United Nations ECE/CES/GE.41/2018/18

Economic and Social Council

Distr.: General
18 July 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twentieth Meeting
Geneva, 26–28 September 2018
Item 6 of the provisional agenda
Dissemination

Plans for disseminating 2021 Census data for England and Wales

Note by the Office for National Statistics of the United Kingdom

Summary

The Office for National Statistics has been working to ensure 2021 Census outputs are more flexible, timely and accessible compared to the 2011 Census. This document outlines our strategic vision for the dissemination of 2021 Census outputs. We also set out the approach we have taken to gather feedback from a spectrum of users on our design and content and how we are planning to incorporate this feedback into our future research.

In early 2018, we held a public consultation to outline our vision for the content and design of 2021 Census outputs. This included our plans to disseminate the majority of census data through a single point of access via the ONS website using a flexible dissemination system. This will be enabled through an innovative combination of statistical disclosure control methods, which include targeted record swapping, and an automated layer of light-touch perturbation and final disclosure checks. We also set out our plans for the design and dissemination of specialist products, including microdata samples and origin-destination (flow) data products.

We would welcome feedback from members on our approach to dissemination and their experiences of engaging with users to seek feedback on proposals.

GE.18-11914(E)
1811914

I. Introduction: Vision for 2021 Census Outputs and Dissemination

1. The Office for National Statistics (ONS) is responsible for the design and delivery of the 2021 Census in England and Wales.

2. Our strategy for the design and content of 2021 Census outputs and overall dissemination approach is driven by statistical and technical methodologies and user need.

3. Following the 2011 Census, feedback from the ONS Census Advisory Group (CAG), representing interests from the main user communities, specified a requirement for greater flexibility in outputs and a need for more timely release of data. We have listened to the user feedback following the 2011 Census. For our 2021 outputs, we intend to maintain or improve on the quality of the outputs we currently produce based on the 2011 Census. We have developed our strategic aims for the 2021 Census based on this feedback and we aim to improve on the dissemination of census outputs in terms of flexibility, timeliness and accessibility.

(a) Flexibility – we aim to disseminate the majority of census data via a web-based interactive flexible dissemination system where users can specify the data they need and build their own tables;

(b) Timing – we are aiming to disseminate national- and local authority-level estimates for England and Wales within 12 months of census day and all other results within 24 months of census day. In 2011, the first release of data was 16 months after census day, with some estimates published up to 4 years later;

(c) Accessibility – we plan to host the web-based interactive flexible dissemination system through the ONS website meaning the majority of census data will be available from one location. We will also follow UK Government Digital Service guidelines and best practice on accessibility.

4. We are also intending to produce ‘Enhanced Census Outputs’. These are census based outputs, where 2021 Census data and administrative data are linked at the record level to enable the production of multivariate, small area outputs. Variables may be considered for inclusion in these outputs where there is an identified user need and we are confident with the statistical quality of the administrative data source. We are currently exploring the possibility of adding variables about income, floor space and number of rooms.

5. The final design and content of 2021 Census outputs and overall dissemination approach will be driven by statistical and technical methodologies, but importantly user needs.

6. This paper outlines our vision for designing and disseminating standard outputs and specialist products for 2021 Census. It will also consider how we are ensuring the census products meet user needs and the research we intend to undertake to ensure we successfully deliver the 2021 Census outputs.

II. Flexible Dissemination System

7. We aim to publish 2021 Census data on the ONS website in a combination of pre-determined tables and a web-based interactive flexible dissemination system where users can specify the data they need. The interactive dissemination system is intended to enable users to define their own queries and build their own datasets, selecting the geography,

population base and variables they need. Users will be able to produce datasets to the same level of detail and comparable with 2011 Census data.

8. These are significant advantages that are intended to enable users to meet earlier deadlines, allow greater time for analysis, and deliver data that are more specific to their needs.

9. In the proposed flexible dissemination system:

- repeating the same query for the same area will always result in the same results;
- the same cell will have the same count when appearing in different tables;
- totals and sub totals for different queries for the same area may be slightly inconsistent, but the differences will be very small;
- the sum of estimates for small areas that constitute a larger area, for example all Output Areas (lowest level geography) in a Local Authority area, will not exactly match data in an identical table for the larger area, but the differences will be very small.

10. We have identified there may be some census data which does not pass through the flexible dissemination system, but will still have the same statistical disclosure control methodology applied, for example data on small population groups derived from topics that are new to the 2021 Census. We are exploring how to provide this data in an alternative format, for example in a pre-determined table.

