

ТЫНЫЧЛЫК

Карауарáан

Isithangami

สันติภาพ

和平

Sülh

Vrede

Béke

שלום سلام

平和

Rauha

Friður

Pace

Peace

Fred

Paix

Paz

ပိပိဒုဝုလုဝံ



The primary objective of the workshop ...

... is to clarify the concepts of scanner data. A roadmap will be drawn up together with the participants, and recommendations on how to meet the challenges ahead will be formulated. The workshop will also address a number of important issues, such as:

- *Quality assurance;*
- *Confidentiality;*
- *Cooperation and agreements with data providers;*
- *Scanner data software.*



The Revolution of Scanner Data and the Challenges Ahead

Muhanad Sammar
Statistics Sweden
Economic Statistics Department, Price Statistics
muhanad.sammar@scb.se





Disposition

- Background
- Practice
- Ensuring data quality
- Other areas
- Challenges ahead



Why the Swedish Price Unit uses scanner data in the CPI

- Improved quality
 - Less volatile data (three weeks)
 - Bigger samples
- [A report by Swedish Consumer Agency](#)
 - For 9% of the items in the survey, the prices were hard to find or could not be found at all.
 - For 6% of the examined products, the prices on the shelves and packages were different from the purchased prices.
- Eliminate human errors

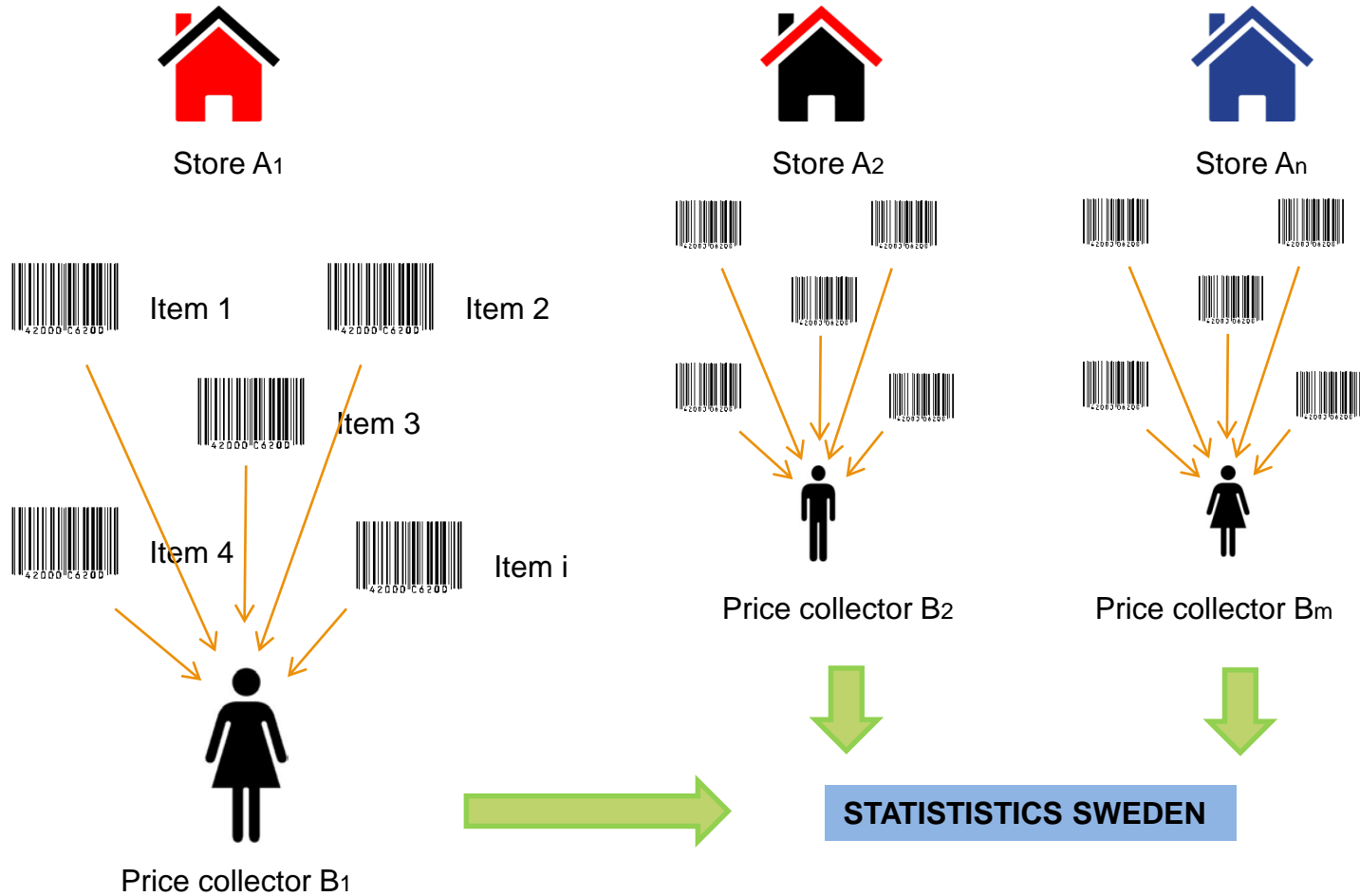


The Swedish Practice - before

- Manual price collection (observed prices)

