UNBigDataLearning

Acquiring Scanner Data for CPI

UN-CEBD Task Team on Scanner Data

Learning Objectives

By the end of the module, participants will be able to:

- Explain the steps in the process of scanner data acquisition and the opportunities and challenges posed when requesting new data at the NSO
- Recognise where and how to obtain new data, and any difficulties associated with this
- Manage the main aspects of the negotiations with the large-scale retail trade companies and the terms of the agreement of scanner data delivery
- Have a clear idea of the resources, knowledge and skills necessary to manage successfully the process of acquisition

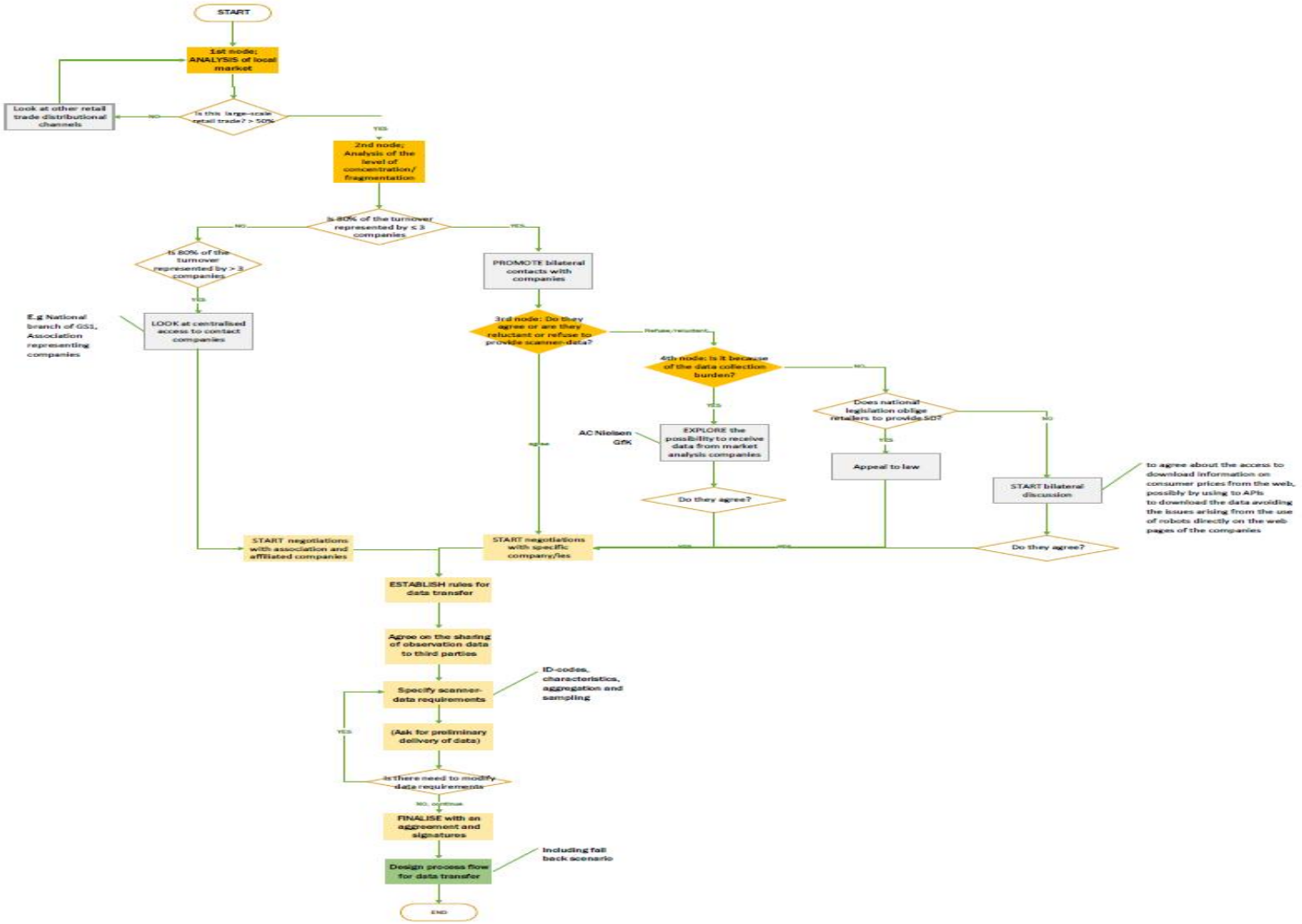
UNBigDataLearning

Contents

1. Main steps to obtain scanner-data

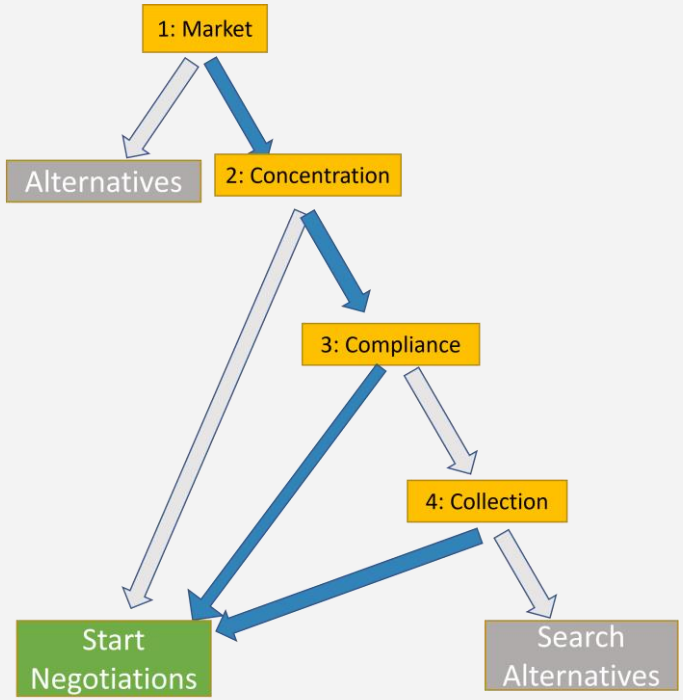
A decision tree to guide the choices

- This section describes step-by-step the pathway from the initial contact with the data provider to the acquisition of scanner data
- A “decision tree” is presented – we recommend this is downloaded
- The graphical representation will helps us to follow the sequence of steps and nodes

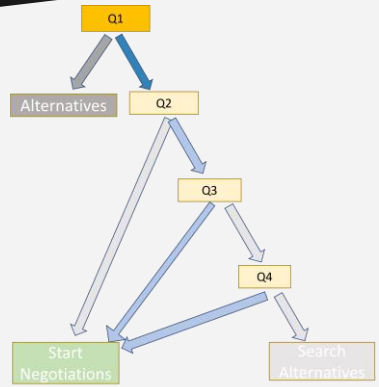

START
1st node: ANALYSIS of local market
Look at other retail trade distributional channels
NO
Is this large-scale retail trade? > 50%
YES
2nd node: Analysis of the level of concentration/ fragmentation
Is 80% of the turnover represented by ≤ 3 companies
NO
YES
Is 80% of the turnover represented by > 3 companies
