Historical Change in Data Source Distribution

Year	Scanner Data	Web Scraping	Field Price Collection
2021	6%		94%
2022	43%	5%	52%
2023	45%	5%	50%
2024	43%	5%	52%

■ SCANNER DATA

■ WEB SCRAPING

■ FIELD PRICE COLLECTION
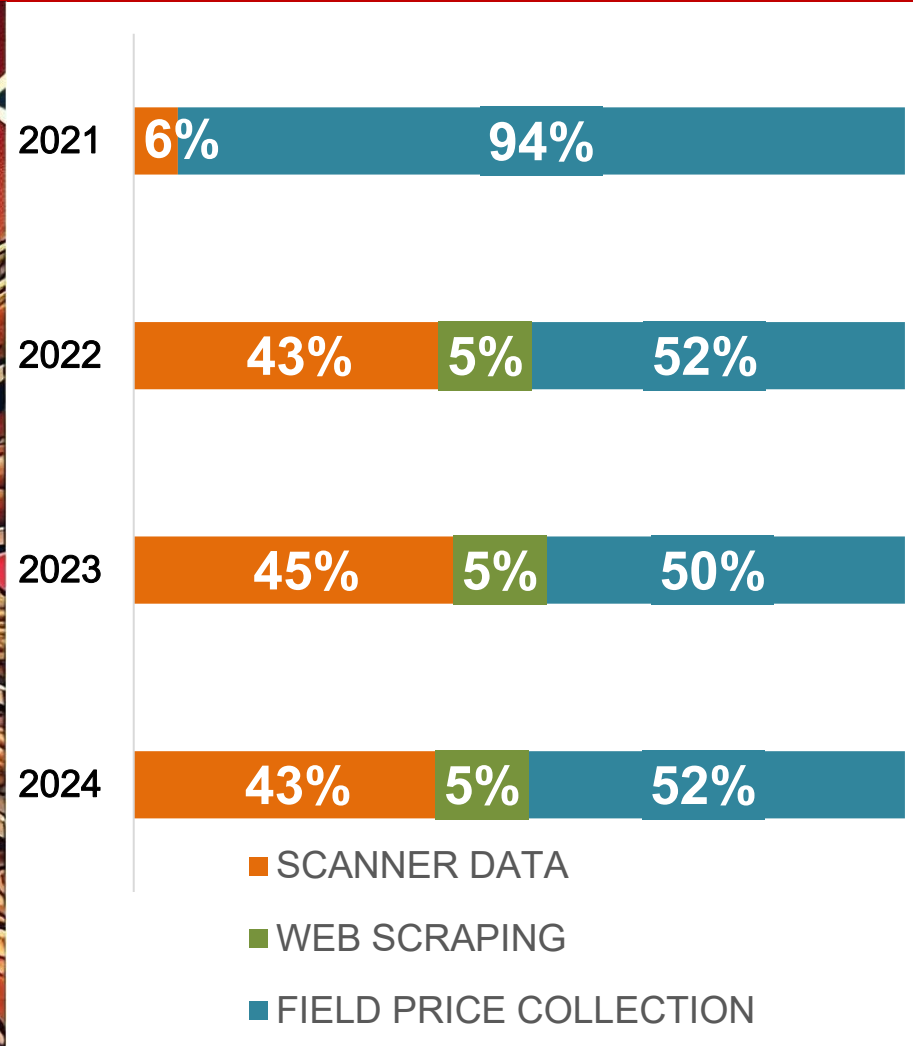
Data Sources used in Turkish CPI (2024)

