 2011 data

- Utilities to enable bulk down load of data via the web

- Data explorer functionality provided jointly with external partners to ensure the requirements of more sophisticated users are met in the most timely and cost effective manner

- An increased set of products ensuring maximum analytical use and therefore value of the data and census outcomes

- CD/DVD products to supplement the online product set with a reduced customised table service

- A limited set of paper products to be sold via publishers

- Microdata products provided via secure mechanisms

- Provision of updated data for inclusion in the ONS Longitudinal Study

- Licensed access for trusted users to more complex outputs that do not satisfy disclosure control requirements for public availability

- Provision of outputs according to European Regulations

This becomes achievable as a result of advances in technology and the implementation in 2001 of a geography which remains stable for 2011. However, delivering these services presents numerous challenges. There are core dependencies and a number of decisions which need to be taken over the coming year to enable us to achieve our aims.
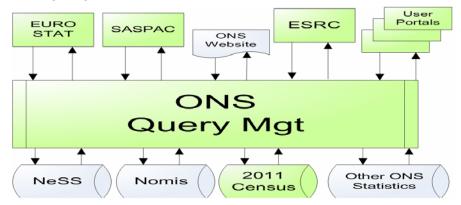
**2011 user experience**

The groupings of users of census data are illustrated in the following diagram:



Each user group brings with it a variety of levels of experience in using census data and a similar variety of requirements for access and uses of the data. Delivering systems and services which meet the full requirements of all would not be practical or financially viable. We will therefore seek to deliver a service which meets the common user needs whilst working with partners to extend functionality to meet the specific needs of certain groups.

**Proposed technical solution**

Our technical approach and vision for partnership working is illustrated simplistically in the following diagram.

ONS will provide a query management interface (an Application Programming Interface - API) which external partners can use to access the various forms of ONS data in a backend data store.This could, for example, be Neighbourhood Statistics (NeSS), Nomis data (containing labour market information), 2011 Census statistics or other ONS statistics.

Users would access the data via an appropriate frontend web interface, which may be the ONS web frontend providing basic data explorer functionality or a more specialist frontend created by external partners such as SASPAC, which provides more advanced functionality for a particular user group. A proof of concept has already been produced by ONS, which has demonstrated the feasibility of the proposed approach. It is expected that API users will be required to register to enable central control and understanding of how the data is used, and potentially for charging depending on the outcome of the licensing policy.

We expect that a number of other, as yet unidentified, organisations will develop alternative front-ends to meet the needs of different types of user. ONS will encourage and support such developments.

Users will also be able to carry out bulk downloads of data via the query management interface which also meets EuroStat Regulations for data accessibility.

Users who require access to more specialist products such as microdata will be able to access these via secure mechanisms such as the Virtual Microdata Laboratory or the Secure Data Service.

Users will also continue to have access to expert staff for queries and information via the Census Customer Service Team.

**Key challenges**

In order to achieve our aim of getting the greatest value possible from the census data we need to ensure it is both easily accessible and provides the right content to meet the user needs, mindful of the need to protect confidentiality and achieve value for money.

The extent of functionality that can be provided and the geographic level and number of dimensions of data that can be supported are dependent upon on the technology used to implement the solution. This applies to backend data processing, tabulation software, web delivery mechanisms and data explorer functionality.

Similarly, determining the type and format of content that can be supported is dependent upon the statistical disclosure control policy and methodology used to protect the published data.

A review of available technology suggests that multi-dimensional tables, or hypercubes, may be the most effective way of transforming data for the user to access. This proposed move to hypercubes 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, increasing the disclosure control challenge.

The balance between the detail contained in the statistics, the functionality of the web tools and the level of disclosure control leads to a set of interrelated decisions, none of which can be made in isolation. The key decisions are:

- The technology to be used to deliver the API functionality
- The technology to be used to deliver the backend data processing
- The technology to be used for web frontend development
- The content of the output products
- The metadata collection and delivery integration mechanisms
- The frontend functionality that will be delivered by ONS and that which will be delivered by external partners
- The number of hypercubes and constituent dimensions
- The number of exact-fit geographies that will be supported
- The degree of Disclosure Control required for each level of geography
- The additional disclosure control mechanisms that may be required for more detailed or more sensitive outputs, such as origin-destination tables

There are many other areas of challenge to be addressed which are less critical than those above. These are detailed later in this document.

**Addressing the key challenges**

To make these decisions, we required an intelligently considered path combining research, iterative prototyping, user consultation, policy formulation and partnerships, culminating in a set of clear, well founded decisions for 2011 Census Outputs.