YES
E.g National branch of GS1, Association representing companies
LOOK at centralised access to contact companies
PROMOTE bilateral contacts with companies
3rd node: Do they agree or are they reluctant or refuse to provide scanner-data?
Refuse / reluctant
4th node: Is it because of the data collection burden?
YES
NO
AC Nielsen GfK
EXPLORE the possibility to receive data from market analysis companies
Do they agree?
Does national legislation oblige retailers to provide SD?
YES
NO
Appeal to law
START bilateral discussion
to agree about the access to download information on consumer prices from the web, possibly by using to APIs to download the data avoiding the issues arising from the use of robots directly on the web pages of the companies
Do they agree?
START negotiations with association and affiliated companies
START negotiations with specific company/ies
ESTABLISH rules for data transfer
Agree on the sharing of observation data to third parties
Specify scanner-data requirements
ID-codes, characteristics, aggregation and sampling
(Ask for preliminary delivery of data)
YES
Is there need to modify data requirements
NO, continue
FINALISE with an aggreement and signatures
Design process flow for data transfer
Including fall back scenario
END

A decision tree to guide the choices

For this next section, we will work through the first 4 nodes of the “decision tree” in detail



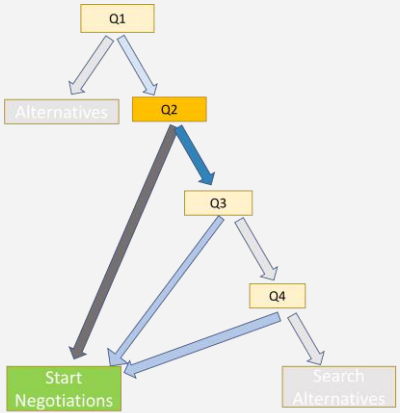
First step/node: Analysis of local the markets



