

Chapter 8

Dissemination

It can be challenging to determine how to present wealth data in a way that is useful to a variety of users, particularly because of the importance of this data in understanding the economic situation in each country. This chapter provides guidance in this field.

8.1. Types of dissemination

When disseminating wealth data, it is important to take into account the needs of the various users of the data. In order to meet this challenge, it is helpful to have a multi-faceted dissemination strategy with a variety of products to meet the needs of the different user groups.

Generally, the data products based on wealth data will fall into three broad types – analysis, data tables and micro-data. Ideally, products from any individual wealth data program would cover all three types (Box 8.1).

Box 8.1. Data products from Australia's 2009-10 Survey of Income and Housing

With the release of wealth data from the 2009-10 Survey of Income and Housing (SIH), the Australian Bureau of Statistics ensured that there was a wide variety of products available to cater to a large variety of users. These included:

- Media release – Short high-level summary (less than 2 pages) of the key findings from the data being released. Ideally suited for the media or the casual user who is looking for a few interesting results. No tables or graphs are included in such a release.
- Summary of findings – Aimed at users looking for more detailed analysis than the media release but still oriented to a general audience. This summary is less than 10 pages in length but covers the broad range of data available, and includes a few graphs and summary tables.
- Feature article – Analysis on a specific sub-population, in this case the *Low economic resource household*, exploiting the unique nature of the wealth data. Of interest to a fairly broad audience and provides an example of the type of in-depth analysis that can be achieved using wealth data.
- Data publication – Contains a number of detailed data tables covering the breadth of the data from the survey. Of interest to users who want to know more about a specific sub-group of the population.
- Micro-data file – A set of micro-data files that are screened for confidentiality (to ensure that an individual respondent cannot be identified) and can be made available to more sophisticated groups of users who want to conduct their own analyses. This is accompanied by a user's guide describing the content of the file in detail.

Readers interested in learning more about the output from the wealth component of the 2009-10 SIH should consult the ABS website (www.abs.gov.au) and search for *Household Wealth and Wealth Distribution, 2009-10*.

8.2. Analysis

Generally, two types of analyses will be produced from a wealth survey: those oriented to a general audience, and more in-depth analyses of interest to academics or policy makers. The more general analysis is often made available at the time of the initial release

of the wealth data set in order to publicise the release and highlight key findings. In-depth analysis usually takes more time and is often conducted by the more sophisticated users once the data are released by the organisation that conducted the survey. These analyses will be produced by a variety of users in various organisations and may be made available to the general public depending on the mandate of the individual or group sponsoring the analysis. In addition, the broad distribution of analysis based on the wealth data can be used to educate users on the correct interpretation of the wealth data.

8.2.1. Data tables

Data tables are one way to make a variety of data available to users who may not have the analytical skills, resources or data access required to produce their own output from the micro-data file of wealth data. Often the tabulated data are produced in a publication or in an on-line database to allow users to browse the data tables and choose those statistics that are of more interest to them. This is a way of providing broad access to a wide variety of data to a large number of users.

As noted in Chapter 3, there are many ways in which household wealth data can be grouped to look at different sub-sets of the population. Table 8.1 provides an example of how data for various groups can be presented in a table to provide a general picture of one aspect of wealth data, in this case net worth.

Table 8.1. **Family net worth by selected characteristics of families in the United States**