Statistics Sweden

Statistiska centralbyrån



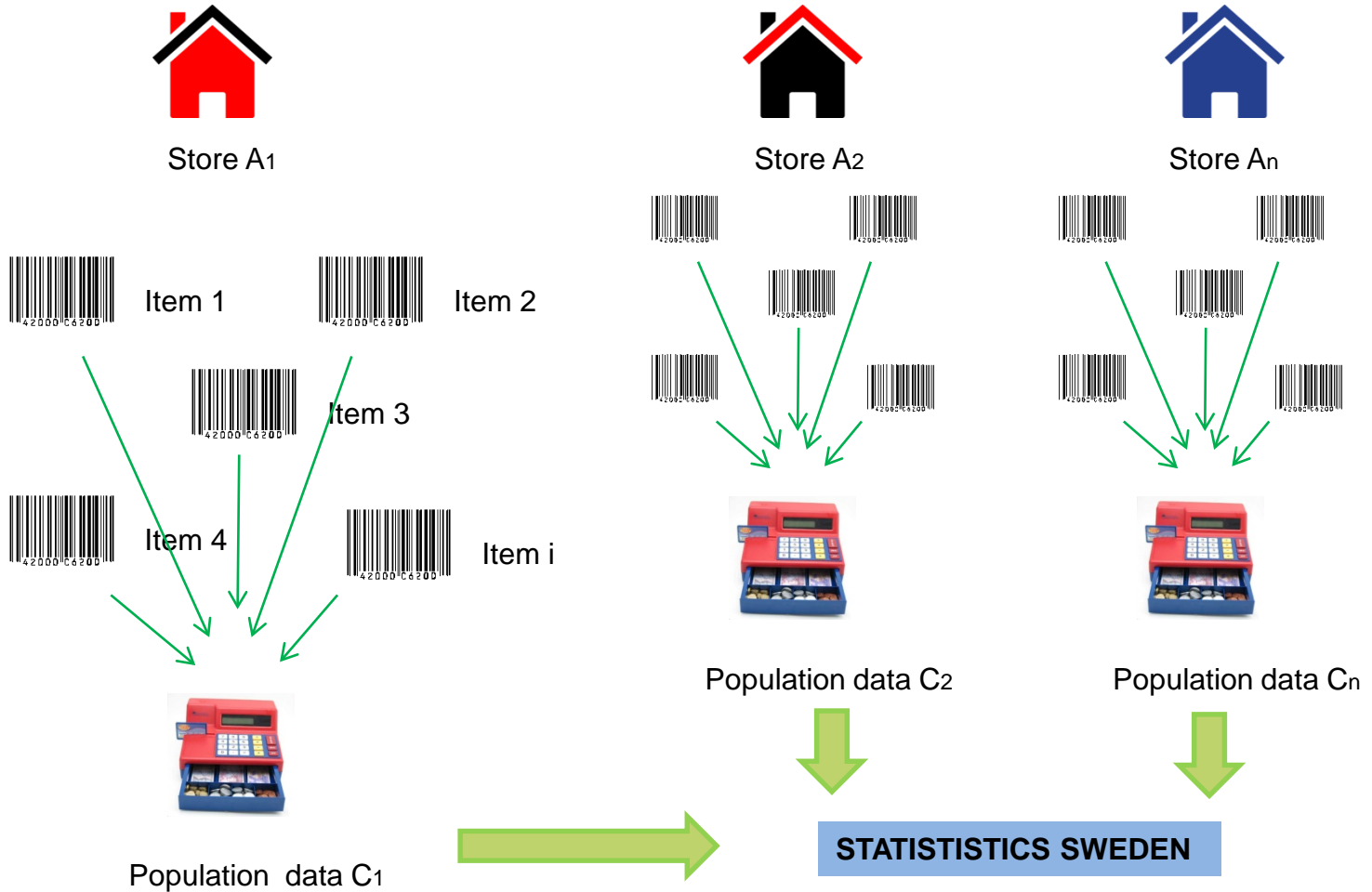


The Swedish Practice - after

- Transaction prices (weekly deliveries)

Statistics Sweden

Statistiska centralbyrån



Steps for ensuring data quality

- scanner data

1. Study how good the information is for each EAN-code. Is the information sufficient to identify a product by its EAN-code? If not, can the retailer provide you with more information (such as brand, package size and product description) just to facilitate the identification of the product?
2. Compare the prices between the manually collected data and the data collected from the retailers (Scanner data). If there are differences, then one should explain them.
3. Study how often EAN codes are reused. Can the same EAN code be used for two different products? Study how common that is.



Steps for ensuring data quality

- scanner data

4. Identify the different discounts that are available for each and every retailer. Which types of discounts are included in the price and which are not? Are deposit fees included in the price?
5. Analyse the scanner data by doing descriptive statistics analysis on the data.
6. Compare the index from scanner data with the index based on the manually collected data. How big is the difference? Try to explain why there are differences.
7. What else? – *Please fill in the questionnaire*



Other areas where scanner data or other big data can be used

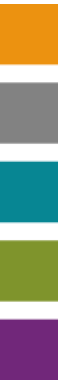
- Labour Force Survey
 - In year 2000: 15 % (non-response rate)
 - In July 2013: 32,5 % (non-response rate)
- Household Budget Survey
 - In year 2000: 48 % (non-response rate)
 - In July 2013: 60 % (non-response rate)
- National Accounts
- Retail trade statistics



Labour Force Survey (LFS)

- External factors

- Difficulties finding relevant telephone numbers
- Increasing reluctance towards answering among sampled units
- Increased competition between different survey organisations and telemarketing companies



Labour Force Survey (LFS)

