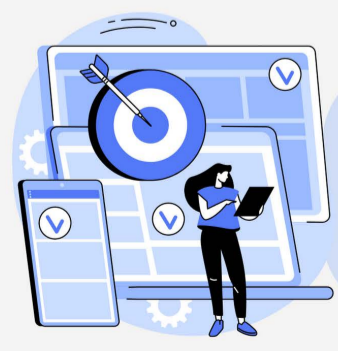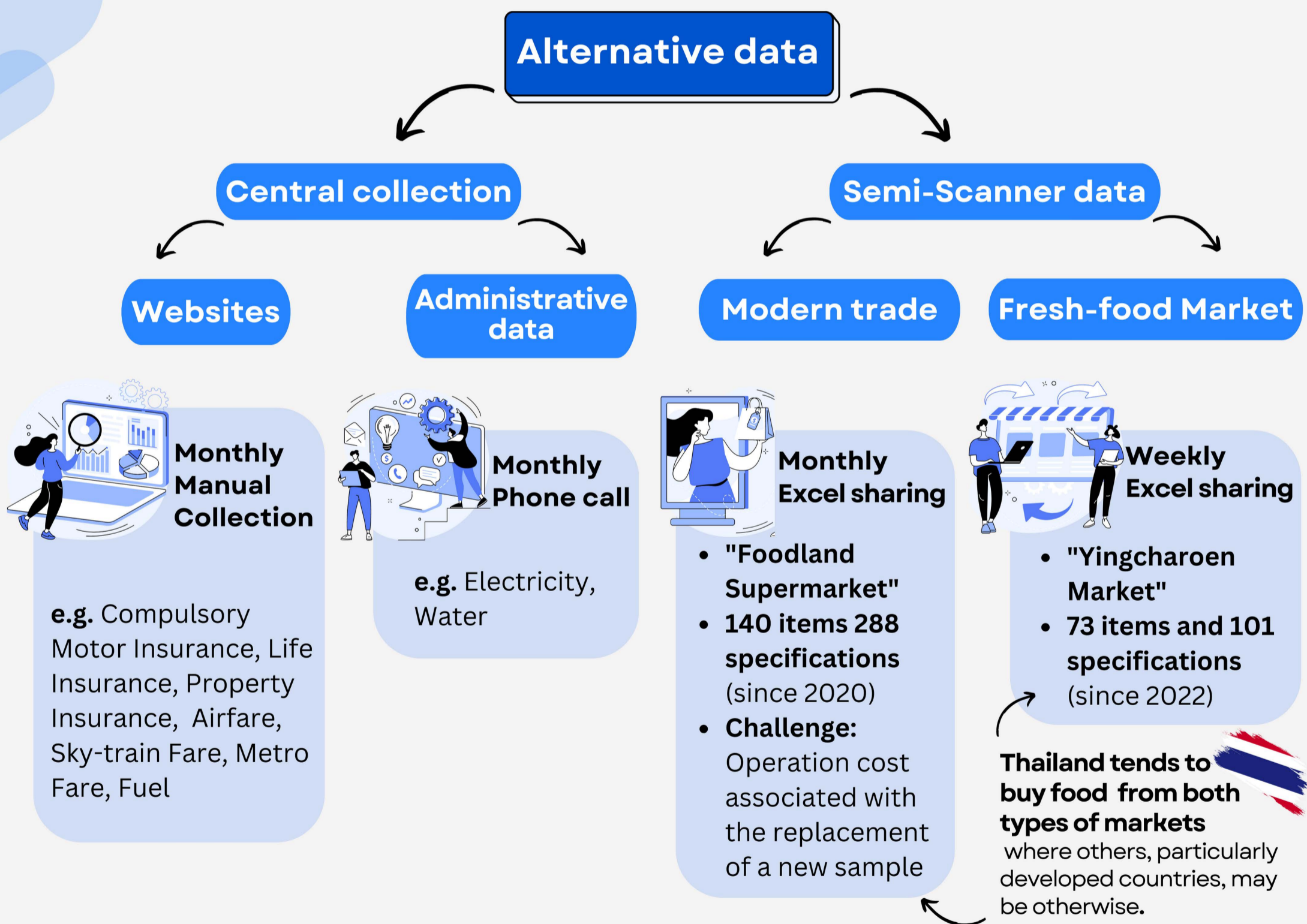# Thailand's progress and future steps for implementing alternative data sources in the CPI

