

# Monitoring Review

Monitoring Review 6/12

7 September 2012

## ***Public availability of official statistics – three ONS datasets***

### **Introduction**

1. This review explores the public availability of the low level data used in the production of some of ONS's statistics. In addressing availability, we have looked at the range of data that are published, and the formats in which they are published. The review was undertaken in response to concerns raised with the Statistics Authority about whether ONS is currently able to comply fully with some elements of the *Code of Practice for Official Statistics*, and the Government's Open Data policy. In order to explore these concerns we selected data about the labour market, consumer prices, and business activity as examples of major datasets. Annex A provides some background. Annexes B, C and D examine the evidence in relation to each of the three datasets.
2. The *Code of Practice* focuses on the importance of identifying and meeting user needs. Some users need access to low level statistical data – aggregate data for small areas or small groups of people, or 'record level' data (often referred to as microdata) about individual people or businesses. These users will usually want to carry out their own analyses and will therefore want the data in formats that support this, together with suitable metadata (information about the data) and advice about interpretation.
3. ONS is required by law, by the *Code of Practice*, and by its published policies<sup>1</sup> to protect the confidentiality of any private data it holds relating to people or businesses. In some cases the legal obligation extends beyond information that the data subject would regard as private. Information that may be readily available on the internet – such as the cost of a tin of beans in a particular supermarket – may be treated as confidential when that same information is collected as part of a statistical exercise and a confidentiality assurance has been given to the respondent.
4. ONS's policy position on protecting data reflects concern that a loss of trust arising from a breach of confidentiality would damage its reputation and potentially undermine the willingness of survey respondents to provide information in future. In practice, this policy not only means that individual records are anonymised but that extensive further steps and checks are employed to ensure that no personal or organisation-specific information can be deduced from aggregate data. Aggregate statistics from which confidential information can be deduced are called 'disclosive', and the procedures for preventing this are called 'disclosure protection'. In general, the more detailed the information that is published, the greater the risk of disclosure.
5. Statistical offices draw a distinction between the *publication* of (non-disclosive) statistical information, and the *release* – under legally binding / contractual arrangements, to researchers

---

<sup>1</sup> See for example the Respondent Charter; <http://www.ons.gov.uk/ons/guide-method/surveys/respondents/household/charter-for-household-survey-respondents/index.html>

– of disclosive information. This review is only concerned with the publication of data, not the controlled release to ‘approved’ researchers who have signed a confidentiality agreement.

6. The *Code of Practice* establishes a presumption that non-disclosive data should be made publicly available. Principle 8.3 of the Code says “Make statistics available in as much detail as is reliable and practicable, subject to legal and confidentiality constraints....” The reference to being ‘reliable and practicable’ is not intended to give the producer body the option of withholding data on the ground that they are not completely ‘reliable’ but to apply only to cases where unpublished data are demonstrably unsuitable for any valid purpose. In effect, the Code seeks to ensure that maximising the utility and use of statistical data is the primary consideration but subject to necessary constraints to protect private information.
7. Recognising the challenges and trade-offs inherent in this requirement, we can nonetheless ask whether everything that should be in the public domain is in fact readily accessible, either directly on the website or after a straightforward request.

### Conclusions

8. ONS is publishing increasing amounts of the detailed low level data about the labour market, consumer prices and business activity that are used to create its aggregate statistics. These data are made available on the web in Excel format - although practice varies between the three datasets due, we are told, to the fact that the implementation of ONS’s Open Data strategy is currently at different stages across the organisation.
9. ONS has told us that it is taking steps to improve data access further. The Consumer Price Index (CPI) case study<sup>2</sup> published in March 2012 and the work towards releasing Census and other datasets in open data formats are two examples. Further plans are contingent on the identification and ranking of users’ priorities, the availability of resources, and the success of the Web Data Access Group.
10. The ONS datasets examined in this review are of 2 ★ status, on the Berners-Lee 5-point scale (see Annex A, paragraph 4). We hope that additional datasets of 3 ★ or 4 ★ status will be made available in future.
11. There are complex trade-offs between avoiding the disclosure of ‘personal’ information and maximising the utility of detailed statistical data. We believe that these issues need to be given greater prominence through the publication of illustrative case studies of risk and benefit.
12. ONS is not fully compliant with the Code of Practice because it does not make available all the many thousands of detailed datasets it holds that contribute to its published statistics. We understand why this is so and we recognise the steps that ONS is taking to making these datasets available. It now publishes records of all datasets made available in response to ‘ad hoc’ data requests (some 13,000 of which are received each year) on its website – either with the requested dataset or with information about how to access the dataset. We see this as an important step towards a more coherent and comprehensive policy underpinned by more sophisticated technology.
13. ONS is currently taking steps to publish substantial volumes of low level data about the labour market, consumer prices and business activity. However, we think that this progress could be supported by further steps:

---

<sup>2</sup> <http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/responding-to-the-open-data-agenda---an-ons-case-study--consumer-prices-index.pdf>

- (a) **More active dialogue between ONS and those users who require detailed data.** In addition to arrangements already in place, ONS could be more proactive in developing an in-depth understanding of the purposes for which users want the low-level data; this should help them to address those needs more effectively. More active dialogue would also help users to understand both the constraints on ONS in terms of the various legal, ethical and practical obstacles to publishing further low level data, and some of the technical issues (Annex A, paragraphs 10, 14 and 16; Annex B paragraph 6).
- (b) **Balancing confidentiality and utility.** Specifically in relation to consumer prices data, ONS could more clearly recognise the user requirement for individual prices and investigate data suppliers' views about publishing these prices, balancing the fact that these data are by definition available to consumers with the need to provide assurances about confidentiality to data suppliers. Because of the wide relevance of this issue – the boundary between confidentiality and utility - in other fields of official statistics, we would encourage ONS also to lead discussions about it with other producers of official statistics, and other stakeholders (Annex A, paragraphs 2, 3 and 14).
- (c) **More flexible IT systems.** Technological limitations are preventing teams in ONS from making available larger sets of non-disclosive low level data for users to manipulate; and are also impeding the implementation of appropriate corporate technical approaches to disclosure control. ONS's website does not currently provide enough functionality for users of low level data; and ONS is currently working to address this. In future we would hope to see the production of 'open datasets' as part of regular statistical developments. When resources permit we would also like to see the development of better processes for secondary disclosure testing for ABS data and a review of the coding of geographies for CPI (Annex A, paragraphs 16 and 17).

## Annexes

<b>Annex A</b>	<b>Background</b>
<b>Annex B</b>	<b>Data relating to the Labour Force Survey</b>
<b>Annex C</b>	<b>Data relating to the Consumer Price Indices</b>
<b>Annex D</b>	<b>Data relating to the Annual Business Survey</b>

## Background

### Research method

1. The research for this review involved the Authority's Monitoring and Assessment team meeting those in ONS responsible for the Labour Force Survey (LFS), Consumer Price Indices (CPI) and Annual Business Survey (ABS) to establish what is (and is not) published, and what barriers exist (or are perceived to exist) to publishing additional, more detailed data. These case studies are summarised in the annexes and the findings of this report are largely based on the information gathered from interviewing Grade 7 staff. As common threads were emerging, the team held further meetings with ONS about its response to the Open Data Agenda, its Web Data Access project and its statistical disclosure control policy, and sought views from ONS' senior management.