First step/node:

- Q1: Is the large-scale retail trade distribution predominant or not?
- Analyze the question by group of products or by COICOP expenditure division
- For each group of products or COICOP division, detect the characteristics and estimate the weights of each retail trade distributional channel in terms of turnover (a summary table could be produced)
- Decision: if large-scale retail trade distribution
  - is > 50% in terms of turnover, go to the 2nd decision node
  - is < 50% in terms of turnover, consider alternatives

Second step/node: Analysis of the level of concentration of the large-scale retail trade distributional channel



Second step/node:

- Q2: What is the level of concentration of the large-scale retail trade distributional channel in the group of products or COICOP division being considered?

- Need to understand the market dominance (% turnover of companies)
- Decision: if 80% of the turnover of large-scale retail trade distribution is represented by:
  - ✓ More than 3 companies – refer to Case 1
  - ✓ Less than or equal to 3 companies – refer to Case 2

Second step/node: Analysis of the level of concentration of the large-scale retail trade distributional channel

Case 1

- 80% of the turnover of large-scale retail trade distribution is represented by > 3 companies
  - ✓ Look for centralized access to companies
  - ✓ Negotiate contacts with national branch of GS1 or with associations representing companies in each market

Second step/node: Analysis of the level of concentration of the large-scale retail trade distributional channel

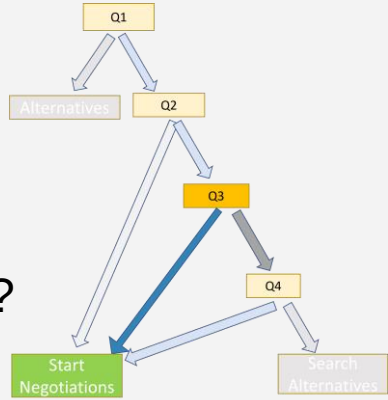
Case 2:

- 80% of the turnover of large-scale retail trade distribution is represented by ≤ 3 companies
  - ✓ Promoting bilateral contacts and meetings

Second step/node: Analysis of the level of concentration of the large-scale retail trade distributional channel

- The approach to the negotiations for scanner data with large-scale retail distribution negotiation in both cases 1 & 2 is aimed at:
  - ✓ Explaining why it is important for NSOs to receive scanner data
  - ✓ Stressing the benefits to the NSO in improving the quality of information for the nation’s inflation measurements
  - ✓ Stressing the benefits to the company as the scanner data enables measurement of the company’s contribution to inflation

Third step/node: Agreement or not to provide scanner data to NSOs



Third step/node:

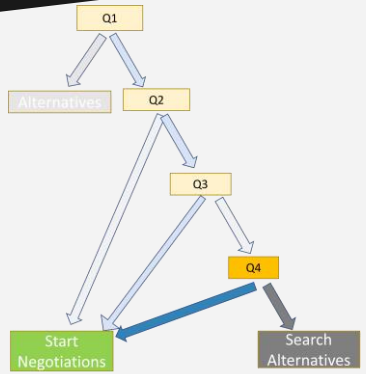
- Q3: Do the companies of large-scale retail trade distribution agree to provide scanner data to NSOs or are they reluctant?

✓ Yes: if they agree, start negotiations with:

- GS1 or sector Association where the market is fragmented (Case 1) or
- Companies in bilateral discussions if concentrated market (Case 2)

✓ No: if they are reluctant, go to the fourth step/node

Fourth step/node: agreement or not to provide scanner data to NSOs



Fourth step/node:

- Q4: Are the companies reluctant to share scanner data due to the burden?

➢ Yes: Explore the possibility to receive the data from market analysis companies that are already recipients of the data. If this is the case, set up a triangle shaped data flow

- ✓ 1. NSO asks the market analysis company for the scanner data
- ✓ 2. The large-scale retail trade company authorizes it
- ✓ 3. The market analysis company provides the NSO with the requested data

Fourth step/node: agreement or not to provide scanner data to NSOs

Fourth step/node:

- Q4: Are the companies reluctant to share scanner data due to the burden?

➢ No: Explore other options for obtaining the scanner data:

- ✓ Look at the legislation and appeal to law if available
- ✓ Start bilateral discussions for approval to web scrape data (no longer scanner data!)