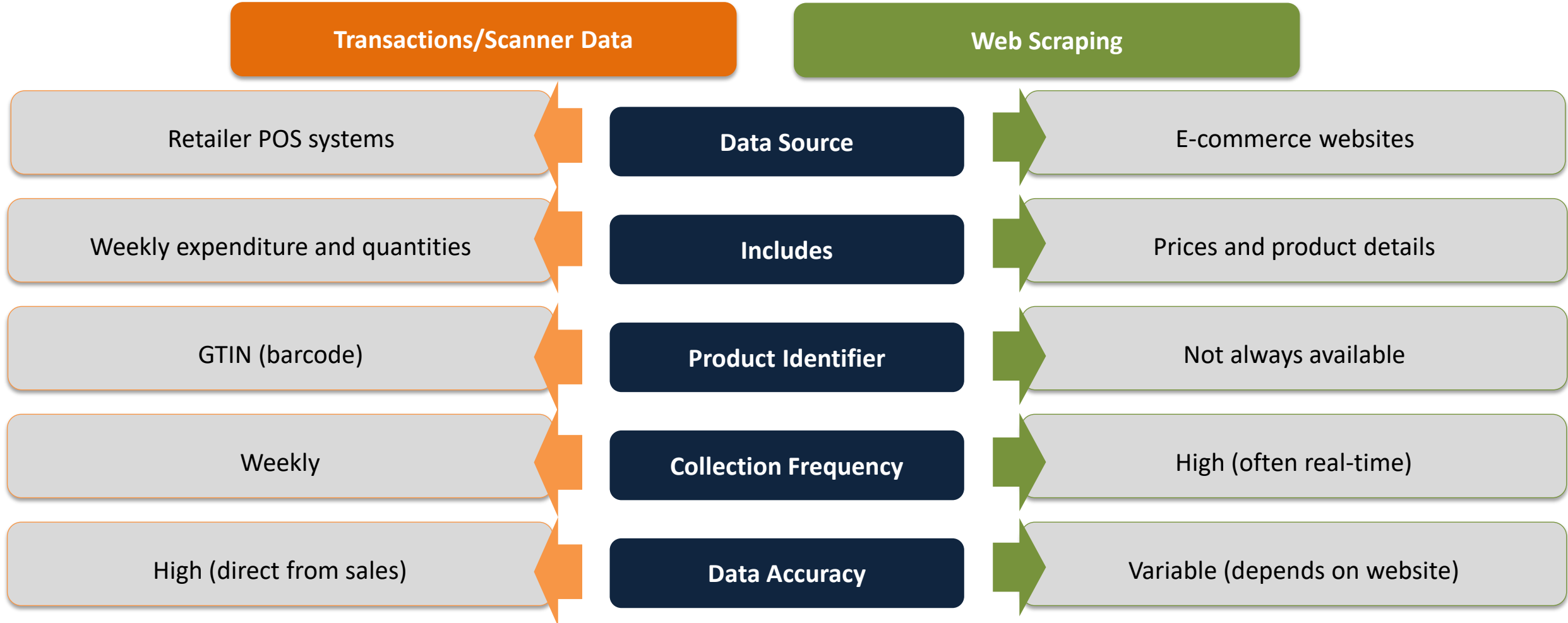| Year	Scanner Data	Web Scraping	Field Price Collection	Total Number of Prices Collected
2021	14.041	-	231.972	246.013
2022	104.968	12.472	126.649	244.089
2023	107.562	12.784	119.866	240.212
2024	109.165	13.017	134.166	256.348

Scanner Data: 43%

01
02
03
05
06
08
09
12

Web Scraping :
5%
03
05
07
08
09
12

Field Price Collection:
52%

All divisions

Scanner Data vs Web-Scraped Data

Transactions/Scanner Data		Web Scraping
Retailer POS systems	**Data Source**	E-commerce websites
Weekly expenditure and quantities	**Includes**	Prices and product details
GTIN (barcode)	**Product Identifier**	Not always available
Weekly	**Collection Frequency**	High (often real-time)
High (direct from sales)	**Data Accuracy**	Variable (depends on website)

Scanner Data Overview

- TurkStat started incorporating scanner data into its data collection methods in 2018.

- TurkStat has been using scanner data to compile CPI since 2021.

- Scanner data is collected from both national and local retail chains.

- Currently, scanner data contribute to the calculation of 322 sub-indices ("aggregate of products") belonging to eight COICOP Divisions (01, 02, 03, 05, 06, 08, 09, 12).

- Scanner data covers 1759 outlets from 33 retails chains.

- For sample selection, TurkStat uses a static approach, where the basket of items is updated each December and monitored throughout the following year.

Scanner Data Overview

- Scanner data is collected weekly using GTINs (barcodes)

- Data is requested cumulatively for weekly periods: (1-7), (1-14) and (1-24)

- Retailers provide the data by t+2 (two days after data collection period)

- For each barcode (GTIN), prices are calculated based on turnover and quantities (weekly price=weekly turnover/weekly quantities).

- Data from the 1-24 period is used for index calculation.

Data Quality Assurance and Processing Steps

- The quality of the received data is evaluated based on including size, number of outlets, etc.

- The Data Transfer and Processing Group checks consistency across data from previous months.

- Missing and incorrect entries are identified.

- If issues are found, data is re-requested for correction.

