

# Responding to the Open Data Agenda:

## An ONS case study

### 1. Introduction

This case study provides details, using the Consumer Prices Index (CPI) as an example, of one way ONS is responding to the open data agenda. The study provides a brief background to the CPI, details of the data previously published and then goes on to discuss the issues that were considered and, more broadly, the steps ONS has taken to respond to user demands for access to more detailed inflation data.

### 2. Summary of key lessons learned

- ensure users know what they want: take time to engage with users to help them understand what they would actually find useful, perhaps even going as far as seeking feedback on the provision of a sample/test dataset
- protect assurances given to data suppliers: it is absolutely key that decisions on additional levels of data published does not contravene confidentiality assurances given to data suppliers
- compliance with legislation and code of practice: it is vital that decisions align to the appropriate legislation and code of practice
- disclosure control: engagement with experts on disclosure control makes the process easier and more robust. It's also important to have a framework in place populated with appropriate individuals in order to deliver efficient, transparent and accountable decision-making
- implementation: take time to consider how best to implement to ensure that the ongoing overhead of additional publication demands are kept to a minimum



### 3. Background

#### 3.1 What is the CPI?

The CPI measures the changes in the general level of prices charged for goods and services bought by UK households and foreign visitors in the UK for the purpose of consumption (rather than investment).

#### 3.2 What data are collected to measure the CPI?

Around 180,000 price quotes are collected each month for around 650 items, from 20,000 retailers and service providers (including those on the internet) in 150 different locations of the UK. The reliability of the index depends very much on retailers' goodwill. ONS gives assurances to retailers that the information collected is treated with the strictest confidence, that no data relating to individual retailers or businesses are ever divulged to or passed to a third party, and that they will not be identifiable within published data.

#### 3.3 How is the CPI used?

CPI uses include:

- as the main UK domestic measure of consumer price inflation for macroeconomic purposes and as the Government's inflation target. The Bank of England's Monetary Policy Committee is charged with meeting this target – CPI annual inflation of 2.0 per cent
- since April 2011, it is used for the indexation of benefits, tax credits and public service pensions
- as a price escalation clause in contracts and is often used when comparing UK inflation against that in other countries

## 4. CPI data previously available to users

### 4.1 All users

Each month, indices, 1-month percentage changes and 12-month percentage changes were (and continue to be) published for total CPI and lower-level components such as bread & cereals, meat, wine, garments, gas, furniture & furnishings, pharmaceutical products, fuels & lubricants, and games, toys & hobbies. However, information below this level such as for sliced white bread, bacon, European red wine, men's suits, leather settees, indigestion tablets, petrol, board games or indeed individual price quotes were not published. The official CPI series starts in 1996, but estimates for earlier periods are available back to 1988.

### 4.2 Approved Researchers

ONS fully acknowledges the importance of microdata to public bodies, research organisations and academic researchers for statistical research and analysis. The data are made available via the Virtual Microdata Laboratory (VML), a government research access facility on-site at ONS, and through the Secure Data Service (SDS), an academic facility which allows secure access to identifiable, but not identified, data for Approved Researchers<sup>1</sup> for statistical research purposes. Users of the VML and SDS must apply to the Microdata Release Unit (part of ONS Legal Services) to gain access as an Approved Researcher.

The CPI microdata (that is, indices for specific items and individual price quotes) are available via the VML and SDS to Approved Researchers for research purposes. These data include prices of goods and services collected across the UK, with the names and addresses of data suppliers removed to anonymise the data.

“ The interest in easier access to more detailed CPI data had been on the increase ”

## 5. User demand for easier access to more detailed CPI data

The interest in easier access to more detailed CPI data had been on the increase when a specific user requested the underlying data (such as individual price quotes and item indices) for the CPI.

This request constituted a precedent in that ONS had not released raw CPI data in this way before – or any other microdata from business surveys. Although the underlying data for the CPI are available via the VML, this request was not for research purposes. ONS, therefore, had to prepare CPI microdata for public use and re-use, for the first time.

<sup>1</sup> the definition of an Approved Researcher is explained in [sub-section 5 of section 39 of the Statistics and Registration Services Act 2007](#)

## 6. ONS response to the request for publication of more detailed CPI data for public use and re-use

ONS took the following steps to reach decisions on this request.

### 6.1 Provision of a sample dataset

The first step was to provide a sample dataset to the user in order for them to consider whether it would prove useful to them and other users. The user did find the dataset helpful but questioned whether the data could be provided at an even finer level of geographical detail. Concerns around disclosure (described more fully later) meant that the data provided were broadly at the geographic level of Nomenclature of Territorial Units for Statistics (NUTS) 1 – for example, South East England, West Midlands, Wales and so on.

### 6.2 The publication of personal information



ONS must not publicly release data that are personal information. The [Statistics and Registration Service Act 2007](#) defines 'personal information' as data which may enable an individual to be identified either directly from the data or by the auxiliary use of published information. In the case of the CPI, personal information is data which could enable the identification of an individual data supplier. The CPI microdata was checked for disclosure risk by ONS's experts in disclosure

control, the Statistical Disclosure Control Branch, to conclude whether the data could be released as a public-use dataset. When undertaking an assessment of the risk of publishing microdata, ONS's Disclosure Control Branch takes account of any assurances given to the data providers as well as the [SRSA](#), the [Code of Practice for Official Statistics](#) and the [Data Protection Act](#).

