

Methodological Paper for the WTO New and Full Subsidy Notifications Database

Mun How MONG*
MAIS, ERSD, World Trade Organization

Shuhui XIANG
MAIS, ERSD, World Trade Organization

* Data Analyst, Market Access Intelligence Section, Economic Research and Statistics Division, World Trade Organization. Rue de Lausanne 154, 1202 Geneva, Switzerland. E-mail: munhow.mong@wto.org.

1 INTRODUCTION AND BACKGROUND

The database of subsidy notifications has been prepared by the WTO Secretariat to present in Microsoft Excel ("Excel") spreadsheet format the information contained in all new and full subsidy notifications submitted by Members in the 2021 notification cycle (G/SCM/N/372... series) that contain substantive information. In other words, the database does not include "nil" notifications. These notification documents in the G/SCM/N/372... series are the exclusive data source reflected in the database. The database was created by using automated data extraction tools to create the Excel file, each entry of which was then verified manually.

The methodology for extracting the information from the original notification documents submitted in Word format into Excel format is devised in a manner that minimizes information loss with only the following changes to the extracted text: i) minor changes to formatting (converting everything to lower case)¹, ii) loss of indenting within responses, iii) renumbering of some headings, and iv) changes to table formatting.² Where necessary, "<tags>" are added to identify additions to the data such as with "<footnotes>". As discussed later, multiple other hurdles were also overcome such as inconsistent wording of questions, merged cells in tables, or differences in presentation of responses. This paper provides examples and guides the user on how to interpret the manner in which the data have been extracted and documents the steps where information might have been lost or when changes to the notifications were necessary for extraction.

The approach was designed with the central aim of minimising changes to the content of notifications submitted by members in the extracted dataset. Only where necessary, such changes are made in order to facilitate extraction and do not change the content itself.³ The objective of this dataset is to facilitate the comparison and analysis of the content inside these notifications and some loss in fidelity is to be expected when extracting the notifications to an Excel database which is more limited in its display and formatting capabilities compared to Word.

Despite our best efforts, some errors may have escaped detection. We very much welcome suggestions or comments that can help improve the quality of the extracted data The dataset user is reminded to refer to the original notification documents for reference or verification purposes. For ease of reference, there is always a direct link to the respective notification document for every extracted programme in the dataset.

The rest of this paper is structured as follows. Section 2 describes the data, section 3 describes the extraction methodology, and section 4 presents a snapshot of the resulting dataset(s). New and full subsidy notifications shall be referred to simply as "notifications" for the rest of this paper.

2 DATA

As of 23 May 2023, the dataset contains notifications from the following members:

Albania; Argentina; Australia; Austria; Belgium; Brazil; Bulgaria; Cambodia; Canada; Chile; China; Costa Rica; Croatia; Cyprus; Czech Republic; Denmark; Dominican Rep.; El Salvador; Estonia; European Union; Finland; France; Germany; Greece; Honduras; Hong Kong; Hungary; Iceland;

¹ The main implication of this is that currencies may appear confusing. For example, the Albanian Lek goes from "ALL" to "all" and may be read as a word instead of an acronym. However, chances for misinterpretation are considered to be low due to the presence of other content providing the context facilitating interpretation.

² There is one other minor loss in extraction that pictures cannot be transferred to Excel, but this affects only two notifications (Romania (N372EUA23) and Dominican Republic (N372DOM)) so far. The next iteration of the dataset will implement a tag to indicate the response has pictorial content.

³ For example, we may modify the formatting of headings and subheadings in order for our script to be able to identify these headings but doing so triggers the automatic numbering feature in Word which adds additional numbers to the front of the heading. However, we stress that doing so does not alter the content or substance of the heading and preserves the hierarchy of the document.

Ireland; Israel; Italy; Japan; Kazakhstan; Korea, Rep. of; Lao, People's Democratic Rep. of; Latvia; Liechtenstein; Lithuania; Luxembourg; Macao, China; Malaysia; Malta; Mauritius; Mexico; Montenegro; Netherlands; New Zealand; Norway; Philippines; Poland; Portugal; Romania; Seychelles; Slovak Republic; Slovenia; Spain; Sweden; Switzerland; Chinese Taipei; Thailand; Türkiye; Ukraine; United Kingdom; United States.⁴

The notifications are all available in Word document format and the responses for individual programmes in the notifications follow the standardised 10-question format agreed by Members (G/SCM/6/Rev.1). The questions as contained in this format are shown in Annex A. Given the standardised format of the Word notification documents, it is feasible to extract their contents into a structured Excel format such that each row represents a single programme with individual columns containing the responses to each question. However, the extraction process is not straightforward because of difficulties such as inconsistencies in question wording, formatting, and structure of responses (free text or tabular form), or even the general formatting of the Word document. The next section discusses the difficulties and technical methods by which these challenges were addressed.

3 METHODOLOGY

The general approach is that we employ a combination of methods from the packages `win32com`⁵, `shutil`⁶, and `beautifulsoup`⁷ in Python programming language to perform various data engineering steps and extract the content from Word documents to an Excel table in a largely automated fashion.