### Code of Practice

2. The most relevant elements of the Code of Practice are as follows:
  - *'ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information'*<sup>3</sup>;
  - *'ensure that arrangements for confidentiality protection are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics'*<sup>4</sup>;
  - *'make statistics available in as much detail as is reliable and practicable, subject to legal and confidentiality constraints, offering choice and flexibility in the format according to the level of detail required by the user'*<sup>5</sup>; and
  - *'ensure that official statistics are disseminated in forms that enable and encourage analysis and re-use. Release datasets and reference databases, supported by documentation, in formats that are convenient to users'*<sup>6</sup>.
3. In observing these practices, public bodies producing official statistics must strike a balance to maximise the utility of the information that they publish consistent with an acceptable degree of risk of disclosure. This is not straightforward and any review of the boundary between confidentiality and utility should be based on a sound dialogue with users and suppliers, and regulators such as the Information Commissioner's Office as well as the Statistics Authority. We understand that the issue of balancing confidentiality and utility has arisen in other areas of the Government Statistical Service, especially stimulated by 'open data' – see below – and we think that ONS is well-placed to provide GSS-wide leadership.

---

<sup>3</sup> Principle 5 Practice 1 of the *Code of Practice*

<sup>4</sup> Principle 5 Practice 4 of the *Code of Practice*

<sup>5</sup> Principle 8 Practice 3 of the *Code of Practice*

<sup>6</sup> Principle 8 Practice 6 of the *Code of Practice*

### Open Data

4. The Government's recent *Open Data White Paper*<sup>7</sup> includes principles, drawn up by the Public Sector Transparency Board, on how to make published data as useful as possible. One of the principles is that 'public data will be published in re-usable, machine-readable form' as defined by the Five Star Scheme that the White Paper states that the Government intends to adopt. The scheme is summarised here:
- ★ \_\_\_\_\_ Available on the web (in any format) with an open licence<sup>8</sup>
  - ★★ \_\_\_\_\_ Available as structured data (e.g. Excel instead of an image scan of a table)
  - ★★★ \_\_\_\_\_ Available in an open non-proprietary format (e.g. CSV or XML)
  - ★★★★ \_\_\_\_\_ All the above plus Uniform Resource Locators (URLs) so that people can 'point' to them directly, rather than just to an area on a website
  - ★★★★★ \_\_\_\_\_ All the above, plus a link from the data to other people's data to provide context.
5. The White Paper does not set any targets but expects that almost all datasets will achieve at least 3 ★ standard. ONS prepared a stock take review for the GSS Transparency Sub-Group of products published via the National Statistics Publication Hub (and with links to data.gov.uk). The review concluded that most products are available to 2 ★ or 3 ★ on the 5 ★ standard, with around 55% of data available to 3 ★ standard.
6. Our conclusions in this review about the standard for publishing data in a re-usable form are based on the LFS, CPI and ABS datasets that ONS publishes (as listed in annexes B, C and D). In some cases we have also highlighted data that could be published but is not. In these cases, it could be argued that even the one ★ standard is not being met.
7. Another of the public data principles<sup>9</sup> articulated by the Public Sector Transparency Board is that 'public data will be available and easy to find through a single easy to use online access point (data.gov.uk)'. Clearly there has been progress made in relation to the availability of public data as over 8,500 datasets are now available through data.gov.uk.

### ONS Open Data Strategy

8. ONS published its Open Data strategy - *ONS Approach to Open Data*<sup>10</sup> - in June 2012. ONS confirmed to us that its plans have not progressed as quickly as it would have wanted, as improvements to the new website have been prioritised in recent months. However, it has started to run information sessions for staff and has advised that more are planned. ONS is also now implementing its new 'ad hoc data' policy<sup>11</sup>, recording information about all data made available in response to 'ad hoc' requests, either with the requested dataset or with information about how to access the dataset, on the ONS Publication Scheme<sup>12</sup>. In the longer term, ONS plans to publish all the responses and is phasing this in starting with responses to requests from the media, Freedom of Information requests and requests for large datasets.
9. ONS is currently developing new detailed census outputs in open data formats. The plan is for these outputs and supporting Application Programming Interface (API) tools to be available to users from November 2012. ONS has identified a further group of 'trailblazers' to explore the

<sup>7</sup> <http://data.gov.uk/library/open-data-white-paper>

<sup>8</sup> Free of restriction on use or redistribution in its licensing conditions

<sup>9</sup> <http://data.gov.uk/blog/new-public-sector-transparency-board-and-public-data-transparency-principles>

<sup>10</sup> <http://www.ons.gov.uk/ons/guide-method/best-practice/open-data/ons-approach-to-open-data.pdf>

<sup>11</sup> <http://www.ons.gov.uk/ons/guide-method/best-practice/open-data/publishing-ad-hoc-statistical-data-and-analysis-requested-by-specific-user.pdf>

<sup>12</sup> <http://www.ons.gov.uk/ons/about-ons/what-we-do/publication-scheme/index.html>

possibilities for making a wider range of 'Open' datasets available from early 2013. ONS envisages this approach being progressively used for more datasets but has cautioned that plans and funding in support of such a roll out are yet to be agreed, and that the future work programme will be based on identified user requirements.

10. ONS has set up a Web Data Access Stakeholder Group to facilitate engagement with interested users. ONS told us that teams involved as trailblazers are already working with this group. Outside this project, there is less evidence of regular dialogue with users about the need for low level data. For example, ONS told us that the ABS User Group does not discuss access to low-level data. Recent assessments of ONS statistics by the Authority include similar examples – for example, *Assessment Report 218 Statistics on Research and Development*<sup>13</sup> identified a user need for detailed industry statistics. ONS could potentially meet this user need from the low-level data that it regularly provides to Eurostat but it was not aware of the demand and so did not publish the data as standard. ONS is now following this up. We suggest that it may be helpful to canvass User Groups to explore their interest in discussing access to low level data.

#### Labour Force Survey (LFS)

11. ONS makes a range of LFS tables, microdata and aggregate datasets available for users via the ONS website, NOMIS and the UK Data Archive. Based on the outputs listed in Annex 1, we conclude that the release of data meets the 2 ★ standard as aggregate data are available to download in Excel format. The assessment of labour market statistics against the *Code of Practice* in March 2010<sup>14</sup> noted that 'ONS publish an impressive breadth of material in relation to the labour market in the UK although the value of this material is substantially reduced because it is difficult for users to find the material they need on ONS's website'. Based on the information collected for this review, the launch of the new ONS website has improved this situation and the LFS team told us that it now has more control over the information it publishes, although we consider the ONS website still to be a barrier to the accessibility of data beyond the publication of standard tables.
12. The LFS team told us it would ideally like to make aggregate datasets more widely available to users, as well as actual microdata, but is restricted by resources and current technology and is looking to the ONS project exploring open data formats to provide the solution. It also said there are a number of technical issues about the datasets that users would need to be able to overcome.

#### Consumer Price Index (CPI)

13. ONS makes a wide range of low level CPI aggregate data and microdata available for users via the ONS website, the Virtual Microdata Laboratory (VML) and the UK Data Archive. ONS also makes available the Personal Inflation Calculator. Based on the outputs listed in Annex 2, we conclude that the release of data meets the 2 ★ standard.
14. ONS has taken steps in the last 12 months to improve access to CPI data. It responded to a request for the publication of more detailed CPI data and detailed item indices and price quotes have been published since September 2011. There is evidence that users have responded positively to the additional data provided, and ONS has benefited from a reduction in resource-intensive bespoke requests. ONS told us that the Government's Transparency Board recognised the publication of the CPI microdata as an example of best practice and

---

<sup>13</sup> <http://www.statisticsauthority.gov.uk/assessment/assessment-reports/assessment-report-218---statistics-on-research-and-development.pdf>

<sup>14</sup> <http://www.statisticsauthority.gov.uk/assessment/assessment-reports/assessment-report-34---labour-market-statistics.pdf>

ONS has published a case study<sup>15</sup> on this to share across the Government Statistical Service (GSS). The case study demonstrates a keen awareness of the remaining challenges. We consider that the improvements have filled a data gap for users and that ONS has published what it can within the constraints of its current IT systems and commitments on confidentiality. However, examining each of these issues in turn:

- The CPI systems apply a local area variable (for example, Windermere and Ambleside) to each record and from that local area variable derives a NUTS 1 geographical code (for example, North West). ONS told us that there is currently no coding of geographies between these 2 levels. This means that as the local area variable is determined to be disclosive, the only alternative is to publish at NUTS1 level. ONS could investigate whether there is a user demand for more flexible coding, for example for local authority level data.
- It can be argued that, unlike other surveys, the data collected for the CPI should not be deemed confidential as it is available to anyone at the point of sale. However, ONS makes a commitment to data suppliers to maintain the confidentiality of the data provided, and it considers this commitment to be important in securing the co-operation of retailers - so it considers that publishing records which could identify the retailer might adversely impact upon the data collection and hence undermine key national indicators.