When the large-scale retail trade companies agree to provide scanner data to NSOs

- Let’s go back to the case when the companies agree to deliver scanner data to NSOs (directly or indirectly via market analysis companies), and try to answer to the question:

How to deal with the negotiations with the companies to establish an agreement about the delivery of scanner data?

2. Negotiation strategy & requirements for scanner-data deliveries

Why an agreement for scanner data delivery

First of all, why is it important to arrive at some kind of formal agreement with the companies?

1. If the CPI compilation is going to be based on the regular acquisition of scanner data, this means, for the products involved, abandoning the traditional data collection moving from a direct to an indirect way to collect the data
2. It implies that the NSOs do no longer have the total control of the production process of CPI given that are a few companies that send a big amount of data to the NSOs that wouldn't be able to detect them through the traditional ways

Why an agreement for scanner data delivery

3. Therefore, it is necessary to arrive at some kind of formal agreement with the data providers (either the large-scale retail-trade or the market analysis companies)
4. Arriving at establishing/signing a formal agreement with the companies, means to secure the production process of CPI also for the components depending on the acquisition of scanner data
5. This concern about how to secure the delivery of scanner data has to be considered also in the terms of the formal agreement (to guarantee the provision in case, for example, of change of the management of the companies)

Requirements for scanner data delivery and terms of the agreements with the companies

- Data requirements and terms of the agreement with the data providers are strictly related

Having clear the first (requirements) allow to well define the second (the terms of the agreement)

Requirements for scanner data delivery

What are the main data requirements to consider to compile CPIs by using scanner data and replace the traditional data collection?

1. Moving from the traditional data collection means moving from a shelf price to a price that is related to the effective monetary transaction occurred and obtained, usually, as the ratio between daily/weekly/monthly turnover and daily/weekly/monthly quantities.

Requirements for scanner data delivery

2. “Scanner data sets contain variety quantities sold and revenue received by the retailer for these varieties for some period of time, usually a week or a month. This information enables NSOs to calculate a price for an individual variety by dividing a variety’s revenue by the quantity sold. This price, referred to as a unit value, represents the average price experienced by consumers over a period of time.” (CPI Manual, 2020)

Requirements for scanner data delivery

3. Therefore, price information derived from scanner data set has to be strictly related to homogeneous products, with common, specific characteristics.

4. It has to be referred to a specific combination of brand, variety and package (elementary item) not to a generic product and, given that it is referred to monetary transaction, it should include possible discounts on the items

Requirements for scanner data delivery

Let’s give an example of the meaning of unit value price.

In the case, for instance, of mineral water, the basic information necessary for CPI compilation is the price of mineral water, sparkling, in bottle by 1.5 liter (unit value price) not the average prices of a liter of mineral water independently on the package proposed (average values).

This is a crucial requirement to define the characteristics of the scanner data provision

Unit values YES, average values NO

Requirements for scanner data delivery

5. As a consequence of what is illustrated before, it is necessary to receive well detailed information on the characteristics of elementary items in order to compare like with like across the time (this is guaranteed by the availability of a unique code of the elementary item with a clear description)

6. Data must be representative of the temporal unit (usually the month for inflation estimates), of national/regional territory (depending on the sample design of the survey and the estimation domains) and of the universe of outlets across the country

Requirements for scanner data delivery

7. The frequency and the timeliness of data delivery has to be coherent with the dissemination calendar of CPI that, in general, is very tight in temporal terms

8. The price data detected has to be transmitted in a format that fulfills the requirements of the NSO's data base and through a secure protocol

Requirements for scanner data delivery

Let's summarize the main CPI requirements to consider for scanner data delivery

- Information on turnover and quantities to obtain unit value price/index
- Information on the specific characteristics of the elementary items (metadata)
- Representativeness of the country/region and of the universe of outlets
- High frequency (weekly at least) and timely data transmission
- Format that fulfills the requirements of the NSO's data base
- Secure protocol for the data transmission

7 keywords:

Unit value, metadata, representativeness, high frequency, timeliness, format, secure protocol

Negotiation strategies and contents of agreement

- The requirements listed before have to be translated into the terms of the agreement (in any form it will be established) to negotiate with the large-scale retail trade companies
- It means to establish:

