Summary

The Office for National Statistics is putting user needs at the heart of preparations for the dissemination of the 2021 Census. Through reviewing feedback from 2011 Census, we have identified our aim to improve the timeliness, accessibility and flexibility of the data delivery for the 2021 Census.

We are investigating the option of a flexible data dissemination online system applying dynamic SDC to enable users to have faster and more flexible access to the data. To ensure we stay customer focussed we are applying AGILE techniques which involve users throughout the process to guide the work we do ensuring we understand and meet our users’ needs.
I. Introduction

1. The Office for National Statistics is putting user needs at the heart of preparations for the dissemination of the 2021 Census. Through reviewing feedback from 2011 Census, we have identified our strategic aims for the delivery of the 2021 Census outputs. This paper outlines the aims and describes how the ONS plans to address them.

II. Background

2. The Office for National Statistics (ONS) is currently running a Census Transformation Programme. The aims of the Programme are to transform the way the census is run and delivered in 2021 and also to look at how the census may develop beyond 2021. One part of this transformation is looking at the way in which census results will be disseminated.

3. As part of the assessment of the 2011 Census, the ONS held a workshop with the Census Advisory Group (CAG) in January 2015 to review the outputs from the 2011 Census and gather feedback to evaluate and inform transformation planning for the 2021 Census. The Census Advisory Groups represent a variety of interests of the main user communities including central and local government, academia, commercial, health and diversity groups. The membership of the CAGs evolves over time reflecting the stage of the lifecycle of census work.

4. At the January 2015 workshop there were a number of presentations from ONS and users of Census data to enable an open and honest discussion reviewing the dissemination of 2011 Census. The content of the presentations and the feedback received in the workshop was collated and published on the ONS website.

5. Although the feedback from the day was wide ranging, a few key themes arose which have significant value in terms of what could be changed to make an improvement to meeting user needs for 2021 Census. These are:

A. Statistical Disclosure Control (SDC)

- Whilst there was support for using record swapping as the Statistical Disclosure Control method, this caused delays in the release of data and was perhaps too restrictive from what was originally planned. This was due to the method requiring each table to be reviewed individually on its own merit and in conjunction with previously published data to avoid disclosure through differencing. In some cases it was also felt that decisions on changing table specifications were too isolated and as a result classifications in a variable were collapsed rather than removing a variable from a table making it less useful. The lack of consistency with 2001 tables was also highlighted.

B. Timetable

- It took too long (16 months) for the first results to be published and there was a long tail to finish the outputs causing the results to become less relevant.
- The more complex products (origin-destination, microdata) were not released until the end. Users felt these are important products that are at risk of suffering from “end of programme fatigue”.
C. Flexibility

- ONS should consider the potential to hold data in a more flexible way than the more traditional approach to create a large number of predefined tables. The thinking should be about constructing data structures in such a way that the information can be used and reused most effectively. This approach was initially proposed for 2011 Census but through consultation with users, the demand for traditional tables took priority.
- A more flexible approach as outlined would bring numerous benefits, more flexibility, improved timing, less need for commissioned outputs, and be more widely accessible. However it would present a challenge to SDC. ONS recognised the need to develop different and enhanced approaches.

D. Other feedback

- When there were changes made due to disclosure issues or changing needs, these changes were not consistently applied across the census offices resulting in inconsistency across UK. In some cases this made it difficult to compare results across the UK.
- Feedback on user engagement and communication was (i) to be clear at all times about the detail of what would be available when and involving users in decisions on changing content and priorities when required, and (ii) to ensure thorough consultation and involvement with smaller communities.

6. As a result of this feedback the strategic aims for the 2021 Census are to improve on the dissemination of Census outputs in terms of flexibility, timeliness and accessibility. We are also seeking to harmonise across the UK census offices where possible and improve our engagement with users.

III. Flexible dissemination to meet user needs

7. In order to achieve these aims, there needs to be an alternative approach to dissemination from that used in 2011 that utilises advances in methods and technologies. Other National Statistics Institutions (NSIs) already use this approach but with trade-offs which are not palatable to users of England & Wales census data. For example our users are not keen on suppression as a form of SDC and prefer to have 1s and 2s available in the data.