Thousands of 2010 US dollars

Family characteristic	2004		2007		2010	
	Median	Mean	Median	Mean	Median	Mean
All families	107.2	517.1	126.4	584.6	77.3	498.8
Percentile of income						
Less than 20	8.6	83.6	8.5	110.3	6.2	116.8
20-39.9	38.8	139.8	39.6	141.3	25.6	127.9
40-59.9	82.8	224.0	92.3	220.6	65.9	199.0
60-79.9	184.0	392.9	215.7	393.9	128.6	294.0
80-89.9	360.9	563.7	373.2	638.1	286.6	567.3
90-100	1 070	2 925	1 172	3 475	1 194	2 944
Age of head (years)						
Less than 35	16.3	84.6	12.4	111.1	9.3	65.3
35-44	79.9	345.2	92.4	341.9	42.1	217.4
45-54	167.1	625.8	193.7	694.6	117.9	573.0
55-64	290.0	976.4	266.2	986.7	179.4	880.5
65-74	218.8	795.1	250.8	1 064.1	206.7	848.4
75 or more	187.7	607.7	223.7	668.8	216.8	677.9
Family structure						
Single with child(ren)	24.0	149.9	24.4	187.4	15.5	143.7
Single, no child, age less than 55	24.2	179.8	26.3	217.2	14.6	117.5
Single, no child, age 55 or more	134.0	405.8	150.7	408.9	102.0	391.6
Couple with child(ren)	140.6	580.5	147.5	629.1	86.7	555.7
Couple, no child	240.2	868.2	236.2	998.6	205.7	864.8
Education of head						
No high school diploma	23.7	157.1	34.8	149.7	16.1	110.7
High school diploma	79.1	227.2	84.3	263.8	56.7	218.1
Some college	79.8	355.7	88.8	384.5	50.9	272.2
College degree	260.2	982.3	298.6	1 54.5	195.2	977.7

Table 8.1. **Family net worth by selected characteristics of families in the United States**
(cont.)

Family characteristic	2004		2007		2010	
	Median	Mean	Median	Mean	Median	Mean
Race or ethnicity of respondent						
White non-Hispanic	162.2	648.3	179.4	727.4	130.5	654.5
Non-white or Hispanic	28.5	176.2	29.7	240.3	20.4	175.9
Current work status of head						
Working for someone else	77.4	310.7	98.5	369.1	55.2	298.8
Self-employed	402.2	1 639.9	407.3	2 057.4	285.6	1 743.7
Retired	160.9	539.8	169.9	569.1	151.1	485.3
Other not working	13.6	186.7	6.0	130.1	11.9	137.5
Current occupation of head						
Managerial or professional	227.3	995.6	258.8	1 174.8	167.3	1 047.0
Technical, sales, or services	51.7	284.8	77.0	325.8	32.6	219.1
Other occupation	65.0	169.8	68.4	201.3	46.6	162.8
Retired or other not working	127.9	485.0	135.6	500.6	93.5	410.4
Region						
Northeast	186.1	655.0	167.1	684.6	119.9	615.2
Midwest	132.4	503.8	112.7	491.2	68.4	399.8
South	73.4	401.0	102.0	525.9	68.2	440.8
West	109.3	605.3	164.1	695.4	73.4	599.9
Urban status						
Metropolitan statistical area (MSA)	120.1	582.0	138.8	652.6	78.4	553.6
Non-MSA	68.2	203.5	82.0	253.9	74.5	236.1
Housing status						
Owner	212.6	720.9	246.0	817.6	174.5	713.4
Renter or other	4.6	62.3	5.4	74.7	5.1	57.2
Percentile of net worth						
Less than 25	2.0	-1.6	1.3	-2.3	..	-12.8
25-49.9	50.2	54.2	56.8	60.9	32.2	35.6
50-74.9	196.7	213.7	230.8	238.6	157.2	168.9
75-89.9	586.7	608.4	601.2	616.7	482.7	527.9
90-100	1 645.5	3 591.1	1 991.9	4 176.9	1 864.1	3 716.4

..: Less than 0.05 (USD 50).

Source: *United States Federal Reserve Bulletin*, June 2012 article, "Changes in US Family Finances from 2007 to 2010", Table 4, pp. 17-18.

As the power of desktop computing increases, so has the potential for statistical organisations to provide users with the ability to customise the tabular output to meet their own specific requirements through the use of self-help web-based table-builder products. Often the starting point of such products is aggregate data at the lowest level of detail possible from the micro-data file. Then users are provided with options on how to build their own tables based on the themes or variables of interest. One of the advantages of starting with aggregate data, rather than a micro-data file, is that less computing power is needed to group aggregate data than to produce tabulations for large micro-data files. Another advantage to this approach is that the detailed aggregate data can be screened for confidentiality prior to being made available to the general user population. An example of this type of product is the OECD.Stat web browser, which provides a single online platform for access to statistical databases from the OECD (<http://stats.oecd.org>).