By the Trade Policy and Strategy Office (TPSO), Ministry of Commerce, Thailand

TPSO — Trade Policy and Strategy Office

1 SUCCESSFUL CASES WHERE ALTERNATIVE DATA ARE USED

Scanner data is limited in Thailand due to the absence of effective legal. The modern trade market is an oligopoly, and large supermarkets cite trade secrets and competition for not sharing data, while smaller companies cite high cost of data acquisition. **Despite the challenges, there are significant progress.**

Alternative data

- **Central collection**
  - **Websites**
  - **Administrative data**
- **Semi-Scanner data**
  - **Modern trade**
  - **Fresh-food Market**

Websites — Monthly Manual Collection

e.g. Compulsory Motor Insurance, Life Insurance, Property Insurance, Airfare, Sky-train Fare, Metro Fare, Fuel

Administrative data — Monthly Phone call

e.g. Electricity, Water

Modern trade — Monthly Excel sharing

- "Foodland Supermarket"
- 140 items 288 specifications (since 2020)
- **Challenge:** Operation cost associated with the replacement of a new sample

Fresh-food Market — Weekly Excel sharing

- "Yingcharoen Market"
- 73 items and 101 specifications (since 2022)

Thailand tends to buy food from both types of markets where others, particularly developed countries, may be otherwise.

Next steps
Automated Price Collection
- API, Web-scraping
- Scanner data

Expand partnerships
Active in dialogue with other fresh-food markets across eight different locations.

---

www.tpso.moc.go.th

fb: สำนักงานนโยบายและยุทธศาสตร์การค้า (Trade Policy and Strategy Office)

คิดค้า.com

narongk.moc.go.th
tanaporns.moc.go.th

Thailand's progress and future steps for implementing alternative data sources in the CPI

By the Trade Policy and Strategy Office (TPSO), Ministry of Commerce, Thailand

2 ONLINE-CPI: MANUALLY COLLECT PRICES FROM ONLINE SHOPS

- Constructed Online-CPI index by compiling weights based on **our Consumer Online Expenditure Survey**, unlike the official CPI, which typically uses weights derived from the NSO survey.

- The weights of both index **differ significantly**, particularly in the Food & Non-Alcoholic Beverages.

1. Source

1.1 E-commerce platforms : Shopee, Lazada
**1.2 Outlets' online shops :** e.g. PIZZAHUT MCDONALDS, KFC, UNIQLO, AIIZ, Pomelofashion, PowerBuy, Watson, JIB

2. Criteria for selecting price representative

Based on number of sales, reviews and trusted stores

3. Discussion

- **Major group Level:** Online-CPI have high volatility, especially for Apparel, Footwears, Medical and Personal care.

- **The price levels of online and offline products often vary:** Interestingly, in the same category, the online price is not always lower than the offline price as commonly perceived. Instead, it can be higher in some instances due to characteristics in consumer's online and offline purchasing preferences.

- **The 2022 online CPI increased by 2.76 %AoA.** All product groups experience an increase of price. Meanwhile, the official CPI rose by 6.08%.

■ Official CPI   ■ Online CPI
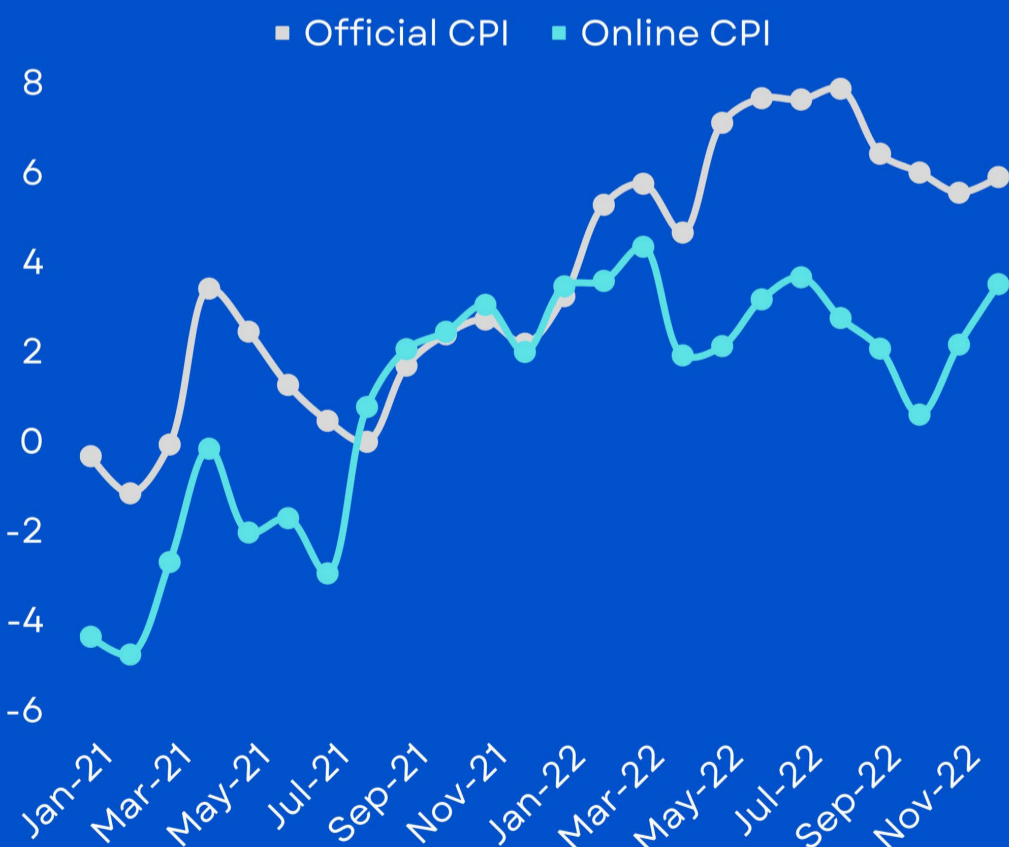
4. Looking ahead