Broadly speaking, we first convert the Word document structure and formatting into tags in HTML, then exploit the tags to differentiate among programmes, questions, and responses. To further simplify the extraction procedure, all content is converted to lowercase which means there may be information loss from formatting; however, this was found not to cause material changes to the extracted information so was deemed acceptable. Thereafter, we performed the steps outlined in the rest of this section to fully extract the data into a structured Excel tabular format.

The remainder of this section first introduces the general layout of programme in a notification document, followed by how we i) identify questions from notifications and assign them the correct question number that follows G/SCM/6/Rev.1, ii) extract different types of responses, iii) extract notifications that have been notified completely in tabular format, iv) handle footnotes and hyperlinks, and v) verify extracted data. The document G/SCM/6/Rev.1 shall be referred to simply as the "questionnaire" for the rest of this paper.

3.1 General layout of a programme in a notification document

For ease of exposition in later sections, it is necessary to first introduce the general layout of programmes in a notification document. The screenshot below is from an actual notification document with our own highlighting⁸:

⁴ G/SCM/W/546/Rev.14, Annex B.

⁵ <https://pbpython.com/windows-com.html>

⁶ <https://docs.python.org/3/library/shutil.html>

⁷ <https://pypi.org/project/beautifulsoup4/>

⁸ G/SCM/N/343/ARG

Figure 1: Example of notification

2 FORESTRY ACTIVITY											
1. Title of the programme	Law No. 25.080 - Investment in Cultivated Woodland										
2. Period covered by the notification	1 July 2016 to 30 June 2018.										
3. Policy objective and/or purpose	The considerable development potential of Argentina's forestry sector has thus far barely been explored. The purpose of the legislation in this area is therefore to promote forestry activity throughout the country, in the interests of fostering medium- and long-term investment and generating employment.										
4. Background and authority	Law No. 25.080: Published in the OJ of 19 January 1999, as amended or supplemented by the following legislation: <ul style="list-style-type: none"> - Law No. 26.432: Published in the OJ of 29 December 2008. - SAGYP Resolution No. 33/2013: Published in the OJ of 8 January 2014. 										
5. Form of the subsidy	Non-repayable economic support and fiscal benefits.										
6. To whom and how the subsidy is provided	Law No. 25.080: Beneficiaries are natural or legal persons making actual investments in the activities covered by the Law. Non-repayable economic support is a money amount that is calculated per hectare and varies according to zone, species and forestry activity, as determined by the implementing authority. Fiscal benefits are granted to natural or legal persons engaged in activities covered by the Law, which provides tax stability and exemptions.										
7. Total amount	The total non-repayable amounts disbursed for economic support in the period 2016 to mid-2018 were as follows:										
	<table border="1"> <thead> <tr> <th>Year</th> <th>Amount (\$)</th> </tr> </thead> <tbody> <tr> <td>2016</td> <td>267,049,743.95⁹</td> </tr> <tr> <td>2017</td> <td>198,531,613.25</td> </tr> <tr> <td>2018</td> <td>87,140,023.67 (estimate)</td> </tr> <tr> <td>Grand total</td> <td>552,721,380.87 (estimate)</td> </tr> </tbody> </table>	Year	Amount (\$)	2016	267,049,743.95 ⁹	2017	198,531,613.25	2018	87,140,023.67 (estimate)	Grand total	552,721,380.87 (estimate)
Year	Amount (\$)										
2016	267,049,743.95 ⁹										
2017	198,531,613.25										
2018	87,140,023.67 (estimate)										
Grand total	552,721,380.87 (estimate)										
	Source: National Directorate of Industrial Forestry Development - Ministry of Production and Labour.										
8. Duration	The benefits granted under Law No. 25.080 apply to all undertakings approved within ten years of the enactment of the Law and any extensions thereof and amendments thereto.										

There are four main categories of information as follows:

1. Highlighted in **yellow**, these are headings⁹ against which multiple programmes may be organised. In this example, "Forestry Activity" is a heading under which one of the programmes is "Law No. 25.080 – Investment in Cultivated Woodland".
 - Hierarchy for Headings is preserved by ">": the way to interpret ">text1>text2" is that text1 is the main heading, followed by subheading text2. An example using actual notification text follows later in the paper.
2. Highlighted in **blue**, these are the questions by which the notification for each programme should be structured.
3. Highlighted in **green**, these are the responses to the respective questions.
4. Highlighted in **red**, responses may be presented in tabular format.

As will be shown below, some Members present responses to entire programmes in tabular format, and we will discuss below how they are extracted.

⁹ The numbering for headings may sometimes be modified as part of the extraction process. However, the sequence in which these headings are numbered in the document remain the same and hence the hierarchy is preserved in the extracted dataset.