#### Annual Business Survey (ABS)

15. ONS publishes a series of Excel tables with the Statistical Bulletins. Based on the outputs listed in Annex 3, we conclude that the release of data meets the 2 ★ standard. A rich resource of microdata is also available to approved researchers through ONS's Virtual Microdata Laboratory (VML) and the UK Data Archive, and the bespoke service provided by the ABS team means that all users potentially have access to any subset of the data, after disclosure control is applied. However, there is a clear imbalance between what is published as standard for all and what is available only on request. We consider that this is partly due to system inadequacies and partly due to historical practices.
16. ONS could do more to make additional ABS data accessible. There seem to be no compelling reason why some standard tables that are advertised as 'available on request' could not be published on the website. We think that the team could also review the bespoke requests received and engage with users to determine what more could be usefully published as standard. Datasets could be published after the main Statistical Bulletin to a pre-announced timetable determined by available resources.
17. There is no question, though, that the resource-intensive manual secondary disclosure procedures used for ABS directly influence decisions on what can be published and when, and that this is a technical challenge that ONS needs to overcome. Current processes for secondary disclosure checking, beyond the production of the core published outputs, require resource that ONS advises can only be funded on a case by case basis from the income generated by the bespoke requests. The secondary disclosure processes are also likely to limit the ability of staff working on surveys such as the ABS to fully exploit the new open data tools that ONS is developing.

---

<sup>15</sup> <http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/responding-to-the-open-data-agenda---an-ons-case-study--consumer-prices-index.pdf>

## Data relating to the Labour Force Survey

### 1. Findings

ONS publishes a wide range of aggregate quarterly LFS data in the monthly Statistical Bulletin<sup>16</sup> and associated tables. The quarterly datasets are deposited at the UK Data Archive and are available to registered users with a few weeks delay; for example results from the Oct-Dec 2011 dataset were published in the Statistical Bulletin on 14 February 2012 and the dataset available at the UK Data Archive on 6 March 2012.

Comparisons with practices in other countries indicates that ONS could do more in terms of releasing microdata from the LFS which can be manipulated by users, rather than focusing solely on tabular outputs.

The LFS team told us they would like to allow users to manipulate LFS data more readily but are limited by the technology available. The team is looking to the ONS project exploring open data formats to deliver some of this functionality.

### 2. Background

The LFS is the primary source of labour market data for the UK. It provides data on employment, unemployment and economic inactivity according to the International Labour Organisation<sup>17</sup> (ILO) definitions. The LFS is conducted according to European Union Regulation<sup>18</sup>, and results are provided to the European Commission. The LFS is a continuous survey. Eurostat core requirements make up about a third of the questions in the survey, which ensures some level of consistency across countries. In addition to the core requirements, an ad hoc module is run each year on a topic defined by Eurostat. For example, the 2010 ad hoc module is *Reconciliation between Work and Family Life*.

### 3. What is made available to all?

*Statistical Bulletin (SB)* – this contains three month averages for the latest quarter and accompanying data tables. Tables are Excel spreadsheets (seasonally and non-seasonally adjusted). Cells are unrounded to allow users to carry out calculations. The SB is published 6 weeks after the date to which the data refer i.e. the latest SB published on the 14th February contains data for the October-December quarter of 2011. From the SB webpage there is an option to link to 'view datasets associated with this release'. It then states that all the data in this dataset are available in spreadsheet format from the labour market data tables. It is then possible to choose to select a series from the dataset and select specific variables i.e. economically active female aged 50-64, and a time series for that variable is given (with option to download). As the 'dataset' only contains aggregate data, it is not possible to create new variables i.e. only the age bands specified are available for analysis.

*Single month estimates (not NS)* – used by HM Treasury and the Bank of England as early indications of movements in headline variables, now made available to all in tabular form.

*Regional Statistical Bulletin* - three month averages, headline data by Government Office Region.

---

<sup>16</sup> <http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/may-2012/index.html>

<sup>17</sup> <http://www.ilo.org/global/lang--en/index.htm>

<sup>18</sup> [http://circa.europa.eu/irc/dsis/employment/info/data/eu\\_lfs/lfs\\_main/lfs/lfs\\_regulations.htm](http://circa.europa.eu/irc/dsis/employment/info/data/eu_lfs/lfs_main/lfs/lfs_regulations.htm)

*Main LFS quarterly datasets, plus household and longitudinal datasets* – made available via the UK Data Archive. Also available are annual Eurostat files and ad hoc Eurostat modules. All non-commercial users are able to register and download LFS data in various formats free of charge. Charge for commercial use is £450. Users are able to run frequencies on the specific variables without registering for the data.

#### **4. What is made available as standard to restricted audiences?**

*UK Data Archive* – more detailed versions of the LFS datasets are also available from the Secure Data Service (SDS)<sup>19</sup>. Researchers are able to register and download these datasets. Both the SDS version and regular versions of the quarterly Oct-Dec 2011 LFS were made available on 6 March 2012 i.e. there is a delay of just under a month from the publication of the headline figures. Seasonally adjusted and non-seasonally adjusted data are available and the sample weights are also made available to users. A Special Licence version of the quarterly LFS is also available to researchers accredited by UKSA as an ‘Approved Researcher’. The Special License version contains further variables and more detailed geographic breakdowns. Age is considered by ONS to be a potentially disclosive variable and so all LFS datasets, other than the Special Licence version, contain only age bands rather than individual age for individuals.

*Virtual Microdata Laboratory (VML)*<sup>20</sup> – LFS and the Annual Population Survey (APS) are listed as available datasets.

*NOMIS*<sup>21</sup> – The local LFS and 4-quarter averages LFS dataset were available on NOMIS prior to the introduction of the APS, which is now the main source of data for local areas information as it includes the local area boost samples previously included in the LFS.

*Neighbourhood Statistics* – aggregate LFS data is available in table format.

#### **5. What is made available to users on a bespoke basis?**

If the customer wants data at a more detailed level than is published, the LFS team told us that they refer them to the LFS Data Service. Such enquiries are not frequent. The published LFS charging policy<sup>22</sup> is as follows:

<b>Database</b>	<b>Annual Subscription</b>
Regional Quarterly Datasets	£250 for all 4 quarters
Household Quarterly Datasets	£100 for 2 quarters
LADB's (1992-1999)	£100 for all years
Annual datasets (1984-1991)	£100 for all years
Annual datasets (1979-1983)	£50 for all years

<sup>19</sup> <http://securedata.data-archive.ac.uk/>

<sup>20</sup> <http://www.ons.gov.uk/ons/about-ons/who-we-are/services/virtual-microdata-laboratory/index.html>

<sup>21</sup> <http://www.nomisweb.co.uk/>

<sup>22</sup> <http://www.ons.gov.uk/ons/about-ons/who-we-are/services/unpublished-data/social-survey-data/lfs/lfs-charging-policy.doc>

## 6. What data could be published but isn't?

The LFS team told us that ONS Social Survey Division (SSD) supplies them with the datasets after having carried out their disclosure checks. The dissemination team do not have access to the dataset prior to disclosure checking. SSD follows ONS guidance in terms of the checks they perform and the cells which are suppressed. The LFS team told us that from their point of view, they could release the entire dataset that they receive as in theory it would be non-disclosive. The barriers to do this are technical, in terms of how to allow access and use of a quarterly dataset with around 120,000 records, and whether this would meet user needs. There are a number of important technical issues to note before analysing LFS data (for example – switch from seasonal to calendar quarters, naming convention for variables, decision of which quarter data to use etc) and the LFS team expressed concern that releasing the entire dataset would lead to greater confusion for the majority of users.