- Internal factors

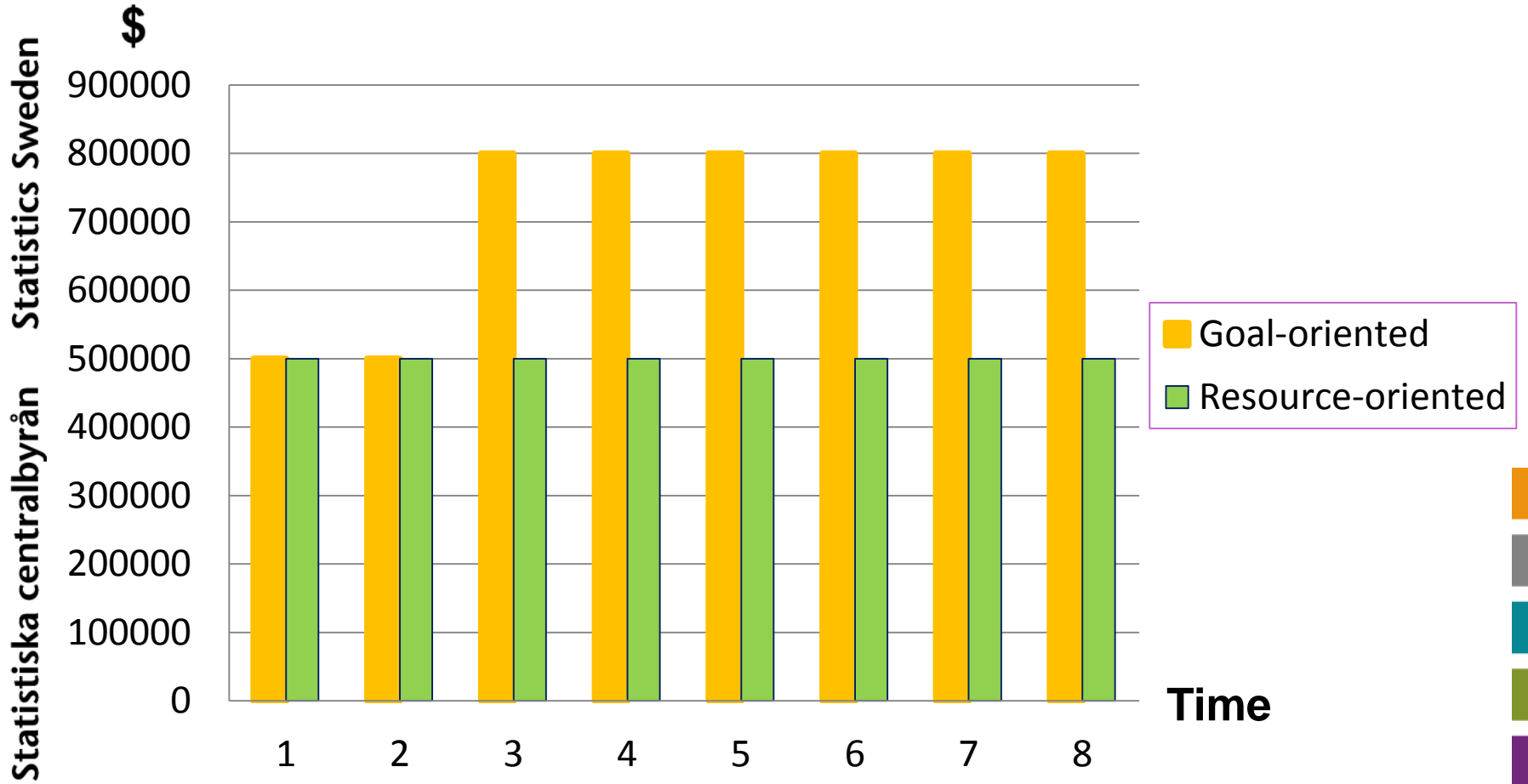
- Increasing workload at the Interview Unit
- Poor interviewer training
- Poor contact strategies
- Inefficient work procedures