8.2.2. Micro-data files

Some users will prefer to conduct their own analysis of the wealth data. In these situations, the main challenge for the organisation that has produced the wealth data is making the data files available with the greatest level of information possible, while still ensuring the confidentiality of the individual survey respondents. Dissemination of wealth data in light of this trade-off can prove to be a challenging task, as wealth distributions are heavily concentrated on specific individuals. However, several solutions in this field enable dissemination with a sufficient level of information. The use of Statistical Disclosure Control techniques, a set of statistical methods that analytically address this trade-off, is one of these solutions, and has been recently explored by Eurostat.¹ The creation of safe centres for micro data access also allows wider dissemination (see Box 8.2). The wide range of solutions available in this field is currently being reviewed by an OECD Expert Group on micro data dissemination that will lead to the publication of some guidelines in this field.

Box 8.2. Statistics Canada's Research Data Centres

Data from the periodic wealth surveys conducted by Statistics Canada are made available as part of the collection of data sets in the Research Data Centres (RDCs). The RDCs are part of an initiative to strengthen Canada's social research capacity and to support the policy research community. RDCs provide researchers with access to micro-data from population and household surveys in a secure university setting. The centres are staffed by Statistics Canada employees, and operated under the provisions of the Statistics Act in accordance with all its confidentiality rules: they are accessible only to researchers with approved projects who have been sworn in under the Statistics Act as "deemed employees". RDCs are located throughout the country, so researchers do not need to travel to Ottawa to access Statistics Canada micro data.

To find out more about the activities of RDCs, please visit the Canadian Research Data Centre Network (CRDCN) website www.rdc-cdr.ca.

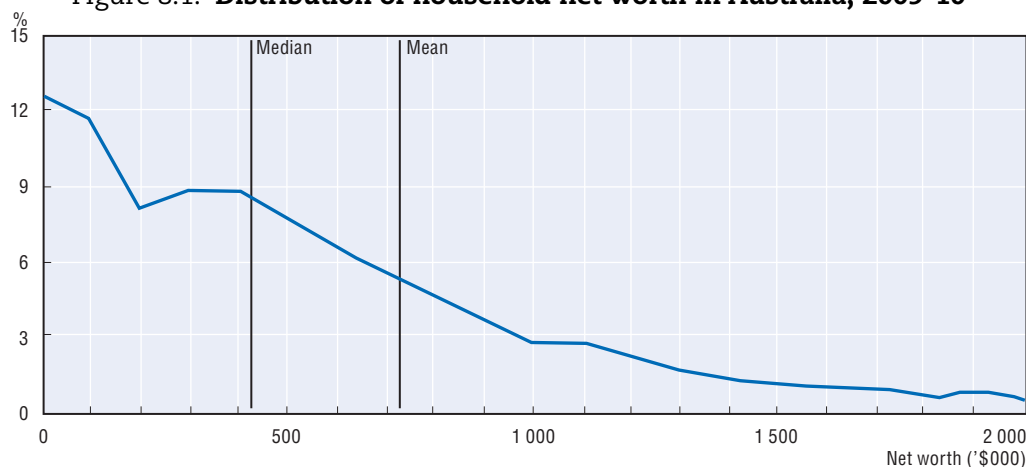
In addition, extensive documentation must be provided to the researchers, including but not limited to the survey questionnaire, detailed description of the survey methodology and detailed description of all variables in the data file (including code sets or ranges of the values of each variable).

8.3. Specific issues concerning the dissemination of wealth data

In this section, some of the specific issues that may arise when disseminating wealth data are discussed.