## 7. What has informed these decisions?

Historical decisions seem to underpin much of the current situation. The LFS dissemination team tend to make additional tables available when they notice an increase in user requests – for example, information on home workers which has become a regular table.

## 8. How far can the barriers to data access be overcome and are any improvement plans in place?

The LFS team feel that they currently make a large amount of aggregate LFS data available in a range of tables on the ONS website, and are continuously trying to make the data more accessible – for example, making tables from the historical quarterly supplement more clearly available. The team told us they would like to make more data available to users, particularly so that users are able to manipulate the data themselves, but are restricted in terms of the technology. They are hoping the ONS project trialling open data formats will deliver this functionality and are acting as one of the group of 'trailblazers' to help develop this.

The LFS team also commented that depending on how a particular month falls, they sometimes wait to publish the quarterly LFS data for around 2 weeks. This is because the publication dates are set far in advance and are largely determined by the availability of claimant count data. ONS could release headline LFS figures earlier than they currently do if that was deemed sensible.

## 9. How does this compare with other countries?

### Dutch Labour Force Survey

A limited set of variables concerning the labour market is published every quarter. This limited set and a large additional set of variables on the labour market are also published on a yearly basis. In addition, three-month averages on the employed and unemployed labour force are available. Users are able to create their own tables using Statline, an electronic databank. Pre-created tables containing headline LFS figures are also available using the databank. Age is grouped and there is a limited selection of variable available.

### Canadian Labour Force Survey

Data from the LFS in Canada is released in a series of monthly tables and a public use microdata file which contains non-aggregated data for a wide range of LFS variables. This microdata file is released monthly i.e. data for February 2012 was released in the dataset available on 9 March 2012.

### **Australian Labour Force Survey**

Data are released each month in tabular format in two waves: a first wave containing preliminary results containing headline figures (employed/unemployed by age/sex/area), and a second wave of final results with more detailed variables (hours worked, industry, occupation etc). The preliminary results are published 31 days after the start of the interviews in that month, and the final results are published 7 days after this release. The ABS also releases a monthly microdata file on CD-ROM and through their Microdata Remote Access Laboratory (RADL). The CD-ROM is available to all users but the file available through the RADL contains more detailed data and is only available to authorised users.

### **Summary LFS tables available each month to accompany the Bulletin**

- A01: Labour market statistics summary data tables.
- A02: Labour force survey summary (employment, unemployment and inactivity).
- A03: Labour Force Survey summary (employment, unemployment and inactivity) for men aged from 16 to 64 and women aged from 16 to 59.
- A04: Labour market summary (employment, unemployment and inactivity) for people below state pension age.
- A05: Labour market status (employment, unemployment and inactivity) by age group.
- A06: Educational status and labour market status (employment, unemployment and inactivity) of people aged from 16 to 24.
- A07: Regional labour market summary.
- A08: Labour market status (employment, unemployment and inactivity) of disabled people.
- A09: Labour market status (employment, unemployment and inactivity) by ethnic group.
- A10: International comparisons of employment and unemployment rates.
- A11: Labour Force Survey Sampling Variability (95% confidence intervals).

### **Employment tables**

- EMP01: Full-time, part-time and temporary workers.
- EMP02: Public and private sector employment.
- EMP03: Public sector employment by industry.
- EMP04: Public sector employment including and excluding financial corporations.
- EMP05: Employment levels by country of birth and nationality.
- EMP06: Employment rates by country of birth and nationality.
- EMP07: Temporary employees: all and part-time.
- EMP08: All in employment by occupation.
- EMP09: Employees and self-employed by occupation.
- EMP10: Part-time and temporary workers by occupation.
- EMP11: Employment by socio-economic classification.
- EMP12: Part-time and temporary workers by socio-economic classification.
- EMP13: Employment by industry (Labour Force Survey).  
Table JOBS02.
- EMP14: Employees and self-employed by industry (Labour Force Survey)
- EMP15: Job related training received by employees.
- EMP16: Employment by Occupation.
- EMP17: Employment rate revisions triangle.

### **Hours of work tables**

- HOUR01: Actual weekly hours worked.
- HOUR02: Usual weekly hours worked.
- HOUR03: Average hours worked by industry.

### **Earnings tables**

#### **Monthly Wages and Salaries Survey**

- EARN01: Average Weekly Earnings (AWE), seasonally adjusted.
- EARN02: Average weekly earnings by sector.
- EARN03: Average weekly earnings by industry.

#### **Labour Force Survey**

- EARN04: Gross weekly earnings of full-time employees (Labour Force Survey).
- EARN05: Gross weekly earnings of full-time employees by region (Labour Force Survey).
- EARN06: Gross weekly earnings by occupation (Labour Force Survey).
- EARN07: Gross weekly earnings by industry (Labour Force Survey).

EARN08: Distribution of gross hourly earnings of employees (Labour Force Survey)

## **Productivity and Labour Disputes tables**

### **Productivity**

PROD01: Labour productivity.

### **Labour Disputes**

LABD01: Labour disputes.

LABD02: Labour disputes by sector.

LABD03: Stoppages of work.

### **Unemployment tables**

UNEM01: Unemployment by age and duration.

UNEM02: Unemployment by previous occupation.

UNEM03: Unemployment by previous industrial sector.

UNEM04: Unemployment rate revisions triangle.

## **Economic Inactivity and Redundancies tables**

### **Economic inactivity**

INAC01: Economic inactivity by reason.

### **Redundancies**

RED01: Redundancies levels and rates.

RED02: Redundancies by industry, age, sex and re-employment rates (not seasonally adjusted).

## **Claimant Count and Benefits tables**

### **Claimant count**

CLA01: Claimant count summary.

CLA02: Claimant count by age and duration.

CLA03: Lone parent claimants of jobseeker's allowance.

CLA04: Claimant count revisions triangle.

### **Benefits**

BEN01: Key out of work benefits.

### **Jobs tables**

JOBS01: Workforce jobs summary.

JOBS02: Workforce jobs by industry.

JOBS03: Employee jobs by industry.

JOBS04: Workforce jobs by region and industry.

JOBS05: Workforce jobs revisions triangle.

### **Vacancies tables**

VACS01: Vacancies and unemployment.

VACS02: Vacancies by industry.

VACS03: Vacancies by size of business.

## Data relating to the Consumer Price Indices

### 1. Findings

ONS makes a wide range of data available for users and positive steps have been taken in the last 12 months to improve data access further. Based on the ONS evidence of increased web hits and reduced requests for bespoke analysis and approved access to the anonymised datasets, we consider that the improvements in 2011 have filled a data gap for users and ONS has published what they can within the constraints of its systems and commitments on confidentiality.

Examining each of these in turn:

- The CPI systems apply a local area variable (for example, Cardiff) to each record and from that local area variable derives a NUTS 1 geographical code (for example, North West, Wales). There is no coding of geographies between these 2 levels. This means that if the local area variable is determined to be disclosive, the only alternative is to publish at NUTS1 level. ONS could investigate whether there is the user demand to support more flexible coding, for example for local authority level data.
- It might be argued, that unlike other surveys, the data collected for the CPI should not be deemed confidential as it is available to anyone at the point of sale. However, ONS do make a commitment to data suppliers to maintain the confidentiality of the data provided and so to publish named records without due consideration would be a serious breach of the Code and if it impacted compliance with the collection, could undermine key national indicators. We therefore consider that if engagement with users indicates a clear demand for individual prices, ONS could investigate the levels of resistance/openness amongst suppliers to publishing these data.