#### 6.2.1 Guarantee of confidentiality to data suppliers (assurances)

Each provider of data to CPI, whether a retailer or other establishment, is given the following written assurance from ONS: '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'.

Protection of this assurance was sacrosanct in the considerations that took place to determine which level of data detail to publish.

#### 6.2.2 Compliance with the Statistics and Registration Service Act

As explained in section 6.2, the SRSA prohibits the public release of personal information. However such information can be made available under certain conditions, which are described in [section 39 of the Act](#).

### 6.2.3 Code of Practice for Official Statistics and the Survey Charter

The [Code of Practice](#) states in Principle 5: Confidentiality 'Private information about individual persons (including bodies corporate) compiled in the production of official statistics is confidential, and should be used for statistical purposes only'. Practice 3 of this principle pledges to 'Inform respondents to statistical surveys and censuses how confidentiality will be protected'.

This is further discussed in the [National Statistician's Guidance](#):

'It is good practice to provide information to the public in general about how confidentiality is protected. The Office for National Statistics, for example, publishes a Survey Charter and advance leaflets for each of its social surveys.

The information provided to respondents should be treated as a binding promise about how their data will be used'.

### 6.3 The recommendation and decision

For many items in the CPI microdata, it would be extremely difficult to identify the retailers involved. However, there are other items for which there is likely to be just one supplier in the local area. Examples include dog kennel fees, carbon fibre fishing rods, smokeless fuel, cinema admission, golf green fees, secondary school cafeteria prices and non-NHS chiropractor rates.

The ONS Disclosure Control Branch assessed that, given the nature of the data, it would be possible to combine the raw data with other published data such as [www.yell.com](http://www.yell.com) to identify some retailers or establishments with a high level of certainty. This would be a breach of the SRSA, the Code of Practice, the ONS Survey Charter, and the undertaking given to the price provider. It would thus be illegal for ONS to make these data publicly available.

The ONS Disclosure Control Branch recommended that the local area variable be suppressed, and the geographical level of detail set at NUTS 1. In addition, six item indices were suppressed as a single (or very small number) of data suppliers contributed to these indices. It was considered that these steps would give sufficient uncertainty that any possible identification of a price respondent would be a correct re-identification.

This recommendation was made to the ONS asset owner of the data who approved release.

## 7. Transformation of the decision into reality

Some of the key issues considered in the lead up to the publication of these detailed data were:

- ongoing maintenance of data publication: this covered the frequency of updating the data and to reach a balance of providing up-to-date data to users, while at the same time not putting the production and publication of the headline CPI at risk. In addition, a reasonable level of investment was made so the data extraction and transformation were developed in a way that helped ensure the minimum level of resource would be required to cope with this additional burden on an ongoing basis
- capability of the publishing tool to handle large data files: this determined how the data were structured in publication. For example, there is no point putting all the data in one file if users cannot then open it
- the format in which the data can be accessed: again a balance had to be reached between adherence to accessibility standards and operational constraints. For example, should the data be made available in excel only or also csv, xml?

- user needs: the CPI team are fortunate in having a good understanding of what users want – the CPI receives the most hits on the ONS website and our users are not shy in engaging with us! Once the fundamentals were resolved, this feedback was used to help shape the publication
- guidance for users: to provide the most assistance to users while at the same time minimising any additional burden to the CPI team (due to an increase in the number of queries) significant thought was given to the information that should be published alongside the data. The conclusion was to provide a detailed glossary (one file) that explained each aspect of the data release

## 8. Outcome

On 13 September 2011, [CPI microdata](#) were released for the first time. The data were announced in the following way:

‘Price quote data and item indices that underpin the CPI are now published for the first time giving users unprecedented access to the detailed data that are used in the construction of the UK’s inflation figures. Data are broadly at the geographic level of Nomenclature of Territorial Units for Statistics (NUTS) 1 (for example, South East England, West Midlands, Wales, etc.). This level of detail was chosen to ensure that no individual retailer or service provider will be able to be identified and therefore is in line with the assurances given to data providers and for the same reason a very small number of item indices (six) are also unavailable.

The data currently covers January 1996 to March 2011. Going forward, these data will be updated once a quarter with around a three month lag [NB. this has already been reduced to two months] compared to the latest CPI publication. For example, the data will next be updated when the November CPI is published on 13 December 2011, at which point the detailed data published will be extended to September 2011.

Publication of these data is new for us as well as for users so we are keen to hear from you on how useful the data are and also how we can improve in the way the data are provided. You can contact us on:

Email: [CPI@ONS.gov.uk](mailto:CPI@ONS.gov.uk)

Telephone: 01633 456900’

## 9. Reaction

The data have been well received and feedback indicates that this information is largely being used by more expert CPI users to improve forecast models, the granularity of inflation analysis, and so on. The data have so far received around 8,000 hits.