Sample Scanner Data

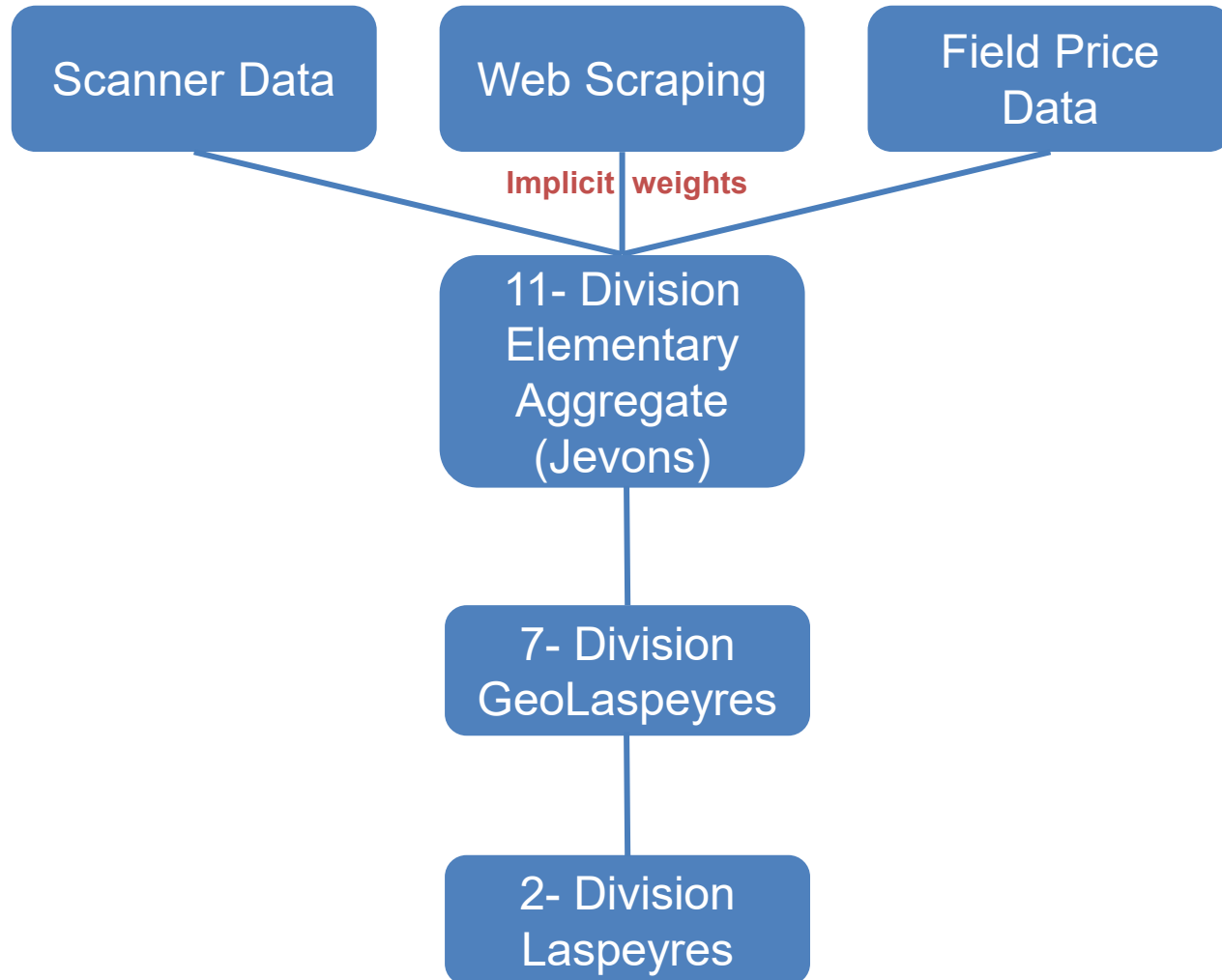| YEAR	MONTH	PERIOD	RETAIL NO	CITY CODE	OUTLET	ITEM NO	BARCODE	BRAND	COICOP	DEFINITATION	QUANTITY	UNIT	SHELF PRICE	UNIT PRICE
2024	9	3	9480423762	10	5907	350740	8699118035015	bendo	01115010100	BENDO MAKARNA BURGU 500 G BENDO	0,5	KG	10,60	21,20
2024	9	3	8140131899	57	2986	402595	2008010903012	coca cola	01222030601	COCA COLA 1 5 LT	1,5	L	34,95	23,30
2024	9	3	8140131899	33	5953	340111	6920354805844	colgate	12130020102	COLGATE MAX BEYAZLIK 50 ML	50	ML	49,50	99,00
2024	9	3	9480423762	2	9539	375439	8699118059943	diyar	01190011801	DİYAR BAHARAT KARABİBER 75 G DİYAR	0,075	KG	45,00	600,00
2024	9	3	1750051846	21	6546	365813	8690637930843	rinso	05611010302	SIVI DETERJAN RENKLİ 3L RİNSO	3	L	122,19	40,73
2024	9	3	9480423762	16	D254	354919	8699118023722	ovadan	01111010101	OVADAN PİRİNÇ BALDO 2500 G OVADAN	2,5	KG	152,94	61,18
2024	9	3	1750051846	70	D228	412547	8695077062765	şafak	01181020100	ŞEKER KÜP 1KG ŞAFAK	1	KG	40,50	40,50
2024	9	3	2120370364	6	185	703214	2392020010359	watsons	12130050303	WATSONS SIVI EL SABUNU ŞEFTALİ SHEA YAĞI 500ML	0,5	L	28,36	56,72
2024	9	3	6320036072	37	146	700960	194252099506	apple	08200010603	APPLE iPhone 11 128 Gb Akıllı Telefon Beyaz	1	PIECE	24.300,00	24.300,00
2024	9	3	6320036072	53	123	700482	8697975600025	tefal	05320021100	TEFAL Tefal Gourmet Minute Izgara Tost Makinesi	1	PIECE	5.149,00	5.149,00
2024	9	3	8360144393	48	2814	700261	6970995787733	roborock	05315010500	ROBOROCK Roborock Q8 Ma Robot Süpürge Siyah	1	PIECE	17.999,00	17.999,00
2024	9	3	8360144393	28	798	702020	194252515747	apple	09130110100	APPLE Apple iPad 9 Nesil 10 2 inch Wi Fi 64GB Space Grey MK2K3TU A	1	PIECE	12.923,38	12.923,38