11. We have worked with an external company, to build a prototype flexible dissemination system and we are now working towards fitting this within the ONS IT architecture.

12. We are working towards providing an application programming interface (API) to allow access to census data via programmable requests as well as through a user interface on the ONS website. The range of data available through such an API would be the same as that provided through the flexible dissemination system.

13. We will also continue to provide a service for creating bespoke or commissioned outputs that have not been included in standard output products following the 2021 Census. We will apply the same statistical disclosure control methodology as all other standard outputs.

III. Statistical Disclosure Control

14. We are planning to adopt an innovative statistical disclosure control (SDC) methodology that will underpin the flexible dissemination system. This approach will enable the data to be protected dynamically and speed-up the process of data release. This has been discussed previously at UNECE Meeting of the Group of Experts on Population and Housing Censuses 2017[1], UNECE Work Session on Statistical Data Confidentiality

[1] https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/WP28_ENG.pdf

2017[2] and agreed as the default method for protecting hypercubes across the EU in the 2020 Census round[3].

15. There will be three main statistical disclosure control techniques to protect the census data:

A. Targeted record swapping

16. This will be the main protection for the census data and is broadly similar to the approach taken in 2011. Record swapping will occur for individuals and households with rare or unique characteristics during the data processing stages.

B. Perturbation – Cell-key method

17. This will be a 'light touch' application of the method. Users will select the population base, geography and variables required via the flexible dissemination system and the tool will then apply a cell-key perturbation method to ensure the table generated satisfies legal requirements on protecting confidentiality. This method will need to ensure:

- there has been sufficient uncertainty created systematically about whether or not the small cell is a true value;
- creating uncertainty does not significantly damage the utility of the data;
- ensure individuals and households cannot be identified by differencing cell totals between tables.

C. Automated table design checks

18. In addition to the record-swapping and perturbation, there will also be automated table design checks to ensure the table meets confidentiality requirements, for example determining the maximum number of variables in a table, percentage of cells that are populated and number of cells with real or fake 1s. These online automated table design checks provide an additional benefit that each geographic area is checked for disclosure separately and considered in its own right. Therefore, some geographic areas for certain variables or combination of variables may pass disclosure checks and others may fail. In 2011, if one geographic area failed a disclosure check, then data would not be available for any area. In 2021, we are proposing that if an area passes the checks, data will be available to the user, regardless of whether another area fails. The cell-key method will also protect against disclosure by differencing in cases where every area but one could be released from a larger area. This should lead to a greater access to the data.

19. We are currently seeking assurance of our methods from a wide range of experts. The UK National Statistician has supported the continued development of the statistical disclosure control approach.

[2] https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2017/7_table_builder.pdf

[3] https://ec.europa.eu/eurostat/cros/system/files/recommendations_for_the_protection_of_hypercubes.pdf

IV. Specialist Products

20. In addition to the standard outputs, we are planning to produce specialist products of microdata samples and origin-destination (flow) data.

V. Microdata

21. Microdata are small samples of census data for whole households and individuals, which include some associated census characteristics but no information that could identify a household or individual. These samples enable users to access data similar to that, which might be collected if users conducted a survey themselves and can be analysed in the same way.

22. In 2011, the following microdata files were created with different access arrangements as can be seen in Table 1 below.

Table 1
Types of microdata

Type of microdata	*Description*
Microdata teaching file	• Accessed via the Office for National Statistics (ONS) website • Random sample of 1% of individuals in 2011 Census records • Limited set of variables
Safeguarded microdata	• Accessible via UK Data Service (UKDS) • Two random samples of 5% of individuals • One sample is a regional-level file • The other sample is a grouped local authority or single local authority-level data where a population threshold is reached
Secure microdata	• Accessed via ONS Secure Research Service at ONS Titchfield, Newport and London • Accessible by ONS Approved Researchers Individual file • 258 variables for over five million individuals (10% sample) • Lowest level of geography – local authority • Includes residents in households and communal establishments Household file • Person- and household-level data • 245 variables for over five million individuals within more than 2.4 million households (10% sample)

23. We anticipate the design of the microdata samples will be broadly similar to 2011 Census. We are working closely with expert microdata users through a Working Group to develop a specification that meets user needs. We are reviewing the 2011 sample design and the variables included within each sample. We also need to ensure the outputs reflect a suitable balance of disclosure risk and utility.