3.2 Identifying questions and coding

While Members submit their notifications based on the questionnaire format, their wording of the question may vary. For example, the standard wording of question 1 of the questionnaire is as follows:

1. *Title of the subsidy programme, if relevant, or brief description or identification of the subsidy.*

However, Members might choose to phrase the question differently, such as¹⁰:

- Argentina: 1. Title of the programme
- Korea: 1. Title of the subsidy programme
- Norway: Skips question 1 entirely
- New Zealand: A. Name of programme
- Kazakhstan: Title of the subsidy program

This gives rise to three pertinent issues that affect the data extraction process:

1. Wording of the questions varies across notifications, and sometimes *even within the same notification document*.
2. Members use different numbering formats to distinguish notification questions. Argentina and Korea use "1" while New Zealand chooses "A". Kazakhstan underlines the questions instead of using numbering to label questions.
3. Members can also skip questions. Therefore, questions and corresponding responses may not be ordered in the same manner as in the questionnaire. For example, Norway's first question in its notification is usually "1. The authority responsible for the subsidy", which actually corresponds to the fourth question in the questionnaire.

There are several implications for data extraction. Firstly, it is not straightforward to assign the appropriate questionnaire number to the question in the notification in an automatic fashion using string matching. For example, assigning question 1 to the string "Title of the programme" as phrased in Argentina's notification will miss out the phrasing in Korea's and Norway's notifications. Also, using numbering format matching is non-trivial; although most members use numerical symbols like Argentina and Korea in the example above, other members may choose alphabet or special font formatting to differentiate between questions and responses.

Additionally, it is not possible to use the order of notification questions as a guide to identifying which question from the questionnaire that the notification question relates to. This is because it is not possible to assume that members provide responses to all 10 questions required by the questionnaire and also arrange them in the same order as set out the questionnaire. In the above example, Norway does not state the programme title as its own question and its first question for each program usually refers to the fourth question from the questionnaire instead.¹¹ Therefore, it would be incorrect to assume that the first question of each programme in any Member's notification always refers to the title of the programme.

Therefore, we settled on identifying and assigning question numbers to questions across all notification documents manually. This ended up being the most laborious part of the process because machine learning methods include a margin of error and more time would be spent correcting errors than if we coded manually. However, as discussed later, we have manually verified whether the questions have been assigned correctly and hence it is unlikely for mistakes to have occurred in

¹⁰ Examples from G/SCM/N/372/ARG, G/SCM/N/372/KOR, G/SCM/N/372/NOR, G/SCM/N/372/NZL, G/SCM/N/372/KAZ respectively.

¹¹ In such cases, headings are assigned as programme titles if members skip question 1. A text "1. title of the subsidy programme" is also added as the question text. Refer to N372GBR for an example.

coding of the questions and escaped detection in the dataset. It is stressed that the questionnaire number is always assigned based on the question wording and not the content of the response.

3.2.1 Mixed questions

Members may sometimes rephrase questions that originally in the questionnaire were presented separately. Take for example the following¹²:

Figure 2: Example of mixed questions

2. Duration of the programme and period covered by the notification
Such income tax allowances are in force since 1993 until the quota is used up. This notification covers period 2018 - 2019.]

In the questionnaire, the duration of the program is question 9 while period covered by the notification is question 2. While it might make sense to combine the responses to both questions since they pertain to time, doing so adds another layer of complexity to the structure of the dataset because the responses in this notification no longer follow the same format as in the questionnaire.

In such cases, we devised a coding formula to assign special question numbers according to two rules as defined below:

1. Choose the smallest question number as the basis for comparison.
2. Assign alphabets to the other question numbers to represent how far apart from the number the other questions are.

The question in the above example is a combination of question 9 and 2 from the questionnaire. We first use question 2 as the basis, and since question 9 is 7 steps away from 2, we assign it the alphabet "g", and hence the question in this example is coded as question "2g".¹³ This example shall be referred to as (question 2+9=2g) for the rest of this paper.

In this way, we preserve the questions and responses exactly as they had been presented in the notification without loss of information, while still being able to assign the responses to the respective questions in the standardised questionnaire format.

3.3 Extracting responses

Responses in the notifications typically fall into four broad categories. Note that responses may contain a combination of these categories.

3.3.1 Null response

Members sometimes leave the response to particular questions blank – a typical example is that question 8 from the questionnaire is a composite of question 3-7 and is very often left blank if it appears in the notification document. In such cases, we replace the blank with "no response". This is not to be confused with the question not appearing at all. For example, if a notification contains question 8 but with a blank response, we will replace it with "no response", but if a notification does not contain question 8 at all, then we will not intentionally add in question 8 and fill in with a "no

¹² G/SCM/N/372/EUA21, Poland

¹³ 3, 4, 5, 6, 7, 8, 9 = 7 counts, and hence it is alphabet g, because A, B, C, D, E, F, G and G is the 7th count.

response". Note this rule only applies when the response is left blank, not when responses are "nil" or "n/a" or something to that effect; in these cases, such responses will also be extracted.

3.3.2 Regular response

Regular responses are extracted but note that they will be converted to lower-case. Because responses are extracted into a single Excel cell, multiple paragraphs in the same response are separated by "//"; in other words, paragraph breaks will be represented by "//".

Note that responses to question 10/10a/10b regarding statistical data are all concatenated into a single response in the extracted database.

