Presentation of results from the Voorburg Group Alternative Data Survey

38th Voorburg Group Meeting
March 5, 2024

Virtual Meeting hosted by Israel

Rohan Draper (Denmark)

ISRAEL
5 | 7 | 19 | 21
MARCH
2024

STATISTICS
DENMARK

Alternative Data Sources in Official Statistics

ISRAEL
5 | 7 | 19 | 21
MARCH
2024
VOORBURG GROUP

Contents

Alternative Data Survey - xlsx

Voorburg Group - Coverage and Alternative Data Survey

2023

Purpose
- To map coverage and alternative data use per index for Voorburg members
- To enable Voorburg members to identify potential development partners
- To highlight trends in contemporary price index data source practises
- Improved decision making regarding SPPI development activities
- This survey is in lieu of previously completed Country Progress Reports

Instructions
- For each industry or product classification select the appropriate data collection method by clicking on the respective cell and choose from the list of available options.
- For definition support see the tab entitled 'definitions'.
- If you do not collect data for a given industry code no action is required (i.e. leave the cell as "<select>")
- "Supplementary information" cells are included if you would like to provide a short description of the alternative data (source, collection, use etc.) and/or include a link to relevant documentation.

Completion
- Only fill-in one survey per country - please coordinate between Prices and Output areas.
- Estimated time for completion is circa 15min.

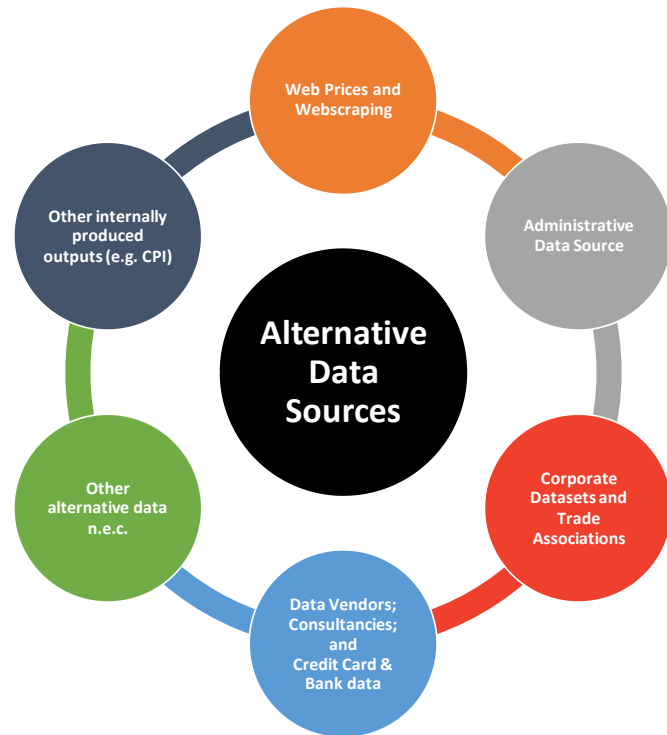| Administrative Questions	
Which country do you represent?	<country>
Which department and/or team do you represent?	<department/team>
Are your PPIs product or industry based (or both)?	<industry/product>
What is your current applied classification (nomenclature)?	<nomenclature>

Tips
If your statistics office does not use the classification in the questionnaire a link is provided below where you can find a correspondence table via UN Stats and Eurostat respectively.
https://unstats.un.org/unsd/classifications/Econ#corresp-isic-un
https://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL

▶ click here to go to *definitions* of alternative data sources   ▶ click here to complete the survey for your country

Web Prices and Webscraping

Administrative Data Source

Other internally produced outputs (e.g. CPI)

Alternative Data Sources

Corporate Datasets and Trade Associations

Other alternative data n.e.c.

Data Vendors; Consultancies; and Credit Card & Bank data

Version: 2023-09-19
Language: English
Resolution: 1920x1080
Excel: 2016 | 32-bit

Alternative Data Sources in Official Statistics

Alternative Data Survey - xlsx

Purpose

o Map alternative data use per index

o Enable Voorburg members to identify potential development partners

o Highlight trends in contemporary price index data source practises

o Enhance decision making regarding SPPI development activities

Voorburg Group - Coverage and Alternative Data Survey

2023

Purpose
- To map coverage and alternative data use per index for Voorburg members
- To enable Voorburg members to identify potential development partners
- To highlight trends in contemporary price index data source practises
- Improved decision making regarding SPPI development activities
- This survey is in lieu of previously completed Country Progress Reports

Instructions
- For each industry or product classification select the appropriate data collection method by clicking on the respective cell and choose from the list of available options.
- For definition support see the tab entitled 'definitions'.
- If you do not collect data for a given industry code no action is required (i.e. leave the cell as "<select>")
- "Supplementary information" cells are included if you would like to provide a short description of the alternative data (source, collection, use etc.) and/or include a link to relevant documentation.