To achieve this we will establish four interrelated workstreams:
1. Technology development
2. User consultation
3. Disclosure control methodology development
4. Metadata system development

Key milestones for each of these are set out below:

**Technology Development**

By March 2010:
- Detail the key user personas (types of user) and their required functionality
- Engage with partners such as SASPAC and Manchester University to determine:
    - Data formats
    - API functionality
    - The frontend functionality to be delivered by ONS
    - The frontend functionality to be provided externally
    - Funding options for partnership working

By May 2010:
- Produce a further prototype, including both ONS systems and external partners' frontends

By June  2010
- Demonstrate prototype systems to key user communities
- Refine key user personas
- Refine backend API system design
- Further develop views on the frontend functionality to be delivered by ONS and that to be provided by external partners
- Refine funding options

By July 2010
- Agree final proposals for partnership arrangements with the associated funding arrangements

By October 2010

- Release a beta version of the backend systems, API and frontends
- Release prototype ONS and partner frontends
- Start system testing

By December 2010

- Release an output system containing a range of 2001 census data for ongoing user review and refinement, and to encourage production of a growing set of externally produced front-end tools

**User consultation**

October 2009:

Launch user consultation through regional consultation events to establish views on:

- Content of output products,
- Population bases to be used
- User persona definitions
- Frontend functionality requirements for each persona
- Geographic products, including:

    - Number of exact fit geographical hierarchies

    - Output Area maintenance

    - Requirements for workplace zones

- Disclosure Control, and the balance between the level of perturbation and the level of detail included in the outputs
- Licensing, funding and charging models
- Microdata samples

By April 2010

- All responses to user consultation received

By October 2010

- Proposed 2011 Census Output product set published for further review

By December 2010

- 2011 Census Output product set agreed

**Disclosure Control methodology development**

By October 2009

- Define user consultation questions and mechanisms in relation to Disclosure Control

By June 2010

- Align Disclosure Control methodology, user requirements for content and functionality and the proposed technical solutions

By October 2010

- Publish decision on Disclosure Control methodology, in relation to content and functionality requirements

**Metadata system development**

By October 2009

- Options for storage and dissemination of metadata identified

By January 2009

- Interim solution for collection and storage of metadata in place

By May 2009

- Trial metadata storage and dissemination system incorporated in backend/API prototype

By October 2010

- Metadata systems implemented in beta version of output systems

**Workstream integration**

In aligning technical decisions with user requirements and disclosure control constraints, we need to manage the balance between user requirements driving the technical solution, and the constraints and possibilities of the technical solution defining the output products and functionality. There are similar issues associated with Disclosure Control.

Our approach to addressing this issue is to base the user consultation on the assumptions that:

- A hypercube solution will be used
- The balance between the level of disclosure control and the level of detail included in publicly available outputs will be similar to that in 2001 output tables
- The 2001 census table definitions provide a good *starting point* for the types of outputs required for 2011

Starting from 2001, table definitions will also provide a fallback, should the hypercube based approach prove unworkable from a technical or disclosure perspective.

A release of a working system with a range of 2001 data from the 2001 census some 18 months is proposed before the first release of 2011 Census outputs.  This is to enable ongoing refinement of backend and frontend systems and the API, to ensure the final system is as mature as possible and tuned to meet the needs of different user groups.

It is also to ensure sufficient contingency in the development process to cater for some of the significant challenges that may arise in this somewhat ambitious strategy.

There is no guarantee that we will be able to address all the challenges identified above, but this strategy, and the close partnership working that is already developing, gives us the best possible chance of success.

As noted above, the fallback is to produce and disseminate information in a similar format and using  similar mechanisms to the 2001 Census data

## Other issues to resolve

As well as the complex interrelated issues to solve what is set out above, there are also a number of less complex areas which need to be addressed by the project. These are:

- Tabulation tools
- Systems hosting
- European requirements
- Microdata/licensing
- Charging and funding
- Wider analytical uses
- 2001/2011 comparisons
- QA of tables
- Output timetable

## Tabulation tools

Tabulation capability is an integral part of the 2011 Census Outputs solution. A number of tried and tested products already exist within the market and ONS will select a tabulation tool in March 2010.

## Systems hosting

The intention to move to a hypercube-based storage model and open API may result in a model which is too large to be hosted under existing ONS arrangements. Over the coming months consideration will be given to the potential scale of the system, and options such as partnership hosting, so that an appropriate solution is in place.