8. We are therefore developing a flexible web-based dissemination system which users will be able to easily access to define datasets they require whilst meeting specific criteria. However the need for disseminating statistics that meet users’ needs must be balanced with the need to avoid the disclosure of personal information. ONS has legal obligations under the Statistics and Registration Service Act (SRSA, 2007) Section 39, and the Data Protection Act (1998) in this respect, and ONS must also conform to the UK Statistics Authority Code of Practice for Official Statistics (2009) that requires arrangements for confidentiality to be “sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics”.

A. Statistical disclosure control

9. To support this, the ONS has been developing innovative statistical disclosure control (SDC) methods.
10. In 2001 the ONS applied a small cell adjustment after random record swapping had been carried out on the microdata. The decision to apply post-tabular small cell adjustment was made at a late stage and was widely unpopular with users since it resulted in inconsistency in tables.

11. In 2011 the ONS moved to using targeted record swapping where every individual and household was assessed for uniqueness or rarity on the basis of a small number of characteristics and every household given a household risk score. A sample of households was selected for swapping. The chance of being selected in the sample was based largely on the household risk score, so that households with unique or rare characteristics were much more likely to be sampled. However every household had a chance of being swapped. Once selected, another ‘similar’ household was found from another area as a ‘swap’. This approach was more popular with users, however the approach meant every table produced had to be manually checked for disclosure issues leading to delays in publication of the data tables.

12. For 2021 the ONS are considering SDC methods that could underpin flexible online dissemination vital in delivering user defined queries. These SDC methods are being discussed in detail at the UNECE SDC meeting in September 2017 in Skopje. The basic principle is that the data would primarily be protected by targeted record swapping, similar to the approach in 2011. However to ensure there is a minimal risk of disclosure through differencing, an added layer of perturbation would be applied in the form of the cell key method, similar to the approach used by the Australian Bureau of Statistics. However, since this is not the primary method of SDC it is possible to make the application of the cell key method ‘light-touch’ limiting the impact on utility of the data.

13. The simplest version of the method is demonstrated in Figure 1. Every record within the microdata is assigned a record key, which is a random number across a prescribed range, typically 0-99. When frequency tables are constructed, each cell has a number of respondents, and the cell key is calculated by summing their record keys. The combination of cell value and cell key is then read from a previously constructed look-up table (termed the ptable) to decide the amount of perturbation that should be used.

![Figure 1. Example of the Cell Key Method](image-url)
14. Where the same cell (or same combination of respondents) appears in different tables, the perturbation will be the same, due to the same cell value and cell key.

15. The main advantages of the method are that it allows tables to be protected without the need for a case-by-case assessment of disclosure risk and that a greater combination of outputs can be produced. This has potential for a step change in the flexibility of outputs. As demonstrated by ABS, the method can be used to systematically protect user defined outputs.

16. This would enable the data to be protected dynamically through a pre-programmed algorithm providing the benefits of:

   - **Timeliness** – through reduced requirement of table checking
   - **Flexibility** – as users can define the dataset they require
   - **Accessibility** – ease of finding the data required

17. There are trade-offs to this approach, the main disadvantage being that although the same cell of data is consistent in all outputs, there may be differences between that cell and the equivalent aggregation of other cells. Hence the number of 20-24 year olds in Southampton will always be the same across different tables but this may not be the same as the sum of 20, 21, 22, 23 and 24 year olds in Southampton.

18. However, the ‘light touch’ of the cell key method should mean that the inconsistencies between different tables are kept to a minimum.

**B. User engagement**

19. Since the ONS are interested in ensuring as far as possible user needs are considered and met, a key component is to involve users as the work develops. This fits well with the overall ONS Census Transformation Programme which has adopted the use of AGILE tools and techniques.