Completion
- Only fill-in one survey per country - please coordinate between Prices and Output areas.
- Estimated time for completion is circa 15min.

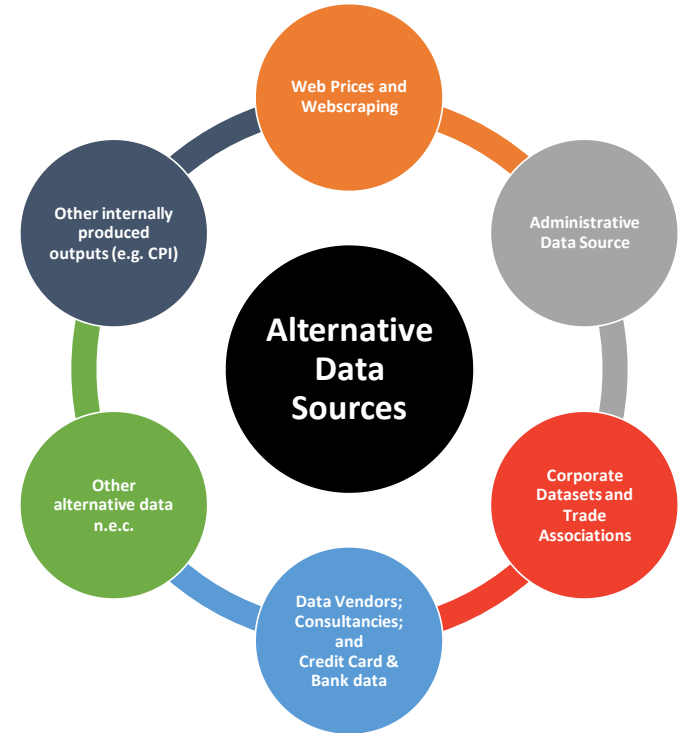| Administrative Questions	
Which country do you represent?	<country>
Which department and/or team do you represent?	<department/team>
Are your PPIs product or industry based (or both)?	<industry/product>
What is your current applied classification (nomenclature)?	<nomenclature>

Tips
If your statistics office does not use the classification in the questionnaire a link is provided below where you can find a correspondence table via UN Stats and Eurostat respectively.
https://unstats.un.org/unsd/classifications/Econ#corresp-isic-un
https://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL

▶ click here to go to *definitions* of alternative data sources   ▶ click here to complete the survey for your country

Web Prices and Webscraping

Administrative Data Source

Other internally produced outputs (e.g. CPI)

Alternative Data Sources

Corporate Datasets and Trade Associations

Other alternative data n.e.c.

Data Vendors; Consultancies; and Credit Card & Bank data

Version: 2023-09-19
Language: English
Resolution: 1920x1080
Excel: 2016 | 32-bit

Alternative Data Sources in Official Statistics

What's new?

o If members have requests and/or ideas on other ways the results can be presented please contact the presenters.

o Attaining increased coverage remains a top priority.

Presentation of Results.

Previous and current results are shown in summary tables to allow for trend assessment (note: impacted by increased coverage). A reference has been added "_2023" to identify what year data was last reported or confirmed as unchanged.

Value Proposition.

Data from the survey continues to provide value regarding investigations into CPI usage by PPI statistical programmes and has been picked up by further constellations of groups like the UN City Groups.

Coverage.

Thank you to the countries that responded last year and those that have sent in responses and increased the coverage of the survey this year! We have increased coverage from 16 to 22 countries contributing to the results.

Alternative Data Sources in Official Statistics

Category Recap

o Categories remain the same since the original survey.

