



EUROPEAN COMMISSION
EUROSTAT

Directorate C: National Accounts, Prices and Key Indicators
Unit C-5: Price statistics. Purchasing Power Parities. Housing statistics

WSMPPS04.2013-02

Workshop

On the use of scanner data

Lisbon, 26-27 September 2013

The use of scanner data as a source for HICP

Author: Network of Price Experts

Table of contents

1. Description and assumptions of the use of scanner data	3
1.1. What is really collected with scanner data?	3
1.2. Potential uses of scanner data	4
2. Compilation procedures with the use of scanner data	5
2.1. The dynamics of EAN codes and their combination with COICOP	5
2.2. HICP computation	6
2.3. Limitations of scanner data	10
3. Potential increase and broader use of scanner data.....	12
3.1. Exploring the possibilities of its use for a monthly HICP	12
3.2. Exploring the possibilities of scanner data for use in PPPs	14
3.3. Exploring the possibilities of scanner data for Detailed Average Price Statistics (DAP)	15
4. Uses of scanner data in Europe	16
4.1. Recent developments in European countries	16
4.2. Motivation and involved NSIs	16
4.3. Definitional issues and Item classification	17
5. Compliance with HICP Legislation	18
6. Compliance of Scanner Data in terms of the PPP regulation	20
7. Scanner Data and Modernisation.....	21
8. Recommendations and Conclusions.....	23
9. Bibliography	25
Annex 1 - practices in compilation procedures with sd and the HICP regulations.....	27
Annex 2 - comparative summary of the experiences made by NSIs using SD.....	37
Annex 3 -international article number (EAN).....	45
Annex 4 - roadmap to use data based on scanner data (SD).	47

1. Description and assumptions of the use of scanner data

1.1. What is really collected with scanner data?

SD is the retail purchase information that is acquired by electronic devices (POS – Point of Sale) that read coded tickets (bar codes) on the products. There are potentially many different ways in which SD could feed into the CPI/HICP database.

The transaction code scheme can be different from NSI to NSI:

- a. using only EAN (formerly the European Article Number, today the International Article Number);
- b. EAN plus some other identifier number, such as PLU (Product Look-Up code) or SKU (Stock Keeping Unit);
- c. EAN and complementary descriptive data, e.g. from the internet, or even
- d. In store numbers.

EAN is an identification product code, used on barcodes, presented as 13 or 8 numeral digits, that is globally managed by GS1 (Global Standards ONE, a non-profit organisation that resulted from the merger of EAN - European Article Numbering Association and UCC - Uniform Code Council in 2005). See Annex 3 for more details on EAN. The EAN codes are formed by a GS1 Prefix, a Company (producer) number, an Item reference and a Check digit. The EAN GS1 Prefix and the Company number are determined by GS1/GS1 national branches. The Item reference is determined by the producer company with guidance from GS1.

A number of fundamental issues need to be clarified prior to considering the adoption of scanner data:

- a. What transactions are covered by scanner data? Exclusively, household expenditure or transactions may include also business expenditure? (page 6)
- b. That part of expenditure, relating to those household transactions that qualify for inclusion in the HICP, which is covered by scanner data (Box 2)
- c. What measure of price is there in SD i.e. the implicit definition applicable to SD? (page 6)
- d. Are those prices with discounts or not? If yes, what information is available about the discounts/reductions that are covered and how they are treated in the SD database? (page 9)
- e. What is the price concept underlying the monthly average prices generated by SD (a real average of transaction prices at detailed level; a ratio between turnover and quantities (unit values) (pag.7); and
- f. What is the data collection frequency (daily, weekly, etc)? (page 11)

Box 2 – Expenditure coverage

The use of scanner data is more appropriate for some parts of the HICP basket than others because of the varying availability and outlet coverage of data sources and the varying complexity of some measurement areas e.g. tariffs for services and utilities are complex and are not covered by SD (although electronic information on sales can be downloaded from the accounting systems of the service suppliers) and sales by small independent shops are not generally included.

In theory, scanner data can also be used for checking the balance of the sample e.g. of brands or varieties using traditional sample regimes e.g. if price collectors are pricing the correct proportion on a given brand of TVs; for probability sampling (whilst still collecting prices in shops) and for expenditure weights

In principle, the compiler of a price index can define an SD dataset according to the purpose and use of the data. Thus, for the HICP, the compiler will want to extract an “HICP centric” data set following as closely as possible HICP conventions and guidelines i.e. a data set consisting of the prices of products available to purchase (or average revenue per item), gathered in one or more data collection waves per month (depending on price volatility), for a fixed panel of products, renewed every year, and will also want to extract the corresponding expenditure data for the weighting scheme, including the identification and treatment of seasonal products. The same data set, if accompanied by detailed product information, can also be used for quality adjustment procedures including the computation of hedonic functions.

1.2. Potential uses of scanner data

SD can be exploited in a number different ways:

- a. Exclusively replacing traditional price collection in the stores, whilst retaining the fixed basket approach and also keeping the international methods for computing the price indices;
- b. As a new powerful source for CPI/HICP computation based on all the possibilities offered by this data source, namely, the coverage for all actual monthly transactions, daily transaction prices, allowing for frequent re-weighting and re-sampling;
- c. As a source of price and sales data for Purchasing Power Parities (PPPs), to address some of the criticisms relating to the sampling dimension (product, geographic coverage and number of price quotations) associated with the traditional approach. This can also provide greater coherence between the HICP and PPPs;
- d. In the computation of detailed average prices for market monitoring concerning price convergence;
- e. As a benchmark for the present HICP;
- f. As a source for computing Cost of Living Indices (COLI's), since SD can incorporate substitution effects due to price changes and superlative indices which when compared

against the HICP computed in the traditional way using weights from an HBS can gauge potential bias in the index¹;

- g. As a source of data to improve the coverage for multipurpose price statistics.

2. Compilation procedures with the use of scanner data

2.1. The dynamics of EAN codes and their combination with COICOP

Although there is a one to one relation between an EAN code and a given product, the former does not identify such products, in the sense that there is not a “take and use” dictionary describing the detailed product characteristics, its connection with EAN code and conversion to COICOP. That is to say, it remains for NSIs to make such a linkage between the characteristic description and the corresponding EAN to COICOP.

There are at least three potential difficulties when using EAN codes:

- Firstly, there is the case associated with price reductions or gifts associated with items bought (e.g. offer of a free DVD when buying a DVD player). With some scanner data systems, each time a volume discount or an associated gift occurs the same product but new “product offer”, acquires another EAN code related with a different packaging, thereby facilitating the easy identification of the new “product offer” (although not necessarily providing an easy solution to how to treat this “discount” in the HICP). Of course, such discounts are not easy to identify where no new EAN code is issued and in both cases such changes in pricing and product packaging will be missed unless the scanner data provider submits separate information on such “discounting” practices. Thus the issue can be solved only with the provision of a list of EANs that have changed within the period together with the reason why;
- Secondly, there is the case of products disappearing and EAN code recycling: an EAN code of an extinct product may be recycled later on for another product, according to some authors, e.g. Norber et al, 2010, when describing the Swedish experience. The same authors note that this can happen after more than a year i.e. after the product has been replaced in the HICP basket and the recycling of the EAN code is not problematic²;
- Thirdly, when a refund from a return happens this may be recorded as a negative price and in some scanner data systems the identification of the item to which it refers may not be straightforward.

-
- 1) Scanner data can provide for some parts of the basket up-to-date weights to deal with substitution bias from older weights from the HBS or based in National Accounts data.
 - 2) Generally, the annually revision of the basket (re-sampling) and also of the weighing scheme are considered a good practices in compiling a CPI that aims to measure “pure” price movements. Currently however, not all MS annually resample their basket of goods or annually update their weights at the detailed level.

Other authors, such as van der Grient, de Haan, 2010) referring to the Netherlands research, report a situation where, during the course of a year, the same EAN is irregularly re-assigned to different products. For example, this can happen when, through a commercial decision, the chain-store re-packs a given product as unbranded. The solution, as referred by the authors, is to exclude those EANs if it is known *a priori*, or exclude the correspondent data in the editing procedure.

2.2. HICP computation

Weight updating

According to HICP regulations, weights from the level of division to the levels of classes (COICOP, 2, 3 and 4 digits) should not vary between months during the year but have to be reviewed at the end of the year, using available data from National Accounts, Household Budget Surveys or other reliable statistics. The updating of weights at lower levels is a matter of subsidiarity but it can be noted that sales figures from scanner data can be used to help update weights at a lower level of COICOP. Scanner data also facilitates the updating of weights at a higher frequency than annually, seasonal products apart where special procedures need to be applied (see Seasonal treatment). However, business expenditure and expenditure incurred by Governments or non-profit institutions on goods or services provided to households free or at prices that are not economically significant such as social transfers in kind, should be excluded unless it can be proven that these are of little consequence. Whether the latter is the case is possible to vary by product group and it is likely that the necessary customer information to identify and exclude the relevant purchases may not be available.

Price measurement

Each scanner data record represents a transaction defined by identifying a quantity and the value of each transaction at a Point of Sale (POS). At this level of detail, the derived unit price can be considered as the transaction price for that item. One should note that this “price” is not a “shelf price”, but a transaction price that exists only if the product is purchased and one which also includes other things such as returned goods as negative prices. Although the price collected at a POS is compliant with the concept of transaction price, according to HICP standards, the overall set may not cover all the HICP basket items available for purchase in the outlet. This occurs for those basket items where no purchase has occurred in the observation period and gives place to “false” missing observations. If the data collection changes focus from POS transactions to Shelf price lists this problem is minimised³.

A transaction price is a market price or the purchaser’s price from the buyer’s point of view. This is the concept underlying a HICP aiming to measure inflation (Council Regulation no. 2494/95, article 3). On the other hand, a unit value is an average value per units. These may not be homogeneous in an economic sense as a unit-value index therefore may be affected by changes in the mix of items as well as by changes in their prices. The impact of this deviation

3) Only in the same situations as for traditional price collection – rupture of stocks

from the conventions of index construction will depend at what level of detail the unit value is taken.

Upper levels, other than transaction by transaction, can lead to biased prices far away from the transaction price concept, namely when the price is derived from weekly values for turnover and quantities.