3.3.3 Tabular response

JavaScript Object Notation (JSON) format is a standard text-based format for representing structured data based on JavaScript object syntax.¹⁴ We converted tables to JSON in the form of key-value pairs. In the example below, the table inside the response to question 7 is presented on the left, and the right side contains the JSON table with the extracted content:

Figure 3: Comparison of original table and extracted content

7. Total amount

The total non-repayable amounts disbursed for economic support in the period 2016 to mid-2018 were as follows:

Year	Amount (\$)
2016	267,049,743.95 ¹
2017	198,531,613.25
2018	87,140,023.67 (estimate)
Grand total	552,721,380.87 (estimate)

Source: National Directorate of Industrial Forestry Development - Ministry of Production and Labour.

JSON format table:

```
{{"year":"2016","amount $":"267049743.95[1]"}, {"year":"2017","amount $":"198531613.25"}, {"year":"2018","amount $":"87140023.67 estimate"}, {"year":"grand total","amount $":"552721380.87 estimate"}}
```

In words, each row of the JSON table corresponds to each row of the table, excluding the first row of the table which is taken to be the "Key". For example, the first row of the table is "Year" and "Amount (\$)", which means that the table has two Keys. Correspondingly, the second row of the table contains the values, where "Year" is "2016" and "Amount (\$)" is "267,049,743.95". This is then repeated for all the rows in the table.

Note that responses that appear to be tables in the Word document may sometimes be extracted as text (i.e. separated by "//" instead of JSON format). The reason for this is that while the response looks like a table in the document it is not formatted as a table and hence it was extracted as regular text. However, these entries are rare and the extracted form does not change the content of the original response. Additionally, some tables might be complex and make the JSON format difficult to interpret, but this is the only way we are aware of in which content from multiple tables can be fit into a single Excel cell. Alternatively, users may refer to the original notification document via the direct link to WTO Docs Online or process the JSON table relatively easily with software such as R or Python.

3.3.3.1 Merged columns in tables

An example best illustrates how we extracted information from tables with merged cells:

Case 1: If the first-row cells are merged, we separate them manually to ensure the "key" is correct and not missing. See below the original and modified versions.

¹⁴ <https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Objects/JSON>

Figure 4: Merged cells in first row (original)

9. Statistical data enabling the effects of the subsidy on trade to be assessed

Production and trade (in hl)

Year	Production		Imports ¹	
	white	red	white	red
2019	497,531	481,914	397,588	1,156,915
2020	413,144	421,091	392,644	1,199,550

1 Custom tariff heading: 2204.21, 2204.22 and 2204.29.

Figure 5: Unmerged cells from first row into individuals (modified)

9. Statistical data enabling the effects of the subsidy on trade to be assessed

Production and trade (in hl)

Year	White Production	Red Production	White Imports ¹	Red Imports ¹
2019	497,531	481,914	397,588	1,156,915
2020	413,144	421,091	392,644	1,199,550

1 Custom tariff heading: 2204.21, 2204.22 and 2204.29.

Case 2: If the merged cells appear in the rows other than the first row, we unmerge them and then duplicate the contents so that we can still apply the same JSON key-value logic in the extraction. In the example below, the original table is Table A and the converted table is Table B.

Table A: There is a merged cell in the first column and second row.

Key1	Key2	Key3
A1	A2	
	B2	B3

Table B: The merged cells are unmerged and contents duplicated.

Key1	Key2	Key3
A1	A2	A2
A1	B2	B3

Note that parentheses, commas, and quotation marks will be dropped from content in JSON format. Some implications include that negative numbers represented by parentheses may appear misleading but such cases are rare. The user is reminded that the original notification document is always easily accessible via a direct link in case of any ambiguity or difficulty in reading the JSON table.

3.3.3.2 Programmes notified in tabular format

Some members also notify entire programmes in tabular format. Take for example the following:¹⁵

¹⁵ G/SCM/N/372/CHN

Figure 6: Example of entire programmes in tabular format

II. SUBSIDIES AT THE SUB-CENTRAL GOVERNMENT LEVEL												
1. BEIJING MUNICIPALITY												
No.	Programme Authority	Programme Title	Policy objective	Legal Basis	Form of subsidy	To whom the subsidy is provided	How the subsidy is provided	Duration of the subsidy programme	Period covered by the notification	Total Amount or Annual Budget (RMB 10,000 Yuan)		Trade Effects
										2019	2020	
1	Beijing Municipality	Fund for Energy Conservation, Emission Reduction and Environmental Protection	To promote energy conservation and emission reduction	Jing Cai Jing Yi No. 1917 of 2017	Financial appropriations	Eligible enterprises	Grant	2017 to present	2019-2020	107143.02	11239.68	Not available
2	Beijing Municipality	Fund for Resolving of Overcapacity in Coal Sector and Resettlement and Relocation of Employees	To resolve overcapacity in coal sector	Jing Cai Jing Yi No. 2276 of 2016	Financial appropriations	Eligible enterprises	Grant	2016-2020	2019-2020	5000	5000	Not available
3	Beijing Municipality	Fund for New Energy Passenger Vehicles	To save energy and protect the environment	Jing Cai Jing Yi No. 1296 of 2018	Financial appropriations	Eligible enterprises	Grant	2017 to present	2019-2020	38975	49520	Not available
4	Beijing Municipality	Fund for Construction of Charging Infrastructures of Electric Vehicles	To save energy and protect the environment	Jing Zheng Ban Fa No. 36 of 2017	Financial appropriations	Eligible enterprises and projects	Grant	2018 to present	2019-2020	Amounts varied		Not available
5	Beijing Municipality	Fund for Promotion and Using of Pure Electric Vehicles in Taxi Industry	To save energy and protect the environment	Jing Cai Gong Yong No. 1325 of 2019	Financial appropriations	Eligible taxi operators	Reward	July 2019 - December 2021	2019-2020	29520	73800	Not available
6	Beijing Municipality	Fund for Promotion and Using of New Energy Light Trucks	To save energy and protect the environment	2020 Plan of Beijing Municipality on Incentives for Operation of New Energy Light Trucks	Financial appropriations	Eligible road freight transport operators	Reward	2020 to present	2019-2020	Not applicable	Amounts varied	Not available