Source	Type	Origin	Code	Description
TRADITIONAL	STRUCTURED	Questionnaire (paper, phone and/or electronic)	QNR	The traditional way of collection price information by asking firms for the information via telephone, paper or electronic questionnaires. This is the default and not an alternative data source.
EXTERNAL	UNSTRUCTURED	Web Prices (manual)	WPR	The price collection data is derived from websites manually. The data source itself is considered an unstructured data source not designed for statistical purposes.
		Webscraping (automated)	WSC	The price collection data is derived from websites automatic systems. The data source itself is considered an unstructured data source not designed for statistical purposes.
	STRUCTURED	Administrative Data Source	ADM	Data which are derived from the operation of administrative systems by public agencies (e.g. data collected by government agencies for the purposes of registration, transaction, regulation and record keeping). Data is often structured for administrative purposes and is highly transferable for statistical purposes.
		Corporate Datasets	COR	Survey respondent provided datasets obtained directly from corporate headquarters in lieu of data collectors collecting data in respondent stores or on their websites. Data pertains to the particular company that is providing said data is often structured for organisational purposes and is highly transferable for statistical purposes.
		Trade Associations	TAD	Industry based surveys that the target industry is producing for themselves.
		Data Vendors (commercially available structured data)	DVS	Data acquired from companies that actively collect and sell data as a business activity. Often such companies provide data on a contractual basis with defined terms and conditions.
		Consultancies (mandated specific task) (transformed data)	CON	Consulting company and/or specialist company is contracted to collect and/or compile data for a specific purpose (mandated or otherwise). Often such companies are utilised on a contractual basis with defined terms and conditions.
		Credit card and bank data	CCD	Financial information collected at the moment of a transfer of funds between a card holder's account and a business account. Data is graded based on the level of metadata available about the transaction. This source is considered a structured data source.
	BOTH	Other alternative data sources n.e.c.	OTH	Other types of alternative data sources not elsewhere classified. For example, transaction-level data from email receipts (like UBER email receipt data). Other special data delivery from third party data collectors not elsewhere classified.
INTERNAL	STRUCTURED	Consumer Price Index	CPI	Data is sourced directly from the Consumer Price Index
		Producer Price Index	PPI	Data is sourced directly from the Producer Price Index
		Structural Business Statistics	SBS	Surveys utilised for benchmarking purposes
		National Accounts	NA	Price indices derived from volume and value data (implicit price indices)

Alternative Data Sources in Official Statistics

CPI Usage

o CPI remains the dominant alternative data source followed by administrative data, PPI, webscraping (manual) and corporate datasets.

o Guidelines for using CPI as an alternative datasource can be found on the Voorburg website.

Class	Name	Frequency	2022
CPI frequency as an alternative data source for a particular 4-digit classification.			
G	**Wholesale and retail trade; repair of motor vehicles and motorcycles**		
4520	Maintenance and repair of motor vehicles	5	4
H	**Transportation and storage**		
4911	Passenger rail transport, interurban	11	9
4921	Urban and suburban passenger land transport	12	9
4922	Other passenger land transport	12	7
4923	Freight transport by road	1	2
5011	Sea and coastal passenger water transport	8	6
5110	Passenger air transport	9	6
5221	Service activities incidental to land transportation	6	4
5310	Postal activities	5	3
5320	Courier activities	2	2
I	**Accommodation and food service activities**		
5510	Short term accommodation activities	12	7
5520	Camping grounds, recreational vehicle parks and trailer parks	6	4
5590	Other accommodation	6	2
5610	Restaurants and mobile food service activities	16	11
5629	Other food service activities	8	4
5630	Beverage serving activities	10	5
J	**Information and communication**		
5811	Book publishing	4	4
5813	Publishing of newspapers, journals and periodicals	4	5
5911	Motion picture, video and television programme production activities	2	2
5914	Motion picture projection activities	7	4
5920	Sound recording and music publishing activities	2	1
6020	Television programming and broadcasting activities	3	2
6110	Wired telecommunications activities	5	5
6120	Wireless telecommunications activities	3	3

Alternative Data Sources in Official Statistics

CPI Usage

o CPI remains the dominant alternative data source followed by administrative data, PPI, webscraping (manual) and corporate datasets.

o Guidelines for using CPI as an alternative datasource can be found on the Voorburg website.