8.3.1. Characteristics of the distribution of wealth data

As was discussed in detail in the previous chapter, the specific features of wealth distribution (e.g. negative wealth, concentration of wealth in a relatively small number of observations) can present a challenge to users (Figure 8.1). It is therefore important that the documentation provided to users includes a description of the distribution of the wealth data so it can be taken into consideration when using the data. Also, as part of the initial release of the data, analysis of the distribution of the wealth data must be front and centre in the material provided.

Figure 8.1. **Distribution of household net worth in Australia, 2009-10**

Note: Households with net worth between \$ 150 000 and \$ 2 050 000 are shown in \$ 100 000 increments.

8.3.2. On trends in wealth data

Most wealth surveys are conducted on a periodic basis, often with a gap of several years between each iteration of the survey. This can cause challenges when determining and explaining any trends in the data. It is important that any trend analysis is accompanied by contextual information that will help the user understand these trends. For example, have there been any economic upturns or downturns that might explain changes in wealth over the time period covered by the data? In addition, it is important when analysing trends in wealth data that any conceptual changes introduced over time which may have affected the final output are explained to users. In some cases, this may result in a break in the time series, where previous data may no longer be comparable to the current data. This case is important so that the organisation producing the wealth data does everything it can to prevent comparisons of data before and after the break in the series. For example, Statistics Canada has wealth data going back to the 1970s, but the wealth data produced since 1999 include an estimate for the value of pensions. As a result, when the 1999 data was released, earlier wealth data was not referenced, since it was no longer comparable with the newer wealth data. In subsequent releases of new wealth data, the trend starts in 1999.

Not all changes in the wealth concepts covered by a survey will be significant enough to cause a break in the time series, but all changes must be documented so that users can decide for themselves if they feel comfortable comparing data over time for their purposes. In addition, it is highly desirable to post revisions to data if errors are detected at a later point in time.

8.3.3. Longitudinal wealth data

Dissemination of longitudinal data presents many challenges. It normally takes multiple iterations of the survey to create a true longitudinal data set (at least 3 waves of data), and the techniques and skills required to analyse the data are often limited to a fairly small group of researchers. Normally, the dissemination of longitudinal data involves making the data set available to researchers. However, survey organisations will often sponsor various analyses of the longitudinal data, since the resulting research papers are a way to signal the availability of the longitudinal data to the research community and to highlight the analytical uses of the data, as a way of inspiring other researchers to begin to use the data.

8.3.4. Dissemination guidelines

Useful guidelines for the general presentation and dissemination of statistical output are available in a number of publications, most notably those by the UNECE (UNECE, 2009a and 2009b). Much of what is provided here is taken from Section 6.11 of the 2011 *Canberra Group Handbook*, which provides best practices for the dissemination of income data, and which are also applicable to wealth data. As with income data, the complexity of wealth data means that the data provider must provide in-depth information on the data (i.e. metadata) and guidance in the appropriate use of the data, such as:

- Definitions of the wealth concepts used, including:
 - ❖ Glossary of all the wealth components included in the output, including clear, detailed definitions of any derived variables. The definitions should include enough detail so that users understand the input to the calculation of the derived variable, allowing them to interpret the survey estimates appropriately.
 - ❖ Description of the measured and non-measured components of wealth. It is important that users know which components were directly measured in the survey and which were not. One key non-measured component could be information on the value of a respondent's pension: if this information is derived based on a combination of survey data and information on the pension plan(s) the respondent is covered by, it is important to provide the methodology on how the value of the pension was calculated.
- Basic information about the sources of the data, including:
 - ❖ Whether wealth data is from a sample survey, administrative data or a combination of the two. If a combination of different sources is used, there must be clear documentation on each source for the various wealth components being disseminated, and an indication of the data source used to produce them. For administrative data sources, it is also important to describe the original reason for collecting the administrative data and any strengths or limitations of the data relevant to the data being disseminated.
 - ❖ Specific information about the data source, such as the statistical units used – individuals, families or households; reference period(s) – the time period(s) covered in the data being disseminated; the survey population – who is included, and equally important, excluded in the survey population. For example, does the data source include the entire population or only a subgroup of the population, such as a specific age group.
 - ❖ Any background information on the survey or administrative data source, such as questionnaires, detailed file layouts of administrative data sources or interviewer instructions.
- The survey methodology, including:
 - ❖ Summary information on the survey methodology, in a form readily available for all users. For example, if there is a general web page for the survey, there should be a link on the main page to information on the methodology.
 - ❖ Detailed documentation on the survey methodology should be made available to users at the time of the initial data release or if that is not feasible, soon thereafter. While important for data users, this documentation will also prove invaluable in the future, considering the often sporadic nature of wealth surveys, which makes it difficult for an organisation to ensure continuity of staff working on subsequent rounds of the survey. Some of the information that should be included are:
 - Sample size and design, including whether a probability sample has been used (or alternatively whether random walk or quota methods were applied); and coverage of