1. The contents and the characteristics of the data set
2. The rules of the data transmission

Negotiation strategies and contents of agreement

- The agreement negotiated has to establish:
  1. The contents and the characteristics of the data set, in terms of:
     - Identification of the elementary items (GTIN, SKU, PLU, Prdkey, others)
     - Description of the elementary items (Dictionary)
     - Availability of the universe of GTINs or of a sample of GTINs
     - Availability of the information referred to all outlets or a sample of outlets
     - Granularity of the statistical unit information (by elementary item and by outlet or by elementary item and geographical area, without information at outlet level)

Negotiation strategies and contents of agreement

Examples of agreements negotiated

- NSO Italy (triangle approach through market analysis company). Data provided refers to the retail outlet and GTIN code
- NSO Finland (bilateral agreements). Data provided refers to the GTIN code, but not the retail outlet

Negotiation strategies and contents of agreement

- The agreement negotiated has to establish:

1. The contents and the characteristics of the data set, in terms of:
    - Geographical granularity of information (it depends on the regional variability of prices)
    - Time granularity (monthly, weekly or daily data)
    - Availability of turnover and quantities sold referred to the elementary item
    - File type and format (csv, txt, xml…)
    - Metadata

Negotiation strategies and contents of agreement

- The agreement negotiated has to establish:
  2. The rules of the data transmission, in terms of:
     - Timing (lag with respect to the reference period of the data, whatever it is the time granularity, day, week or month)
     - Data delivery and transfer method
     - Frequency of data transfer
     - Rules for update of the data in case of changes (revisions policy) and in case of unforeseen circumstances (mistakes) with a named contact to manage possible issues or mistakes

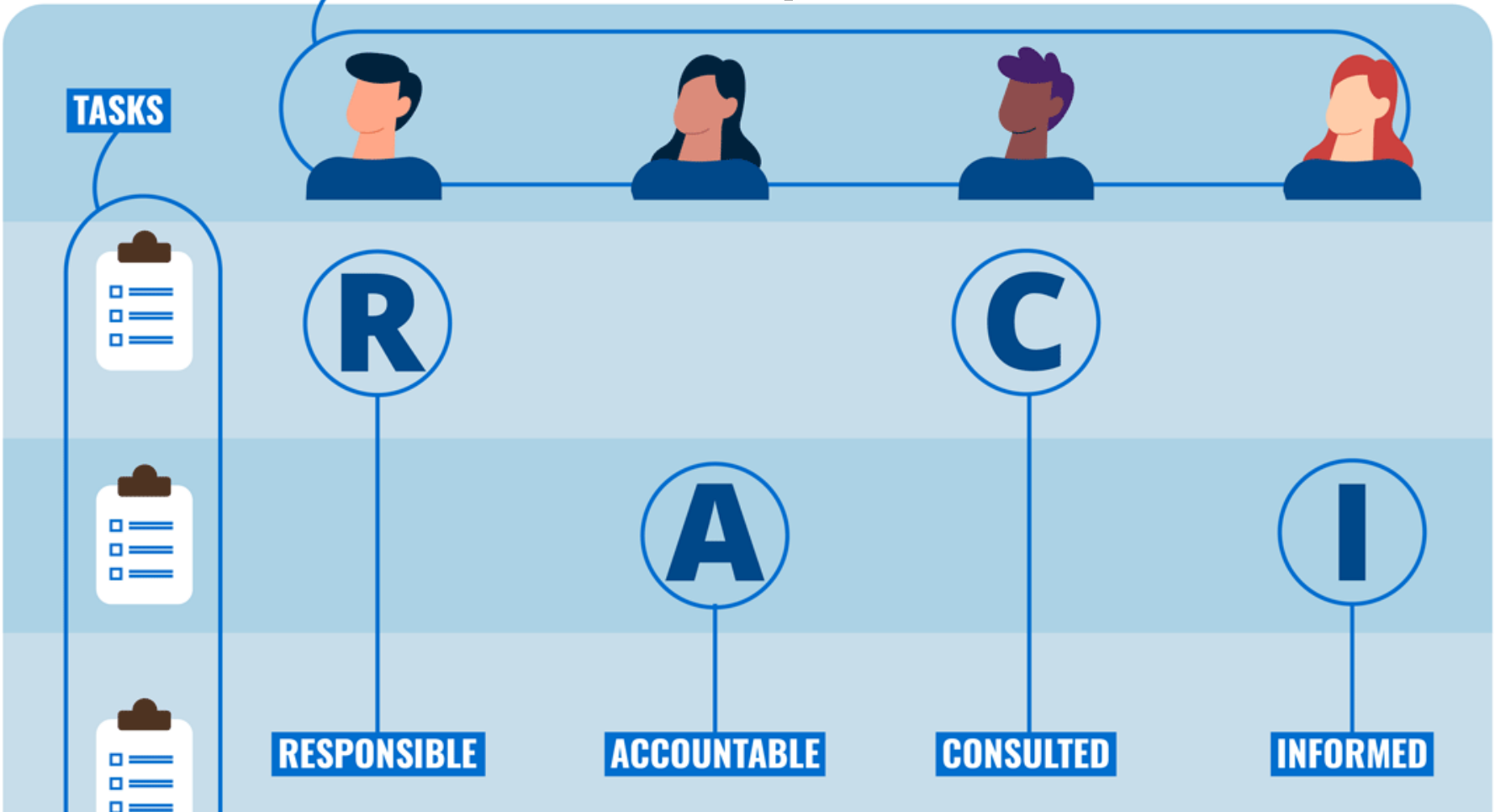
Negotiation strategies and contents of agreement