CPI frequency as an alternative data source for a particular 4-digit classification.			
Class	**Name**	**Frequency**	**2022**
K	Financial and insurance activities		
6419	Other monetary intermediation	2	3
L	**Real estate activities**		
6810	Real estate activities with own or leased property	4	3
6820	Real estate activities on a fee or contract basis	7	4
M	**Professional, scientific and technical activities**		
6910	Legal activities	4	3
7120	Technical testing and analysis	3	2
7420	Photographic activities	3	2
N	**Administrative and support service activities**		
7710	Renting and leasing of motor vehicles	4	1
7729	Renting and leasing of other personal and household goods	3	2
7911	Travel agency activities	2	2
7912	Tour operator activities	4	3
P	**Education**		
8510	Pre-primary and primary education	3	2
8521	General secondary education	3	2
8522	Technical and vocational secondary education	4	3
8550	Educational support activities	3	1
R	Arts, entertainment and recreation		
9311	Operation of sports facilities	5	4
9312	Activities of sports clubs	2	2
9321	Activities of amusement parks and theme parks	3	1
S	**Other service activities**		
9601	Washing and (dry-) cleaning of textile and fur products	4	3
9602	Hairdressing and other beauty treatment	5	5
9603	Funeral and related activities	3	4
9609	Other personal service activities n.e.c.	2	2

Alternative Data Sources in Official Statistics

Multi-source Usage

o This table shows the frequency of multisource statistics per industry and class. For example, a result/count of six means that six unique statistical institutions have reported using multiple sources of data to produce a particular index.

