

# ONS Web Data Dissemination Strategy Proposal

## Dear User

Please consider and respond to the enclosed enquiry regarding your usage of Nomis for accessing Labour Market Statistics.

We appreciate that this enquiry may seem similar to previous requests for information and may appear to ask for some similar information to the recent ONS Consultation. However it is important to note what is different in this particular enquiry.

Based on information that we have regarding general usage of Nomis, both from usage statistics and previous discussions and consultations, ONS is proposing an alternative solution to allow you to access the data that you currently access from Nomis – The ONS Data Explorer. More information about the proposed ONS Data Explorer is given in the attached note.

For many users, we anticipate that the ONS Data Explorer will meet their data needs.

We are also aware that some information is downloaded from Nomis that would not be met by the ONS Data Explorer. However, currently we do not know what the importance of this data is to users, how they make use of it and the impact on the general public. It is these users that we particularly want to hear from to better understand this usage of the more detailed data currently available.

Please take the time to consider the proposal in the attached enquiry and to provide us with the requested information if you believe that you would be affected by the proposal.

Responses should be sent by 20 May 2011 to:  
Bob.Watson@ons.gov.uk

Thank you

Bob Watson  
Sub-National Labour Market  
Office for National Statistics

# ONS Web Data Dissemination Strategy Proposal

The ONS currently has a wide range of data disseminated through a number of different formats and portals; from columns of figures in pdf files to Neighbourhood Statistics, from spreadsheets to Nomis. The ONS intends to streamline this array of different ways of disseminating data to provide a smaller number of solutions with greater functionality.

This document attempts to explain what these solutions will mean to users and in particular what the consequences might be for users who currently access data through Nomis.

## PART 1 - WEB STRATEGY

The proposed ONS web strategy is that datasets will be disseminated through a new architectural solution, the ONS Data Explorer, with access to disclosive micro and unit level data available through a separate secure service, the Virtual Microdata Laboratory, limited to authorised users only.

### ONS Data Explorer

The core way in which ONS is planning to disseminate data is through a new tool, the ONS Data Explorer.

ONS will load datasets into a Content Repository, which will be accessed through the Data Explorer. These datasets will store aggregate only non-disclosive results, giving access to pre-defined outputs and categories within variables. Within the limitations of ensuring protection of individual level information, ONS will be able to make available much larger, multi-dimensional datasets, than have previously been available through the ONS website. Users will be able to download datasets into their own systems or interact with and customise the data through the Data Explorer.

### Virtual Microdata Laboratory

The VML contains copies of many of the Microdata sets that are used to produce the statistics published on Nomis and in other ONS releases. It allows the user to access these datasets to run analysis using a limited range of statistical tools available within the VML environment.

Currently the VML can be accessed at ONS sites and remotely by people within the Government Secure Internet (GSI), following an approval procedure for the work they are doing.

## PART 2 - LIMITATIONS

### ONS Data Explorer

The ONS Data Explorer will be designed to give users easy access to a wide range of datasets stored within a Content Repository.

The datasets for the Content Repository, will need to fit a number of criteria. They will contain aggregate, non-disclosive results, for pre-defined outputs and categories. The Content Repository will also only host ONS datasets.

For datasets to be created that contain non-disclosive data, it will be necessary to carefully consider how detailed an analysis is sensible for each source. It will only be practical to create aggregates that will work across the dataset for the majority of categories. For example, it will not be practical to produce a breakdown that will work for the largest local authorities, if it will not work for a large number of smaller ones. It will not be practical to produce certain breakdowns for detailed geographic levels. It will not be practical to produce breakdowns for detailed categories such as some age breakdowns.

So it will be important for us to design breakdowns that give users as much of the information they need, while not pushing the datasets too far. However, it is important to note that the dissemination will be limited to the breakdowns that we eventually design, rather than allowing users the flexibility to tailor their own outputs.

For example, if we decide that a dataset will have age breakdowns for 16-24, 25-49, 50-64, 65+, it will be these breakdowns and combinations that are available. However, if this became the eventual design of the dataset, a breakdown such as 16-60 would not be published or available.

Another example of where this pre-defined output approach would change current flexibility would be in the availability of geographic breakdowns. We would be able to support a limited number of pre-defined geographic breakdowns, rather than allowing the range and flexibility currently available. Based on current usage, this would mean we would be likely to output statistics for countries, government office regions, local/unitary authorities (both tiers) and parliamentary constituencies. Beyond that core geographic set, published outputs would be limited, e.g. possibly headline employment/unemployment/inactivity levels and rates for travel-to-work areas. Further, for those core geographies, it may not be practical to produce all the breakdowns that we define for all the geographies, e.g. we might produce a wider range of variables for region than for parliamentary constituency. This would mean that user defined geographic areas would not be supported in the way that they currently are.

Obviously, these limitations will have more impact when a combination of factors is involved, such as looking at a combination of age, duration and geography.

It should also be noted that the criteria to only include ONS datasets would mean that datasets hosted on behalf of other departments, such as DWP Benefits and JobCentre Plus Vacancies, would no longer be available.

More detail on what this may mean for specific datasets is available in Annex A.