In the example above, each row contains the information for one programme, and each column corresponds to a question from the questionnaire. In such cases, we extract the information from these tables in a similar way as described in section 3.3 into JSON format, and further extract information from the JSON table into individual rows in Excel such that each programme from these tables is represented similarly as programmes that had not been presented in tables.¹⁶

Note that because parentheses, commas, and quotation marks will be dropped from the content in JSON format, these punctuation marks will also be absent from programmes that have been notified in tabular format.

3.3.4 Other extracted content

3.3.4.1 Footnotes

Responses with footnotes will contain tags "<response>" and "<footnote>". Take the following example¹⁷ extracted response from the dataset:

<response>the brønnøysund register centre on behalf of the ministry of trade, industry and fisheries.[4]/<footnote>[4] Previously The Norwegian Tax Administration on behalf of The Ministry of Finance ("Compensation Scheme 1").

The <response> tag denotes the start of the response content, and "[4]" indicates that this response contains the fourth footnote in the notification document. The <footnote> tag is followed by [4] which indicates that what follows is the content of the fourth footnote. Note that responses without

¹⁶ Note that China's 2021 notification (G/SCM/N/372/CHN) had an attachment called "List of Documents Referred to in Column "Legal Basis" in "Subsidies at the Sub Central Government Level"". We extracted these entries and added them to the response of question 4 for sub central government level programs. These can be identified with the tag <Documents Referred to in Attachment>.

¹⁷ G/SCM/N/372/NOR

footnotes will not have the <response> nor <footnote> tags; also note that these tags should appear together or not at all.

An example¹⁸ below shows how footnotes in headings are represented:

<heading>>1 agriculture>1.2 other agricultural programmes[1]>1.2.1 expensing of multi-period livestock and crop production costs <heading footnote> [1] all the programmes in this section are tax-related, authorized under the u.s. internal revenue code.

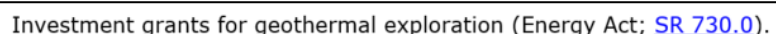
The representation for footnotes in headings is similar to that of footnotes in responses, except that the tags are named as "<heading>" to identify the wording and <heading footnote> to identify the relevant footnote, with "[1]" to indicate the content of the first footnote. Again, headings without footnotes will not have these tags and that they should appear together or not at all.

3.3.4.2 Hyperlinks

Responses with embedded hyperlinks will contain tags "<response>", "<hyperlink text>", and "<hyperlink>". The reason for an additional tag "<hyperlink text>" is because there is no "[X]" denoting the position of the embedded hyperlink as there is in footnotes. Take the following example¹⁹ extracted response from the dataset:

<response> investment grants for geothermal exploration (energy act; sr 730.0). <hyperlink text> SR730.0 <hyperlink> <https://www.fedlex.admin.ch/eli/cc/2017/762/fr>

Figure 7: Actual text from document



Investment grants for geothermal exploration (Energy Act; [SR 730.0](https://www.fedlex.admin.ch/eli/cc/2017/762/fr)).

Compare the extracted response with the screenshot from the actual document. As before, the <response> tag identifies that this response content has additional information, and "<hyperlink text>" indicates the part of response content that contains an embedded hyperlink. This is then followed by the actual link itself denoted by <hyperlink>.

3.3.5 Verification of extracted data

Lastly, we conduct a manual verification of the extracted content against the original notification documents in order to ensure as far as possible that i) content have been extracted without change in meaning, ii) the questions have been assigned/coded correctly, and iii) there is no material loss of information in the extracted data compared to the original documents (e.g., through formatting or otherwise).

3.3.6 Corrigendums

Corrigendums may contain corrections or clarifications by Members regarding their original notification. In such cases, the affected content of the original notification will be updated accordingly in the extracted dataset.

4 RESULTS

The extracted dataset can be presented in three formats as described below.

