

Census Output Dissemination Events

16 May 2011, GLA, London

17 May 2011, Manchester Museum, Manchester

1. London event attendee Questions & Answers

Q. Will commissioned tables be available via the API? I assume not.

A. Yes, it is intended to release via the API and website all commissioned tables that meet disclosure criteria and have been released to the commissioning customer.

Q. Did you chose to implement SDMX because it is what Eurostat wants rather than what customers want?

A. SDMX is an internationally established format for the description and transmission of structured data and metadata. Although used to describe the dataset structures and associated descriptions, and enhance the discovery and exploration of data, it is likely that the API will provide a number of optional file formats for the actual transmission of data in response to API queries. These formats are currently being discussed within ONS, and will be defined as development and testing of the beta API progresses.

Q. Are you saying there will be no large relational datasets at OA level?

A. The statistical disclosure control constraints applied mean that as the level of geography detail increases, the level of detail in classifications, and the number that may be cross-tabulated reduces. The potential to publish more classification combinations at low levels of geography has been explored, but in order to avoid implementing post tabular disclosure measures whilst still ensuring the confidentiality of information, it is unlikely that more detail than was published in 2001 is possible. This means that while it may be possible to publish large relational datasets at local authority and higher levels of geography, at Output Area and other low levels of geography the datasets published will include cross-tabulations and detail very similar to those tables published from the 2001 Census.

Q. At previous events ONS have said that bulk data would not be delivered in an offline way – so has this changed?

A. The primary means of dissemination of data from the 2011 Census will be via the ONS API and website. However, it is the intention to make all data published in this manner available, in bulk, via portable media. Further details about the format for the provision of data in this form will be made available from September 2011 as detailed plans for the production and specification of data and formats progresses. If you would like to be involved in further discussion of data format specifications then please register an interest in joining a proposed Census Bulk Data Working Group by contacting census.outputs@ons.gov.uk.

Q. Does the delivery of the API rely on the delivery of the website? The new website is already 4 months late, will similar delays effect the API?

A. The new ONS website is scheduled to launch at the end of August 2011, postponed from the original date of 30 April. After the initial launch, there will be further developments including an online data explorer tool to allow users to customise, interact with and download datasets, and an API enabling re-use of ONS data by others. Although complimentary, the initial website launch and subsequent API and website developments in preparation for the release of 2011 Census results are being managed separately. The delay to the initial launch of the new website is not anticipated to have an effect on development of the beta API throughout 2012 or the final API and enhanced website in time for the release of census results.

Q. People just want table layouts with cell numbering and the same numbering used to identify the counts when delivered in bulk. This is what most people want.

A. ONS intends to initiate workshops in September with consumers of bulk data to ensure all needs are understood and data is provided in the most usable formats and structures and with appropriate supporting information.

Q. Will there be consultation on the bulk output format?

A. Further details about the format for the provision of bulk data will be made available from September 2011 as detailed plans for the production and specification of data and formats progresses. If you would like to be involved in further discussion of data format specifications then please register an interest in joining a proposed Census Bulk Data Working Group by contacting census.outputs@ons.gov.uk.

Q. Will there be a delay between web release and bulk delivery release?.

A. The availability of bulk delivery data will be coincident with the same data published on the ONS website and made available via the API. The only difference in the timing of receipt of the data should be that necessitated by the transportation (from ONS to the customer) of the physical media on which bulk data is supplied.

Q. Do ONS have any plans to incorporate any aspects of linked data in their handling of metadata?

A. ONS is taking account of linked data in its development plans and will utilise this approach as appropriate.

Q. Are any internet completion rates known yet and can they be communicated?

A. Early data about completion rates is known, but as completions are still progressing, and follow up activities are still live, these cannot yet be published. This information will be available over time, and will be incorporated into the metadata released on the ONS website.

Q. Instant Atlas have an application which accesses NeSS NDE. Will it work with the WDP/ONS API?

A. That will be a matter for Instant Atlas, but we hope they, and many others, will build applications to access the API.

Q. Will there be any licensing issues with the use of ONS boundaries or those from other sources or services either directly or used to create the ONS boundaries?

A. ONS are working to an agreement that boundaries will be made freely available to all parties with no licensing restrictions on use, including commercial re-use, in line with OS Open Data terms. This is subject to agreement between ONS, Ordnance Survey and Royal Mail.

Q. (In response to NOMIS presentation) Is there any documentation that might help LADs or others wanting to create or work with similar applications either individually or in groups?

