

**13th Meeting of the Advisory Expert Group on National Accounts,
1-3 October 2019, Washington D.C., USA**

Agenda item: 2.3.3

RECORDING OF DATA IN THE NATIONAL ACCOUNTS

Introduction

This paper provides a preliminary review of issues related to the valuation of data in the SNA. Section 2 reviews the treatment of data in the 2008 SNA. Section 3 examines the economic uses of data and the growth in data-based business models. The following two sections discuss two approaches for the valuation of economic transactions in data. Section 4 discusses a framework for measuring data as an asset produced by firms while Section 5 examines an approach to treating data as a household asset that is exchanged for free services implications of valuing barter transactions related to data. Section 6 summarizes the discussion.

Main issues to be discussed:

The AEG is invited:

- to comment on the paper;
- to discuss the options presented in sections 4 and 5;
- in general, to indicate directions for future work.

**13th Meeting of the Advisory Expert Group on National Accounts,
1-3 October 2019, Washington D.C., USA**

Agenda item: 2.3.3

Recording of data in the national accounts

Table of Contents

1. Introduction.....	3
2. Recording data in the SNA	3
3. Economic uses of data (under construction).....	4
4. Data assets produced by firms	4
5. Data as household assets.....	6
6. Summary.....	7

1. Introduction

1. The 2008 SNA specified that databases should be valued on a sum of costs basis and only reflect the cost of preparing data in the appropriate format but not the cost of acquiring or producing the underlying data. In line with business accounting rules, data are considered non-produced assets and appear residually under goodwill only when a market purchase of a firm occurs. However, recent growth in the digital economy and in the monetization of data holdings by firms has raised concerns regarding the lack of visibility of data in the national accounts.

2. This paper examines the issues related to the valuation of data in the SNA. Section 2 reviews the treatment of data in the 2008 SNA. Section 3 examines the economic uses of data and the growth in data-based business models. The following two sections discuss two approaches for the valuation of economic transactions in data. Section 4 discusses a framework for measuring data as an asset produced by firms while Section 5 examines an approach to treating data as a household asset that is exchanged for free services implications of valuing barter transactions related to data. Section 6 summarizes the discussion.

2. Recording data in the SNA

3. The 2008 SNA explicitly references the recording of data in the context of the discussion on databases in paragraphs 10.112-10.114:

“10.112 Databases consist of files of data organized in such a way as to permit resource-effective access and use of the data. Databases may be developed exclusively for own use or for sale as an entity or for sale by means of a licence to access the information contained. The standard conditions apply for when an own-use database, a purchased database or the licence to access a database constitutes an asset.

10.113 The creation of a database will generally have to be estimated by a sum-of-costs approach. The cost of the database management system (DBMS) used should not be included in the costs but be treated as a computer software asset unless it is used under an operating lease. The cost of preparing data in the appropriate format is included in the cost of the database but not the cost of acquiring or producing the data. Other costs will include staff time estimated on the basis of the amount of time spent in developing the database, an estimate of the capital services of the assets used in developing the database and costs of items used as intermediate consumption.

10.114 Databases for sale should be valued at their market price, which includes the value of the information content. If the value of a software component is available separately, it should be recorded as the sale of software.”

4. The value of investments in databases excludes the costs of acquiring or producing the data content. In line with business accounting rules, when the market sale of a database occurs, such as when a firm is purchased, the value of the data content of the database appears residually under goodwill (Ahmad and van de Ven, 2018). Thus, the value of data will not appear under fixed capital formation and it is treated as a non-produced asset under goodwill only when a market purchase of a firm occurs and a revaluation of its assets is required to realign them with market transactions.

5. In general, data are used in the production of other goods and services within the producing unit. Since data is not considered a produced asset, its production is not recorded as output when the activity is internal to the producing unit. In this context, expenditures related to the production of data are treated as industry inputs: intermediate consumption and generation of income (COE, etc.).

6. In the absence of distinct product classifications for data, revenues associated with the market production of data are captured in a variety of industry outputs and are treated as intermediate consumption by the buyers.

3. Economic uses of data (under construction)

7. Data has become ubiquitous in economy and society...

There are three different ways in which data can impact production (Wixom and Ross 2017):

- Improving internal business processes and decisions
- Wrapping information around core products and services
- Selling information services to new and existing markets.

Ahmad and van de Ven (2018) list the following data-driven business models:

- Using internal data for enhancing efficiency of production processes and decision-making
- Collecting user data for selling targeted advertising
- Creating new services based on analyzing big data
- Provision of data-related services by providing fee-based access

4. Data assets produced by firms

4.1 Definition of data and databases

8. Statistics Canada (2019) has developed a measurement approach which clearly identifies the boundaries around data by delineating its differences from general information on the one hand and from databases on the other. Data are defined as “observations that have been converted into a digital form that can be stored, transmitted or processed and from which knowledge can be drawn” (Statistics Canada 2019, p. 7). Since databases are explicitly measured in the 2008 SNA, a second boundary avoids an overlap in definitions by excluding the normalizing process which organizes data into a definite database structure from the definition of data.

9. The 2008 SNA includes the cost of preparing data in the appropriate format in the cost of databases. Ahmad and van de Ven (2018) explicitly interpret this recommendation to include the costs of digitizing data. This is a logical interpretation in the context of valuing the costs of producing databases.

10. However, the costs of digitizing data apply to a very wide array of data capture and storage activities, well beyond those required for preparing data for databases. The requirement to separately record the value of data requires delineating the costs of data preparation in general as a separate activity. The digitizing of observations can be undertaken by one unit and supplied to another, thus it should be possible to distinguish data digitization transactions.