¹⁸ G/SCM/N/372/USA

¹⁹ G/SCM/N/372/CHE

4.1 Long format

In the long format, each row contains the response to an individual question for each programme. The advantage is that the exact question wording is preserved, but it is difficult to use for comparison or analysis. A screenshot of an actual notification text²⁰ is first provided to assist the reader in visualising how the data is transposed from Word document to Excel:

Figure 8: Example of notification

10 RENEWABLE ENERGY PROGRAMS

10.1 Advancing Renewables Program

1. Title of the subsidy

Advancing Renewables Program.

2. Period covered by the notification

Financial years 2019/2020 and 2020/2021.

3. Policy objective and/or purpose of the subsidy

The Advancing Renewables Program supports a broad range of development, demonstration and pre-commercial deployment projects that can deliver affordable and reliable renewable energy for Australian families and businesses. The Program Objective is to fund Activities that contribute to one or more of the following. Program Outcomes:

- (a) reduction in the cost of renewable energy;
- (b) increase in the value delivered by renewable energy;
- (c) improvement in technology readiness and commercial readiness of Renewable Energy Technologies;
- (d) reduction in or removal of barriers to renewable energy uptake;
- (e) increased skills, capacity and knowledge relevant to Renewable Energy Technologies".

4. Background and authority for the subsidy

The ARP was launched in July 2015.

The Australian Renewable Energy Agency (ARENA) was established by the Australian Government on 1 July 2012 under the *Australian Renewable Energy Agency Act 2011* (as amended) (ARENA Act) with a remit to improve the competitiveness and increase the supply of renewable energy in Australia.

5. Form of the subsidy

Funding is provided against specific milestones in accordance with individual funding agreements.

6. To whom and how the subsidy is provided

Funds were committed to 74 projects through individual funding agreements. The dollar value for each funding agreement varied depending on the specific project. Applicants were assessed against merit criteria and specific eligibility requirements. Applications were assessed by a technical advisory group.

7. Level of subsidy per unit

Funding ranged from AUD 136,000 to AUD 47 million per project. The total commitments made for 2019/2020 and 2020/21 amounted to AUD 291 million.

8. Duration

Duration varies according to individual funding agreements. The last funding agreement expires on 30 June 2020.

9. Statistical data permitting an assessment of the trade effects

No statistical data permitting an assessment of trade effects is available.

The programme in the above notification is extracted into Excel and presented as follows:

²⁰ G/SCM/N/372/AUS

Figure 9: Example of notification after extraction

	A	B	C	D	E	F
1	FileName	Heading	Question	qn_num	Text	DOL
128	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	1. title of the subsidy	q1	advancing renewables program.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
129	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	2. period covered by the notification	q2	financial years 2019/2020 and 2020/2021.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
130	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	3. policy objective and/or purpose of the subsidy	q3	the advancing renewables program supports a broad range of development, demonstration and pre-commercial deployment projects that can deliver affordable and reliable renewable energy for australian families and businesses. the program objective is to fund activities that contribute to one or more of the following. program outcomes://(a) reduction in the cost of renewable energy;//(b) increase in the value delivered by renewable energy;//(c) improvement in technology readiness and commercial readiness of renewable energy technologies;//(d) reduction in or removal of barriers to renewable energy uptake;//(e) increased skills, capacity and knowledge relevant to renewable energy technologies".	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
131	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	4. background and authority for the subsidy	q4	the arp was launched in july 2015.//the australian renewable energy agency (arena) was established by the australian government on 1 july 2012 under the australian renewable energy agency act 2011 (as amended) (arena act) with a remit to improve the competitiveness and increase the supply of renewable energy in australia.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
132	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	5. form of the subsidy	q5	funding is provided against specific milestones in accordance with individual funding agreements.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
133	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	6. to whom and how the subsidy is provided	q6	funds were committed to 74 projects through individual funding agreements. the dollar value for each funding agreement varied depending on the specific project. applicants were assessed against merit criteria and specific eligibility requirements. applications were assessed by a technical advisory group.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
134	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	7. level of subsidy per unit	q7	funding ranged from aud 136,000 to aud 47 million per project. the total commitments made for 2019/2020 and 2020/21 amounted to aud 291 million.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
135	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	8. duration	q9	duration varies according to individual funding agreements. the last funding agreement expires on 30 june 2020.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS
136	N372AUS	>10 renewable energy programs>10.1 advancing renewables program	9. statistical data permitting an assessment of the trade effects	q10	no statistical data permitting an assessment of trade effects is available.	https://docs.wto.org/dol2fe/Pages/FE_Search/FE_S_S006.aspx?SymbolList=G/SCM/N/372/AUS

The individual components of the extracted dataset are as follows:

1. Column A contains the filename of the notification. The filename is N372AUS in this example, where N372 corresponds to the document code for 2021 notifications and AUS is the code associated with Australia.
2. Column B contains the headings in the document. As noted previously, ">" in column "Heading" indicates the level of the heading:
 - a. Cell B128 shows heading ">10 renewable energy programs>10.1 advancing renewables program". This means that "10 renewable energy programs" is the main heading, and "10.1 advancing renewables program" is a subheading under the main heading.

Note that headings are repeated for all following programmes and their respective questions-responses. In this way, the content hierarchy is preserved in the database as it appears in the original notification.