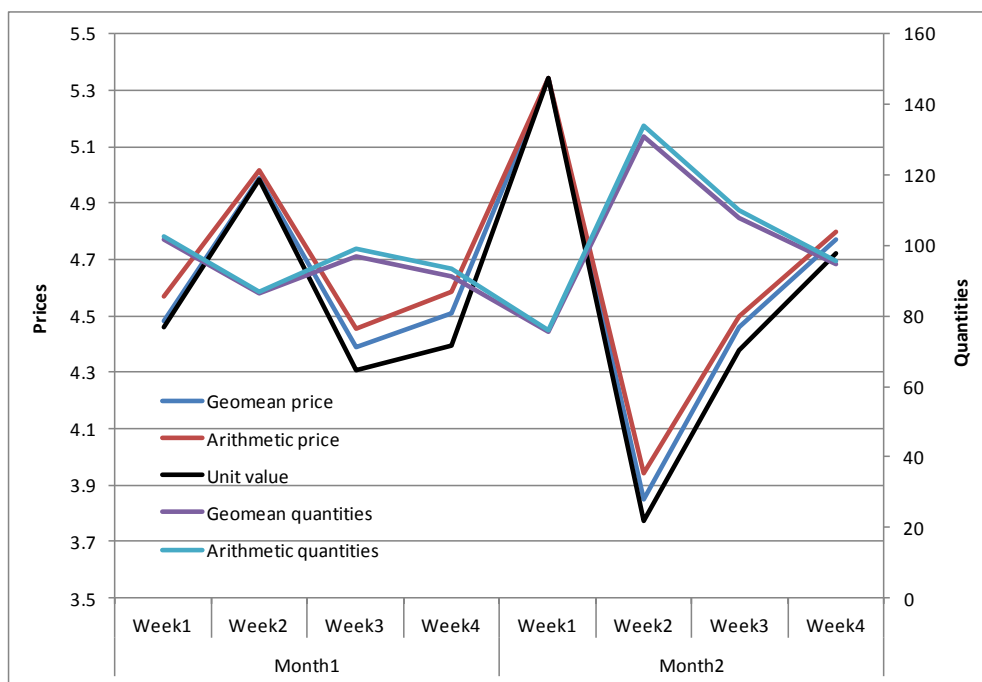
It seems that the unit price issue, and the possible anomalies described above, are generally ignored by those NSIs who use scanner data. The difference in the nature of “price” and, more particularly, the difference between a transaction price and a unit price is not fully considered. SD is viewed as an opportunity to gather useful information on the quantities and turnover of purchased items at the same time as obtaining the (unit) price. These solutions raise problems that have the potential to put the resulting price concept further away from that of the HICP: for instance, the unit price compiled with the weekly aggregation of quantities and turnover can result in a different estimate of inflation compared with one based on individual transaction data and biased from a weekly average of daily transaction prices per unit since there might be a negative correlation between price level and quantities transacted. The following table and related chart provides a simple example. Using data of daily prices and quantities for two months, weekly average price and weekly average quantities were calculated, both arithmetic and geometric means, and also the weekly unit value.

$$\bar{P}_{Arithmetic} = \frac{1}{n} \sum_{n=1}^7 P_n$$

$$\bar{P}_{Geomean} = \left(\prod_{n=1}^7 P_n \right)^{1/7}$$

$$UValue = \frac{\sum_{n=1}^7 Q_n \times P_n}{\sum_{n=1}^7 Q_n}$$

		Geomean quantities	Geomean price	Arithmetic quantities	Arithmetic price	Unit value
Month1	Week1	101.475	4.480	102.429	4.571	4.462
	Week2	86.345	4.988	86.714	5.014	4.981
	Week3	97.001	4.387	99.000	4.457	4.309
	Week4	91.075	4.510	93.571	4.586	4.397
Month2	Week1	75.647	5.343	75.714	5.343	5.342
	Week2	130.966	3.850	133.857	3.943	3.772
	Week3	107.735	4.458	110.000	4.500	4.377
	Week4	94.496	4.770	95.571	4.800	4.722
Month 1		93.798	4.585	95.429	4.657	4.524
Month 2		100.214	4.573	103.786	4.646	4.438
Monthly rate of change		6.8%	-0.3%	8.8%	-0.2%	-1.9%



In some cases, weekly unit values used as a proxy for actual “prices” can under-estimate inflation and, in other cases, can over-estimate inflation. In a simple example, consider the following case: in the first day a product costs 10. One item is sold so the transaction price is 10 and the turnover is also 10. In the second day the effective price falls 50% from 10 to 5 and two items are sold. The transaction price falls from 10 to 5. But the two-day “price” computed as a unit value is $(10+2 \times 5)/3=6.67$. The arithmetic price mean is $(10+5)/2=7.50$. The geometric price mean is $(10 \times 5)^{(1/2)}=7.07$. In this case, the unit values used as “prices” underestimate the actual fall in prices.

Replacement products

The introduction of replacement products will be constrained by the necessity to preserve the stability of the consumption segment by purpose, since “Consumption segments shall form the fixed objects in the index basket to be followed by the HICP” (amending Regulation (EC) No. 1334/2007 of Regulation No. 2494/95 article 2a, p.4). There is also the issue relating to the treatment of seasonal items (see f.). It goes without saying that the Laspeyres-type formula continues in force (*ibidem*, p.3).

Elementary indices and product level

Many approaches can be considered for the elementary level. The one which is nearer to the HICP is defined by the pair (retailer x EAN code). However, the product boundaries remain with some ambiguity: should it be defined by a composition of different EANs, on the assumption that there is a sort of identity (homogeneity) among items with “similar” EANs? If this is the case then product replacement as well as seasonal treatment procedures would be easier and more flexible. So far, all investigation and readings about a possible hierarchical organisation of the EAN identification code, leads to the conclusion that this assumption is not universally applicable. This leads to the need of mapping EANs complemented by other auxiliary identification codes associated to item description to the COICOP classification at the elementary level.

Quality adjustment

An HICP compiled from scanner data when the variables collected are not at the level of transaction cannot include explicit quality adjustment unless additional product information is trawled from, for example, the internet as has been done by some countries such as the UK. However, this can be resource intensive if products change at different times in different outlets.

Seasonal treatment

According to Regulation (EC) No. 330/2009 seasonal, weights of seasonal products are required not to vary over the year, except when an approach based on class-confined seasonal products is followed (in this case a seasonal product that is out-of-season has a weight of zero). In the former case, when HICP strictly follows an annual weighting scheme then weights for seasonal products are fixed, which imposes one constraint on the variability of the segment product by purpose that contain this (or these) seasonal product(s). Frequent tranches of SD facilitate the recognition and determination of out-of-season and in-season periods.

Treatment of price reductions

There are two kinds of price reduction, one which has a universal scope, in the sense that the reduction is available to all consumers, and the other where availability relates to a specific subgroup of the population, such a pensioners, or depends on a specific connection or relationship between the retailer and the consumer.

Universal in scope: Regarding straightforward transparent price reductions which are shown by an amended price tag, if the EAN does not change, the item it is correctly entered into the index with the new lower price. However, if the EAN changes this does not constitute an un-resolvable problem providing that the elementary product level is located above the level of the each particular EAN and the price collection is by transaction over a week for each EAN, or by day per EAN, or by transaction over a day for each EAN, respectively. In those cases there is no problem with the computation of an average price at elementary level. Of course, if it is possible to match the previous EAN item with the EAN discount item, then the elementary level can be taken into consideration.

Constrained scope: this includes those discounts by retailer card or by ticket/coupon, and volume discounts which should not be included in the HICP framework, since they do not have universal application to all potential customers. The exclusion of specific EANs from data collection can be easily tackled in the framework collection described above, i.e. collection by transaction over a week for each EAN, by day per EAN, by transaction over a day for each EAN, respectively):

EAN level: these discounts have a different EAN Prefix from the original, so its exclusion is immediate if the product is defined at the EAN level;

Products defined at a higher level than the EAN code: if it is possible to identify these specific EANs, then they could be simply excluded from the collection procedure. If no action is taken then the assumption is that few transactions of this sort take place and that they have no significant impact on measured inflation. If this assumption is correct, ignoring them does not break the comparability rule.

In the HICP new legal Framework “price reductions available to consumers with special conditions attached (discriminatory price reductions), may be disregarded unless they are significant. Were such reductions are included they shall be taken into account (sampled) according to their relevance in the HICP”. Also the Task Force HICP Quality Improvement on the treatment of discounts concluded the following: A certain concern has been expressed that discriminatory price reductions can be overestimated in the HICP when NSIs are using scanner data, because it is only representative for part of the population. This argument was considered as most relevant for certain special cases like discriminatory discounts. It is proposed that NSIs using SD must secure with proper sampling procedures that these kind of price reductions are not under- or over-represented in the all items HICP/ and any given HICP sub-index.

In general, items subject to a price discount maintain the original EAN code and price on the Point of Sale data (scanner data), while the discount values are aggregated and presented as a single item by the Point of Sale. In these circumstances the discount value for each item is not included in scanner data collection. On the other hand, other SD systems follow another procedure, namely when each item is registered with value net of the discount, and then the discount may be correctly included in collected data. But it should be noted that in some instances unit values may include information not relevant to a CPI e.g. the inclusion of “negative prices” for returned goods.

Summary of key issues on discounted items

Volume Discounts	Can be processed as a quantity adjustment if an EAN correspondence is put in place, or processed as ordinary sale if transaction or daily variables are used.
Price Discounts	Either included in the transaction price (as a reduced price) or identifiable with a specific EAN.
Subscription Discount	There are some prefixes for the payments from the subscriber system (cumulative values discount from previous purchases). These items should be ignored.
Coupon Discount	There are some EAN prefixes that identify the transaction as a coupon – They must be excluded from SD processing routines.

2.3. Limitations of scanner data

Outlet and geographic coverage

Scanner data may be limited in terms of outlet coverage for example small shops and markets etc. which possibly do not use scanner data equipment. Small outlets and markets may have different price changes and levels. Disregarding these outlets can lead to miscalculation of inflation or price levels for PPPs and multipurpose price statistics. The magnitude of the problem depends on their weight. Consumer habits in certain countries and locations are likely to favour shopping in small outlets rather than supermarket chains. These considerations lead to the issue of geographic coverage. It makes sense that supermarket chains are located primarily in the city centres while in the villages and country side small shops prevail. Computing an HICP from data exclusively obtained from SD may prejudice the HICP’s required outlet and geographic coverage.

Incomplete basket coverage

It is unrealistic to suppose that scanner data can be used for all expenditure categories in the HICP due to the complexities of the pricing structures or the transactions not being scanned.

In general, services are not likely to be covered in sufficient depth so the use of SD may be largely limited to goods. Such cases include tariffs, insurance, health, education, social protection, and charges proportional to transaction values and some other services etc. Also some goods such as new and second hand cars, motorcycles are unlikely to be covered. These items can make up a significant part of the weight of the HICP basket.

The main COICOP categories that are being covered by the countries are: 01 Food and non-alcoholic beverages (all classes); 02 Alcoholic beverages and tobacco (all classes); 04.3.1 Materials for the maintenance and repair of the dwelling; 05.5 Tools and equipment for house and garden; 05.6 Goods and services for routine household maintenance; 06.1 Medical products, appliances and equipment; 09.3 Other recreational; 12.1.2/3 Electrical appliances for personal care; other appliances, articles and products for personal care.

