

Article

# Introducing alternative data sources into consumer price statistics: May 2020

Our plans to include alternative data sources into the production of consumer price statistics.

Contact:  
cpi@ons.gov.uk  
cpi@ons.gov.uk  
+44 (0)1633 455171

Release date:  
27 May 2020

Next release:  
To be announced

## Table of contents

1. [Introduction](#)
2. [What are alternative data sources?](#)
3. [Benefits of using alternative data sources](#)
4. [Current methods of data collection](#)
5. [Implementation plan](#)

# 1 . Introduction

Accurate measures of inflation play a vital role in business, government and everyday life. From rail fares to taxes to pensions, financial transactions in every area of our lives are regularly adjusted to reflect the change in prices over time. It is crucial to measure these changes in price as accurately as possible.

We are currently undergoing transformation across many areas of our inflation statistics, including identifying new data sources and improving methods.

As part of this transformation, we are updating the way we collect and process price information to reflect our changing economy and produce more robust, timely and granular inflation statistics for businesses, individuals and government.

## 2 . What are alternative data sources?

We are introducing two data sources to help us transform our consumer price statistics collection. These are scanner data and web-scraped data.

### Scanner data

Scanner data are collected by retailers at the point of sale. Scanner data will provide us with significantly more information on the number and type of products sold, allowing us to more accurately reflect changing consumer spending patterns.

Our ambition is to collect hundreds of millions of prices every month from the UK's leading retailers, which will improve our understanding of how prices of products, as well as the number of each product sold, are changing in the UK economy. Over the last year, we have made significant progress towards achieving these plans. We are now receiving data from six prominent high-street stores and are in discussions with several others. Importantly, these data only show total weekly figures regarding what retailers have sold, rather than what individual people have bought.

We are also working with GS1 UK – the provider of industry-standard product identifiers (barcodes) – whose catalogue of product descriptions may help streamline the data collection process.

### Web-scraped data

Web-scraped price data are collected from retailers' websites and can provide a wealth of additional product information about online prices, such as product descriptions. For example, as well as obtaining the price of a laptop, we can collect information such as the laptop's RAM and processor speed, which help us make sure the products are comparable over time.

Since November 2018, we have been receiving regular web-scraped data. These data are for different types of product covering areas such as clothing, electronic items and package holidays. There are no historical series available with these data (unlike scanner data) so we will need to build up a sufficient time series of high-quality data before a final impact assessment can be completed.

## 3 . Benefits of using alternative data sources

Alternative data sources provide many benefits compared with our current sources, including improved product coverage, high frequency of collection, as well as potential cost savings.

Scanner data can also provide additional information such as expenditure per item, while web-scraped data contain a rich source of product information that is useful for things like accurate classification and determining quality. Scanner data also have the potential to allow us to provide greater regional coverage of prices and expenditure such as, for example, regional inflation measures.

However, a significant period of research and analysis is being undertaken before these new sources are fully integrated into the headline estimates of consumer price inflation.

## 4 . Current methods of data collection

We will continue to use traditionally collected data when they cannot be replaced by scanner or web-scraped data (for example, from independent shops who do not have a website or where it is not practical to collect scanner data).

In addition, we will continue to [use administrative data sources](#) that are already currently used for some areas of the consumer basket, such as for owner-occupiers' housing costs and private rents. We envisage that in future our consumer price statistics will be a mix of scanner, web-scraped, administrative and traditionally collected data.

## 5 . Implementation plan

Our ambitious plan is to include data from alternative sources in our headline consumer price statistics by Quarter 1 (Jan to Mar) 2023.

We have split our delivery programme until 2023 into three phases, each lasting a year. The phases are:

- research
- application
- engagement

### Research

The first phase (research) will run until the end of 2020 and involves further developments of the systems for web-scraped and scanner data to enable research and impact analysis, alongside research into the methods needed to produce high-quality indices using web-scraped and scanner data (a full research programme is outlined in our [Consumer Prices development work plan](#)).

### Application

The second phase (application) will run throughout 2021 and involves the application of research to our targeted item categories (subject to data availability) within the inflation basket, notably:

- groceries
- clothing
- tech goods
- used cars
- package holidays
- air fares
- rail fares
- chart collected items

These items have been discussed and prioritised in line with advice from our [Stakeholder Advisory Panel for Consumer Prices \(APCP-S\)](#).

## Engagement

The third phase (engagement) will run during 2022 and involves the quarterly publication of aggregate experimental indices, including web-scraped and scanner data in conjunction with traditionally collected data.

Throughout each phase, we will also be liaising regularly with our [advisory panels on consumer prices](#), our users, and the Office for Statistics Regulation, to ensure that our future plans for consumer price inflation measurement are appropriate for improving the quality of our statistics and meeting our ongoing user requirements.

## Timeline

2020: Research into the methods required to process alternative data sources.

2020 and 2021: Development of systems for processing alternative and traditional data sources.

2021: Application of methods and impact analyses for priority items.

End of 2021: Recommendations of methods for each item and data source.

2022: Parallel run to produce experimental aggregate measures (quarterly publications).

Quarter 1 (Jan to Mar) 2023: Alternative data sources used in aggregate measures of consumer price statistics.

2024 and beyond: Roll-out use of alternative data sources to new items within inflation basket.