

1st Meeting of the 2011 Census Bulk Data Working Group

Friday 14 October 2011, City Hall London

Attendees:

Chris Ashford	ONS Chair
Brendan Davis	ONS
Phil Starling	ONS
Craig Smith	ONS
Alan Lewis	SASPAC
Dudley Murrell	SASPAC
Richard Price	Experian
Sheena Murchie	CACI
Justin Hayes	MIMAS
Rob Dymond-Green	MIMAS
Mark Ireland	NOMIS
Ed Turnball	NRS
Dick Caulton	Callcredit Marketing Solutions
Simon Whalley	Beacon Dodsworth

Apologies:

Ian Shipley	WAG
-------------	-----

1. Welcomes and introductions

Chris Ashford welcomed everyone to the inaugural 2011 Census Bulk Data Working Group.

2. Terms of Reference

The Terms of Reference were agreed with some minor changes. References to GROS to be replaced with NRS and in Section 3, reference to "sets of 100 tables" to be removed.

Action 1: Chris Ashford to make amendments.

3. Review of 2011 Census dissemination events held in May 2011

Two Census dissemination events were held in London and Manchester on 16-17 May 2011. The events provided an overview of Census dissemination plans set in context with the ONS Web Strategy to deliver, in stages, a new website and access to ONS data and web content via an application programming interface (API). These events marked the start of the engagement process with users to address the needs for 2011 data in bulk form.

Before addressing this directly, both events first provided some background on potential alternatives to traditional bulk supply with demonstrations of applications designed to query and retrieve data via live access to an ONS API. They were developed by Census partners (MIMAS, Nomis and SASPAC) primarily for the purpose of testing an alpha version of the ONS API in December 2010, but provided an opportunity to demonstrate both what could be achieved in a very short period of time, and the potential for future, more mature developments.

Slides used at the events are available with a record of every question raised and full written responses at

<http://www.ukcensusoutputs.net/cswsgm:census-diss-events-2011>

4. Possible data structures and formats - advantages and disadvantages

Phil Starling presented an indication of the direction being explored by ONS for the provision of the standard output, in bulk, using a combination of SDMX and CSV in a bespoke format. This format should, by preserving description of data structure, allow for reconstruction and presentation of the datasets in the same form ONS intends to disseminate them via non-bulk output methods.

The slides are available at

http://cswsg.wikidot.com/local--files/cbdwgm:meetings-and-events/CBDWG_meeting_1_slides_14_10_11.ppt

During the discussion that followed Richard Price re-iterated a shared user preference for 2011 data to be provided in the same form as 2001, with table outlines that feature cell numbering matched to a flat data structure in custom bulk CSV format.

Justin Hayes explained a potential alternate means of cell identification that more closely aligns with the planned ONS format and might also provide a compromise that suits the needs of traditional bulk output users. Justin agreed to circulate a paper of his proposals to the group, via ONS.

Action 2: Justin to supply ONS with proposed cell identification and data structure.

Ed Turnball updated the group with NRS progress in developing an output production system and the options being considered for output formats. Ed confirmed that that NRS are considering options similar to those of ONS, and that an SDMX component has not been ruled out.

It was reiterated that user expectation is for all three statistics offices to use a common bulk output format for 2011 to improve on the discontinuity in 2001 output formats between the ONS and GROS data format.

Further discussion concerned the potential to add additional cell values to the tables as specified in the latest round of consultation. Mark Ireland raised the issue of completing the intersection of existing cross tabulations in tables where in some instances they are not – this does not mean providing additional cross tabulations, but does mean explicitly making visible values which can be derived from existing values. For example, in an age by sex table which presently includes males by age, females by age and a total figure for all people, this would mean including figures for all people by age.

ONS confirmed that because no new data is being provided, there should be no disclosure concerns with this proposal. But, this will represent significant change to the tables as currently specified in consultation with substantial changes needed to the underlying classification and data structure specifications. Without commitment, ONS will consider the implications of these changes within the context of work currently progressing to deliver to the planned output schedule, and will report a decision at the next meeting.

5. Special user requirements.

This discussion was held over to the next meeting.

6. Next steps & close

ONS will consider the clarified user preference for cell numbering and 2001 style CSV output, and the potential alternative suggested by Justin Hayes, and report on progress with output production development at the next meeting.

7. Date of next meeting

January 2012