Data availability and the correct specification of data characteristics

When requesting scanner data from a potential data provider, special care must be taken regarding the specifications and the variables asked for:

- Weekly unit values that have been used by some NSIs in the past, may not give reliable results. The geometric mean of the transaction unit values for a given period of time may be a more accurate option. The appropriate time period will depend on the volatility in the prices being observed;
- One possible drawback of scanner data is the uncertainty regarding the inclusion in data collection of universal reductions in the form of price discounts. Information about discounts and the method of treatment by the provider is crucial for this point. Promotions and price reductions must be clear and it must be possible to identify these in the data to ensure their correct treatment and also for analytical purposes;
- The timing of the consumption, in particular for services (for example, the case where tickets are bought for an event that will occur in a later date) needs to be identified. Along with any compulsory additional charges such as booking and or delivery fees.
- Product descriptions are often not as detailed as required to match EANs to COICOP, to overcome this issue a straightforward communication between HICP compilers and someone responsible in the providers is fundamental.

SD collection

The collection of the records within the observation period can represent a huge quantity of data records, a scale that NSI's resources may not be capable of dealing with, particularly with regard to effective quality control and editing of the information obtained. NSIs can be overwhelmed by the sheer size of the data set, which may be beyond the capacity of the CPI team particularly in smaller NSIs. This could be a real issue for NSIs with limited human and IT resources.

The data transfer scale is dealt with by NSIs using differing processing techniques. Concerning the detailed report of the transactions and the consequent huge scale of data transfer, the adopted solutions to avoid it, involve using:

Some degree of aggregation (temporal aggregation at elementary item level according to HICP rules) performed by SD providers prior to delivery of the data to the Statistics Office.

Use the same frequency of the traditional model of price collection for which prices for products with stable prices are collected with a lower frequency than for products with volatile prices. Items with stable prices (e.g. cookies, shampoos ...) may be collected monthly at a pre-defined point in time. On the other hand more volatile prices (e.g. fresh fruits and vegetables) price collection can have a daily frequency.

In compliance with the regulation stating the fixed basket approach the request to the providers can be based on predefined list of EAN codes.

For monthly/weekly data collection it should be requested price list data, instead of POS transaction data. Although not being strictly SD (but available from the retailer information system), those price lists are compliant with the HICP price concept (shelf price), i.e. prices of products available to be purchased.

Only once a year data on aggregated transactions (quantities and transaction turnover at elementary level for twelve months) should be transmitted in order to update the basket and the weights.

3. Potential increase and broader use of scanner data

Looking forward, SD offers the possibility of contributing to the re-engineering of consumer price statistics production. It could be possible to make both efficiency savings and quality improvements to HICPs, PPPs and detailed average price for multipurpose price statistics using SD for a large part of the data collection.

3.1. Exploring the possibilities of its use for a monthly HICP

Different objectives or motivations may be defined for adopting scanner data. Clearly, implementation approaches can be a “step by step” or a “once and for all” process.

Both traditional shop based price collection and those using SD are potentially costly. There is no clear evidence that the use of SD will be less expensive than the traditional price collection method. Some experiences developed so far reveal a significant cost increase from adopting scanner data. The cost analysis must be subject to further research. The use(s) of scanner data should be evaluated not only on its costs, but also on the quality improvements and efficiency gains that may occur through its use (and, conversely, the dangers of being overwhelmed by large quantities of data). The current legal framework is not clear in respect to suppliers of SD as to whether they must supply their data in the requested format for free, although experience suggests, from other surveys conducted by National Statistical Institutes, that whilst it may be possible to use statistical law to enforce the provision of data it may not be possible to apply

such laws to insist on the provision of such data in a standard highly-specified form. Thus, in reality most suppliers are likely to charge for the data and its manipulation before transmission to NSIs.

In the future, there is a potential increase of SD coverage not only regarding the store types but also specific products and services. General modernisation and fiscal awareness may lead to a broader use of electronic invoices leading to even more precision in the data that is registered. In Portugal at least (though is not currently the case in other countries), the willingness to transmit to fiscal authorities all these data from all data providers regarding all products and services may facilitate collection directly from administrative sources, instead of protocols with a few chain store thus decreasing the coverage of retailers’.

SD has the advantages of extensive coverage within the population of outlets using scanners and of providing data on quantities sold, as well as on prices and/or average revenues generated. The latter is generally available separately for each model of the variety of goods sold, with the potential for direct input into index construction.

At the very least, the provision of better detailed weighting information and the identification of representative products is the minimum which could be expected from the use of SD.

New technology for processing transactions, together with increased capability to store and process information, leverage the potential use of bar-code SD from retailers for the measurement of inflation.

Traditional price collection involves visiting or contacting outlets via telephone or email on a regular basis and at a pre-specified time to obtain isolated price quotations with no expenditure information, using sampling methods which are generally purposive and can involve the price collector in some element of sample selection e.g. when asked to identify the most representative brand sold in a particular outlet. SD sampling is likely to be more accurate leading to a reduction of sampling errors⁴. The extremely large size of SD databases and the amount of detailed information they contain can be used to argue for the use of such data on a sample basis as this is likely to be more efficient, easier to quality assure and less prone to processing errors. Also, aggregation could be made over stores, or a monthly chained index may be constructed or other alternative methods for price index compilation could be tested such as index formulation and the use of superlative indices, monthly chaining, and specific indices for categories of retailers’ among others.

Scanner data can also be used to compile an index which is close to a Cost of Living Index, computed either in its own right or as a benchmark for comparative analysis, regarding that with the monthly/weekly transmission of transaction data (prices and quantities) the computation of a superlative index, i.e. using current weights, which can express the price increase according to the present consumption structure is a possibility. No significant problems in the manipulation of SD arise that cannot be overcome at least in theory. If the issue is the measurement of the cost of each consumption segment, according to its purpose, it

4) The calculation of the sampling error is based on the standard deviation of our sample. The greater the sample standard deviation, the greater the standard error (and the sampling error). The standard error is also related to the sample size. The greater the sample size, the smaller the standard error because the greater the sample size, the closer the sample is to the actual population itself.

is sufficient to identify the set of segments to be followed and estimate an index given by the ratio of unit values. In addition, no methodological constraints would be relevant for the sampling frequency. Of course, there might be issues relating to the cost and price collection burden. It should be noted that a COLI is not the target index for the HICP. However, this may be the approach adopted for some national CPIs which have differing national uses.

Relative weights can be attributed to price changes at a highly detailed level, which facilitates the estimation of cross-elasticities of substitution. In addition the regular supply of SD information on sales volume enables the compiler to estimate the substitution bias effects to inform strategic decisions on the computation of the CPI/HICP and how much the HICP differs numerically from other target measures of consumer inflation. Thus, the availability of such a huge amount of data on a regular and timely basis can significantly increase the potential for monitoring and improving the quality of a CPI/HICP against the target index. Of course, the official HICP should be compiled according to the requisites of the legal framework so any deviation from the latter would need changes to the HICP regulations. This also applies to PPPs.

3.2. Exploring the possibilities of scanner data for use in PPPs

PPP surveys aim to produce reliable and representative average prices for tightly specified products which can be found in at least two (but normally a lot more) countries within the Eurostat / OECD comparison program. A long-standing general criticism of current PPP surveys is that the derived average prices are based on very small sample sizes, which could affect their reliability in terms of representativeness. SD possesses the potential for vastly enlarged sample sizes with the advantage (for the products that they cover) to improve the reliability of PPP estimates. This would be dependent on the ability to efficiently identify matching products from the data sets, particularly where variables such as size can vary within set ranges, also for some goods such as TVs and kitchen appliances, identical products may be sold with different model numbers in different countries, all of which would require some post-sampling, checking and data manipulation. The internet, used as a supplementary source of information to the EAN, combined with the internal code of the retailer (giving specific characteristics), can provide more product details, particularly for hi-tech goods, to identify through hedonic regression the price-determining characteristics to ensure a close-match of price-determining characteristics or some quality-adjustment where there is some mismatch. Whilst the use of hedonics is a valid approach; resources may limit the practicality of this approach.

More serious problems arise with the coverage of data sets. The advantages of using SD for PPPs will be reduced if the outlets covered by the data sets are not representative of the price levels of the chosen sampled items or the universe of items covered by the price comparison exercise. The same would apply if data was absent for a major retailer.

PPP surveys are “snap shot surveys” and as such temporary price reductions, such as those available during seasonal sales, discount days or as “special offers” and also discounts that are offered only to a selected group of purchasers, such as store account holders or holders of

certain credit cards etc, should be ignored as these can occur at different times in different countries⁵. But the challenge is the difficulty in separately identifying these in SD.

Since PPP results are used to adjust National Accounts annual aggregates for differences in price levels, PPP survey data are annualised using detailed CPIs. This in theory takes into account the effect of sales. SD, being available on a continuous and timely basis, can be compiled to cover a year to give annual average prices so the use of CPIs to annualise PPP data – which is not the best methodology - can be by-passed.

SD has the potential to vastly increase data quality in terms of both sample size and methodology, as past investigations suggest that annualising data is not always reliable. As PPPs are primarily used for spatial comparisons of GDP, the use of SD has the potential to improve PPP estimates for some COICOP categories, mainly goods. Also SD is potentially timelier than the traditional surveys with annualisation using CPIs. Services, which are among the weaker areas of PPP data collection, are less likely to benefit from SD.

3.3. Exploring the possibilities of scanner data for Detailed Average Price Statistics (DAP)

The approach to Multipurpose Price Statistics has been to produce detailed average price statistics (DAP), with the aim of responding to new policy needs for specific purposes. . Bulk SD containing as the main product key the EAN's (see 3.1.) is not required. Instead micro data should be transmitted with both COICOP and EAN classification for products.

SD would give more scope to the detailed average prices sample. A key requirement for the use of SD for detailed average prices is timeliness, product coverage and the availability of strict comparability characteristics, as is required for PPP product definitions.

In general SD may serve this aim well, and will benefit from the technological developments for the use of the same data collection procedure for the HICP, PPP's and MPS, thus improving efficiency.

Instead of MPS trying to match the same products as the HICP and PPP's in order to avoid more costs, the use of SD would allow for the possibility that MPS use a different sample. For instance, suppose that for a specific purpose the users request a product price index that has very few observations/varieties in the coverage of the regular HICP and PPP's, resulting in a lack of accuracy of the indicator due to a small or unrepresentative sample. The large availability of data provided by SD allows the compilation of that product price index with more data potentially overcoming the problem of accuracy and bias that may result from use of few observations for an index. However, there is the danger, of course, of being overwhelmed by a large number of data requests and large quantities of data leading to issues of inadequate data handling and editing.

5) Eurostat-OECD Methodological Manual on Purchasing Power Parities, page 109, European Union / OECD, 2012.