o At least two or more countries using a multi-source approach increased from 27 classes to 57 classes between 2022 and 2023 (results in the table show three or more)
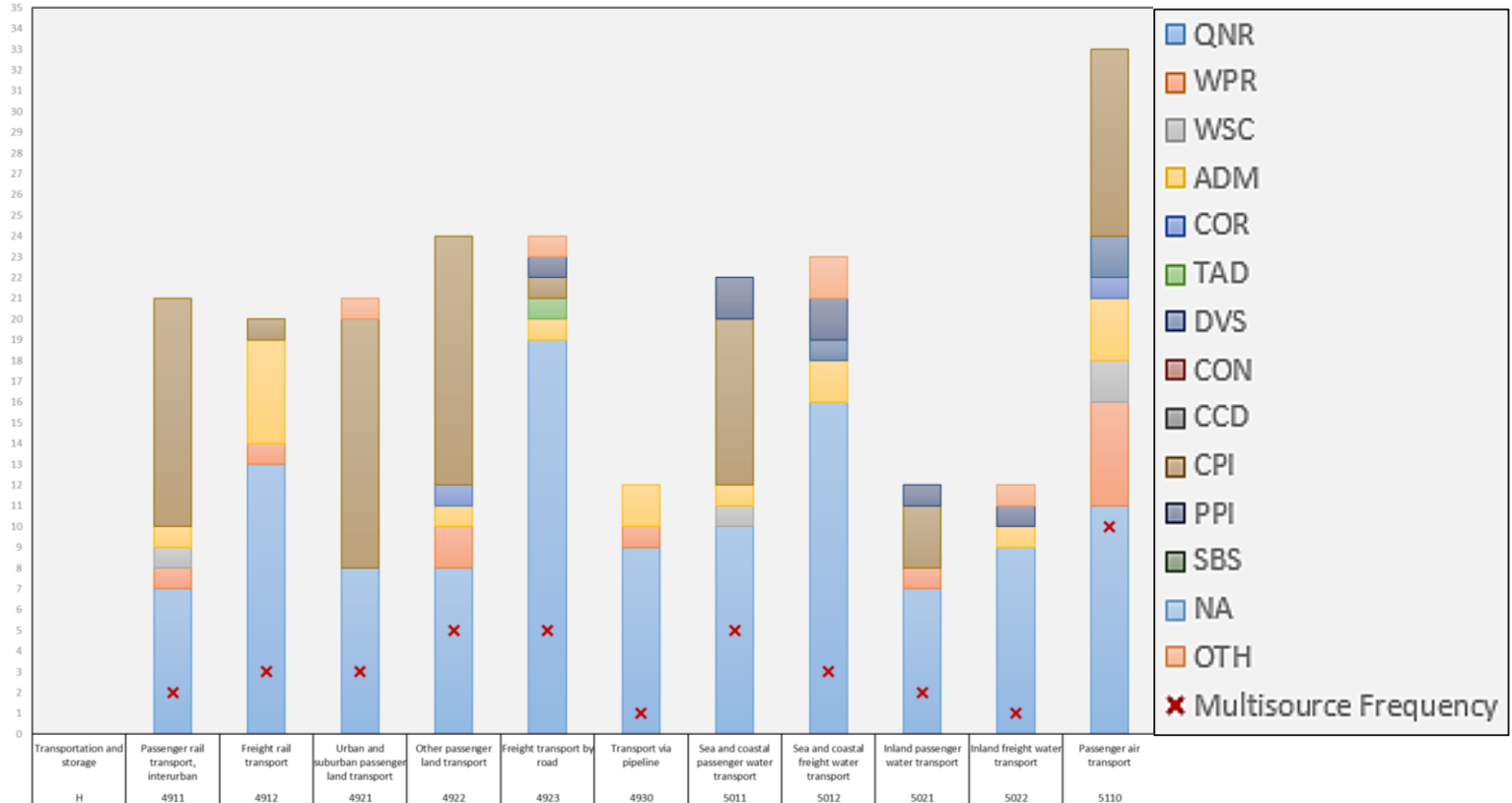
Occurance of institutions using multiple data sources for a particular 4-digit classification.			
Industry	Class	Name	Count
H	5110	Passenger air transport	10
J	6110	Wired telecommunications activities	8
H	5310	Postal activities	7
I	5510	Short term accommodation activities	7
J	6120	Wireless telecommunications activities	7
M	6910	Legal activities	7
L	6820	Real estate activities on a fee or contract basis	6
H	4922	Other passenger land transport	5
H	4923	Freight transport by road	5
H	5011	Sea and coastal passenger water transport	5
H	5221	Service activities incidental to land transportation	5
I	5610	Restaurants and mobile food service activities	5
H	5320	Courier activities	4
N	7710	Renting and leasing of motor vehicles	4
N	8129	Other building and industrial cleaning activities	4
M	7310	Advertising	4
N	7730	Renting and leasing of other machinery, equipment and tangible goods	4
G	4520	Maintenance and repair of motor vehicles	3
H	4912	Freight rail transport	3
H	4921	Urban and suburban passenger land transport	3
H	5012	Sea and coastal freight water transport	3
I	5590	Other accommodation	3
I	5630	Beverage serving activities	3
J	5813	Publishing of newspapers, journals and periodicals	3
J	6190	Other telecommunications activities	3
K	6419	Other monetary intermediation	3
L	6810	Real estate activities with own or leased property	3
M	7120	Technical testing and analysis	3
N	8121	General cleaning of buildings	3