3. Column C contains the exact question text as it appears in the notification, and column D contains the corresponding question number from the standard questionnaire format.

- a. Because the programme in this notification provided responses for questions 3-7 individually, question 8 from the questionnaire format²¹ becomes irrelevant and may have been the reason it was omitted from the notification. Therefore, question 8 for this notification refers to "duration", which corresponds to question 9 of the questionnaire. It follows that question 9 for this notification refers to "statistical data...", which then corresponds to question 10 of the questionnaire. This is a very common reason why question numbers in the notifications often do not correspond directly to the numbers in the standard questionnaire format.
4. Column D contains the question number from the questionnaire that corresponds to a particular question from the notification. For example, the question in cell C135 is "8. duration" which actually corresponds to question 9 in the questionnaire.
5. Column E contains the responses to the respective questions in the notification. As noted previously, "/" represents paragraph breaks and using the response in cell E130 as an example, each of the bullet points in the response are separated by "/". This enhances readability and preserves the paragraphing of the original response in the notification. However, note that indentation within responses is not preserved in the extracted dataset.
6. Column F contains the weblink to the original notification that can be downloaded from the WTO Docs Online portal and is repeated for all rows pertaining to the particular Notification document from which the information had been extracted.

4.2 Wide format

In wide format, each row contains all the responses to each programme which makes it easier for comparison and analysis but loses the column with actual wording of each question. Note that compared to the long format, all rows that do not have an identified question number will also be removed from this version of the dataset.

A screenshot is presented as follows, with pink highlighted cells as an example of the question appearing in the notification but for which there was no response and hence "no response" had been added:

Figure 10: Screenshot of dataset in wide format with "no response"

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	FileName	Heading	q1	q2	q2g	q3	q3abcd	q3c	q4	q5	q5a	q5ab	q5b	q6	q7	q8	q9	q10	DOL
1032	N372EUA10	>2 regional schemes>2.3 oversea	2.3.3 transpo	2019 and 2020.		the objective of the measure is t	law no. 20	direct subsidy.						beneficiari 2019: eur 17.42 milio	the schem	unavailabl	https://do		
1033		>2 regional schemes>2.3 oversea	2.3.4 call for	2017 and 2018.		to support the purchase of capa	aid schem	subsidy.						beneficiari 2019: eur 3,898,259	/no respon	no respon	https://do		
1034		>2 regional schemes>2.4 urban	2.4.1 tax con	2019 and 2020.		to revitalize economic activity ar	- article 44	article 44	quaterdecies of the general taxat					beneficia - article 44 octies://tc	article 44	- article 44	https://do		
1035		>3 research and development	(i 3.1 innovatio	2019 and 2020.		to help enterprises develop a re	aid schem	subsidy, repayable advance, loan.						beneficiari amounts paid://-	201	no respon	2,854 beni	https://do	
1036		>3 research and development	(i 3.10 energy t	2019 and 2020.		to create leading technological i	aid schem	subsidy.						no respon	eur 17.5 million comr	no respon	no respon	https://do	
1037		>3 research and development	(i 3.11 young ir	2019 and 2020.		to help create innovative enterp	article 44	: tax and social benefits.						beneficiari amount of tax exemp	enterprises created b	https://do			
1038		>3 research and development	(i 3.12 competi	2019 and 2020.		to increase the competitiveness: aid schem	subsidy.							- beneficia	amount for 2019: eur agreement-	over 10.0	https://do		

Another screenshot is presented as follows, using the same example of special question number for combined questions discussed previously (question 2+9=2g), with the affected response highlighted in yellow:

²¹ Question 8 from the questionnaire reads as: *For the information cited in items 3 to 7 above, the notification does not necessarily have to have an independent heading corresponding to each item, and may provide information on multiple items in one heading (e.g. provide information on items 3 and 4 under one heading). In this case, the notification must clearly specify what items are covered by which heading.*

Figure 11: Screenshot of dataset in wide format with special questions

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	FileName N372EU02	Heading >1 income tax allowances for joint venture enterprises	q1 income tax allowances are established for joint venture enterprise and granted by the minister of finance, taking individual decisions in this respect. in case of positive decision of the minister of finance, the enterprise has the possibility of exercising the right to the tax allowance in the first year when it gains profit.	q2	q2g such income tax allowances are in force since 1993 until the quota is used up. this notification covers period 2018 - 2019.	q3 the creation of incentives for foreign investments.	q3abcd	q3c	q4 the authority responsible for the program is the minister of finance. the legal basis for implementation of the program is art. 23 of the act of 14 June 1991 on companies with foreign shareholdings (journal of laws of 1997, item 143), read in conjunction with art. 93 of the act of 19 November 1999 on business activities (journal of laws 1999, item 1178).	q5 income tax allowance s.	q5a	q5ab	q5b	q6 in 2018 income tax allowances were granted to 2 legal entities. in 2019 income tax allowance s were granted to 1 legal entity. no data available for 2020.	q7 in 2018 the assistance granted amounted to 8,000,000. in 2019 the assistance granted amounted to 3,000,000. no data available for 2020.	q8	q9	q10 the program does not directly impact the foreign trade.	DOL https://d ocs.wto.o rg/dol2fe /Pages/FE _Search/F E_5_S006. asp?Sym bolList=G /SCM/N/3 72/EU/Ad d.21

4.3 Standardised wide format

This format is similar to the wide format, except that the columns for special question numbers have been removed and the responses repeated for the combined questions. Using the previous example of mixed questions "2g", which refers to the response in the original notification combining responses to both questions 2 and 9, question "2g" has been removed from the wide format and the response is duplicated in columns for questions 2 and 9 for the standardised wide format.