4. Uses of scanner data in Europe

4.1. Recent developments in European countries

Most of the large retail chains and even medium retailers use IT and Information Systems based on electronic transaction management operations. NSIs have become increasingly interested in benefitting from SD usage for the reason that this could bring potentially big savings on their CPI/HICP, PPP, detailed average prices data collection and operation costs, whilst also delivering potential value-added from acquiring new kinds of data (both in a qualitative and quantitative sense), for which there was no previous access. The facts lead to uncertainty concerning the potential savings and to question whether it may perhaps be more costly.

4.2. Motivation and involved NSIs

Some countries have started research projects on the use of SD, several in the context of a new powerful data source and others are exploring alternative methodologies to produce indices (monthly re-sampling, weighted Jevons formula, weighted average of outlet prices etc.).

A number of MSs have investigated the potential use of scanner data for compiling their CPI/HICP but only a handful use SD for the direct production of their indices, mainly contributing to HICP compilation.

For example, some of the countries are aiming to reformulate their CPI/HICP compilation streamlined with the exploitation of SD:

- a) Statistics Netherlands introduced supermarket SD into the CPI/HICP in June 2002 (Schut, 2002) initially for two chains but now just for one;
- b) Statistics Norway has been exploiting SD since August 2005 to compute the sub-index for food and non-alcoholic beverages (Rodriguez and Haraldsen, 2006).
- c) In Norway and in the Netherlands both prices and expenditure weights (for a large sample of items from each product group) are derived from the SD;
- d) Since January 2011, Statistics Denmark (DST) has received SD from the largest supermarket chains on a weekly basis for food and beverages for research purposes. DST still produces a regular CPI using traditional price collection methods.

Conversely, other NSIs have chosen a different approach, aiming to fine-tune the way SD could complement traditional price collection procedures.

- a) In France, INSEE engaged research efforts using SD to solve representativity problems in their regular CPI/HICP production arising from the turnover in the market place of product makes and varieties, (Faivre, 2012);
- b) The Swiss Federal Statistical Office follows a similarly pragmatic approach. SD from some major retail chains is used as an additional source for price collection; prices taken from SD simply replace some of the prices formerly collected in the outlets

without changing the underlying principles of computing price indices, (Muller et al, 2006);

- c) Statistics Sweden undertook a comparative study on the feasibility of SD use for CPI/HICP purposes, including a comparative analysis between SD and annually collected prices during the full years 2009 and 2010, taking into consideration the EU regulations on prices and items for the CPI/HICP, (Norberg et al, 2010).

Finally, there is still a group of NSIs that are making their first steps on this subject, and trying to get the best they can from the experience of each other and to benefit from the best practices.

- a) In the UK, ONS has been awarded a Eurostat grant to undertake research on the exploitation of SD. The research will conclude with the production of a short guide on the use of scanner data in the compilation of price statistics. The ONS had previously used scanner data for local probability sampling to improve sample representativity amongst hi-tech and high turnover goods, however, this proved to be very expensive;
- b) In 2011 Statistics Portugal was awarded with a Eurostat grant to undertake the initial research on the exploitation of SD from the period of 2011-2013. Although the expressed aim was to increase price collection efficiency, it is not yet clear if this is the orientation for subsequent research, (Saraiva, 2012).

Annex 1 gives a comparative summary of the experiences of the above NSIs highlighting the following aspects: “implementation status”, “objectives”, “price framework”, “sample management” and “index compilation”.

4.3. Definitional issues and Item classification

Some NSIs reported that, in rare cases, EANs may be re-used causing the occurrence of a identification product code which isn't unique and exclusive to the product in question and leading to two products having the same EAN at two different points in time. This kind of anomaly can occur for EANs where the item reference part is attributed by the retailer itself when packaging unbranded or own-branded products. These products may have prefixes in the special ranges 200-299 and 400-499.

Additionally, it is a general practice of NSIs to request regularly to the retailer-chain a list of characteristics and full descriptions for each EAN.

The mapping between item identification codes and some detailed levels of COICOP classification is achieved by Statistical Offices using a variety of differing methods:

- a) buying the mapping from market research companies;
- b) performing the matching in-house;
- c) using other sources of dictionary lists (PLU, SKU, in store numbers), or even
- d) using characteristic descriptive sets, and the appropriate IT application (describe-and-match program working on a semantic description), to automate the classification process.

There are substantial numbers of EANs that vanish and others that are newly introduced to the monthly collection. The “churn” in the different products in the retail market that are available to purchase can lead to a considerable number of changes in the composition of the periodic collection of scanner data from any one shop. Due to these circumstances, the mapping stage is an important process and can be very time-consuming when carried out properly. The solutions adopted by countries range from:

- a) manual mapping;
- b) automate the classification using extra codes or/and characteristics, or even
- c) the use of some of the characteristics of EAN numbering to group similar products.

Some countries ignore the issue but this does not, of course, solve the problems arising that can undermine the statistical integrity of the CPI/HICP basket if ignored.

It is interesting to note that some NSIs work with a static basket of items (EANs) and between basket updates use a set of procedures (missing price estimates⁶) to maintain continuity. Other NSIs maintain a “fixed basket” at a higher level of COICOP combined with a dynamic basket at lower level.

5. Compliance with HICP Legislation

SD specific characteristics are the following:

- a) Its intrinsic connection with effective sales of products;
- b) Its high degree of availability;
- c) Its universality, in the sense that, in principle, the EAN code gives an unequivocal identification worldwide.

The first two characteristics, essentially positives, can nevertheless lead to problems arising from different approaches by MS (with impacts in harmonised results, sampling procedures, price concept, computation methods for average prices, index formulation, among others) and be a twofold factor of miss coordination and loss of comparability of the HICP. Undoubtedly, the implementation of procedures for using SD, if done without clearly defined rules that are compatible with the criteria for HICP index construction, can generate calculation procedures that do not match the standard methodologies of the HICP, namely those concerning the maintenance of the Laspeyres formula, the nature of the collected variable (price concept), and representativeness of chosen products. Such methodological deviations from HICP rules can clearly be a source of non-comparability and also represent a move away from the target index.

Scanner data “prices” are obtained from quantities and volume sales and can be a unit value, geometric mean or arithmetic mean. The geometric mean is the most suitable with the HICP standards. The unit value is not strictly compliant either with the transaction price concept or with the HICP standards.

6) According to the article 6 of Commission Regulation no. 1749/96.

In the experiences reported by countries no references were made to how unit price averages were calculated, for instance whether a geometric mean was used.

If the current HICP regulations and guidelines are retained and adhered to, the following would be the approach to adopt:

- a) Use a sample of representative items from the whole database;
- b) Maintain the same weighting scheme for all months of a given year;
- c) Use the same scanner database for re-sampling of product offers and to help derive the weights in the following year;
- d) Use information from the same database to perform research activities, namely monitoring CPI/HICP quality standards, substitution bias, and production of alternative indicators (e.g. a COLI);

One of the key issues in using SD in HICP compilation is compliance to collection and compilation methodologies of the HICP.

The conclusions drawn from a comparison between the legislation and the possibilities of using SD point to a reasonable level of compatibility for the main HICP/CPI indices and according to the main regulated issues the use of SD will not break the HICP comparability rule, subject of course, to some appropriate procedures being put in place.

Annex 1 identifies the main issues that could bring serious problems of non-comparability and/or non-harmonisation (i.e. that could lead to systematic differences of more than one tenth of one percentage point on average over one year).

- a) Price concept;
- b) Production of results, the “unit value” issues;
- c) Timeliness – Risks related to delays from a retail chain;
- d) Expenditures incurred by business, by Governments or by non-profit institutions;
- e) Second-hand goods inclusion;
- f) Coverage of services;
- g) Tariff prices;
- h) Treatment of insurance;
- i) Treatment of products in the health, education and social protection;
- j) Timing of entering purchase prices
- k) Discounts treatment
- l) Changes on item quality;
- m) Treatment of service charges proportional to transaction values.

These can be addressed by:

- a) Amending the current HICP protocols (provided it does not jeopardize the HICP target measure and compromise the aim of the index) and/or;
- b) Designing and fine-tuning the SD input so that variables are conceptually closer in line with the objective of measuring inflation.

According to article 7 of Council Regulation 2494/95, all statistical units are required to cooperate in the collection or provision of price data and shall be obliged to allow observation of the prices actually charged and to give honest and complete information at the time it is requested. In this context, it could be argued that retailers are obliged to supply SD data at micro-level, although in fact the regulation refers to “allow observation” that might be different from requesting data with a specific format and variables. Also it does not necessarily oblige them to give information in a standard format provided by Eurostat. It is interesting to note that many countries have specific regulations for this purpose.

Comparability at the level required for the HICP can be achieved if some work regarding the harmonisation of processes is developed by 1) amending some regulations e.g. (price concept definition) or 2) by standardising across EU countries the definitions and format for using scanner data. The former is likely to be the easiest to execute but could involve some methodological compromises. It is important to note that any changes proposed would affect comparability with those countries which are not using scanner data. Any changes to regulations would need to ensure that comparability is not questioned for countries that are at different stages in implementing SD or are using different kinds of data sources / collection methods etc.

6. Compliance of Scanner Data in terms of the PPP regulation

Regulation No 1445/2007 of the European Parliament and of the Council setting up common rules for the provision of basic information on Purchasing Power Parities and for their calculation and dissemination, establish in Point 1 of article 2, the minimum frequency for data collection. If SD is used, the frequency does not have to be constrained to this standard considering that the increase in the outlay on processing may be at marginal.

Point 2 of the same article, which addresses to the concept of a PPP average price, will significantly benefit from the use of SD. Temporal adjustment factors mentioned in (i) of article 3, which addresses to the PPP definitions and concepts, and Point 3 of the Annex I describing “National Average Price” will not be considered necessary since PPPs can rely on real annual data. The same for spatial adjustment factor described at (j) and Point 4, given that SD provides prices for the whole country. Although there are outlet and geographic coverage limitations by SD (see 3.3 a). Through the use of the same information the consistency between spatial and temporal dimensions (PPP and HICPI) is improved.

Representative items referred at (l) of article 3 of the Annex I are easier to identify from the transacted volume information that SD can provide.

Article 3 (c) describes the price concept of PPPs which is different from that used in the HICP which refers to prices of products available to purchase i.e. includes non-discriminatory discounts and sales prices, in the PPP regulation the price concept is the purchase price paid by consumer, excluding sales prices such prices can obtained from SD transactions. Discounts whatever their nature (quantities or price reductions, universal or specific groups) are not to be considered in PPPs computation. SD can comply with this requirements regarding that the price discounts are easily identified through the procedures explain in this text.

Gains in timeliness obtained from SD, as a result of plugging into readily available data collected by outlets electronically at the time of a purchase, can facilitate the earlier publication of PPPs thereby reducing the 36 months delay referred in article 9.

New forms of provisional data are foreseen in 1.3 and 2.2 of Annex I, thus no issue arises from the fact that EUROSTAT may establish the use SD instead of traditional price collections.

The required pattern of consumption (mentioned at 5.1.9 of Annex I) is likely to be more accurate when computed from complete transaction data⁷.