Scanner Data Handling and Processing Pipeline
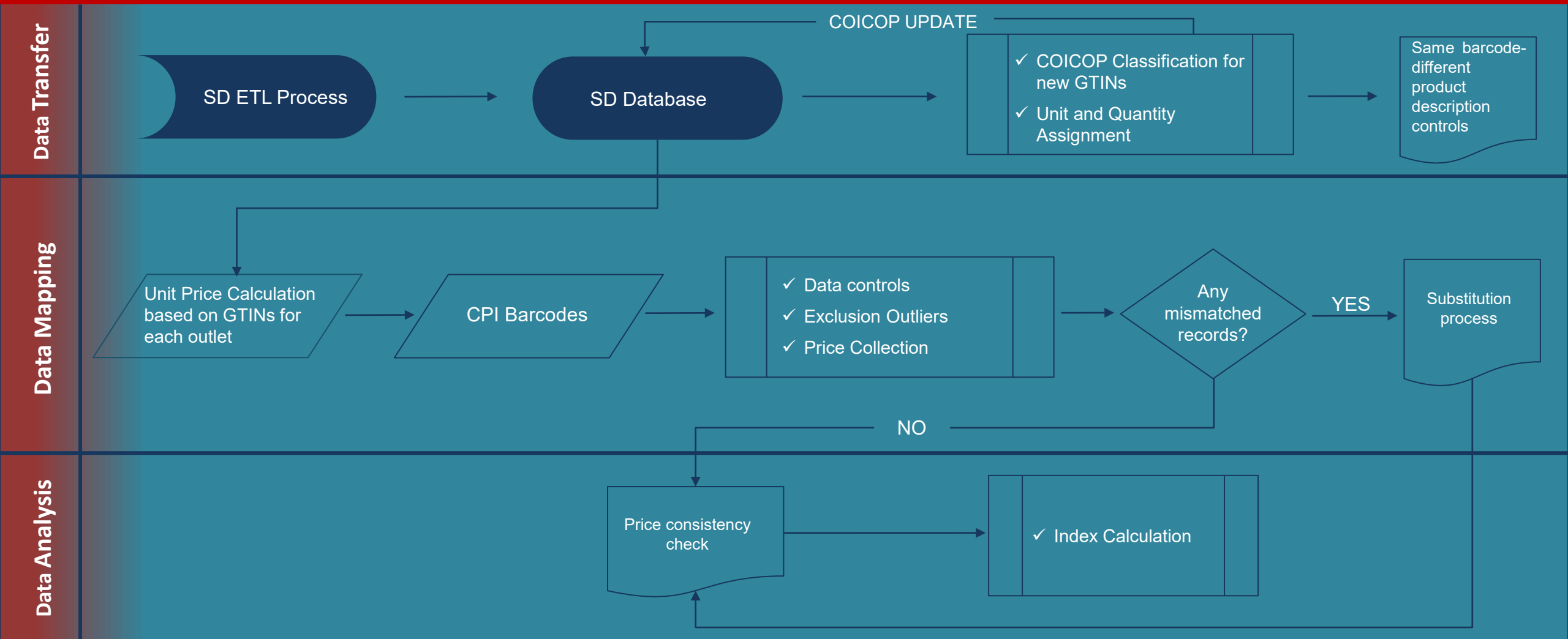
COICOP Classification Process