ISRAEL
5 | 7 | 19 | 21
MARCH
2024

H – Transportation and Storage (example)

Visualisation

o This visualization shows the frequency that each price collection method is being utilised on a per class basis (4-digit level)

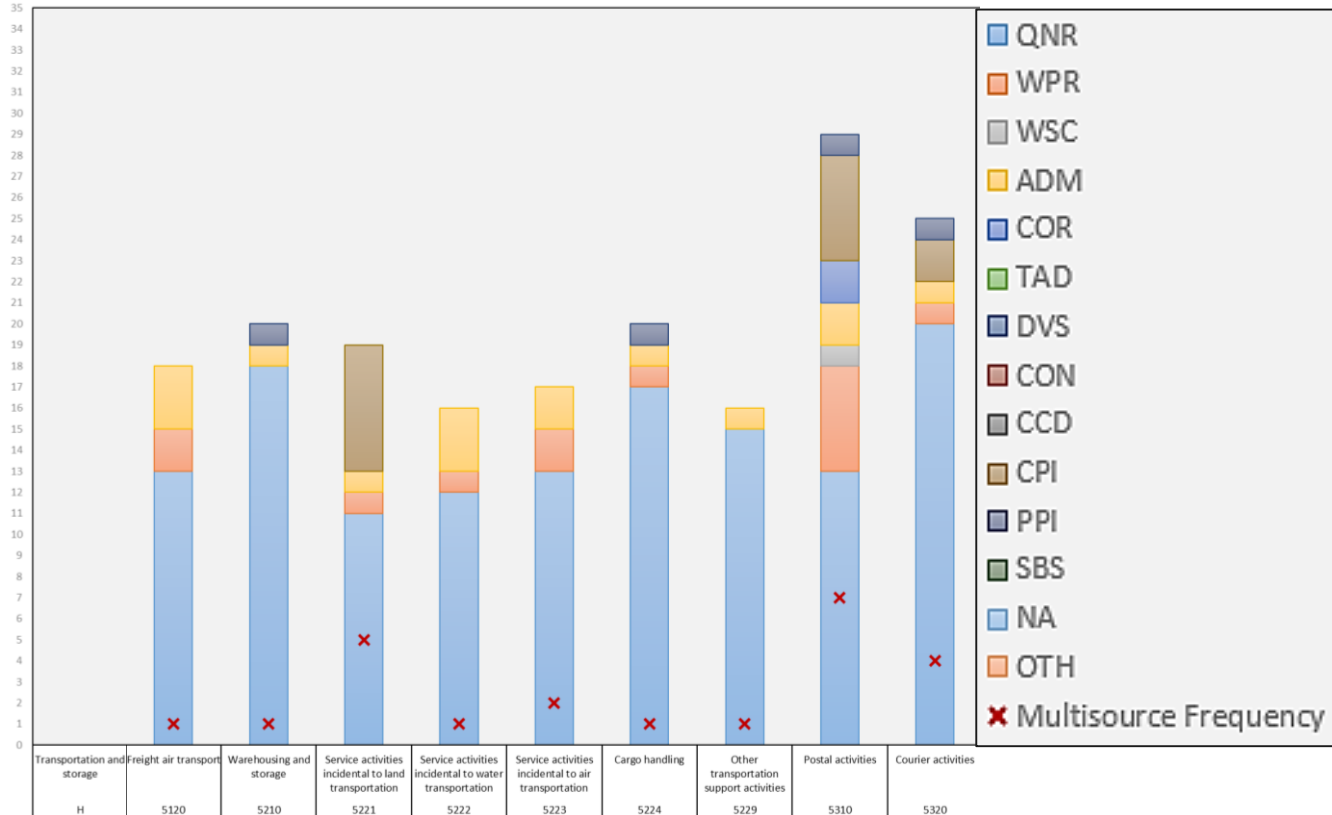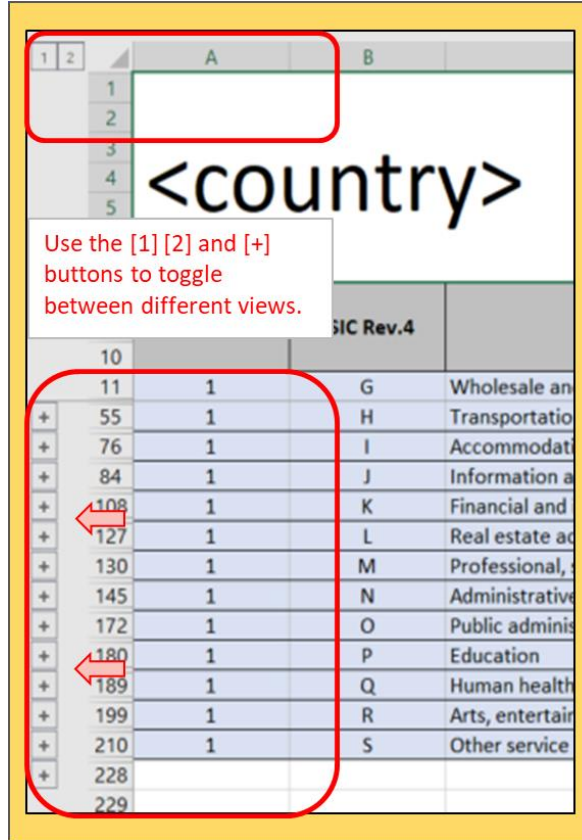o The **x** shows the frequency that unique NSIs are using multiple sources to produce a particular 4-digit index

H – Transportation and Storage (example)

Visualisation

o This visualization shows the frequency that each price collection method is being utilised on a per class basis (4-digit level)

o The **x** shows the frequency that unique NSIs are using multiple sources to produce a particular 4-digit index

Alternative Data Sources in Official Statistics

Survey and Results Guidelines

o Please remember to expand the industries via the '+' icon or [1] / [2] buttons

o This enables completion of information at the 4-digit level

Alternative Data Sources in Official Statistics

Survey and Results Guidelines

o In the "Results_Per_Country" sheet, data can be sorted by 4-digit level classification via "Column2"

o For those developing or reviewing an index this provides a good overview of coverage per country

Note: A visualisation of this information has been suggested for a future version of the results

ISRAEL
5 | 7 | 19 | 21
MARCH
2024

What's next?

o If members have requests and/or ideas on other ways the results can be presented please contact the presenters.

o Attaining increased coverage remains a top priority. Please contact the Voorburg Bureau via Rohan with contributions and/or queries.

Administrative Data Visualisation.

Other.

Feedback and/or ideas from participants can be sent to Rohan for consideration by the Voorburg Bureau. If time permits, additions to the excel summary document will be actioned by the 15th March.

Rohan Draper (Denmark)
rjd@dst.dk

STATISTICS
DENMARK