Consumer price survey procedures, as prescribed by regulation and guidelines (see Annex I), would need to be replaced with new procedures / guidelines, when using SD⁸.

The need for validation procedures by Member States, and reporting assessment, remains and may vary according to the level of data manipulation/aggregation that is required from NSI's.

SD as a source of data for PPPs may overwhelm the sampling dimension issue. A huge amount of data for all countries could be available at a EUROSTAT data centre warehouse allowing for the inclusion of more items, stores and geographic regions, increasing the coverage and reducing the bias associated with estimations based on few data entries, assuming EAN numbers are standard and relate to exactly the same product with exactly the same features (this would need to be tested, especially for electronic goods).

PPPs require a more exhaustive and precise product specification and this goal can also be facilitated with SD collection. In most cases EANs in different countries are the same for the same item. When this is the case EUROSTAT could easily match equal items in different countries. In other cases when the same products have different EANs in different countries, as long as items specifications for each EAN are available, EUROSTAT can perform the matching relying on characteristics specifications.

At present each price item is collected once every three years. The prices not observed in a particular year are estimated using consumer price index data. Since CPI/HICP basket is not exactly the same as the PPP basket, this raises consistency problems. The use of a huge database provided by SD could overcome these issues.

7. Scanner Data and Modernisation

Using scanner data to compute price statistics can be in accordance with the present organisational model, whereby Member State's compute their own HICP. Alternatively, a new organisational model can be constructed based on the idea of a central data warehouse managed by Eurostat. In theory, the latter option leads to a choice between two ways for data transmission:

7) Only for those outlet-types represented by SD.

8) Only for some outlets types – SD would need to be supplemented by prices data from other sources and for some COICOP groups such as many services.

1. Statistical units provide micro-data to Member States Statistical Offices, that perform most or all of the index computation procedures and transmit final outputs (macrodata) to Eurostat, together with any micro-data required by Eurostat for quality assurance purposes;
2. Statistical units provide micro-data to Member States, who perform some manipulation of the data, including basic quality assurance, prior to transmitting it to Eurostat to compute the HICPs for each Member State and aggregates across countries.
3. Statistical Units submitting SD for individual Member States to Eurostat, which then cleans the data and computes the all the relevant indices.

In reality (2) and (3) are not seen as feasible options. Both Options would raise fundamental issues of governance and Option 3, additionally, would over-burden Eurostat with data, which the individual Member States are in a much better position to handle.

OPERATIONAL OPTIONS

	PPP	HICP
ADVANTAGES	Higher collection frequency; Improved geographic coverage and temporal adjustment; Efficiency gains from the data and computation centralisation; Representative items identification; Timeliness; Pattern of consumption.	Efficiency gains in the tools for the data treatment and index compilation.
DISADVANTAGES	SD will not cover all outlet types especially markets which means some tradition collection will still be required	Index must be released monthly two weeks after the reference date. Faced to this restrict deadline there is a risk of delaying the release of the index when this huge process is centralised; Quality adjustments and substitutions require a precise knowledge regarding each country market place and economic reality, that NSIs experts have more access to than EUROSTAT experts SD will not cover all outlet types especially markets which means some tradition collection will still be required
	Many services not covered Statistical Units may not be willing to transmit data to EUROSTAT	

Regarding the possibility of PPP computation being centralised in Eurostat, the potential advantages may outweigh the potential disadvantages. The same may not be true for the HICP. At present it is not feasible to centralise HICP computation as the HICP is often produced as part of national CPI production system, the uses, and compilation procedures of which, are not harmonised. Further work and experiences based on pilot exercises should be developed in order to explore a feasible solution.

8. Recommendations and Conclusions

For the time being and in order to overcome incomplete basket, outlet and geographic coverage a combination of SD with traditional methods of price collection is the most appropriate approach.

Further work is recommended in developing the **benchmarking** of SD and HICP results over a period of time (1 to 2 years, say) in order to review compliance and for determining the effectiveness of some HICP regulations/guidelines. For example, for the same periods where HICP prices are collected, what would be the results on measured inflation if, instead of a discrete sampling of HICP price data as per current practices, scanner data were taken or sampled on a continuous basis to reflect prices and expenditures corresponding to actual transactions over a temporal period of, say 4 weeks. Checking differences in elementary indices between one approach and the other would isolate, for example, the following errors and discrepancies:

- a) Errors in the prices collected.
- b) Errors of data entry.
- c) Bias due to discrete instead of continuous sampling.

For such a use to be effective a set of control procedures would need to be applied, namely the use of a representative subset of retailers and items. It would also be important to ensure that the recorded transaction prices in scanner data conform to the HICP definition of a “transaction” price and overtime to the principle of a fixed basket i.e. the identical products in terms of all its characteristics are priced each month or an appropriate replacement strategy is in place with quality adjustment methods that comply with HICP regulations and guidelines. This kind of test could be developed as many times as needed for the same or different sample subsets.

The possible improvements achieved with the use of SD depend on requesting the appropriate data in the appropriate format. Special care must be taken regarding the data characteristics requested from the provider, as these will determine the compliance or non-compliance with the Laspeyres concept, HICP regulations, comparability, accuracy, timeliness and efficiency.

Consumers and retailers in scope are not always households; some are companies or social institutions, etc. Even though these do not constitute any constraint regarding price movements but at the end of the year for weights and basket update these expenses must be excluded. To perform this task it is possible to ask for a list quantities of products sold to consumers with those types of VAT numbers or according to each country fiscal and economic scenario other solutions can be found.

Although SD can provide data on quantities sold, allowing for a continuous basket and weights updating, compliance to the **fixed basket** approach must be retained.

The annual weights and basket revisions requires requesting to the providers once a year all data on transactions at elementary In order to solve issues related to IT resources, price lists (used by POS) can be used, disregarding SD for weekly/monthly collection and use the SD only annually for the updating. Another option that is compliant with a saving in IT resources is, to use the same procedure as in traditional collection methods, i.e. request SD more

frequently for a list of Items with high volatility prices, and less frequent SD for another list of items where prices are more stable.

Several countries reported the use of a unit price computed from data on volume and quantities. In order to comply with HICP guidelines, a **geometric mean** must be computed (see page 7) and issues relating to the reimbursement for goods returned and for items being supplied free with other purchases (e.g. free DVDs with a DVD player) would need to be addressed.

EUROSTAT can gather much of the data required to produce PPP's, HICP and MPS but, as well as data-overload, would face problems such as a lack of outlet coverage across countries and the lack of standardised MS databases both in structure and contents. Country reports identified a constraint to produce MPS and to constitute and maintain a micro data warehouse since the data is provided in different formats. In the light of these findings no change in organisational model would be envisaged with the take-up of scanner data.

Problems related with **EAN's compatibility with COICOP** could benefit the most from the development, and application of an automated process for mapping EAN identifiers to the appropriate COICOP level classification. The development and application of a software tool to match descriptive characteristics is possible. These could also facilitate the achievement of a micro data warehouse managed by EUROSTAT and fulfil the aims of statistic modernisation. Additionally, **sharing** between NSIs, EAN to COICOP mapping lists could speed up and reduce overall costs for classification. The automation process is in line with the statistics modernisation process, supported by a European data centre warehouse.

Due to their specific characteristics, PPPs is the price index that can mostly benefit from the use of SD and a central data warehouse. Although centralised HICP production may also be a possibility that requires further research work based namely in dedicated pilot studies.

The NSIs and the suppliers of SD need to ensure a secure and speedy supply of data. Difficulty in the supply of price data for the computation of a monthly HICP to a tight timetable is a risk. One way of mitigating the inherent **operational risks** associated with scanner data is for the index compiler to use several sources/suppliers. This will spread the risk of failure in data transmission maintaining an acceptable coverage of market transactions. It is not replacement, is assuring that enough data sources are available to allow missing data estimation according to HICP standards.

Notwithstanding the potential benefits from the use of SD in HICP, PPP and MPS computation, its limitations (see 3.3.) should not be disregarded. These limitations require that traditional price collection will have to be a practice along with SD collection. Costs associated with the use of SD have to be considered. These will depend on the options taken relating to the manipulation of a huge amount of data and the possible need of IT tools, as well as the prices charged by data providers to supply the relevant data in usable form.

9. Bibliography

- Barcellan, R. (2011). Multipurpose Price Statistics. 13th Meeting of the International Working Group on Price Indices. Copenhagen.
- Bradley, R., B. Cook, et al. (1997). An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI. Third Meeting of the International Working Group on Price Indices. Copenhagen.
- Dalén, J. (1997). Experiments with Swedish Scanner Data. Third Meeting of the International Working Group on Price Indices. Copenhagen.
- FAIVRE, S. Would scanner data improve the French CPI?, INSEE, Consumer Price Statistics Division.
- Fenwick, D. and A. Ball (2001). Sampling in Consumer Price Indices: what role for scanner data? Sixth Meeting of the International Working Group on Price Indices. Canberra.
- Grient, H. v. d. and J. d. Haan (2010). The use of supermarket scanner data in the Dutch CPI, Statistics Netherlands.
- Grient, H. v. d. and J. d. Haan (2011). Scanner data Price Indexes: The “Dutch Method” versus Rolling Year GEKS. Twelfth Meeting of the International Working Group on Price Indices. Wellington, Statistics Netherlands.
- GS1 (2006). An Introduction to the Global Trade Item Number (GTIN).
- Haan, J. d. and E. Opperdoes (1997). Estimation of the Coffee Price Index Using Scanner Data: the Choice of the Micro Index. Third Meeting of the International Working Group on Price Indices. Copenhagen.
- Hawkes, W. J. (1997). Reconciliation of Consumer Price Index Trends with Corresponding Trends in Average Prices for Quasi-Homogeneous Goods using Scanning Data. Third Meeting of the International Working Group on Price Indices. Copenhagen.
- [https://en.wikipedia.org/wiki/International_Article_Number_\(EAN\)](https://en.wikipedia.org/wiki/International_Article_Number_(EAN)).
- Jain, M. and R. Abello (2001). Construction of Price Indexes and Exploration of Biases Using Scanner Data. Sixth Meeting of the International Working Group on Price Indices. Canberra.
- Jain, M. and J. Caddy (2001). Using Scanner Data to Explore Unit Value Indexes. Sixth Meeting of the International Working Group on Price Indices. Canberra.
- James, S. and R. Campbell (2012). UK Scanner Data Project – Obtaining Scanner Data. Workshop on Scanner Data for HICP. O. f. N. S. Prices Division, United Kingdom. Stockholm.
- Johansen, I. and R. Nygaard (2011). Dealing with bias in the Norwegian superlative price index of food and non-alcoholic beverages. Twelfth Meeting of the International Working Group on Price Indices. New Zealand.
- Johansen, I. and R. Nygaard (2012). Various data collection methods in the Norwegian CPI. Consumer Price Indices. UNECE/ILOs. Geneva.
- Mikkelsen, J. (2012). Extending the Danish CPI with scanner data - A stepwise analysis. Prices and Consumption, Statistics Denmark.
- Müller, R., H. M. Herren, et al. (2006). Recent developments in the Swiss CPI: scanner data, telecommunications and health price collection. Ninth Meeting of the International Working Group on Price Indices. Suisse. London.
- Netherlands, C. (2011). Price indices from Scanner data Frequently asked questions. Working Group 'Harmonisation of Consumer Price Indices', Luxembourg.
- Norberg, A., M. Sammar, et al. (2010). A Study on Scanner Data in the Swedish Consumer Price Index. Twelfth Meeting of the International Working Group on Price Indices. Wellington.