11. Treating data as a separate product implies the need to change the 2008 SNA recommendation for valuing databases to avoid overlap with the definition of data. The costs of databases will have to explicitly exclude the costs of digitizing data and to restrict the cost of data preparation solely to the data normalizing function. Does this present a measurement difficulty; should databases be an exception?

4.2. Capitalization of data in the SNA

13. Even though recent years have seen an explosion in the generation and use of data, there are limited market transactions directly related to the sale of data as a separate product as most data are generated within the producing unit in the process of producing other goods and services. In term of uses, purchased data, or alternatively its production costs, can be consumed as an input in a single production cycle (as intermediate consumption, compensation of employees, etc.), sold as final consumption, or contribute to production over several production cycles as capital formation.

14. While some data are likely being consumed in production and as final consumption, most business models listed in the previous section reflect the uses of data repeatedly over several production cycles. In this context, data, defined as the cost of digitizing observations, has all the characteristics of the own-account production of an asset.

15. Where data is produced for own use and contributes to production over more than one year, it should be valued on a sum of costs basis, including the cost of the capital used, similarly to other own account intellectual property products. As with software for example, market sales should distinguish between the sale of copies, in which the original remains on the books of the original producer and the purchase is also capitalized by the acquiring firm, and the sale of the original where an ownership transfer occurs. Own-account production of data increases GDP.

16. Should the costs of digitizing data when undertaken in support of ancillary activities be considered out of scope for own-account data production? These are expenditures in support of the normal activity of an enterprise; do they have the characteristics of an asset?

17. In measuring the own-account production of data by sum of costs is there possible overlap with other forms of own-account capital formation that are similarly measured such as databases, software, R&D, and maybe even mineral exploration?

18. Similar to the treatment of knowledge-capturing products in the 2008 SNA, the household production boundary will exclude the production of data.

4.3 Measuring the value of data

19. Statistics Canada (2019 b) has generated experimental estimates of the value of data based on the sum of costs approach. Estimates are based on measuring the labour costs of the relevant occupational groups plus an additional estimate for intermediate inputs (50%) and a return to capital (3%). Employment and earnings are derived from the quinquennial Census of Population and from the monthly Labour Force Survey for post-census years. In 2018, estimates showed a lower bound of \$9 billion and an upper bound of \$14 billion of investments in data. These figures represented between 2% and 3% of total of gross fixed capital formation (\$498 billion) in Canada that year.

20. While the subjective assumptions relating to time use and even the type of occupations selected could be refined through further inquiries and surveys of enterprises, these figures provide a rough gauge of the expected size of the value of data given the currently proposed measurement framework. Has digitization become so efficient or so embedded into other economic and social processes that its direct costs have become relatively minor?

5. Data as household assets

21. Ahmad and van de Ven (2018) examine the implications of valuing barter transactions related to the free acquisition of data. Their framework is constructed around the assumption that a large amount of data is being acquired by firms, mainly from households, through non-monetary transactions where access to a platform or a social media site is exchanged for acquiescence by users to providing their private data.

22. However, not all data being collected by firms is part of an exchange and this presents a difficulty. For example, when stores capture customer information there is no free service provided in exchange. This may lead to valuing data that is being bartered but not recording a value for the same data if it is captured through other mechanisms.

23. The researchers examine three possible approaches to measuring barter transactions: 1) market equivalence; 2) correlation to advertising revenues; and 3) user-based valuations created from information on the willingness to pay for the bartered services.

24. There are two possibilities for the treatment of data as an asset: it can be treated either as a non-produced asset or as a produced asset.

25. Treating data as a non-produced asset whenever a barter transaction occurs requires imputing an output value for the freely provided service and an equivalent value for uses, typically household consumption, thereby increasing GDP. The treatment requires recording the value of data as an 'other changes in volume' of the household units in the capital account. The acquisition of the asset by the industry producing the free service would correspond to a disposal of an asset by the household generating the data

26. Treating data as a produced asset would have a larger impact than treating it as a non-produced asset. Own-account production of data by households increases GDP. If subsequent copies of the data are used for capital formation, such as the generation of a database, this increases GDP a second time. Finally, the use of the assets to generate a market output would increase GDP a third time.

6. Summary

27. The 2008 SNA treats data as a non-produced asset and records it residually as goodwill only when a market purchase of a firm occurs and a revaluation of its assets is required to realign them with market transactions. The recent growth in data-based business models has raised concerns regarding the lack of visibility of data in the SNA. This paper reviews two approaches to the valuation of data.

28. The first approach to the valuation of data treats the digitization of observations by firms as own-production of a data asset. This method increases GDP. Measurement is based on sum of costs and is anchored to labour inputs for a set of relevant occupations. Concerns include the possibility of overlap with other forms of own-account production and the definition of an appropriate scope for delimiting relevant digitization activities and excluding run of the mill enterprise activities that do not produce data assets that can possibly be monetized.

29. The second approach reviewed treats data as typically generated by households and bartered in exchange for free services by a platform or a social media site. If data is treated as a non-produced asset, GDP increases by the value of the bartered transaction, roughly. However, if data is treated as an asset produced by households, then it can impact GDP in several ways: first, as own-account production, second as it potentially enters the production of other assets, and finally as the derived assets are used in production.

30. Imputing a value for non-market transactions in the core accounts requires the development of a rigorous approach that does not appear to be fully elaborated yet.