24. We are also going to be assessing the impact of the UK Digital Economy Act 2017[4] on microdata samples. The Digital Economy Act provides legal rights to access data held

[4] http://www.legislation.gov.uk/ukpga/2017/30/contents/enacted

by Government departments, other public bodies, charities and large/medium sized businesses. This Act may enable us to make more record-level census data available to researchers with the ability to join to administrative data in a secure research environment.

25. We are also making sensitive census data more easily available to users via our Secure Research Service[5]. This means Approved Researchers will be able to access sensitive data from more physical locations across England and Wales than was previously possible.

VI. Origin – Destination (flow) data

26. Origin-destination (flow) data show the flows of people from one place to another. We produced four types of origin-destination statistics following 2011 Census on migration (internal and international), workplace (commuting patterns), second residence, students. Origin-destination data were also accessible via the ONS website and secure research settings. Due to the complex nature of these data, some tables were not published until four years after census day.

27. Historically, it has been difficult to find a suitable statistical disclosure control methodology that protects the confidentiality of the data, but allows data to be available to meet user needs.

28. In 2001, origin-destination tables were produced using a post-tabular small cell adjustment statistical disclosure control methodology. This allowed wide and easy access for users, but adversely affected the utility of outputs. In 2011, the statistical disclosure control policy using targeted record swapping required that different levels of access were provided for origin-destination data.

29. For the 2021 Census, we have been exploring the option to protect data by applying the cell-key perturbation method and supressing the remaining disclosive values. This methodology is being presented at the Privacy in Statistical Databases 2018 Conference in Valencia[6]. We hope this methodology will make more data publicly available to users.

30. We have already created a Working Group to ask for feedback on our proposed statistical disclosure control methodology and the content of the product. We anticipate origin-destination data are unlikely to be available via the flexible dissemination system tool, due to the nature of the dataset – flows from location A to location B.

VII. Metadata

31. Metadata are produced alongside census outputs to provide contextual information about the data and aid understanding and interpretation. Prior to the 2011 Census, users identified the requirement for metadata located alongside the outputs and clear signposting to supporting information. We provided a range of documentation about the information contained within the data tables, and supporting documentation, for example about the quality of outputs and definitions used.

32. In 2021, we intend to produce metadata including definitions, details of variables, classifications and quality indicators to support the use of statistical outputs. These are likely to include confidence intervals and other quality measures and information from

[5] https://www.ons.gov.uk/aboutus/whatwedo/paidservices/virtualmicrodatalaboratoryvml
[6] https://unescoprivacychair.urv.cat/psd2018/index.php?m=soon

post-census coverage and quality surveys such as local coverage rates, item non-response and imputation rates.

33. We plan to extensively test the 2021 Census online dissemination platform with a range of data users. We will use this testing to identify the most appropriate way to provide consistent and clear metadata to users.

VIII. Ensuring quality outputs - National Statistics Accreditation

34. The ONS has a responsibility to ensure the 2021 Census results are correct, accurate and comply with the UK Government Statistical Service Code of Practice for Official Statistics[7] as determined by the UK Statistics Authority (UKSA).

35. The Code of Practice for Statistics promotes the production and dissemination of official statistics that inform decision-making. It outlines the necessary principles and practices producers of statistics must commit to when producing and releasing official statistics. The Code, when complied with, gives users of statistics and citizens confidence the published government statistics are of public value, are high quality and trustworthy. The three 'pillars' that constitute the Code of Practice are:

- Trustworthiness – confidence in the people and organisations that produce statistics and data. Trustworthiness is a product of the people, systems and processes within organisations that enable and support the production of statistics and data;
- Quality – data and methods that produce assured statistics. This means statistics fit their intended uses, are based on appropriate data and methods and are not materially misleading;
- Value – statistics that support society's needs for information. This means the statistics and data are useful, easy to access, remain relevant and support understanding of important issues.

36. Official statistics assessed as fully compliant with the Code of Practice are designated as National Statistics by the Office for Statistics Regulation, the regulatory arm of the UKSA, in line with the Statistics and Registration Service Act 2007.

37. In 2011, Census statistics were awarded National Statistics status. We have already started research to identify the processes and procedures required to ensure we retain National Statistics accreditation before or in readiness for publishing 2021 Census outputs.

38. These processes are likely to include, but are not limited to:

- Notifying users at least four weeks in advance of the release of census results;
- Providing supporting metadata, including a report on quality and methodology, to accompany and support the release of the main statistical outputs;
- Providing users with our assessment of the quality of census results, based on our rigorous quality assurance of outputs.