- the survey population in the sampling frame and whether the design was single stage or involved some clustering.
- Imputation, including how any imputation was done, the percentage of imputed information, including the number of incomplete units for each wealth aggregate and the percentage of the wealth amount that was imputed. This is particularly important for wealth data, since non-response varies by wealth categories. The results of any studies conducted on the impact of imputation on the data should be included.
 - Procedures taken to protect data confidentiality. Wealth is often concentrated in a relatively small number of observations, which can cause problems in protecting the confidentiality of the individual survey respondents when disseminating the data. It is important that users are informed of any special procedures implemented to protect confidentiality and of the impact of these procedures on the estimates.
 - Comparability of the data over time. When time-series data are being disseminated, it is important to inform users of any changes to the data that may have affected the data for the time period covered. For example, if the data source is administrative records, it is important to provide users information about any changes in the administrative systems that might affect the data. In the case of a survey, if there were changes to the way in which the data were collected over the time period or changes in survey concepts, then it is important to mention these changes in the documentation accompanying the release of the data. Ideally, data in a time series should be adjusted to ensure that the data are comparable over time, but often it is not possible to quantify the precise effect of these changes.
 - Guidelines on the use of the wealth data. In addition to detailed descriptions of the variables available, survey organisations should make available any information that could help users in using and interpreting the survey data. For example, if there are data manipulations that are often performed by some users, such as creating different aggregates or custom-derived variables, the survey organisation should endeavour to make the computer code for these manipulations available to all users. Making available the various analyses done using the data in the past is another way of educating users on how to interpret the data.
- Information on data quality. Information about the quality of the wealth data being disseminated should be made available to users, including, but not limited to, information on:
 - ❖ Sampling errors. Where information is from probability samples, an indication of sampling error should be provided. As a minimum, the relative standard error (RSE), i.e. the standard error expressed as a percentage of the estimate for which it is calculated, should be provided for the key variables being disseminated.
 - ❖ Suppression of unreliable data. While it is recommended that data for which the relative standard error exceeds a certain limit should not be published, the thresholds for suppression should be based on the professional's judgment of the "fitness for use" of the estimates.² In the case of complex designs or indicators, the standard errors may not be readily available for all breakdowns. In this case it can be appropriate to use the number of underlying observations instead. For example, if it was found that estimates with an acceptable standard error normally were based on at least 30 observations, then any estimates based on fewer than 30 observations would be suppressed.
 - ❖ Response errors. These may be due to many factors, including faulty design of the questionnaire, interviewers' or respondents' misinterpretation of questions, or

respondents' lack of knowledge/records, or faulty reporting. If there is information available on the type of response errors that may have occurred in a survey, this should be provided in the user documentation.