The dataset presented in this manner loses the information that questions 2 and 9 had been combined for a particular programme in a particular notification but duplicating the responses in this manner produces a truly standardised dataset with responses for all notifications harmonised across the common 10-question questionnaire.²²

A screenshot is presented as follows, using the same example of special question number for combined questions discussed previously (question 2+9=2g), with the now-duplicated responses highlighted in **yellow**:

Figure 12: Screenshot of standardised wide format

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FileName N372EU02	Heading >1 income tax allowances for joint venture enterprises	q1 income tax allowances are established for joint venture enterprise and granted by the minister of finance, taking individual decisions in this respect. in case of positive decision of the minister of finance, the enterprise has the possibility of exercising the right to the tax allowance in the first year when it gains profit.	q2 such income tax allowances are in force since 1993 until the quota is used up. this notification covers period 2018 - 2019.	q3 the creation of incentives for foreign investments.	q4 the authority responsible for the program is the minister of finance. the legal basis for implementation of the program is art. 23 of the act of 14 June 1991 on companies with foreign shareholdings (journal of laws of 1997, item 143), read in conjunction with art. 93 of the act of 19 November 1999 on business activities (journal of laws 1999, item 1178).	q5 income tax allowance s.	q6 in 2018 income tax allowances were granted to 2 legal entities. in 2019 income tax allowances were granted to 1 legal entity. no data available for 2020.	q7 in 2018 the assistance granted amounted to 8,000,000. in 2019 the assistance granted amounted to 3,000,000. no data available for 2020.	q8	q9 such income tax allowances are in force since 1993 until the quota is used up. this notification covers period 2018 - 2019.	q10 the program does not directly impact the foreign trade.	DOL https://d ocs.wto.o rg/dol2fe /Pages/FE _Search/F E_5_S006. asp?Sym bolList=G /SCM/N/3 72/EU/Ad d.21

The database that will be made publicly available is in the standardised wide format.

4.4 A note of caution when using the database

Even though it is structured as such, it would be misleading to count each row of the dataset as a unique subsidy programme because the coverage year of each program varies. Additionally, Members may still structure their notifications differently even with a standardised question

²² In some cases, mixed questions may overlap with responses to other questions. For example, some programmes in G/SCM/372/MUS have responses to q5a (a combination of q5 and q6) and also q6. In such cases, the response for q5a will be duplicated for q5, and the responses for q5a and q6 will be concatenated and replaces the response for q6.

format; for example, we sometimes observe that the same subsidy programme may appear twice in the same notification document simply because the name of the programme had changed slightly between the two years.

Again, this database does not replace the content of the original document and the user should always refer to the original notification document via the direct link for each notification on the WTO Docs Online portal where necessary.

Annex A: List of questions in G/SCM/6/Rev.1

Qn. Num.	Question
1	Title of the subsidy programme, if relevant, or brief description or identification of the subsidy.
2	Period covered by the notification. The period to be covered by the notification should be the most recently completed calendar or fiscal year. In the latter case, the start and end dates of the fiscal year should be specified.
3	Policy objective and/or purpose of the subsidy.
4	Background and authority for the subsidy (including identification of the legislation under which it is granted).
5	Form of the subsidy (i.e., grant, loan, tax concession, etc.).
6	To whom and how the subsidy is provided (whether to producers, to exporters, or others; through what mechanism; whether a fixed or fluctuating amount per unit; if the latter, how determined).
7	Subsidy per unit, or in cases where this is not possible, the total amount or the annual amount budgeted for that subsidy (indicating, if possible, the average subsidy per unit in the previous year). Where provision of per unit subsidy information (for the year covered by the notification, for the previous year, or both) is not possible, a full explanation.
8	For the information cited in items 3 to 7 above, the notification does not necessarily have to have an independent heading corresponding to each item, and may provide information on multiple items in one heading (e.g. provide information on items 3 and 4 under one heading). In this case, the notification must clearly specify what items are covered by which heading.
9	Duration of the subsidy and/or any other time limits attached to it, including date of inception/commencement.
10	Statistical data permitting an assessment of the trade effects of the subsidy. The specific nature and scope of such statistics is left to the judgement of the notifying Member. To the extent possible, relevant and/or determinable, however, it is desirable that such information include statistics of production, consumption, imports and exports of the subsidized product(s) or sector(s):
10a	for the three most recent years for which statistics are available;
10b	for a previous representative year, which, where possible and meaningful, should be the latest year preceding the introduction of the subsidy or preceding the last major change in the subsidy.