### 2. Background

Consumer price indices measure the change in the general level of prices of consumer goods and services in the UK over time. They are important indicators of how the UK economy is performing and are widely used by the government, business and society in general. They also show the impact of inflation on family budgets which is of considerable interest to the general public.

The CPI was developed as a comparable measure of inflation across European Union (EU) Member States. Internationally, it is known as the Harmonised Index of Consumer Prices (HICP). It was created by an EU Council Regulation<sup>23</sup> passed in October 1995.

The Authority's assessment report<sup>24</sup> detailed a range of uses of the CPI including: the Bank's Monetary Policy Committee (MPC) using the CPI to assess inflationary pressures, taking it into account in setting interest rates to achieve the inflation target; as a deflator for expenditure measures when the effect of price changes is eliminated to allow the series to be presented in real terms; and in wage bargaining and linking contract payments.

---

<sup>23</sup> Council Regulation (EC) No 2494/95 of 23 October 1995 concerning harmonized indices of consumer prices, *Official Journal L 257*, 27/10/1995 P. 0001 - 0004

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995R2494:EN:HTML>

<sup>24</sup> <http://www.statisticsauthority.gov.uk/assessment/assessment/assessment-reports/assessment-report-79--consumer-price-indices.pdf>

A simple analogy for a consumer price index is that of a 'shopping basket' full of goods and services on which people typically spend their money. The content of the basket is fixed but, as prices of individual products vary, so does the total cost of the basket. A consumer price index measures how the total cost of that basket changes over time.

The consumer price indices are based on the prices of over 650 items - specified goods and services. Price collectors visit 20,000 shops each month, in around 150 places throughout the UK, and record about 110,000 prices. A further 70,000 prices are collected, mostly via the internet by ONS, although some are collected by external organisations or market research companies. Weights are attached to the different items to reflect their importance in the typical household budget. Weights are based on information about spending taken mainly from the Living Costs and Food Survey and from the National Accounts. These weights are combined with the monthly price information to produce the consumer price indices.

### 3. What is made available to all?

*Consumer Price Indices Statistical Bulletin*<sup>25</sup> – a monthly release that presents the latest headline estimates of consumer price inflation. It presents estimates for the CPI and RPI for the most recent three years. It also presents other inflation estimates based on these two indices and tables on price indices, percentage changes and weights for the groups of goods and services that make up the CPI.

*Consumer Price Indices Briefing Note* – published at the same time as the Statistical Bulletin and presents background briefing material on the statistics that lie behind the headline results.

*Personal Inflation Calculator* – ONS makes this tool available on its website<sup>26</sup>. This is an online tool that enables users to input their personal spending patterns to obtain an estimate of their own personal inflation rate and to see how it differs from the national estimates.

*Time series*<sup>27</sup> – data can be downloaded from the ONS's website – The tables available through this means are listed at the end of this report. Users can download complete tables or create bespoke datasets from across the tables.

*Detailed CPI and RPI Reference Tables* – these are described by ONS as 'providing a one-stop-shop which aims to complement the Consumer Price Indices Time Series data sets'. 50 tables are made available for users. ONS advise the make-up of these tables was determined by: a desire not to withdraw any data from users that were currently available unless it was clear the data were no longer used; a review of the number of web hits that the various data received; a review of e-mail and telephone requests to identify the most popular data requested; and the need to remove duplication. The CPI team did not engage users directly when reviewing the tables but told us they have just launched a new correspondence database to better document and understand user needs.

*Detailed item indices and price quote data* – since September 2011, ONS has made this detailed data available at the NUTS1 geographic level (for example, North West, Wales) in Excel workbooks on the ONS website. These are published every three months.

---

<sup>25</sup> <http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/january-2012/index.html>

<sup>26</sup> <http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc14/index.html>

<sup>27</sup> <http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/january-2012/cpi-time-series-data.html>

#### 4. What is made available as standard to restricted audiences?

Anonymised CPI microdata are made available through the VML and the SDS. Academics and researchers must apply for 'approved researcher' access to these services. Specific item indices and individual price quotes are made available with the names and addresses of suppliers removed.

#### 5. What is made available to users on a bespoke basis?

Since ONS made the detailed item indices and price quotes available on its website in September 2011, the producer team told us that bespoke analysis has been minimal. In general, users can be pointed to published data. The team is mainly called upon to help users construct their own chainlinked indices. The team makes a template and supporting guidance available and resource permitting, offers assistance in constructing the indices. ONS advised that it has not charged for any bespoke analysis since 2008.

#### 6. What could be published but isn't?

Data suppliers are given a written assurance by ONS that 'All of the information that you provide is kept strictly confidential. It is illegal for us to reveal your data or identify your business to unauthorised persons'. This has been informed by ONS's interpretation of the requirement in the Statistics and Registration Service Act 2007 not to release 'personal information' and the principle in the Code of Practice to 'ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information'.

ONS advised us that they have assessed the level of detail to publish which ensures that no individual retailer or service provider will be able to be identified. The CPI team told us that they had hoped to publish all detailed anonymised price quote data and item indices that underpin the CPI with the local area variable (for example, Cardiff or Windermere or Ambleside) but following a review in 2011, ONS methodologists assessed that the local area variable must be suppressed to protect the identity of the establishments. CPI has therefore published the anonymised price quote data and item indices coded only with the NUTS1<sup>28</sup> geography – for example, North West England or Wales.

#### 7. What has informed these decisions?

In the last 18 months, there appear to have been two key drivers of change. Firstly, the Authority published its Monitoring Brief *Communicating Inflation* in December 2010.<sup>29</sup> The Brief considered issues raised in public debate about official statistics on inflation, including aspects of the way in which the official measures of inflation are communicated. A conclusion of the Monitoring Update<sup>30</sup> published by the Authority in January 2012 was that:

*'ONS reviewed the arrangements for releasing these statistics and made some significant changes. ONS now releases these statistics through two statistical releases and a single data file, that are all published on the same day.... ONS also publishes a complete set of data tables (and)...the new tables provide users with access to the underlying data much sooner than had previously been the case, and in a much more user friendly format.'*

<sup>28</sup> <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/eurostat/index.html>

<sup>29</sup> <http://www.statisticsauthority.gov.uk/assessment/monitoring/monitoring-briefs/monitoring-brief-7-2010---communicating-inflation.pdf>

<sup>30</sup> <http://www.statisticsauthority.gov.uk/assessment/monitoring/monitoring-briefs/monitoring-brief-1-2012---monitoring-update---january-2012.pdf>

Secondly, ONS advised us that in response to a user request for access to all the CPI microdata, the CPI team undertook a review in 2011 to establish what more could be made available for public use and re-use. ONS told us that it was this review that resulted in the publication of the detailed item indices and price quotes at NUTS 1 level for the first time in September 2011.

## **8. How far can the barriers to data access be overcome and are any improvement plans in place?**

We consider that ONS is correct to suppress the name of the establishment and the local area variable whilst they continue to make a clear written confidentiality commitment to data suppliers. However, unlike for other business surveys, the argument might be made that the price data collected from retailers is 'public' information as it could be accessed by any consumer directly at the point of purchase. ONS could therefore review with their data suppliers whether this is a necessary commitment. Examples exist where ONS has worked with data suppliers to make more information available and ONS methodologists have confirmed that they could advise on this. Engagement with users on how far the latest release of detailed data has fulfilled their needs would determine whether such a review is justified and ONS would need to approach such a review with caution to avoid risking reduced compliance with a data collection which underpins such important National Statistics.