- If a COICOP has previously been assigned to a barcode, unit and quantity codes are directly assigned using a reference table from the database.

- For new products, one of the top 3 most similar previously assigned COICOPs is selected based on description similarity and sent for verification.

- Additionally, machine learning projects using various algorithms are being developed to enhance this process, which is still in progress.
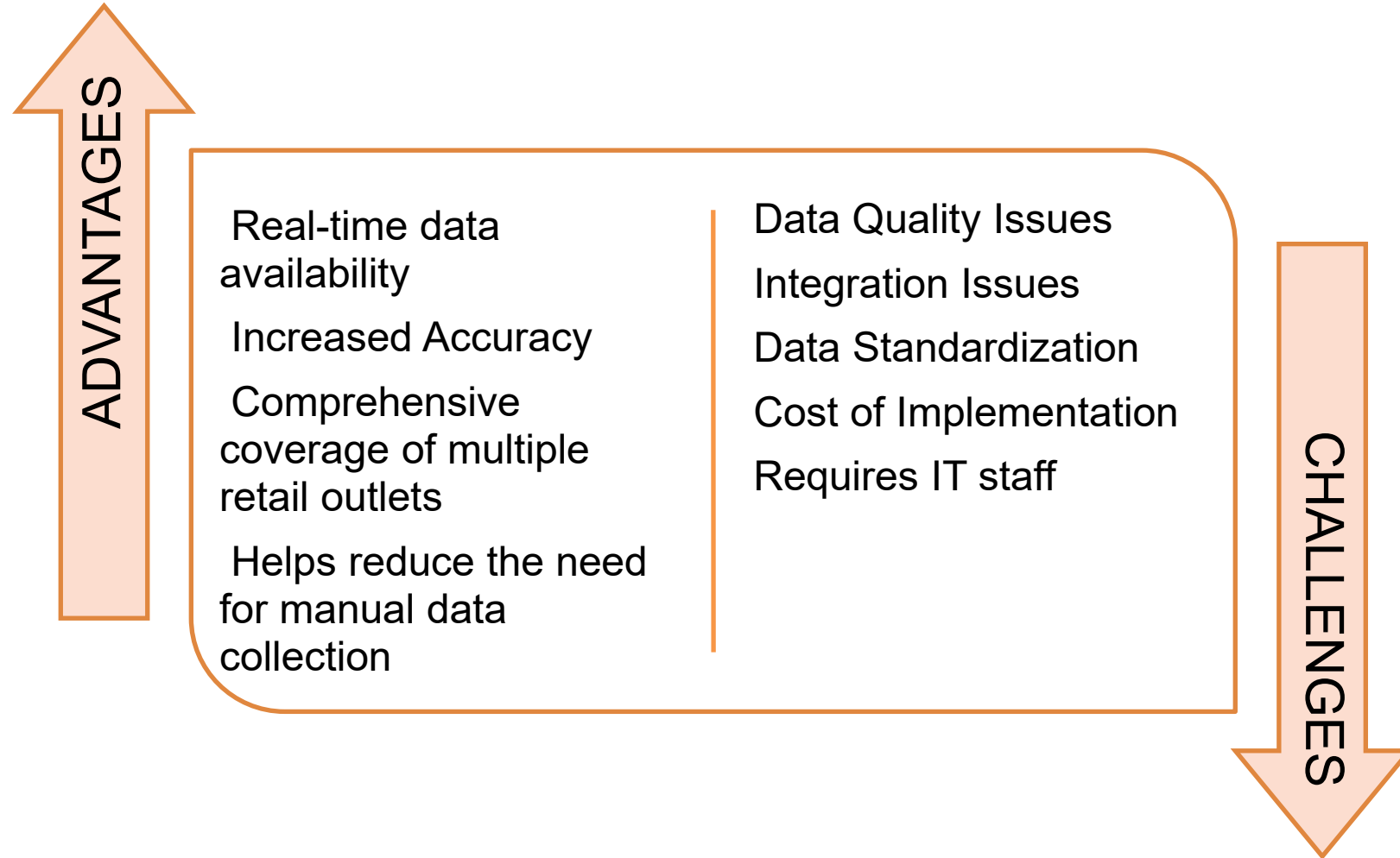
Integrating Data Sources into Index Calculation