20. AGILE is defined as an iterative, incremental method of managing the design and build activities of engineering, information technology and other business areas that aim to provide new product or service development in a highly flexible and interactive manner.

21. The Agile approach is centred on four values:

   - Communication with parties is more important than standard procedures and tools.
   - Focus on delivering a working application and less focus on providing thorough documentation.
   - Collaborate more with clients.
   - Be open to changes instead of freezing the scope of the work.

22. AGILE is particularly of value in working towards a flexible dissemination solution. This approach ensures users can view the work as we progress and be involved in steering the design. Ultimately it helps us understand user needs and it helps achieve empowerment and buy-in for the users.

23. The first phase of the development involved working with an external company on a 3 month project to develop a prototype system. The aims of this initial prototype were to demonstrate that i) it is possible to programme the SDC into a flexible dissemination system, ii) that the algorithm can be processed fast and iii) that there are trade-offs between timely flexible outputs and inconsistencies in the data.

24. In accordance with AGILE principles we formed a user group which covers representation from a variety of user types. We held an event with the user group in April 2017 to demonstrate the application of the cell-key algorithm in real time using the
prototype. We demonstrated the system, and allowed the users to test it before gathering feedback from them on the system and their views on priorities for the next stage of development. We also sent a link to the prototype to the CAGs along with a questionnaire to encourage a wide range of user feedback.

25. In general, users were enthusiastic about the prospect of gaining access to census data in a more timely fashion than ever before and that they will be able to define the tables they require and are keen to see how the approach develops. They were accepting of the trade-off that due to the perturbation there will be inconsistencies when data are extracted in a different format, but they did highlight some concerns. In particular they highlighted the need for clear guidance on the differences in totals resulting from the perturbation, and the need for good metadata enabling repeatability of data queries. We demonstrated to users that these differences should be marginal since the cell key method is “light-touch” to alleviate risk of disclosure through differencing. The targeted record swapping will still be the main approach for the disclosure control.

C. Next steps

26. The first prototype was based on data from a small public microdata sample with a limited number of variables for individuals with little geographical information. Since we have received positive feedback from users we are progressing the development to a more sophisticated prototype which includes more complex data and business rules. The data will be based on a microdata sample with more variables, households, and more detailed hierarchical geography. The aims of the next phase are to develop the implementation of methods, provide more refined prototype for users to continue to build their confidence in the system, and to enable us to have a means to iteratively trial variations of business rules.

27. To provide flexible dissemination using an innovative SDC method is not without risk. To mitigate this risk we are planning comprehensive assurance of the methods from a wide range of experts and users. The evidence gathered will be used to show due diligence in assessing methods and risk and should help the National Statistician and Registrars General to make a decision confident that the data are sufficiently protected.

IV. Other outputs

28. Although flexible dissemination would replace the need for hundreds of multivariate tables, there remains a requirement for other outputs which cannot be met through a flexible dissemination system. In order to address this, an overall package of what the 2021 Census outputs might include is being developed and will form the basis of a consultation with users in early 2018. This is to get an early view on what the key priorities are for the outputs and should help us to understand well in advance of any decisions if users have concerns that we could address. This consultation will take on the more traditional approach of asking users to comment on written documentation in formal responses. Following from the consultation, a number of working groups will be formed with key stakeholders to ensure user engagement as the work develops.

29. The use of the cell key method to protect the tabular outputs enables flexible dissemination but complicates the dissemination of other types of data such as flow data and microdata. We are currently considering how this might be addressed.

30. As well as information received from census respondents, the ONS is committed to the use of administrative data to provide “enhanced outputs”. Defining the scope of the enhanced outputs is in progress and is likely to consider where administrative data may replace a census question or where it potentially provides additional information not previously asked. In both cases the priorities will be where there is a strong user need.
There is still work to be done on ensuring the protection of these data when the scope is more clearly defined.

31. This work is being carried out by the ONS for the 2021 Census in England and Wales, however we are working closely with our colleagues in Northern Ireland and Scotland to harmonise where possible to meet the user need for UK harmonised statistics.

---