39. The interests of users of statistics are integral to the Code of Practice. We are going to continue to involve users throughout the design and development of 2021 Census outputs to ensure we retain the National Statistics status.

[7] https://www.statisticsauthority.gov.uk/code-of-practice/

IX. Stakeholder Engagement

40. We recognise there are a broad spectrum of census data users, with many different interests and needs. It is not feasible for us to consult with all census data users to the same extent and level of detail.

41. Figure 1 illustrates the key stakeholders we seek to engage with, but we recognise there may be other users we have not identified.

Figure 1
2021 Census Outputs and Dissemination stakeholders

Abbreviations:

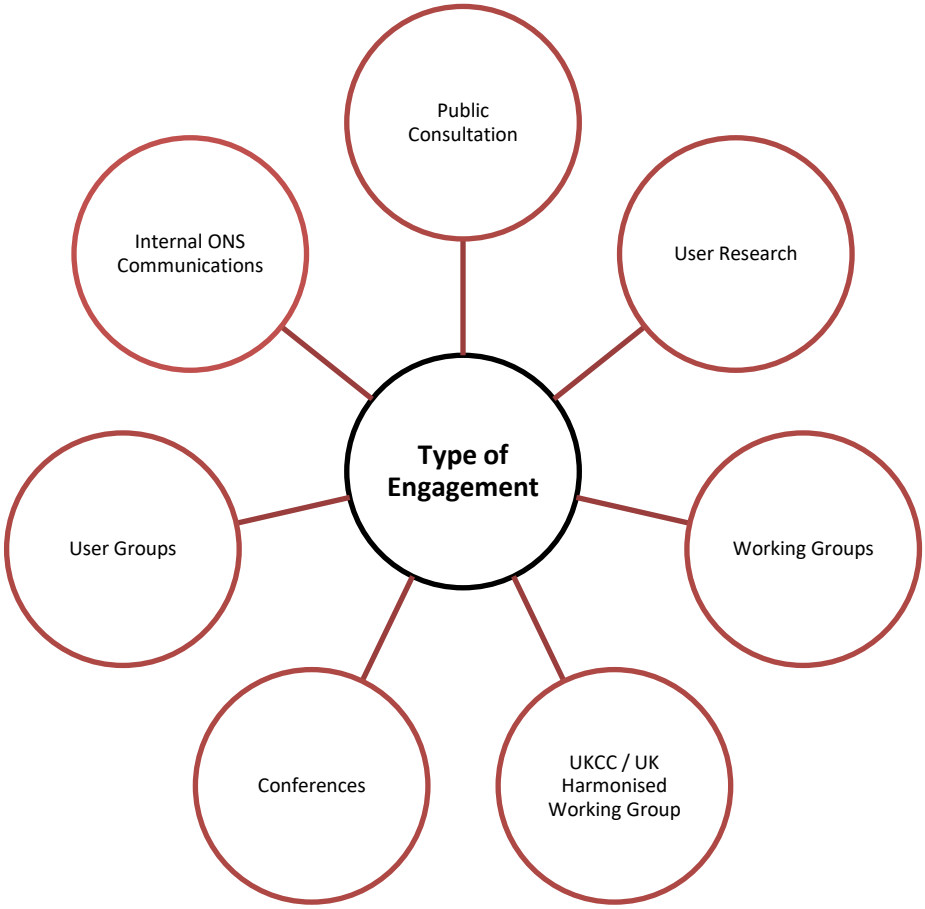
NRS	National Records of Scotland
NISRA	Northern Ireland Statistics and Research Agency
ONS	Office for National Statistics
GSS	Government Statistical Service
UKSA	United Kingdom Statistics Authority
Long. Study	Longitudinal Study

42. Therefore, we intend to engage with users in a variety of ways throughout the duration of the design and dissemination of outputs. This includes inviting all potential users to participate in our public consultations and then actively engage with interested users from a range of sectors to ensure the outputs meet their needs. We have also formed Working Groups to inform the design of our Microdata and Origin-Destination products.

Stakeholders from national government, local government, academia and the commercial sector are represented in the Working Groups. We plan to continue to test the functionality of our proposed dissemination mechanism and make decisions informed by extensive user research. User research involves engaging with users, often on a 1:1 basis to identify their specific needs and pain points, before collating these and feeding back to the project team to inform future design.

43. Figure 2 illustrates the type of stakeholder engagement proposed throughout the lifespan of the outputs and dissemination project.