Scanner Data

Web Scraping

Field Price Data

Implicit weights

11- Division Elementary Aggregate (Jevons)

7- Division GeoLaspeyres

2- Division Laspeyres

- Currently, TurkStat uses scanner data as a replacement for existing field surveys.

  – The methodology remains unchanged

- Index calculation is conducted by combining data collected from three sources: field data, scanner data, and web scraping data.

Scanner Data Advantages and Disadvantages

ADVANTAGES

Real-time data availability

Increased Accuracy

Comprehensive coverage of multiple retail outlets

Helps reduce the need for manual data collection

Data Quality Issues

Integration Issues

Data Standardization

Cost of Implementation

Requires IT staff

CHALLENGES
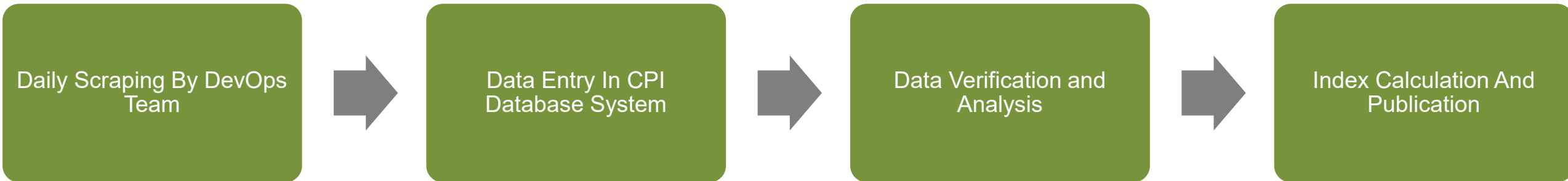
Web - Scraping Overview

- Web Scraping has been in use since 2022.

- Web Scraping codes was developed by TurkStat using Python.

- The IT department is responsible for developing and maintaining the scripts, while the Price Department handles the data processing.

- It is applied by monitoring predefined links.

End-to-End Process for Web Data Collection and Integration into Production Systems

Contact Companies for Data Collection (Data Collection Department)

Based on the Statistical Law, an official letter is sent to the companies to which the websites belong.

Sign Standardized Protocol (Data Collection Department)

A protocol is signed, specifying the nature of the data to be collected, the period, the addresses from which connections will be made, confidentiality, and other legal points.

The contacts for both parties in case of issues are clarified.

Write Procedures for Web Scraping
Procedures are written for collecting desired prices and other information from pages identified by the Price Statistics Unit via web scraping.

Automate Procedures (Data Transfer and Processing Group)

After trials, procedures are automated.

Transfer Data to Production Database (Data Transfer and Processing Group)

After a certain observation period, items decided to be used in the index are transferred to the production database.

Conduct Control Procedures (Price Statistics Unit)

Control procedures are conducted to ensure the accuracy of the data transferred to the production database.

Summarize Prices by Item Code (Price Statistics Unit)

Prices are summarized by item code.

Transfer Summarized Data for Index Calculation (Price Statistics Unit)

The summarized data is transferred for index calculations.