- Nygaard, R. (2010). Chain Drift in a Monthly Chained Superlative Price Index. Workshop on scanner data. UNECE/ILOs. Geneva.
- Okamoto, M. and T. Sato (2001). Comparison of Hedonic Method and Matched Models Method Using Scanner Data: The Case of PCs, TVs and Digital Cameras. Sixth Meeting of the International Working Group on Price Indices. S. B. o. Japan. Canberra.
- Ribe, M. (2011). On the RGEKS price index formula for scanner data. Twelfth Meeting of the International Working Group on Price Indices. Wellington.
- Richardson, D. H. (2000). Scanner Indexes for the CPI. Sixth Meeting of the International Working Group on Price Indices. Canberra.
- Rodriguez, J. and F. Haraldsen (2006). The use of scanner data in the Norwegian CPI: The «new» index for food and non-alcoholic beverages. Economic Survey.
- Triplett, J. E., Ed. (2003). Using Scanner Data in Consumer Price Indexes. Some Neglected Conceptual Considerations, University of Chicago Press.

Annex 1 – Practices in compilation procedures with SD and the HICP regulations

The following table compiles HICP legal framework and match its compliance with scanner data features.

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
Council Reg. No. 2494/95	HICP	Art. 1	Aim	not applicable			
		Art. 2	Definitions	not applicable			
		Art. 3	Scope			X	SD compares the prices of the products bought instead of comparing the prices of the products offered (transaction vs. prices of products available to purchase).
		Art. 4	Comparability			X	Comparability is unlikely to be fully achieved unless all Member States compile their HICPs using similar scanner data identified by EAN codes.
		Art. 5	Timetable and derogations therefrom	not applicable			
		Art. 6	Basic information	X			As long as comparability is not affected other sources may be used
		Art. 7	Sources		X		Statistical units are obliged to allow price observation and not to transmit data.
		Art. 8	Frequency	X			SD has the potential to be used both for constructing (un-weighted) elementary indices and for computing weights and also for updating the weights of the relevant elementary indices on a regular basis, with a minimum time-lag. Frequency of updating can be established on a store/store chain basis.

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
		Art. 9	Production of results			X	Indices compiled from scanner data are usually “unit value” indices where there is not a one-to-one relationship between the numerator and the denominator of the price-relative i.e. not matched pairs. However, this problem does not occur if the assessment is made to a transaction level data.
		Art. 10	Transmission of results		X		Delays may happen if one of the sources fails.
		Art. 11	Publication				
		Art. 12	Comparability of Data				
		Art. 13	Funding	not applicable			
		Art. 14	Procedure	not applicable			
		Art. 15	Review	X			

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
Commission Reg. No. 1749/96, art. 2 and art. 5 amended by Commission Reg. No. 1687/98, 1688/98 and 1334/07.	Initial implementing measures	Art. 1	Aim				not applicable

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
		Art. 2	Definitions (1334)			X	Will include some business expenditure and expenditure incurred by Governments or non-profit institutions on goods or services provided to households free or at prices that are not economically significant such as social transfers in kind. Supplemented data collection will be required for the purchases of second-hand goods. Coverage of services is likely to be limited. Scanner data facilitates compliance with the definition of newly significant goods. If unit values are extracted then these may contain 'conditional offers'. If ticket prices are extracted then scanner data is compliant, but ticket prices might not include some legitimate discounts or other offers e.g. three for the price of two, extra 20% free (i.e. special larger tin of baked beans – the EAN code may not change.
		Art.2 (a)	Principles (1334)		X		SD takes into account all the data or sample statistics; consumption segments are easily identified; quality change is not always considered.
		Art. 3	Initial coverage (1687+1688)	X			Coverage could be enhanced on the product domain although coverage of services by SD is likely to be limited.
		Art. 4	Newly significant goods and services	X			Will be easier to identify expenses in newly significant goods.

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
		Art. 5	Minimum standards for procedures of quality adjustment (1334)			X	One of the difficulties when using SD is to access detailed information about the product characteristics, hence increasing the difficulties on performing quality adjustments. Depending on specific circumstances replacements and quality adjustments are in general less precise.
		Art. 6	Minimum standards for prices	X			It is possible to comply with these standards.
		Art. 7	Price indices for elementary aggregates	X			Such ratios can be calculated <u>but</u> the populations covered in the numerator & denominator will differ because of the lack of individual price quotes.
		Art. 8	Minimum standards for sampling	X			SD allows for reduction in sampling errors.
		Art. 9	Quality control	X			keeping record of archives and decisions made during the production process in order to be provided to Eurostat at request.
Commission Reg. No. 2214/96, art. 2, art. 3 and art.4 amended by Commission Reg. No. 119/13, 1749/99 and 1708/05	Transmission and dissemination of sub-indices	Art. 1	Aim				
		Art. 2	Definitions	X			
		Art. 3	Production and transmission of sub-indices	X			Weights for indices and sub-indices to be transmitted are annually revised. Values including and net VAT can be transmitted in two different data bases.
		Art. 4	Quality control	X			
Commission Reg. No. 2454/97	Repealed by Com. reg. N. 1114/10						

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
Commission Reg. No. 2646/98	Treatment of tariffs	Art. 1	Aim			X	It is unlikely that SD could be used as source for Tariff prices.
		Art. 2	Definitions			X	
		Art. 3	Basic information			X	
		Art. 4	Data sources			X	
		Art. 5	Procedure			X	
		Art. 6	Comparability			X	
		Art. 7	Quality control			X	
Commission Reg. No. 1617/99	Treatment of insurance	Art. 1	Aim			X	Faced to the complexity in identification of the product detailed characteristics SD is not a feasible data source.
		Art. 2	Definitions			X	
		Art. 3	Treatment of insurance weights			X	
		Art. 4	Treatment of insurance prices			X	
		Art. 5	Comparability			X	
		Art. 6	Quality control			X	
Council Reg. No. 2166/99	Treatment of products in the health, education and social protection	Art. 1	Aim	not applicable			Some complexity in definition of product. The prices net of reimbursements and payments of claims are available. Small providers are not considered in SD. Not all the requirements can be fulfilled. Unless if products priced zero are included in the SD, when a product raises its price from zero to some amount it will be treated as a new product instead of a price raising.
		Art. 2	Definitions	X			
		Art. 3	Coverage			X	
		Art. 4	Prices			X	

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
		Art. 5	Basic information			X	Not all purchaser prices are available in SD since several small providers are not covered.
		Art. 6	Data sources			X	
		Art. 7	Comparability	not applicable			
		Art. 8	Quality control	X			
		Art. 9	Implementation	not applicable			
		Art. 1	Aim				
Commission Reg. No. 2601/00	Timing of entering purchase prices	Art. 2	Timing		X		Scanner data will not always indicate when the consumption of a service can commence e.g. an air ticket, theatre ticket. The use of scanner data will have to be restricted to goods.
		Art. 3	Implementation	not applicable			
		Art. 1	Aim				
Commission Reg. No. 2602/00 art. 6 amended by Commission Reg. No. 1921/01	Treatment of price reductions	Art. 2	Purchaser prices	X			SD provides purchased prices as they are obtained from shops registers.
		Art. 3	Inducements			X	It is very unlikely that these will be transparent in the scanner data as markers or different product codes are not usually given.
		Art. 4	Specification change		X		Specifications are not clear it may be implemented by change in codes.
		Art. 5	Implementation	not applicable			
		Art. 6	Revision	not applicable			
		Art. 1	Aim				

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
Commission Reg. No. 1920/01	Treatment of service charges proportional to transaction values	Art. 1	Aim	not applicable			
		Art. 2	Definitions	X			Electronic data transmission could comply with this.
		Art. 3	Treatment of service charges	X			Electronic data transmission could comply with this.
		Art. 4	Financial services n.e.c.		X		Electronic data transmission could comply with this for some types of service charge. However many of these services are not in any case processed in electronic format.
		Art. 5	Other services n.e.c.		X		Electronic data transmission could comply with this for some types of service charge. However many of these services are not in any case processed in electronic format.
		Art. 6	Basic information		X		It is not possible for all cases to obtain information at the necessary level.
		Art. 7	Comparability		X		Comparability is possible if electronic data is available according to the concepts of art 4 and 5.
		Art. 8	Quality control	X			
		Art. 9	Implementation	X			
Commission Reg. No. 1921/01	Revisions	Art. 1	Aim	not applicable			
		Art. 2	Definitions	X			Revisions can be carried out with the same rules and standards.
		Art. 3	Revisability	X			
		Art. 4	Mistakes	X			
		Art. 5	New or improved information	X			
		Art. 6	Changes in the system of harmonised rules	X			

Regulation	Issue	Article N.	Article title	Compliant			Comments	
				Fully	Partially	Not		
		Art. 7	Estimates of the impact	X				
		Art. 8	Release of revisions	X				
		Art. 9	Provisional results	X				
		Art. 10	Quality control	not applicable				
		Art. 11	Amendment	not applicable				
		Art. 12	Implementation	not applicable				
		Commission Reg. No. 1708/05	Reference period	Art. 1	Aim	not applicable		
Art. 2	Definitions			X			Index and price reference period are not affected. Weighting is at Com. Reg. N. 2454/97	
Art. 3	Implementation measures			not applicable				
Art. 4	Amendment			not applicable				
Council Reg. No. 701/06	Temporal coverage of price collection	Art. 1	Aim					
		Art. 2	Representation					
		Art. 3	Minimum standards for price collection	X			Scanner data can facilitate this.	
		Art. 4	Implementation					
Commission Reg. No. 1334/07	Amending							
Commission Reg. No. 330/09	Minimum standards for the treatment of seasonal products	Art. 1	Subject matter	not applicable				
		Art. 2	Definitions	X				
		Art. 3	Scope	X			EANS for seasonal products must be identified ex-ante.	
		Art. 4	Minimum standards	X			EANS for seasonal products must be identified ex-ante.	