- The agreement negotiated has to establish:
  2. The rules of the data transmission, in terms of:
     - Security, data protection, unforeseen circumstances
     - Confidentiality and access Rights -- who is allowed to process the data -- and in which statistics they will be used together with the possibility to share data with third party
     - What service the NSO can provide in return for scanner data
     - Duration and termination of the agreement
     - Applicable law and competent courts

Negotiation strategies and contents of agreement

- ✓ Request of some preliminary data, for example, request one year of monthly data (best granularity possible) to investigate the weights and prepare the basis of the indices, to:
  - ➢ Better determine the granularity available
  - ➢ Better define the details within the agreement
  - ➢ To prepare the procedures to treat scanner data as far as they will be regularly delivered

4. Skills required



Managing scanner data acquisition for CPI

- Obtaining scanner data as such is a project to be managed by NSOs by identifying stakeholders and assigning responsibilities
- This approach is necessary to deal with the different issues that are in the path
- RACI matrix appears as a good tool to clear identify the different roles for the different stakeholders

What is a RACI matrix

- What is a RACI matrix?
- It is a specific Responsibility Assignment in project management that specifies all the stakeholders and the different levels of their involvement
- RACI is the acronym for
  - ✓ Responsible
  - ✓ Accountable
  - ✓ Consulted
  - ✓ Informed
- Look at the theoretical scheme of a RACI matrix

What is a RACI matrix

Responsible. Who carries out the work to complete the task and release the deliverables. Each task should have at least one Responsible (or more) and usually he/she has to be in the project team

Accountable. Who reviews the work involved in the project and makes sure the responsible completes the work on time. Only one accountable person usually in a leadership or management role

Consulted. Who contributes before and throughout the work providing information useful for the project and waits for bidirectional communication. May be individuals in the project or outside but whose activity is affected by the outcome of the project. Not necessary for all the tasks

Informed. Parties and persons interested who have to be informed about the state of play of the project without being overwhelmed by the details. They should be out of the team and may be heads of affected teams or senior leadership in the company

The stakeholders in the project of acquiring scanner data

- Economists, Statisticians, Data Analysts, Data Scientists, Project Managers, Retailers are all stakeholders that should be involved in each step of the scanner data acquisition process
- Their role and position in the RACI matrix is different depending on the task
- The following RACI matrix is a comprehensive and general scheme that has to be tailorized to the country specificities and to the concrete availability of resources, keeping a clear distinction in the roles

The RACI matrix of the project of acquiring scanner data

Tasks	Economists	Statisticians	Project Manager	Data Analysts	Data Scientists	IT experts	Retailers
Analysis of the market	R	R	A	I	I	I	C
Analysis of the level of concentration/fragmentation of the large-scale retail trade market	R	R	A	I	I	I	C
Promote bilateral contacts and meetings with companies/national branches of GS1/Associations of companies	C	C	A/R	I	I	I	C
Start negotiations with specific companies/national branches of GS1/Associations of companies	C	C	A/R	I	I	I	C
Specify scanner data requirements	C	R	A	C	C	C	C
Finalize with agreement and signatures	I	C	A/R	I	I	I	C
Data treatment and data cleaning	I	C	A	R	R	R	-