Web - Scraping Work-Flow

| Daily Scraping By DevOps Team | → | Data Entry In CPI Database System | → | Data Verification and Analysis | → | Index Calculation And Publication |

Web – Scraped Items

03
Clothing and footwear

Men's jacket
Men's pijamas
Men's pullover
Men's shirt
Men's socks
Men's sport shoes
Men's suit
Men's sweatshirt
Men's tracksuit
Men's trousers
Men's t-shirt
Men's underwear
Overalls for baby
Scarf
Skirt
Slipper for woman
Tie
Women's anorak jacket
Women's boots
Women's cardigan
Children's trousers
Children's t-shirt
Women's coat
Women's dress
Women's footwear
Women's jacket
Women's leggings
Women's pijamas
Women's pullover

05
Furnishings, Household Equipment and Routine Household Maintenance

Air conditioner
Bed clothes
Bed cover
Bed pique set
Blanket
Blender
Built-in Cooker
Built-in Microwave Oven
Built-in Oven
Carpet
Chair
Combi boiler
Dish washing machine
Double bed
Freezer
Iron
Nesting table
Pillow
Toster
Washing machine

07
Transport

Automobile
Bus fare
(inter-urban)

08
Communication

Mobile phone
Mobile phone
spare parts and
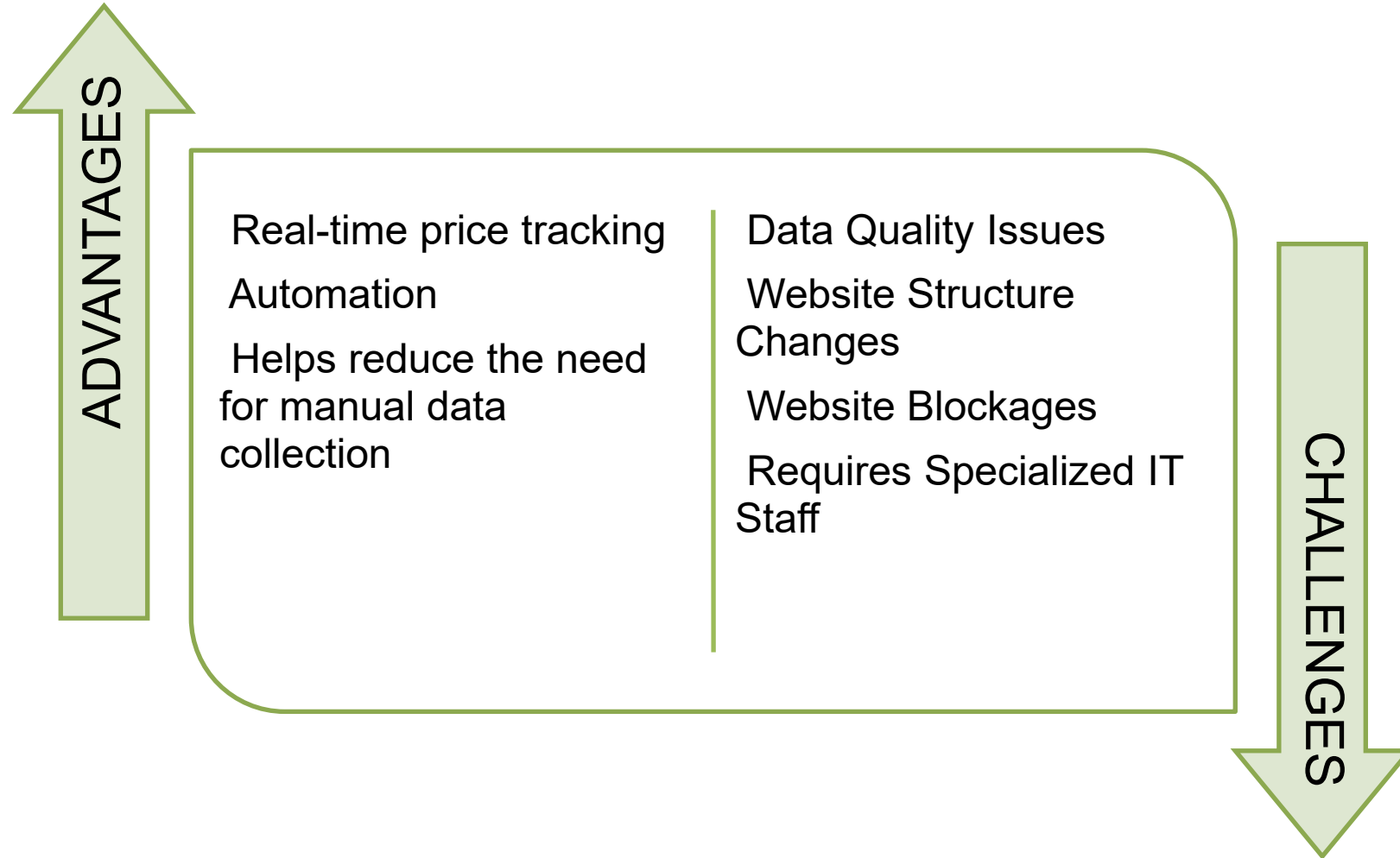accessories