Regulation	Issue	Article N.	Article title	Compliant			Comments
				Fully	Partially	Not	
Commission Reg. No. 1114/10	Quality of weightings	Art. 5	Comparability		X		Would be advisable the use of SD as a benchmark of HICP during a certain period of time to test if comparability remains unaffected.
		Art. 6	Application	not applicable			
		Art. 1	Subject matter	not applicable			
		Art. 2	Definitions	not applicable			
		Art. 3	Minimum standards for HICP weightings	X			Weights for HICP Index and sub-indices are annually revised as usually. National accounts data uses many sources of data to overcome reporting issues etc. Mention potential of incorporating scanner data to improve national accounts? For Elementary Product Groups SD can facilitate frequent revisions.
		Art. 4	Quality control	X			
		Art. 5	Application	not applicable			
		Art. 6	Repeal	not applicable			

Annex 2 – Comparative summary of the experiences made by NSI's using SD

Notes to the table:

Differences in description may mirror just differences in how countries have described what they do.

	DK	FR	NL	NO	SE	SW
Implementation status						
	Test phase	Test phase Daily SD will be collected from retailers. For a couple of years, this system should be used as a shadow system (with "traditional" data collection going on meanwhile), until it be assessed that CPI indices produced with SD meet all requirements to be incorporated in the official CPI	Using a combination of traditional methods and SD in HICP production with the first adopted model and testing a new model (RYGEKS)	Using in HICP current production	Using in HICP current production	Using in HICP current production
Objectives						
Usage of Scanner Data (SD)	Substitute the traditional CPI computation by SD based methods.	Integrate the SD on CPI as a price source / price collection method.	Partial substitution of fundamental indices by those processed using SD methodology.	Partial substitution of fundamental indices by those using SD methodology.	New price collection method to be used wherever convenient. A step by step process replacing traditional price collection by SD is undergoing.	New price collection method to be used wherever convenient. A step by step process replacing traditional price collection by SD is undergoing.

	DK	FR	NL	NO	SE	SW
Notes on the objectives	The variables collected lead to some deviation from HICP computation according to the regulations.	Maintaining a Laspeyres index with a yearly updated fixed basket in line with HICP regulations.	Increase in price quality with reduction of operation cost and respondent burden are the main drivers for the change. This objective was defined in order to implement a new methodology for CPI&HICP compilation which does not fully comply with the actual HICP legal framework.	Operation costs reduction (Paper questionnaires sent by regular mail) and modernisation; respondent burden reduction and raise on price quality are the main drivers for The change. Note that NSIs CPI is basically a COLI.	Some expectations on sampling error reduction with the use of SD. The method was only implemented for data collection purposes. The variables collected lead to some deviation from HICP computation according to the regulations.	Data collection was implementing in a way which does not compromise actual HICP legal framework.
Price framework						
Price concept	Aggregated turnover and quantity over a week. A unit value is derived as the variable for elementary index computation.	Price is defined as an average of transaction prices.	Aggregated turnover and quantity over a week. A unit value is derived as the variable for elementary index computation.	The “price” is the average price of the midweek for of the month. This concept is similar to a unit value.	The price concept is defined as a unit price taken on daily data. This concept can be considered as a proxy for actual transaction prices.	The price as an average price at transaction level.
Item selection (Fixed/Dynamic SD Basket)	Dynamic monthly changes - 4 months of selling and aggregated turnover representing 50% of COICOP 01 in each supermarket chain	Fixed basket - with regular overlap substitutions	Dynamic basket. A product stays on sample if between 2 consecutive month matching it has a share over a threshold (share taken from the matched pairs)	Replacements made using the matching method based on item characteristics for COICOP 01 and 02, but not used for COICOP 03.	Fixed basket	The sample is the same as for CPI; the chosen collected prices are the ones that show the higher quantities sold by the shop. This information is used for yearly re-sampling according to a chained Laspeyres-type price index.
Data provider	Supermarket Chain	Supermarket Chain	Supermarket Chain	Retail Chain	Supermarket Chain/Outlet (?)	Supermarket Chain/Retail Chain (?)

	DK	FR	NL	NO	SE	SW
Statistical Unit	Supermarket x EAN (item)	Supermarket x EAN (item)	Supermarket Chain x EAN	Retail outlet x(EAN+PLU)	Retail outlet x(EAN)	in store numbers
Variable type	Unit value for the week (weekly turnover/weekly quantity)	Unit price by transaction in each outlet (daily turnover/daily quantity) \approx transaction prices	Unit value for the week (week turnover/week quantity)	Unit value for the week	Transaction prices	Transaction prices
Variable acquired	Turnover and quantity for the week and EAN	Daily turnover and quantity by transaction	Turnover and quantity for the week and EAN	Turnover and quantity for the week and EAN	Average of daily transaction prices by outlet and EAN	Average of daily transaction prices by in store numbers
Collection Periodicity	Weekly unit values. Two per month on SD.	Weekly prices	Weekly unit values. Three for each month	Weekly unit values. One (mid month week) for each month	The first 3 full weeks are observed. Each price has a predefined week and a week day to be collected.	Monthly. Actual transaction price collection takes place during first 14 days of the month.
Sample Management						
Geographic scope	All country	All country	All country. Besides, a stratification "branch" was added jointly to CIOCOP. This allows for the inclusion of supermarket chains as they are considered as having nation-wide pricing policies.	All country. Stores grouped in retail chains. A single consumption profile over the country.	All country	It includes 11 regions and 5 supermarket chains (at same level)
Product scope	COICOP 01 Division - Food and non alcoholic beverages	COICOP 01 Division – Food and non alcoholic beverages	COICOP 01; 02; 05.5; 05.6.1; 06.1.1; 06.1.2; 09.3.4; 12.1.2/3	COICOP 01 Division - Food and non alcoholic beverages and progressive inclusion of Divisions 06 and 07	COICOP 01 Division - Food and non alcoholic beverages excluding perishable fruits, vegetables and meat and 02.2 Tobacco	01* Food and non-alcoholic beverages; 02.1* Alcoholic beverages; 05.6.1 Non-durable household goods; 09.3.4 Pets and related products; and 12.1.3 Other appliances,

	DK	FR	NL	NO	SE	SW
						articles and products for personal care.
Product Coverage	60% of the sales of COICOP Division 01	8 Product Families covering 1050 items (1306 observations). 55% of overall EANs corresponds to 72% of sales and 100 EANs covers 56% of sales.	All possible EANs on this field	All transactions are considered - accounting for 98% of the turnover of those chains covered, except for fish and meat that is found to be sold on supermarkets	84% of national consumption for the covered products and 45% of the total outlet turnover	
Supermarket/ Store Chains	1 supermarket chain	6 Supermarket chains (aggregate share of 30%, about 1000 units)	6 Store Retail chains	4 Retail chains (all of them) + Pharmacies chain + Petrol chain. In total 158 stores (from 3078) are capable to transmit SD for the CPI.	20 outlets from one retail chain	5 supermarket chains
Starting Basket	Based on representativeness with the following criteria: items must have been sold in the 4 previous months; the item basket adds up to 50% of the COICOP 8-digit turnover for each supermarket chain. This could put a problem with seasonal goods that are in the market less than 4 months.	Random sample on selected product groups proportional to sales.	The approach is similar to a "census". Basket renewal in a monthly basis with possible impact in the weighting scheme which differs from month to month at elementary level.	The approach is similar to a "census". Basket renewal in a monthly basis with possible impact in the weighting scheme which differs from month to month at elementary level.	Every Item showing in the base year less than 50% difference between the "actually observed" CPI price (traditional collection method) and the SD collected prices is retained.	Initial CPI/HICP standard basket

	DK	FR	NL	NO	SE	SW
Basket Maintenance	Monthly substitution. If no substitution available, procedures to treat missing values are applied.	Monthly EAN chain with automatic substitution.	Automatic matching with a threshold mechanism. Substitutions and imputation of missing prices. For each product it should be maintained 50% of the items (in the same retail chain) or assured 80% of the expenditure.	All Items found on supermarket sales are candidates to be included	The basket is static until next annual revision.	Monthly substitution for disappearing products. A product (in store number) is considered in a COICOP elementary item of the index using the turnover and quantity transacted.
Substitution process	Substitution of an EAN by an equivalent in the same supermarket. To select the replacing EANs the following criteria is used: the 3 EANs with the higher turnover for a COICOP 8 digit level complemented with those in order to cover at least 45% of the total turnover.	Automatic based on the "nearest EAN".	Not applicable (see starting basket).	Not applicable (see starting basket).	Do not see any option to adjust for missing values. Not see this as a big problem.	After a period of continuous missing values (up to 6 months) a substitution can be decided based on the turnover analysis.
Procedures for estimating missing prices	Same as for normal CPI/HICP	No estimated price. Price replacement by the "nearest EAN".	Based on chain variations of "similar" at product level (COICOP elementary index level).	Same as for normal CPI Missing are rare occurrences.	Using standard procedure. The rate of SD missing is higher than traditional collection method.	Missing prices are estimated according to the HICP legal framework.
Data cleaning	Not clear from the available texts if this procedure applies	Not clear from the available texts if this procedure applies	The prices (monthly unit values) are subject to two automatic data cleaning procedures. Firstly, month-to-month	Same method as Netherlands but base in factor 3. Thus, items for which the current price is 200% higher of 50% lower	Outliers' detection no decision on the method to remove them.	Clear distinction between transaction data sets and Item characterisation data set.

	DK	FR	NL	NO	SE	SW
			<p>price changes of a factor greater than 4 are considered implausible and declared invalid. Thus, items for which the current price is 300% higher of 75% lower than the price in the previous month will be deleted. dumping filter, has been developed to</p> <p>exclude items from the computation which exhibit a strong price decrease in</p> <p>combination with a strong decrease in expenditures (stock clearances)</p>	than the price in the previous month will be deleted.		
Indices						
Elementary index level	Standard Elementary COICOP Level - 8 digit level	Standard Elementary COICOP Level.	Equivalent to COICOP 6 digit level. The elementary level which is still supermarket-specific in most cases is comparable to the six digit COICOP level.	At a higher level than EAN, at which EAN constitute a homogeneous product. (COICOP 6 digit level)	Standard Elementary COICOP Level	Standard Elementary COICOP Level
Elementary Index Formulation	Weighted Jevons. The weighting at the store level is based on the supermarkets turnover	Jevons	Weighted Jevons Monthly chained index Turnover is used to aggregate supermarket chains	Monthly chained Törnqvist Expenditures taken from SD	Jevons	Jevons
Aggregate	Elementary indices at	Chained Laspeyres	Chained Laspeyres	Chained Laspeyres	Chained Laspeyres	Lowe

	DK	FR	NL	NO	SE	SW
Formulation of elementary indices	COICOP 8-digit level are aggregated based in weights derived from HBS.					
Aggregation Chaining granule	Yearly	Yearly	Monthly + Yearly	Monthly + Yearly	Yearly	Yearly
Other issues						
Discount Items	There is no distinction between non-volume discounts and ordinary sales price.	-	-	-	Different kind of discounts can be identified. The deposit fee cannot be deducted from the full price.	Manually possessed
Seasonal Items	No use of a fixed basket and a dynamic basket applies to seasonal items.	Seasonal food items were not selected for SD collection.	Dynamic basket applies to seasonal items.	Seasonal items are excluded	-	Few seasonal items were considered. The decision was to ignore them in SD transmission.
New Processes	Maintenance of a correspondence mapping between COICOP and EAN; Monthly evaluation of the sample representativeness (turnover based).	-	Data cleanup procedure: identification and deletion of unit-price outliers (monthly variation over 300% and below 75% and use of "Dumping Filter" to identify products that are being phased-out of market (low price+ low turnover).	Renewing of observed items	-	Pre-aggregations made by retailers. Substitutions similar to "dumping filter" similar to NL procedure.