The output strategy could potentially result in many hundreds of frontend web interfaces to the ONS API, some with very large user communities requiring high volumes of data. To meet or manage this demand, ONS may need to consider funding arrangement with the highest volume partners to enable ONS to build systems of the required scale.

## European requirements

Work has been under way for some time 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 proposed here for our 2011 Census Outputs (i.e. hypercubes, with an API). The statistical content covered by the regulations will also in the most part be delivered via the 2011 Census. Where topics are not covered, alternative sources will be identified to fulfill the duty laid upon the ONS as far as possible.

## Microdata/licensing

Work is already under way to engage with key users of census microdata and the Microdata Working Group is the primary forum for decisions on what is needed in this area and how best to provide it. The user consultation round will investigate and evaluate the complex requirements for microdata samples and how these may be impacted 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 work to resolve issues of microdata samples and the technologies to deliver these will

progress over the coming 12 months and will be fed by the outcomes from work being undertaken to refresh the wider ONS Microdata Strategy. An outline timetable is as follows:

- The Census Microdata Working Group agreed the programme of work required to develop microdata samples in consultation with CCSR – July to August 2009
- Users will be consulted on general requirements for microdata and kept informed of developments during the consultation round October 2009 to April 2010
- Consultation with a subset of key users and interested parties will take place during July and August 2010
- The final strategy for microdata will be produced in October 2010

**Charging/funding**

During the coming year we will work towards decisions around the funding of, and charging for census outputs. The project faces key dependencies in terms of funding allocation from HM Treasury, Royal Mail/OS licensing policy and the extent to which partnerships with other organisations can be achieved. The overall goal is to meet the census aims and provide standard census outputs free at the point of delivery, with some cost recovery by charging for more specialist products. However this approach will only be confirmed once the dependencies are resolved. These are summarised as follows:

- Discussions with HM Treasury in November 2009
- Analysis of the potential scale of systems to host the service between January 2010 and March 2010
- Engagement with Partners for joint funding bids between January 2009 and March 2010
- User views sought on charging options from October 2009 to April 2010
- Partnership working proposals to be finalised in July 2010
- Central funding allocation made in mid 2010
- Funding and charging policy finalised in Autumn 2010

As in 2001, the 2011 Census will continue to work with value added resellers. The extent of this relationship will be determined during 2010 along with the associated charging policy.

**Wider analytical uses**

A key aim of the 2011 Census is to ensure as wide a possible use and exploitation of the census outcomes. In 2001, a variety of 'Focus On….' reports were produced which took the outcomes from the census, expanding on them using other sources to provide informative reports about what the outcomes tell us about trends within the UK. For 2011, we are seeking to expand on these and have started work with other parts of ONS to establish how this should be taken forward, with requirements already being gathered from a number of sources.

We plan to continue linking census data with the ONS Longitudinal Study, and will explore linkage (in a safe ONS setting, for statistical purposes only) with other ONS and GSS survey and administrative data sources, such as:

- The Annual Survey of Hours and Earnings (ASHE)
- School pupil data

We will also explore the use of the census as a sampling frame for further surveys, and quality assurance of other survey sources.

**2001/2011 comparisons**

2011 Census Outputs will include a set of products which provides comparisons between some aspects of the 2001 and 2011 data. This has been made possible by the introduction of a geographic base in 2001 which has remained stable[1] for 2011. Work is currently under way to establish which areas will be comparable between the censuses given changes to questions, classification and methods. This should be completed in Summer 2010.

**QA of tables**

The 2011 Census 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. Work is under way to establish what procedures, processes and resources will be required to achieve this. It is expected that by March 2010 we will have a fully costed plan of activities.

**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 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 start until the key challenges set out above are addressed, so that we can be sure-footed when presenting dates. It is expected that the timetable for release will therefore be available in October 2010.

---

[1] Approximately 5% of OA will have changed between censuses.

## Next steps

Following agreement of the strategy the next step is to develop and finalise delivery plans in conjunction with external partners to ensure the strategy can be met and the key issues addressed, and to conduct detailed user consultation.

It is inevitable that the activities and dates set out in this strategy will change as understanding develops and issues arise.  The strategy has been designed to allow at least 18 months contingency time for further refinement and development of systems beyond the proposed release of some 2001 data using 2011 output systems in December 2010.  Clearly, the sooner the ONS system can be finalised, the more time there will be for external front ends to the ONS system to develop and mature.

This strategy therefore provides a framework for planning and development that enables ONS, users and partners to move forward over the next 12 months, but will need to be refreshed as the project progresses.