A. ONS intends to set up forums and galleries to illustrate best practice and offer guidance.

Q. What consideration has been given to future proofing geographical metadata by aligning with initiatives like INSPIRE?

A. The production of geographic outputs for the 2011 Census is being considered in the context of INSPIRE. However, as the data specification for the theme of Statistical Units is not yet agreed, it will not be possible to release the outputs in an INSPIRE compliant format in 2012. Once the specification for INSPIRE compliance is available, the outputs will be reproduced in line with both the INSPIRE data specifications and the INSPIRE deadlines.

Q. Will the bulk data format also be a format that is available for download from the website?

A. It is likely that the API will provide a number of optional file formats for the transmission of data in response to API queries. These formats are currently being discussed within ONS, and will be defined as development and testing of the beta API progresses. One of the formats is likely to be the census bulk data format, but this cannot be confirmed until the bulk data format is specified and agreed later in 2011, and further progress has been made with the development and testing of the beta API.

Q. Will attachment of reference metadata to web published data cause a delay in release, and therefore a delay in bulk data release?

A. One of the aims stated in the Registrars General agreement is for "Concurrent publication of appropriate metadata with associated statistical outputs." It notes that not all quality measures and other supporting information will be available at the time of publication, but ONS intends to publish as much as possible, in phases, beginning with key definitions, classifications and similar material prior to the release of first detailed results.

The release of data will not be delayed or compromised by the availability of reference metadata, or the time taken to attach supporting information to structured data. Releases will be accompanied by as much reference metadata, in consistent sets, as can be processed and quality assured in time for publication. Supplementary information will then be attached in phases between releases as they become available.

Q. Will there be a cap on API access in terms of maximum individual download or overall use?

A. Caps are being considered, and limits on file sizes for download are likely, but no decisions have been made at this stage. These decisions are likely to be system dependent, and ONS needs to test the capacity and performance of the system as part of development.

Q. Will there be a single source for UK data?

A. ONS, GROS, and NISRA are proposing a single UK database at ONS from which UK output at all geographies could be produced and published. Alternatively, an application that queries all three sites might be feasible, should all three offices build open APIs. This need not be developed by the Census Offices but by an external organisation if it wished.

Q. What will the role of NeSS be in the future in relation to the new WDP API? Will NeSS continue?

A. NeSS is a separate part of the ONS website, and the integration of NeSS into the main ONS website is part of ONS web strategy; consolidating infrastructure, software, and support. It is unlikely that this will occur before 2011 Census results are released and therefore some Census data will appear in NeSS.

Q. What will the learning curve be like for those moving from using NDE to the new WDP API?

A. The NDE currently offers output in some of the formats proposed for the new API so a steep learning curve is not anticipated. ONS plans to provide guidance and examples of best practice in support of the new API.

Q. What are the plans for parish amalgamation in 2011? Will amalgamation be the same or similar to 2001?

A. Amalgamation should occur in much the same way as for 2001 Census – further details will be available as more information about plans for outputs and geography become available later in the year.

Q. Reference metadata is not planned for inclusion with bulk data – is this because of file storage? Is this the only concern? If size or volume is not an issue, why not include the reference metadata with bulk supply?

A. ONS intends to provide bulk output from its data production system. The association of reference metadata with datasets will be managed within the publication system as part of the final processes to prepare releases for publication via the ONS website and API.

All reference metadata will be available via the website and API - this includes reference metadata that is available at the time of each release, and ongoing updates and supplementary content that will added over time to further enrich the data.

We do not currently have plans for a means to provide a static day of release snapshot of reference metadata to accompany bulk data. It may be possible to include with bulk data products some level of reference to the ONS website from which the latest metadata content can be accessed, but this will depend on the final format of content and the systems to manage it.

2. Manchester event attendee Questions & Answers

Q. In 2001 postcode allocation to OAs was inconsistent compared to the way gridlink was used for other data sources. Will that happen again for 2011?

A. The intention is for ONS to release a set of postcodes allocated to Output and Super Output Areas that is consistent with the method of allocation in Gridlink, and the National Statistics postcode directories. Allocations are by point-in-polygon using the postcode's geometric centroid.

Q. When will we know the results of OA redesign?

A. The OA Maintenance will have completed the redesign by December 2012.

Q. How will the move to rationalize and reduce the number of government websites effect ONS plans?

A. The ONS Web Strategy is to consolidate all its web sites into one, in line with government initiatives.

Q. What plans are there to guarantee website and service resilience around the peak usage periods for new census data releases?

A. ONS is working with its web host to ensure resilience.

Q. How will the API effect the provision and availability of UK wide data – will it improve upon the separate availability from 2001? Will Scottish and Northern Irish data be available via the ONS API?