Based on what the CPI team told us, it would also seem that the way the CPI systems are coded means that there are only two options for coding indices, either with the detailed local area variable or the NUTS1 geography. This means that if the local area variable is deemed disclosive, the geographical detail available to users is very limited. ONS might consider whether more flexibility could usefully be built into the production systems, for example the coding of NUTS2 (for example, Merseyside, Cheshire, Lancashire) or NUTS3 geographies.

Before investing in such a development, ONS would again need to assess whether the user need is sufficient to support this. Since publishing the individual indices and price quotes in September, ONS told us that there have been around 8,000 additional hits on this area of the ONS website and that the demand for bespoke analysis has dropped away. Also, whilst the anonymised records continue to be available from the VML, the administrators report that there have been no applications to access the CPI datasets since December 2011. This may indicate that this latest public release of detailed data is now meeting the majority of user needs.

## **9. How does this compare with other countries?**

*The Authority reported how inflation statistics are reported in other countries in the Monitoring Brief Communicating Inflation<sup>31</sup>. The Brief did not cover data access arrangements. Instead, we have identified what is available from the websites of the relevant producer bodies.*

### **Dutch CPI**

Statistics Netherlands publishes its headline CPI inflation rate each month together with a measure based on the European harmonised (HICP) method which allows it to present international comparisons. Pre-created tables containing the headline CPI and HICP, and CPI by product group are available using the electronic databank, Statline. Users are also able to create their own tables at a similar level of detail using Statline. An interactive personal inflation calculator is provided for users.

---

<sup>31</sup> <http://www.statisticsauthority.gov.uk/assessment/monitoring/monitoring-briefs/monitoring-brief-7-2010---communicating-inflation.pdf>

**Canadian CPI**

The CPI for Canada is published monthly with a summary of price movements, inflation rates and the factors underlying them. Tables are published for the latest price index movements for eight major components (e.g. food, shelter, transportation) as well as more detailed components and product groups. The publication presents national and regional indices. A series of further summary tables are available and the CANSIM (Canadian Socio-Economic Information Management System) database enables the user to manipulate tables for download, for example, average prices by commodity.

**Australian CPI**

The CPI for Australia is published every 3 months. A series of summary tables is made available for users with indices for groups, sub-groups, expenditure classes and cities.

## **Detailed CPI and RPI Reference Tables**

These tables are advertised by ONS as providing a 'one-stop-shop' pulling together the numerous tables that were published prior to the launch of ONS's new website. The tables are made available every month with the Statistical Bulletin. The following is an example from February 2012 with the time periods covered quoted in order to give an illustration of the time series available.

- Table 1: CPI: Detailed figures by division
- Table 2: CPI, RPI and other selected indices: the latest three years
- Table 3: CPI: Detailed figures by divisions, groups and classes
- Table 4: A breakdown of the differences between CPI and RPI
- Table 5a, 5b, 5c: CPI: All items index, 12month % change and 1month % change
- Table 6: CPI: Detailed indices: the latest 13 months 12 months % change
- Table 7: CPI: Detailed indices: the latest 13 months
- Table 8: CPI: Detailed indices: annual average 1996-2011
- Table 9: CPI: Detailed indices: annual average 12 month % change 1997-2011
- Table 10: CPI: Weights: 1996-2012
- Table 11: CPI: Contributions to changes in all items
- Table 12: CPI: Detailed goods & services indices: the latest 13 months
- Table 13: CPI: Detailed goods & services indices: the latest 13 months 12 months % change
- Table 14: CPI: Detailed goods & services indices: 1997-2011 12 months % change
- Table 15: CPI: Detailed goods & services indices: 1996-2011
- Table 16: CPI: Detailed goods & services: weights 1996-2012
- Table 17: Constant Taxes (CPI-CT)
- Table 18: Difference between CPI and CPI-CT rates of change
- Table 19a, 19b: CPI: All items excluding indirect taxes (CPIY)
- Table 20: RPI: All items: 1947-2012
- Table 21: RPI: All items: 12 months % change 1948-2012
- Table 22: RPI: All items: 1 month % change 1947-2012
- Table 23: RPI: Detailed figures latest month
- Table 24: RPI: Detailed indices: the latest 13 months
- Table 25: RPI: Detailed changes: the latest 13 months 12 months % change
- Table 26: RPI: Detailed annual average indices: 1988-2011
- Table 27: RPI: Detailed annual average changes 12 months % change:1989-2011
- Table 28: RPI: Weights: 1987-2012
- Table 29: RPI: Contributions to change in all items
- Table 30a, 30b: RPI: Goods and services: weights and indices: 1987-2012
- Table 31: RPI: Goods and services: 12 months % change: 1989-2012
- Table 32: RPI: Long run series: 1947-2012
- Table 33: Internal purchasing power of the pound (based on RPI): 1977-2011
- Table 34: Composite Price Index: 1800-2011
- Table 35: Composite Price Index: 12 month % change: 1800-2011
- Table 36: Tax and Price Index (TPI): linked back to 1974
- Table 37: Tax and Price Index (TPI): 12 months % change
- Table 38: RPI: All items excl. mortgage interest payments (RPIX): 1975-2012
- Table 39: RPI: All items excl. mortgage interest payments (RPIX): 12 months % change:1976-2012
- Table 40: RPI: All items excluding Mortgage Interest Payments and indirect Taxes (RPIY)
- Table 41: RPI: All items excluding Mortgage Interest Payments and indirect Taxes (RPIY): 12 month % change
- Table 42a, 42b, 42c: RPI: All items excl MIPs & indirect taxes: seasonally adjusted (SARPIY) Index and 12 & 1 month % change: 1987-2012
- Table 43: RPI: Pensioner indices & all items RPI excl. housing: 1987-2011

- Table 44: RPI: Pensioner indices group annual averages: 1988-2011  
 Table 45: RPI: Pensioner indices: 1 person pensioner household weights: 1992-2011  
 Table 46: RPI: Pensioner indices: 2 person pensioner household weights: 1992-2011  
 Table 47: HICP - International comparisons: EU countries: 1997-2012  
 Table 48: HICP - International comparisons: EU countries:12 months % change: 1997-2012  
 Table 49: Average retail prices of selected items: for 14 February 2012  
 Table 50: Average retail prices of selected items: latest 13 months

### Time Series datasets

These are downloadable datasets for the CPI and RPI made available on the same day as the Statistical Bulletin. The user can download the complete dataset select series from across a range of tables to download as a single dataset. Series can be selected from the following tables (illustrative example based on February 2012):

- 1.1: CPI indices 1988 to 2012: 2005=100
- 1.2: CPI annual percentage change 1989 to 2012
- 1.3: CPI monthly percentage change 1988 to 2012
- 1.4: CPI weights 1988 to 2012
- 1.5: Contributions<1> to changes in all items CPI
- 1.6: CPI & RPI differences (old method Jan 1997 to May 2010)
- 1.6A: CPI & RPI differences (improved method Jan 2005 onwards)
- 1.7: CPI goods and services indices 1988 to 2012: 2005=100
- 1.8: CPI goods and services annual percentage changes 1989 to 2012
- 1.9: CPI goods and services weights 1996 to 2012
- 1.10: CPI Indices - Constant Taxes (CPI-CT) December 2005 onwards
- 2.1: RPI indices 1987 to 2012: Jan 1987=100
- 2.2: RPI annual percentage changes 1948 to 2012
- 2.3: RPI monthly percentage changes 1988 to 2012
- 2.4: RPI goods and services 1987 to 2012
- 2.5: RPI weights 1987 to 2012
- 2.6: Contributions<1> to changes in All Items RPI
- 2.7: RPI seasonally adjusted RPIY 1987 to 2012: Jan 1987=100
- 3.1: RPI average retail prices of selected items
- 3.2: RPI pensioner indices and annual change 1987 to 2011: Jan 1987=100
- 3.3: RPI pensioner indices weights 1992 to 2011
- 3.4: EU HICP indices and annual percentage changes 1996 to 2011: 2005=100
- 3.5: RPI all items long run series 1947 to 2012: Jan 1974=100
- 3.6: Composite Price Index and annual change 1800 to 2011: Jan 1974=100
- 3.7: Internal purchasing power of the pound based on the RPI 1976 to 2011

### Item indices and price quotes

Every three months ONS publishes detailed CPI datasets which include the individual price quotes and item indices that underpin the CPI. The data are published at the geographical level of NUTS1. At this level no individual retailer or service provider will be able to be identified.