Figure 2
2021 Census Outputs and Dissemination methods for engaging with stakeholders



X. Public consultation – Initial View on 2021 Census Output Content Design

44. In Spring 2018, we held a public consultation which outlined our strategy for 2021 Census outputs, including our initial proposed design of outputs and the dissemination channels for England and Wales[8]. This consultation was the first step towards

[8] https://consultations.ons.gov.uk/census/initial-view-on-the-2021-census-output-design/

understanding user needs for 2021 Census outputs. The consultation closed at the end of May 2018.

45. The consultation covered all aspects of 2021 Census outputs, including our plans for a flexible dissemination system to access the majority of census data and the approach for statistical disclosure control for 2021 Census outputs. We also informed users of our plans for specialist products (including microdata and origin-destination products), geography and the use of administrative data to produce enhanced census outputs.

46. We consulted with users to:

- obtain feedback on our proposed design and method of dissemination for 2021 Census data;
- help us prioritise our future research and development of 2021 Census outputs;
- inform the order in which we release our products after 2021 Census;
- understand user requirements for a variety of census products.

47. We received 196 responses to the consultation. Over half of the responses were received from local authorities/councils. We also had feedback from academics, national government and charities.

48. We are currently analysing the results and plan to publish our response to the consultation in autumn 2018.

49. The consultation was accompanied by five interactive roadshows across England and Wales, attended by approximately 250 users, from national and local government, academia and the commercial sector. These allowed us to further promote the consultation and outline our vision for 2021 Census outputs, but also enabled users to try out a prototype of the flexible dissemination system, read poster displays about our work and discuss this with members of ONS staff.

50. Overall, the majority of feedback from the roadshows was positive. Users welcomed the opportunity to use the flexible dissemination system and they provided useful insights into the design of the user interface and impact of the cell-key perturbation method on results. For example, in what format they wanted to download the data, how the data were displayed in the table they created and whether they wanted hover-over displays to describe the variables they were selecting.

XI. Analytical Programme

51. We wish to ensure the widest possible awareness and use of 2021 Census outputs and add value by helping users to understand and interpret census data. This includes our expert users of census data, but also individuals who may have a personal interest. We aim to provide commentary and analysis to support the release of census data. As well as using census data, alongside administrative and survey data to describe the society we live in.

52. We intend to consider the different types of census data users and provide analysis and commentary in a range of formats, including stories, summaries and interactive data visualisations. We aim to publish this supporting information alongside the release of 2021 Census data according to a pre-released timetable.

XII. Conclusions and next steps

53. This paper has set out our vision for designing and disseminating standard outputs and specialist products for 2021 Census and how we are ensuring the products meet user needs.

54. We are continuing to focus our research on ensuring that we can meet our strategic aims. In 2019, the ONS is undertaking a dress rehearsal of the collection, data processing and dissemination components of the census. We are going to be assessing the interfaces and dependencies, processes, methods and systems relating to outputs and dissemination to ensure we are prepared and co-ordinated to deliver the 2021 Census outputs. We will not be publishing any data externally from the 2019 Rehearsal. This will include testing the functionality of the flexible dissemination system with dynamic statistical disclosure control operating on the ONS technical platform, as well as the creation of microdata and origin-destination products.

55. We are also developing an application programming interface (API) to allow users access to 2021 Census data via programmable requests as well as through the ONS website. The range of data available through the API would be same as provided through the flexible dissemination system. We have many users who already use APIs in their work and would value this functionality to reduce the time and money required to import data into their systems.

56. We are also aiming to have a fully functioning version of the flexible dissemination system on the ONS website prior to the release of 2021 Census estimates. To increase the timeliness of 2021 Census data, we would like users to become familiar with the online system and have the relevant APIs imbedded in their systems prior to the release of census data.

57. For more information about our plans, please contact the 2021 Census Outputs and Dissemination Team at census.outputs@ons.gov.uk or visit https://www.ons.gov.uk/census/censustransformationprogramme/census2021outputs

58. We would welcome feedback from members on the following:

- Our statistical disclosure control methodology for standard outputs and specialist products (microdata and origin-destination);
- Experience of designing and disseminating metadata to accompany statistics for a range of audiences in a useful and easy-to-use style;
- Experiences of engaging with stakeholders and the best techniques to get meaningful interaction and feedback with them;
- Experience of disseminating census estimates alongside variables derived from administrative data (enhanced census outputs).