### **Virtual Microdata Laboratory**

The Virtual Microdata Laboratory contains copies of many of the Microdata sets that are used to produce the statistics published on Nomis and in other ONS releases. It allows the user to access these datasets to run analysis using a limited range of statistical tools available within the VML environment.

For those external users who are able to access VML, there is a requirement for them to apply for approval to use the Microdata for a specific purpose, e.g. a specific research project. However, a number of test cases are being prepared to allow certain government departments access for more general research functions rather than specific purposes.

It is available for use at ONS sites. This availability has recently been extended to other government departments being able to access the VML remotely across the GSI.

A final limitation is that not all datasets that are currently used by ONS will be available. For example the claimant count Microdata is not owned by ONS and may not become available through the VML.

## **PART 3 – WHAT WE NEED FROM USERS**

We need users to think about what impact these changes may have on them.

We are aware that many users access their data through Nomis. However we are also aware that for many users it is the ease of access to a limited amount of data that is important to them, rather than the wider range of flexibility available through Nomis. Therefore for many users the ONS Data Explorer should meet their requirements.

On the other hand, for some users it is the range of flexibility that is important to them. Some users are interested in geographic areas that would fall outside of the normal administrative and electoral hierarchy. Some users rely on information for very detailed variable definitions or a wide range of options. These users are far less likely to have their needs met through the ONS Data Explorer.

If you are one of these users that might not have their needs met through the ONS Data Explorer, then we want to hear from you. Specifically we want to know what the impact of the replacement of Nomis with the Data Explorer might mean for you.

- Are there direct financial implications involved?
- The potential loss of which data affects you?
- What work will it prevent you from doing?
- Will you still be able to achieve your aims without this data?
- Is this data useful or essential?
- What impact would you not being able to carry out your work have on society?

We are looking for specific examples to better understand what data people really need and exactly what impact the loss of that data, and the move from Nomis to an ONS Data Explorer as described above, would have. Obviously the more evidence we have, the better informed ONS will be in its decision making process.

Please, where possible put a monetary value on the impact, both to describe the type of wider impact, e.g. the size of funds that would be affected, and to describe the direct financial implications that you might incur.

## ANNEX A

### IMPACT ON SPECIFIC DATA SETS

#### Annual Civil Service Employment Survey

The ACSES dataset on Nomis is currently very flexible allowing users to cross-tabulate any combination of 9 different variables against one another.

The ONS Data Explorer would only be able to host a subset of these covering the most useful pre-defined cross-tabulations. ONS would work with users to ensure that as many of the most used cross-tabulations are included as possible.

#### Annual Population Survey

The Annual Population Survey dataset on Nomis is one that covers the widest range of information. However, the current level of dissemination available means that many definitions that are possible produce answers that are either disclosive, or of a very poor quality. Consequently the datasets that will be disseminated will require significant aggregation and therefore careful design in order to produce the most appropriate range of outputs.

Given the size of the sample available within typical authorities, it is unlikely that we could produce statistics for measures that represent much less than 4% of the population aged between 16 and 64 years. So, for example, we could produce figures for unemployment, however, it would not be sensible to split the unemployed between more than two categories. Therefore we would produce unemployment by gender, but an age breakdown of unemployed for local authorities would not be sensible.

#### Annual Survey of Hours and Earnings

The ONS Data Explorer would be able to host all of the data currently disseminated for this data set through Nomis.

#### Business Register and Employment Survey

The ONS Data Explorer will only host the type of aggregate results from BRES that are currently available on the ONS website. Separate arrangements will be put in place to provide users of BRES access to the kind of information currently available from Nomis.

#### Census of Population

The ONS Data Explorer should be able to host all of the data currently held on Nomis for the Census of Population.

#### Claimant Count

The claimant count is one of the most flexible outputs on Nomis, with one of the widest range of data available. Although the nature of the dataset, i.e. being a large dataset of administrative data, means that a large range of outputs are possible, there will still be practical limitations on what is likely to be produced.

It is likely that a range of statistics will be produced for the main geographic areas of interest, such as local authorities, government office regions and electoral geographies.

For other common geographic areas only a limited range of statistics would be likely, such as male and female stock levels, along with some age bands and possibly some duration bands.

Geographies that might fall within this category would include travel to work areas, wards and middle layer super output areas.

Other geographic areas, such as postcode based geographies, historic geographies and Jobcentre Plus geographies would be unlikely to be supported.

For those geographies where a larger range of outputs are supported, this may include stocks and flows with a limited range of age and duration bands. We may also produce statistics for levels of claimants by occupation.

### **DWP Benefits**

The ONS Data Explorer would only host ONS Data. Consequently this dataset would no longer be available. Dissemination for this data would be the responsibility of DWP.

### **JobCentre Plus Vacancies**

The ONS Data Explorer would only host ONS Data. Consequently this dataset would no longer be available. Dissemination for this data would be the responsibility of DWP.

### **Job Density**

The ONS Data Explorer would be able to host all of the data currently disseminated for this data set through Nomis.

### **Population Estimates**

The ONS Data Explorer would be able to host all of the data currently disseminated for this data set through Nomis.

### **Workforce Jobs**

The ONS Data Explorer would be able to host all of the data currently disseminated for this data set through Nomis.