Item Indices YYYYQt Q  
 Price Quotes YYYYQt Q

### Personal Inflation Calculator

This interactive application allows the user to enter their own spending patterns to generate a personal inflation figure.

## Data relating to the Annual Business Survey

### 1. Findings

A rich source of microdata is available to approved researchers and the professional bespoke service provided by the ABS team means that all users potentially have access to any cut of the datasets, after any disclosure control is applied. However, there is a clear imbalance between what is published as standard for all and what is available only on request. We consider that this partly due to system inadequacies and partly due to historical practices.

ONS could do more to make additional ABS data accessible. There is no reason why the standard tables advertised as 'available on request' could not be published on the website. We think that the team could also review records of the bespoke requests received and engage with users to determine what more could be usefully published as standard by redirecting resources within the current constraints. Datasets could be published after the main Statistical Bulletin to a pre-announced timetable determined by available resources.

There is no question, though, that the resource-intensive manual secondary disclosure procedure directly influences decisions on what can be published, and when, and this is a challenge which ONS needs to overcome for the following reasons:

- The ABS team needs to apply secondary disclosure to any tables they make available to users. This requires staff time and ONS advises that, beyond the production of the core published outputs, this time is currently funded on a case by case basis from the income generated by the bespoke requests.
- As this is a manual process, ONS methodologists also advised that there is likely to be a tendency by those applying the disclosure rules towards over-suppressing datasets rather than risk breaching confidentiality agreements.
- The secondary disclosure processes are likely to limit the ability of staff working on surveys such as the ABS to fully exploit the new open data tools that ONS is developing.

Ideally, ABS would have large multi-dimensional datasets to which automated secondary disclosure controls have already been applied. Users would then be able to extract from this the data they need. It is not clear that the ABS infrastructure could support this without investment.

### 2. Background

The Annual Business Survey (ABS), formerly the Annual Business Inquiry - Part 2 (ABI/2), collects financial data from businesses end year accounts. The survey approaches an annual sample of approximately 62,000 businesses, covering almost three quarters of the UK economy<sup>32</sup>.

The ABS is the primary source of data used in the compilation of the Input-Output Annual Supply and Use tables and provides a major contribution to the UK National Accounts 'Blue Book'. ABS is used for benchmarking monthly and quarterly ONS surveys such as Capital Expenditure and Stocks. HM Treasury and the Bank of England use the statistics to inform and monitor economic policy and the Department for Business, Innovation and Skills (BIS) uses the statistics to monitor change in industry sectors.

The ABS underpins ONS compliance with the European Structural Business Statistics.

---

<sup>32</sup> In terms of Gross Value Added

Variables collected include turnover, wages and salaries, purchases of goods and services, value of stocks, capital expenditure and international trade in services. To minimise burden on business, ONS use a short and long questionnaire. On the long questionnaire there is a more detailed breakdown of the main variables whilst the short questionnaire only requires totals. For example, total employment costs are broken down into gross wages and salaries, employers' National Insurance contributions, contributions to pension funds and amounts payable through redundancy and severance.<sup>33</sup>

### 3. What is made available to all?

ABS statistics are first published in *ABS Provisional*<sup>34</sup> in November, 11 months after the reference period. Revised statistics, which incorporate later returns from businesses are published in *ABS Revised*<sup>35</sup> the following June. ONS then publishes the first regional statistics in July, in *ABS Regional*<sup>36</sup>.

A series of reference tables<sup>37</sup> is published in support of the three Statistical Bulletins. These tables take the form of Excel workbooks. The content is the same for the provisional and final statistics with a reduced set of variables made available by country and Government Office Region (GOR).

The latest published statistics only cover time series back to 2008. This represents a break in the series when the ABS moved from SIC 2003 to SIC 2007. The historical time series are available on the ONS website but are presented separately under the 'Annual Business Inquiry'.<sup>38</sup>

### 4. What is made available as standard to restricted audiences?

When ONS publishes the revised ABS statistics, the producers make a dataset available to VML administrators to enable approved researchers access to the restricted microdata. ONS advised that loading the 2009 ABS dataset has been delayed due to the recent relocation of VML operations but it would ordinarily be made available in October, 22 months after the reference period. The published dataset contains individual anonymised records from the ABS for an extensive range of collected and derived variables. VML administrators also forward the dataset to the SDS<sup>39</sup>. There are currently 40 VML users registered for ABS and 173 SDS users.

To comply with the European Structural Business Statistics Regulation, ONS supplies an ABS dataset to Eurostat which includes an added dimension beyond that published directly by ONS. Eurostat publishes a subset of the structural business statistics by size of enterprise and so require ABS data broken down by employment sizeband.

---

<sup>33</sup> More information on the ABS sampling frame and the use of long and short questionnaires is documented in the Quinquennial Full Review of the Annual Business Survey. <http://www.ons.gov.uk/ons/guide-method/method-quality/quality/reviews/quinquennial-full-review-of-the-annual-business-survey-2010.pdf>

<sup>34</sup> <http://ons.gov.uk/ons/rel/abs/annual-business-survey/2010-provisional-results/index.html>

<sup>35</sup> <http://ons.gov.uk/ons/rel/abs/annual-business-survey/2009-revised-results/index.html>

<sup>36</sup> <http://ons.gov.uk/ons/rel/abs/annual-business-survey/2009-provisional-regional-results/index.html>

<sup>37</sup> <http://ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-249421>

<http://ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-249334>

<http://ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-249520>

<sup>38</sup> <http://www.ons.gov.uk/ons/rel/abs/annual-business-inquiry/1995-2007-national/index.html>

<sup>39</sup> <http://securedata.data-archive.ac.uk/>

## 5. What is made available to users on a bespoke basis?

The Statistical Bulletins make reference to ‘additional standard extracts available on request’. ONS advised that these are excel datasets which extend to more variables and include more detailed regional and industry breakdowns than ONS publishes in the reference tables made available on the day of publication. ONS told us it makes these tables available to users who request them as soon as they are ready. The producers advised that there is no specific pre-announced date of release for these tables as the disclosure checking procedures take some time and implementation is resource dependent. ONS has taken the decision not to hold up the main publication whilst these tables are prepared.

In addition, ONS offers a bespoke service and told us that they can potentially supply any cut of the ABS dataset on request, subject to disclosure testing. ONS charges for this service, for which they receive regular and repeat requests from the likes of central government, local authorities and city analysts. The charge reflects the time taken to produce the bespoke datasets. One-off requests come from a range of users including market researchers and students examining specific issues. ONS told us that it helps users to identify the best data to meet their needs, provides what they can from published datasets and carries out bespoke analyses as required.

At the time we talked with ONS, responses to bespoke requests had never been published but the production team were aware of the imminent changes to ONS policy and told us that they were working towards this, investigating any potential further disclosure risk presented by publishing bespoke responses alongside each other.