- Examining the integration of online and offline CPI
- Update weights to reflect trends of online shopping after COVID-19

www.tpso.moc.go.th
fb: สำนักงานนโยบายและยุทธศาสตร์การค้า (Trade Policy and Strategy Office)
คิดค้า.com
narongk.moc.go.th
tanaporns.moc.go.th

Thailand's progress and future steps for implementing alternative data sources in the CPI

By the Trade Policy and Strategy Office (TPSO), Ministry of Commerce, Thailand

3 REVISITING THE POTENTIAL OF WEB SCRAPING FOR COMPILING CPI

Five years ago, we attempted to use web scraping to collect prices. However, when dealing with Thai text, we faced the challenge that the scraped data lacked **word delimiters**. This poses a difficulty in cleaning data and extracting useful information. Moreover, we faced limitations in accessing detailed product pages on certain **dynamic websites**.

น้ำส้ม **VS** น้ำ / ส้ม

Orange juice   *Water*   *Orange*
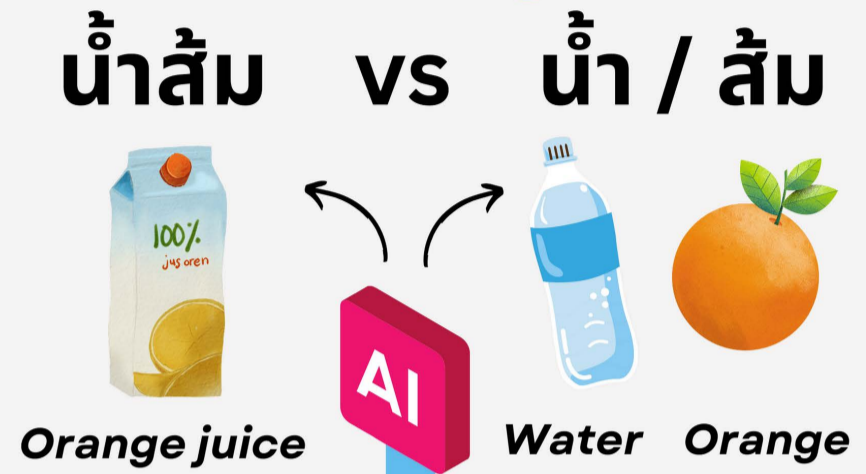
New Experiment

Electronic products from 4 websites
(Lazada, Shopee, Banana IT and PowerBuy)

- **Structured presentation** of product characteristics
- **Extensive variety of brands and models** can be effectively captured
- **Limited number of websites and platforms** dominate the consumer electronics market

Classification

- **Keywords Extraction:** e.g. brand, electric power, volume
- **Website Category:** maped to index category
- **Product Comparison:** matched with structured websites
- **Expert Classification:** manually categorized

Challenges

- **Number of page restrictions and Legal Considerations**
- **Missing Data:** Sales numbers and transportation fees
- **Organizational Management :** CPI staffs prefer to scrape data autonomously, however, our expertise in this area still are limited.