09
Recreation and Culture

Items for sport
and recreation
PC and Laptop
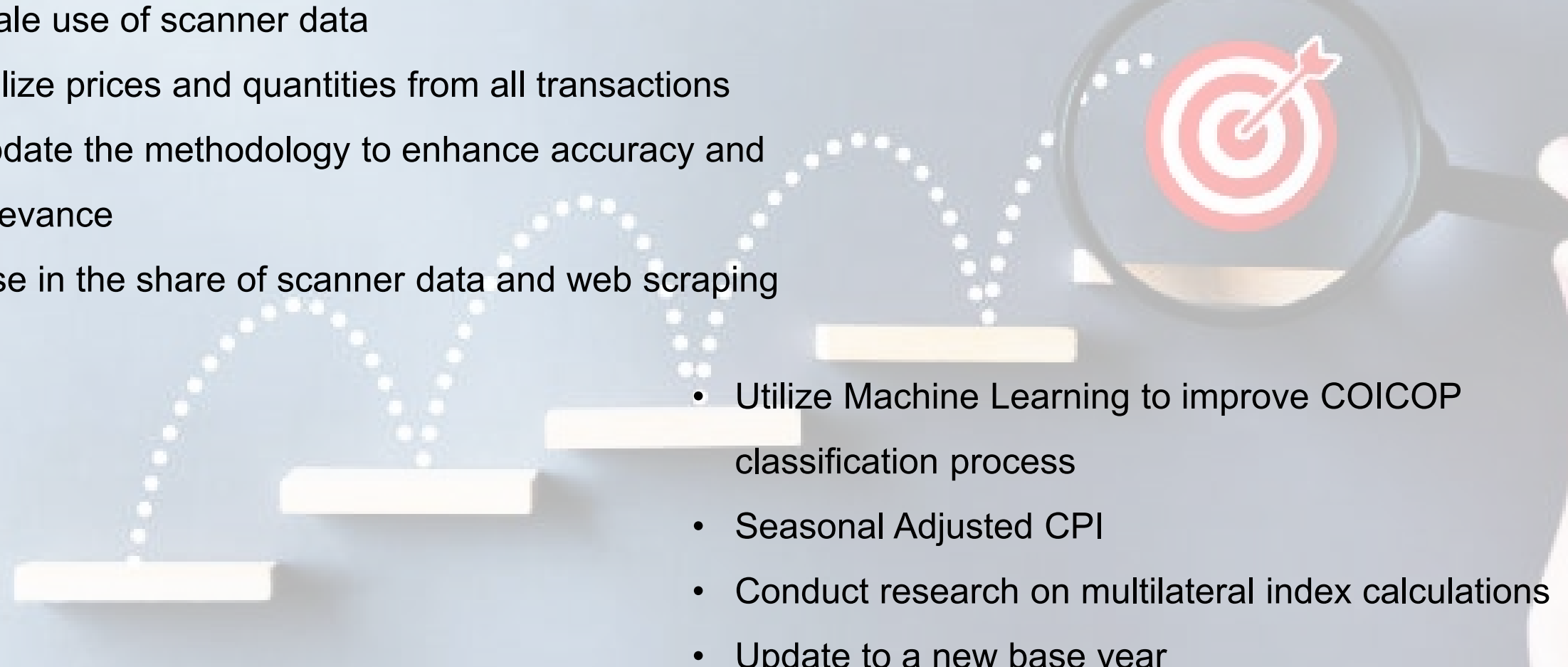PC equipments
Tablet (PC)
Television

12
Miscellaneous Goods And Services

Bag for women
Hair care
appliances
School bag
Travel goods
Umbrella
Watch
Smart watch

Web Scraping Advantages and Disadvantages

ADVANTAGES

Real-time price tracking

Automation

Helps reduce the need for manual data collection

CHALLENGES

Data Quality Issues

Website Structure Changes

Website Blockages

Requires Specialized IT Staff

Future Plans

- Full scale use of scanner data
  - ➢ Utilize prices and quantities from all transactions
  - ➢ Update the methodology to enhance accuracy and relevance
- Increase in the share of scanner data and web scraping data

  - Utilize Machine Learning to improve COICOP classification process
  - Seasonal Adjusted CPI
  - Conduct research on multilateral index calculations
  - Update to a new base year

Thanks for your attention!

Questions & Answers