## 6. What could be published but isn't?

Effectively, because of the nature of the bespoke service provided, ONS advises that any analysis can potentially be performed on the ABS dataset at the request of a user. ONS has confirmed that the vast majority of variables (including those specific to certain industries) are available on their Special Analysis System. The only constraints are the protection of confidentiality through the application of ONS disclosure rules or a lack of available resource to perform the analysis and related disclosure checking.

Named individual records are never made available.

## 7. What has informed these decisions?

The key factors which appear to have influenced the data made available to users are:

- *Disclosure testing, systems and resources* – The influences on data access of disclosure testing, the functionality of the ABS production systems and the finite level of resources available to the team are intertwined. At the most basic level, the commitments to survey respondents to protect their data impacts what can be published. In addition to this basic principle, however, is the practical application of the ONS disclosure rules. Primary disclosure is applied as part of the automated production process but secondary disclosure is a manual process. The producers told us that secondary disclosure is a highly resource intensive process, meaning that decisions on the level of detail to be published on the day of release is determined by what it is possible to test within that limited timeframe. More resource is then assigned to secondary disclosure testing after the main publication date to make some additional standard datasets available but, beyond this, ONS told us that disclosure testing can only be resourced where it is applied in response to funded bespoke requests.

- *History* – The content of published datasets has evolved over time with any changes introduced primarily as a reaction to significant events such as the development of the ABS from ABI/2 or the introduction of the new industry classification, SIC 2007. There is no regular review of data accessibility as part of the annual production cycle, and whilst ONS meets regularly with an ABS User Group, this is not a regular item on the agenda.

## 8. How far can the barriers to data access be overcome and are any improvement plans in place?

ONS has confirmed there are no barriers to making both the standard datasets which are currently made available to users on request, and the datasets returned to Eurostat, available via its website as soon as they are ready. We consider that this would be a quick way to make more detailed data available to all users. Formally incorporating into the production cycle an annual review of what is published and how this might be extended within current resource constraints may also yield some easy wins. A review of common and repeat bespoke analyses may inform this, though resource constraints are likely to mean a staged release of data would be necessary.

Ideally, the ABS team would have the infrastructure to support efficient and robust secondary disclosure testing, allowing them to publish much more comprehensive datasets from which users could extract what they need. This would minimise the need for any bespoke analyses and have benefits for both users and producers.

In practice, this would mean a review of disclosure rules and an investment in automating processes. ONS told us it has made a package of tools available called Tau Argus (developed by Statistics Netherlands) which can apply optimal or near optimal secondary suppression to tabular outputs. However, it is not integrated into production systems and ONS advised that it has proved difficult for ABS and some other ONS surveys to implement, meaning only a limited number of producers use it. The producers told us that they are open to working with methodologists to review disclosure practices but are constrained by systems and resources.

However, whilst inadequate systems definitely seem to be a key barrier to development we consider that the imbalance between standard publication and the bespoke service is also the result of historical practice.

Finally, it should be noted that the producers have been preparing plans to deliver on the requirements of the recent assessment<sup>40</sup> of the ABS statistics against the *Code of Practice for Official Statistics*<sup>41</sup>. Requirements include evaluating the user need for a longer consistent time series, reviewing the allocation of resources taking into account user needs, improving accessibility of the statistics on the ONS website and investigating whether the statistics can be published more quickly.

## 9. How does this compare with other countries?

Countries take different approaches to producing structural business statistics, drawing upon surveys, administrative sources or a combination of both. The data collected support a wide range of economic statistics.

### French Structural Enterprise Statistics

INSEE provides a service called ALISSE which provides on-line access to Structural Enterprise Statistics to enable users to access to more comprehensive, detailed data than that available in the

<sup>40</sup> <http://www.statisticsauthority.gov.uk/assessment/assessment-reports/assessment-report-180--statistics-from-the-annual-business-survey.pdf>

<sup>41</sup> <http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html>

key figures or in the detailed data accessed by theme or region. They advise that data are often presented on more local geographical levels that are not present in the central part of the site and users can construct their own tables.

**Swedish Structural Business Statistics**

The Structural Business Statistics survey is based on annual accounts from all the enterprises in Sweden and detailed sector data are available in Sweden's Statistical Databases. Users can construct tables by selecting variables, industries (approximately 600 listed) and time series.

### Annual Business Survey tables

The same set of reference tables is released in excel format at provisional and final results stage and are published with the Statistical Bulletin. Data are published in provisional form 11 months after end of reference period, then in final form after 18 months. Regional datasets are available 19 months after the reference period.

### Annual Business Survey YYYY - Section A-S: Whole Economy

This table provides a summary breakdown by SIC07 section or 'letter' for 3 years<sup>42</sup>. The variables presented include number of enterprises, turnover, approximate gross value added at basic prices, purchases of goods, materials and services, employment - point in time and average during the year, employment costs, net capital expenditure – total, acquisitions and disposals, stocks and work in progress - value at beginning and end of year.

### Industry breakdowns

Each of the following tables presents the same information but with a more detailed industry breakdown (2-digit, 3-digit and 4-digit SIC07):

- Annual Business Survey YYYY - Section A: Agriculture (part), Forestry and Fishing
- Annual Business Survey YYYY - Section B: Mining and Quarrying
- Annual Business Survey YYYY - Section C: Manufacturing
- Annual Business Survey YYYY - Section D: Electricity, Gas, Steam and Air Conditioning Supply
- Annual Business Survey YYYY - Section E: Water Supply, Sewerage, Waste Management and Remediation Activities
- Annual Business Survey YYYY - Section F: Construction
- Annual Business Survey YYYY - Section H: Transport and Storage
- Annual Business Survey YYYY - Section G: Wholesale and Retail Trade
- Annual Business Survey YYYY - Section I: Accommodation and Food Service Activities
- Annual Business Survey YYYY - Section J: Information and Communication
- Annual Business Survey YYYY - Section K: Financial and Insurance Activities (part)
- Annual Business Survey YYYY - Section L: Real Estate Activities
- Annual Business Survey YYYY - Section M: Professional, Scientific and Technical Activities
- Annual Business Survey YYYY - Section N: Administrative and Support Service Activities
- Annual Business Survey YYYY - Section P: Education (excludes public sector)
- Annual Business Survey YYYY - Section Q: Human Health and Social Work Activities (excludes public sector)
- Annual Business Survey YYYY - Section R: Arts, Entertainment and Recreation
- Annual Business Survey YYYY - Section S: Other Service Activities
- Annual Business Survey YYYY - Division 45: Wholesale and Retail Trade and Repair of Motor Vehicles and Motorcycles
- Annual Business Survey YYYY - Division 46: Wholesale Trade except of Motor Vehicles and Motorcycles
- Annual Business Survey YYYY - Division 47: Retail Trade except of Motor Vehicles and Motorcycles

### Regional Annual Business Survey Tables

Annual Business Survey YYYY - Regional Data: Region by Division – presents turnover, approximate gross value added, purchases of goods and employment costs broken down by industry (2-digit SIC07) for the Government Offices for the Regions.

Annual Business Survey YYYY - Regional Data: Division by Region – similar to that outlined above but includes tables for each Government Office for the Regions.

<sup>42</sup> There was a break in the time series with the introduction of SIC2007. The earliest year for which data is available on this basis is 2008.

Annual Business Survey YYYY - Regional Data: Whole Economy by Region – This table presents the same information as above but with no industry breakdown.

### **Quality Measures**

Annual Business Survey YYYY - ABS Quality Measures presents standard errors for the key variables.

### **Annual Business Inquiry**

The data relating to years prior to 2008 is published under the Annual Business Inquiry (which preceded the Annual Business Survey and Business Register Employment Survey). The content of these tables is similar to those published from the Annual Business Survey tables but it is presented on the basis of previous industry classifications. A full list of the tables is not provided here.

Annual Business Inquiry 1995-2007 National Results.  
Annual Business Inquiry 1998-2007 Regional Results.