A. ONS, GROS, and NISRA are proposing a single UK database at ONS from which UK output at all geographies could be produced and published. Alternatively, an application that queried all three sites might be feasible, should all three offices build open APIs. This need not be developed by the Census Offices but by an external organisation if it wished.

Q. Will we be able to get everything - both geography and data - in a single query? This was an issue with NeSS which was a reason some people did not use it.

A. API responses can contain structure, data, or both. Structure includes details of dimensions and their items, and geography is one such dimension. Additional geographical information will be available as metadata sourced from ONS Geography and available on the ONS website and via the API.

ONS is also supplying boundaries to Ordnance Survey to make available via their Openspace API service which we would recommend as the source of boundary data.

Q. Will bulk data supply be available from day one?

A. Yes

Q. Will we be able to specify our own age bands for the data/query we want? Last time, the age band required was often available but not in combination with the other variables required. Will this change at all, or for example, specifically for LAD level and above?

A. Proposed content of datasets below LA is described in the recent consultation; user defined age bands will be compositions of existing age bands, e.g. 0-5yrs and 6-10yrs being used to create 0-10yrs. New constructions, such as 3 – 7yrs will not be possible unless part of the existing datasets. The content of datasets for LA and above is not yet specified, but any data associated with a single year of age can be grouped according to a user requirement.

Q. Will details of the API be available in advance to enable users to prepare their systems and applications in advance?

A. ONS intends to issue details and guidance, along with best practice examples, and establish user forums

Q. There are issues with people being able to find the geography that they want – is this an issue that is handled in the API or the application which accesses the API?

A. The API will enable users to access all published information, but how that information is served to users depends on the application. However, the requirement to search hierarchical geographies is being addressed in API development. It is too early to say exactly what geographic functionality the API will have, but we do have a number of advanced geographic features in the requirements list. The test API which NOMIS and others tried out, and the improved version planned for September 2011 both only have simple geography navigation capabilities, based on SDMX codelists.

Q. What data is planned to be served via the API and in what form? Will there be cubes, tables, a build a table service, commissioned tables – some or all of these?

A. All published ONS data and associated supporting information will be available; this will include large multi dimensional datasets, tables (including commissioned), and structural and reference metadata. The ONS web site will use the API to support its own data explorer, that is expected to enable some manipulation and user definition of tables(using aggregate data). It should be noted that there are no plans to develop a table building solution that uses unit record (raw) data, and the API will not provide such data.

Q. Have the cubes that will be released already been specified, or if not, when will they be specified?

A. The specifications of multidimensional datasets at LA level are expected to be made public from January 2012.

Q. What skills are needed to be able to access, develop for and exploit the API?

A. Some basic programming skills are necessary, though some software is available that requires virtually no programming. Knowledge of the data is very important.

Q. Will there be any volume/size limits on downloads, and if so what?

A. Caps are being considered, and limits on file sizes for download are likely, but no decisions have been made at this stage. These decisions are likely to be system dependent, and ONS needs to test the capacity and performance of the system as part of development.

Q. Will we be able to have the information needed to edit the 2001 OA boundaries ourselves to create the new boundaries?

A. The redesign of Output Areas and Super Output Areas will be done through an automated process which splits and merges only those OA/SOAs that have breached agreed population and household thresholds, using postcode building blocks. This therefore means that there will not be an opportunity for users to redesign their own OAs as this would create an inconsistent and subjective set of geographies.

Q. What guidance will be available to help people create the sort of NOMIS profile style report?

A. ONS intends to issue details and guidance, along with best practice examples, and establish user forums.

3. Results of the ad-hoc poll of attendees

	London	Manchester
How many attendees plan to use the API	Approx 2/3	8
How many attendees plan to use bulk data	Approx 1/6	6
How many attendees are familiar with SDMX	None	None
How many attendees use SASPAC	Approx ½	
How many attendees plan to use a service that is accessing the API	Not asked	12

4. Results of the metadata prioritization exercise – compiled and summarized results from both days

	Priority 1	Priority 2	Priority 3	Priority 4
Definitions, Glossary	15	10	10	8
Accuracy and Reliability	4	8	8	9
Index of linked variables	6	4	4	2
Comparability	3	6	4	5
Index of linked source questions	4	5	3	5
Statistical population, bases, subsets, filters	4	4	2	1
Corrections/Revisions/Version details and dates	5	3	1	1
Documentation of methodology	1	2	4	3
Comments, See also, Related data/variables	2		5	2
Confidentiality	2	2		1
Use restrictions, licensing		1		2
Dataset intent, use or purpose	1			
Contact name and details (stable)	1			
Timeliness		1		1
Coherence				1

Other responses included those relating to discovery, metadata being attached to data, geographic look-up and boundary files, and metadata provided in the Welsh language.