- ❖ Non-response errors. In surveys, non-response errors occur because some sample units do not respond to the survey. Response rates should be provided to users, including any information available on the units who did not respond (e.g. if specific geographic areas or income groups had higher non-response rates) and, in the case of time-series data, if the non-response pattern is different now than in the past. This is particularly important for wealth surveys, since it is difficult to convince high-wealth households to respond to these. For correct interpretation of response rates, it is useful to provide information on whether substitutions were allowed.
- ❖ Effect of large values. Wealth data can be particularly affected by the presence or absence of extreme values. An explanation of any procedures applied to the data to account for extreme values should be included in the documentation. At a minimum, users should be informed of the fact that the results may include extreme values and that some estimates may be influenced by the presence or absence of these extremes.
- ❖ Comparability of data to other sources. As part of the validation of the wealth estimates produced, the estimates should be compared with other sources of wealth information, often the data available from the System of National Accounts. The results of any such comparisons, including explanations of any differences, should be provided to users.
- ❖ Comparability of the data over time. When time-series data are being disseminated, it is important to inform users of any changes to the data that may have affected the data for the time period covered. For example, if the data source is tax records, it is important to provide users information about any changes in the tax systems that might affect the data. In the case of a survey, if there were changes to the way in which the data were collected over the time period or in survey concepts, then it is important to mention these in the documentation accompanying the release of the data. Ideally, data in a time series should be adjusted to ensure that the data are comparable over time, but often it is not possible to quantify the precise effect of these changes.

In addition, the documentation should include a description of any procedures put in place to adjust the estimates for data quality issues that may have arisen.

8.4. Summary

The key highlights from this chapter can be summarised as follows:

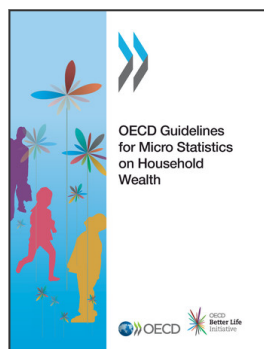
- When disseminating wealth data, it is important to take into account the needs of the various users of the data. In order to meet this challenge, it is helpful to have a multi-faceted dissemination strategy with a variety of products to meet the needs of the different user groups. Generally, the data products based on wealth data will fall into three broad types – analysis, data tables and micro-data. Ideally, products from any individual wealth data program would cover all three.
- The more general analyses produced from a wealth survey are often made available at the time of the initial release of the wealth data set in order to publicise the release and highlight the key findings. A more in-depth analysis usually takes more time and is often conducted by more sophisticated users once the data is released by the organisation that conducted the wealth survey. These analyses will be produced by a

variety of users in various organisations and may be made available to the general public, depending on the mandate of the individual or group sponsoring the analysis.

- Data tables are one way to make a variety of data available to users who may not have the analytical skills, resources or data access required to produce their own output from the micro-data file. Statistical organisations may provide users with the ability to customise tabular output to meet their own specific requirements through the use of self-help web-based table-builder products.
- Some users will prefer to conduct their own analysis of the wealth data. Often survey organisations will need to provide various ways for researchers to access the micro-data, for example by producing two versions of the data set. The first is a general micro-data file suitable for wide distribution after extensive screening to ensure confidentiality. The second is a more detailed micro-data file, with the results of analysis based on this file vetted to ensure confidentiality.
- At a minimum, the documentation provided to users of wealth data needs to include a description of the wealth distribution so that it can be taken into consideration when using the data. Also, as part of the initial release of the data, analysis of the distribution of the wealth data must be front and centre in the material provided.
- Any trend analysis should be accompanied by contextual information that will help the user understand the trends shown. For example, there may have been economic upturns or downturns that might explain changes in wealth over time. In addition, it is important to explain any conceptual changes introduced over time that may have affected the final output.
- Disseminated data should be accompanied by metadata to help users understand the data. The metadata should include definitions of concepts used, information about data sources, survey methodology, guidelines on use of the data, and data quality information including sampling error rates, non-response rates, etc.

Notes

1. http://neon.vb.cbs.nl/casc/SDC_Handbook.pdf.
2. Estimates can be divided into three groups: those with a low RSE, which can be used without restriction; those with a higher RSE, where the data should be used with caution; and the third group, where data with a very high RSE are suppressed.



From:

OECD Guidelines for Micro Statistics on Household Wealth

Access the complete publication at:

<https://doi.org/10.1787/9789264194878-en>

Please cite this chapter as:

OECD (2013), "Dissemination", in *OECD Guidelines for Micro Statistics on Household Wealth*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264194878-11-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.