	DK	FR	NL	NO	SE	SW
Some of the difficulties	<p>To build an association between COICOP products (CPI) and EANs or other item identifiers from Retailers;</p> <p>The large number of EAN items in the basket;</p> <p>Not being able to identify volume discounts.</p>	-	<p>EAN volatility;</p> <p>At EAN level no quality adjustments are performed using the present software version. The possibility to implement QA still exists for the new software.</p>	<p>Vulnerability to a price provider failure in reporting, because of the reduced number of providers.</p> <p>It was decided to maintain the possibility to use the traditional interview in case of SD failure.</p>	-	

Annex 3 –International Article Number (EAN)

An **EAN-13** barcode (originally **European Article Number**, but now renamed **International Article Number** even though the abbreviation **EAN** has been retained) is a 13 digit (12 data and 1 check) [barcoding](#) standard which is a [superset](#) of the original 12-digit [Universal Product Code](#) (UPC) system developed in the [United States](#). The EAN-13 barcode is defined by the standards organization [GS1](#).

The EAN-13 barcodes are used worldwide for marking products often sold at retail [point of sale](#). The numbers encoded in EAN-13 bar codes are product identification numbers, which are also called [Japanese Article Number](#) (**JAN**) in [Japan](#). All the numbers encoded in UPC and EAN barcodes are known as [Global Trade Item Numbers](#) (GTIN), and they can be encoded in other GS1 [barcodes](#).

The less commonly used **EAN-8** barcodes are used also for marking retail goods; however, they are usually reserved for smaller items, for example [confectionery](#).

2-digit ([EAN 2](#)) and 5-digit ([EAN 5](#)) supplemental barcodes may be added for a total of 14 or 17 data digits. These are generally used for periodicals (to indicate the current year's issue number), or books and weighed products like food (to indicate the manufacturer suggested retail price or MSRP), respectively.

Components

The **GTIN-13** encoded in the bar code has four components:

- The **GS1 Prefix**, the first three digits, usually identifying the national GS1 Member Organization to which the manufacturer is registered (not necessarily where the product is actually made).^[4] When the EAN-13 symbol encodes a conversion of an [ISBN](#), the GS1 Prefix will be either 978 or 979. Likewise the prefix will be 979 for [ISMNs](#) (whose prefix is shared with ISBN) and 977 for [ISSNs](#).
- The **Company number**, consisting of three to eight digits depending on number of GTIN-13s required by the manufacturer to identify different product lines (in ISBN and ISSN, this component is used to identify the language in which the publication was issued and managed by a trans-national agency covering several countries, or to identify the country where the legal deposits are made by a publisher registered with a national agency, and it is further subdivided any allocating sub-blocks for publishers; many countries have several prefixes allocated in the ISSN and ISBN registries).
- The **Item reference**, consisting of two to six digits (in ISBN and ISSN, it uniquely identifies the publication from the same publisher; it should be used and allocated by the registered publisher in order to avoid creating gaps; however it happens that a registered book or serial never gets published and sold).
- The **Check digit**, a single [checksum](#) digit. The check digit is computed modulo 10, where the weights in the checksum calculation alternate 3 and 1. In particular, since the weights are relatively prime to 10 the EAN system will detect all single digit errors. But, since the difference of consecutive weights is even, the EAN system does not detect all adjacent transposition errors.

The complete number is used as a reference key to look up information about the product line held on a database; the number is never normally broken down into its components within users' systems.

Prefixes⁹

Each country has one or several blocks of GS1 Company Prefixes to assign. Some special ranges exist. The usage for some is GS1 Member Organisation (GS1 MO) specific. The GS1 Company Prefixes that begin with the 3 digits listed below are used to construct what are considered RCN's (Restricted Circulation/Distribution Numbers) or are specific to a particular industry, such as the publishing industry:

020 – 029 Restricted distribution (GS1 MO defined)

040 – 049 Restricted distribution (GS1 MO defined) N

050 – 059 Coupons

200 – 299 Restricted distribution (GS1 MO defined)

977 Serial publications International Standard Serial Number (ISSN)

978 – 979 Bookland International Standard Book Number (ISBN)

9790 International Standard Music Number (ISMN)

980 Refund receipts

981 – 982 Common Currency Coupons

990 – 999 Coupons

Note from the Authors:

GS1 Prefixes reserved for distribution purposes correspond to the ones used by supermarket chains when make their own packages, own branded products or unbranded products.

The other interesting ranges are the coupon ranges, and the refund receipts range that may be used for the use of "subscriber discount" non universal discounts.

Sample:



GTIN-13 number encoded in EAN-13 barcode.

The first digit is always placed outside the symbol; additionally a right ">" indicator is used to indicate a "Quiet Zone" that is necessary for barcode scanners to work properly.

9) Extracted from Global Trade Item Number - Wikipedia, the free encyclopedia:
<https://en.wikipedia.org/wiki/GTIN>

Annex 4 – Roadmap to use data based on scanner data (SD).

Context

The context of this example is the regular production of the CPI/HICP mainly by using SD, in an efficient way and following the present HICP regulations. This can also apply to the other areas in price statistics (PPPs, Multipurpose Price Statistics).

Overview of major actions

1. Preliminary actions

Before contacting potential respondents (data providers) the team should:

- a. Have a clear picture of the aims and objectives;
- b. Build enough knowledge about both HICP regulations and scanner data related issues;
- c. Establish the boundaries for the request to or negotiate with the data providers (respondents)
- d. Define the criteria to determine the success of a trial.

2. Data Request/Negotiation phase

Each retail chain should be considered as an independent data source, regardless of providing data from a set of outlets.

Within the predefined boundaries an agreement on:

- a. Data content to be transferred,
- b. Data structure and transfer protocol,
- c. Timetable (what, when and with what frequency),
- d. Transfer media;
- e. Retail units to be included (a sample or exhaustive),
- f. Provision structure (data provided by each retail unit or by central chain),
- g. The conditions and timetable for the trial period, should

be obtained.

3. Trial actions

Regarding the provider:

- a. Regularity and timeliness in data transmission well-matched with CPI/HICP production timetable.
- b. Accordance to the established protocols (both structure and content).

Regarding the data processing:

- a. Put in place a method to guarantee product item classification.
- b. To put in place a "simulation" environment to compile a CPI/HICP based both in traditional data collection and scanner data.
- c. Evaluate missing data; possible substitutions; price estimates and other related aspects.
- d. Compare the compiled indices traditional data collection with those from scanner data.

Regarding internal resources:

- a. Evaluate IT needs for a regular production cycle.
- b. Evaluate and correct issues related to transmission storage and data integration of scanner data.
- c. Evaluate resources needed to maintain the EAN classification task.

Detailed aspects

In this section we made an exercise that intends to help and guide on the section 2.1 below No. 1, 3 and 4 and section 2.2, No. 1.

For the major actions we explore the pertinent topics showing the preferable (according to the legal regulation) and the acceptable options.

Theme: To be remembered			Recommended	Acceptable
For a monthly index compilation				
Price Concept				
Concept	Prices of products available to purchase (see note 1)		If it is not possible to collect lists of prices of products available to purchase (data provided to de POS devices), then, it is advisable to request "daily unit values" based (see notes 3 and 4). The turnover and quantities are provided by POS. Any average of these daily figures is acceptable (arithmetic, geometric, weekly or monthly aggregated)	
Collection Period	Specific collection day(s) for specific sets of item products. Preferably synchronized with traditional collection agenda.		If it is not possible to get filtered data on the specified periods, the advisable alternatives are averages: weekly average of daily prices of products available to purchase or weekly average of daily unit values (POS) (see note 5).	
Collection Frequencies	Monthly. Chose 3 or 4 weeks for price collection - 1 day for regular items; 3 days for fresh and vegetables products.		See alternatives for the collection period below	

Theme: To be remembered	Recommended	Acceptable
Product Item Subset	Should follow closely the COICOP coverage of HICP/CPI	Try other ways of filtering data (see note 6) otherwise try to get the whole set of items.
Name	Product item name for each item product.	
Classification Enumeration (see 7)	Target defined upon the HICP basket needs. Choice made under COICOP Classification (see note 2).	
Item Enumeration	Product item to be monthly collected on the specific outlet - derived from SD identifiers mapping to COICOP classification (a filter definition that the provider applies to price reporting to the NSI)	
Set of Characteristics - Description	If possible ask for product item descriptions	
Set of Identifiers	If used ask for other used complementary product item identifiers (e.g. SKU, PLU) (see note 8)	
Redundancy	Redundancy should be considered in order to allow for possible item substitution without asking for changes on filter definitions	
Price reductions	All kind of reductions should be reported Clearly inform provider that NSIs should be alerted for the presence of reductions.	
Typology	Exemplify with some typology (e.g. quantity, price, coupon,), Request agreement on the way the provider will inform about the reductions applied on the requested period.	
Data transfer	See 2.2	

Theme: To be remembered	Recommended	Acceptable
Structure		
Dimension		
Media		
For a yearly processing		
Turnover and quantities	Request total turnover and quantity sold during a yearly period - all products sold during the period should be in scope	At least the coverage should be extended in order to be able to assess possible basket changes.

Notes to the table:	
(1)	Each month, the meaningful price list subset is defined as the "list of all prices for which a POS transactions or positive stock quantity is found"
(2)	For a trial or for an evaluation of the new basket, it is recommended trying to get all items from some types of products (for instance: all-item products for COICOP Division 01,
(3)	The daily unit value corresponds to the ratio turnover / quantity sold for the specified item product during a day. This roughly corresponds to the item shelf price since it is assumed that retailers do not change item prices during the working day.
(4)	Note that the daily unit price option cannot cover all the selected prices of products available to purchase since only products effectively sold are covered.
(5)	It is assumed that if the provider cannot supply weekly averages; then it will not be able also to provide monthly averages.
(6)	If it is not acceptable for the provider to filter the requested item product in a detailed way, It is important to find an alternative acceptable way for filtering data in order to avoid excessive useless price data. For instance, the use of some retail chain internal product group classification could be very helpful for this purpose,
(7)	This is the targeted COICOP list for HICP or CPI. For the PPP the most important is the item list.
(8)	Access to other complementary identifiers or to internal classification schemes helps on achieving a higher degree of accuracy on the automated mapping process (EAN->COICOP)