Cost chart

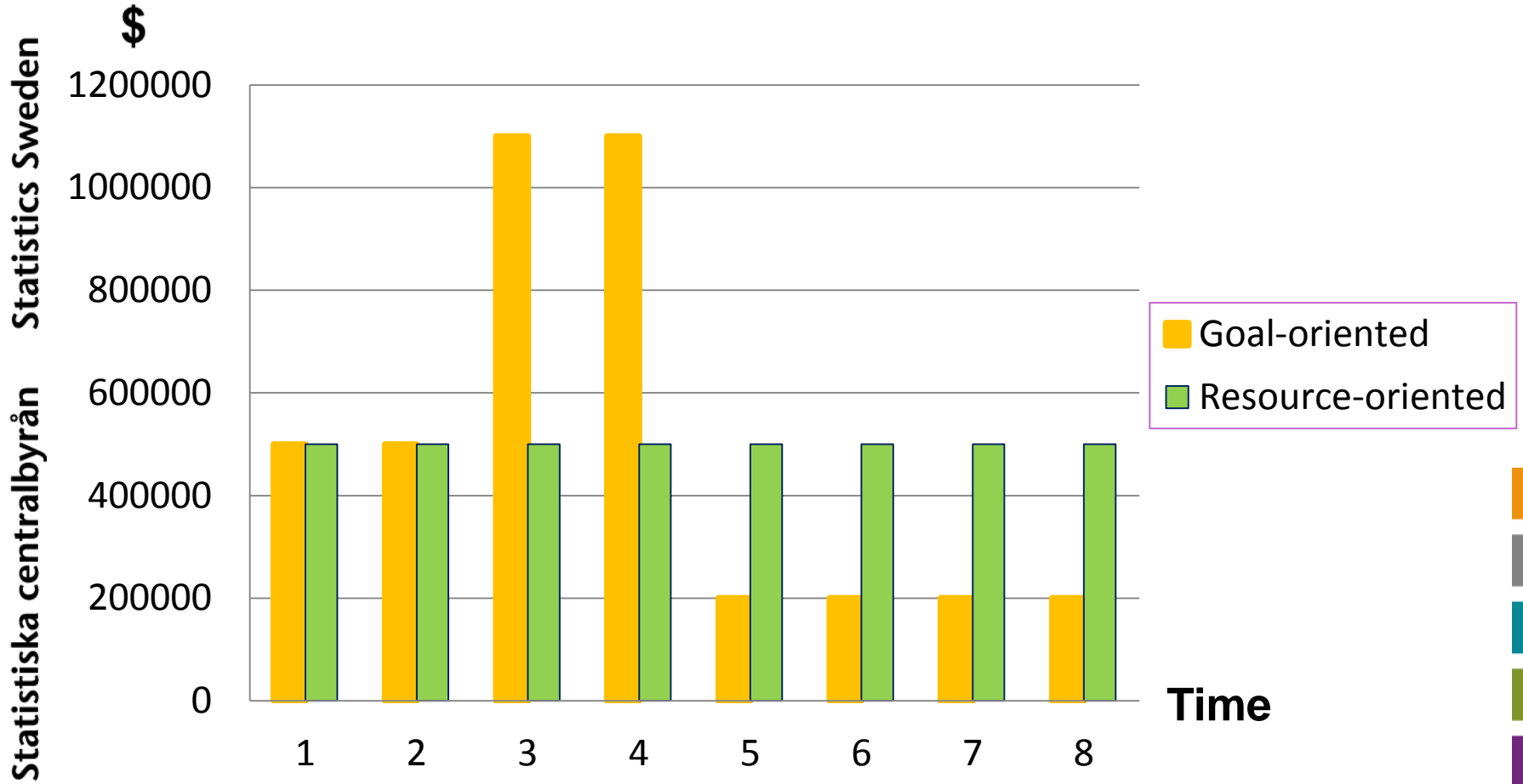
- the wrong aim or the wrong method





Cost chart

- the right aim or the right method



Scanner data as a tool?

- Hedonic regressions

Regression coefficients of the explanatory variables in the reference period (0) shall be defined by the formula below:

$$\ln(P_{jy}^0) = a_{jy} + \beta_{1jy}X_{1jy} + \beta_{2jy}X_{2jy} + \beta_{3jy}X_{3jy} + \dots + \beta_{njy}X_{njy} + \varepsilon_t$$

where,

P_{jy} = the transaction price of the product in the base period

a_{jy} = the coefficient of the constant

X_1 = first feature

X_2 = second feature

X_3 = the dummy variable for third feature (1=Yes, 0=No)

X_n = the dummy variable for n:th feature (1=Yes, 0=No)

ε_t = disturbance term



Scanner data as a tool?

- Hedonic regressions

The regression coefficients should then be used to estimate base prices, \hat{P}_{jy}^0 , for each product in the following way.

$$\hat{P}_{jy}^0 = \exp(\hat{\alpha}_{1jy} + \hat{\beta}_{1jy}X_{1jy} + \hat{\beta}_{2jy}X_{2jy} + \hat{\beta}_{3jy}X_{3jy} + \dots)$$



Challenges ahead – next step

- Institute new regulations that facilitate for NSI:s to get scanner data
- Build up a unified production system for scanner data processing (including data collection, data entry, data editing and data processing)
- Appoint a review panel
- Create a key between EAN and COICOP classifications
- Create a key between the CN8 and COICOP classifications
- What else? – *Please fill in the questionnaire*





What is the solution for the future?

- Scanner data?
- Might not be the optimal solution for all time, but for some surveys it is the most suitable method today!





Thank you for your attention!

If you want to learn more, please don't
hesitate to contact Statistics Sweden

muhanad.sammar@scb.se

