

**STATISTICS AND DATA DIRECTORATE
COMMITTEE ON STATISTICS AND STATISTICAL POLICY**

Working Party on National Accounts

Recording and measuring data in the System of National Accounts

Meeting of the Informal Advisory Group on measuring GDP in a digitalised economy

9 November 2018

OECD Conference Centre

Paper prepared by N. Ahmad and P. van de Ven for the Meeting of the OECD Informal Advisory Group on measuring GDP in a digitalised economy on 9 November 2018, agenda item3.c.

Contact: Mr Peter van de Ven (peter.vandeven@oecd.org)

JT03438732

Recording and measuring data in the System of National Accounts

Nadim Ahmad (OECD) and Peter van de Ven (OECD)

Abstract

The 1993 SNA introduced the notion of databases, with further clarifications provided in the 2008 SNA that specified that databases should reflect only the value of the underlying database management systems and the costs associated with the digitisation of data. This recommendation reflected the view that the underlying value (information content) associated with the data itself was de facto a non-produced asset (because to do otherwise would indirectly open the door to the capitalisation of knowledge), with outright purchases of databases recognised in the accounts as goodwill, and as such, their contribution, as a factor of production, is de facto invisible in the accounts. Recent years have seen an explosion in the generation of data, and the use of data, notably in advertising based business models, raising questions about the 'invisibility' of data in the accounts. This paper and presentation attempts to address these issues as a way of encouraging further debate, both conceptually and in the field of actual measurement.

Table of contents

Recording and measuring data in the System of National Accounts	2
1. Introduction	4
2. The recording of data in the 2008 SNA	4
3. A taxonomy and classification of data-driven business models.....	7
4. Options for recording and measuring data in the SNA	9
5. Summary and conclusions	14

Tables

Table 1. Maximum user data values based on Financial Times Simulator (USD), 2016.....	12
Table 2. Relative share of user's data to the global economy, for selected providers, 2016.....	12

1. Introduction

1. The 1993 System of National Accounts (SNA) introduced the notion of databases, with further clarifications provided in the 2008 SNA that specified that databases should reflect only the value of the underlying database management systems and the costs associated with the digitisation of data. This recommendation reflected the view that the underlying value (information content) associated with the data itself was de facto a non-produced asset, with outright purchases of databases that included the intrinsic value of the underlying data recognised in the accounts as goodwill.
2. However, recent years have seen an explosion in the generation of data, and the use of these data, in for example advertising based business models. But because the underlying value of acquired data is not captured as production in the SNA, it is de facto invisible. The recent developments in the role of data have put this earlier recommendation in the spotlight, and have led to a renewed debate on the recording and measurement of data flows and data assets.
3. This paper attempts to shed light on the issue. To frame the debate, Section 2 provides an overview of discussions held during the update of the 1993 SNA and the 2008 SNA. Section 3 looks again at data, and, in particular, its heterogeneity, by providing a more detailed taxonomy and classification of data-driven business models to feed the discussion in Section 4, which considers the potential implications, pros and cons, of alternative treatments to the 2008 SNA recommendation. Section 5 summarises and concludes.

2. The recording of data in the 2008 SNA

4. In the 2008 SNA, the recording of databases is described, albeit in a very concise form, in paragraphs 10.112 – 10.114, as follows:

10.112 Databases consist of files of data organised 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 data base 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.

5. Importantly, from the above, the coverage of databases is limited only to the investments, purchased or produced on own-account, in making data available and accessible. It is thus constrained to the costs of preparing data in the appropriate format, including the costs of digitising existing data. The costs related to the development of database management systems (DBMS) are instead treated as investments in computer software, and not as part of investments in databases. Perhaps one of the most important aspects of the 2008 SNA recommendation, especially given the increasing tendency for firms to purchase data (for example from consumers), is that the costs related to the acquisition and/or the production of the knowledge embodied in the data (in other words, the value of the data itself) should not be included as gross fixed capital investment.

6. It is important to note in this respect that the decision not to treat the data, in and of itself, as produced does not mean that data has no value. It clearly does, as recognised in the discussions¹ preceding the 2008 SNA recommendation. Future benefits can very clearly be derived from data, either through the sale of a database (including the value of the data), or in creating additional value added in support of the production of other goods and services, such as advertising. In the former case the 2008 SNA captures the value of data as goodwill when a market transaction occurs (which de facto means that data are treated as a non-produced asset), whilst in the latter, although data remains in and of itself invisible, its contribution to production is accurately reflected.

7. Although the contribution of data to production is always captured, data itself is only valued when a market transaction occurs (recorded as a transaction in non-produced assets). In this sense, data in the SNA, as a non-produced asset, is similar, at least in an accounting sense, but still different to other non-produced assets, such as land. Like data, land is also used in production, but like data, land, as a non-produced asset, means that it cannot be readily identified as a separate factor of production (which it obviously is). However, whereas with land it will be possible (at least in theory) to always identify its value on the balance sheets of the owner, and so indirectly, especially when the owner is also the user,² to estimate the contribution that land makes to production, with data, value will only ever be captured when an observable market transaction takes place (and in practice, at least until recently, this would only have been possible with respect to large transactions.³

8. This decision to only recognise data in the accounts when a monetary transaction occurs reflects the fact that the underlying value of data reflects its information or knowledge content. Valuing all data as a non-produced asset therefore, whether purchased or otherwise, would de facto require that all knowledge, including human capital, be treated as a non-produced asset. That is not to say that, conceptually, this shouldn't be done; there has been a long discussion over the years on human capital and indeed on other knowledge

¹ For more information on the main arguments, reference is made to documents SNA/M2.04/04, SNA/M1.05/19.1 and SNA/M1.05/19.2 (including the summary conclusions), discussed at the meetings of the Advisory Expert Group (AEG) on National Accounts held on December 2004 and July 2005; see <https://unstats.un.org/unsd/nationalaccount/aegm.asp>.

² Note that where owners of land make the land available to others they generate rents (i.e. primary income), not rentals (i.e. output).

³ One issue that will need to be considered as this discussion develops is how national statistics authorities currently value small scale (monetary) transactions in data, which may also include cumulatively large sums spent by firms acquiring data for the purpose of developing a database for subsequent sale or to generate revenues.

based assets, and whether these should be recognised in some form (including as produced assets), in the accounts. But to do so would require approaches to be developed that were internationally comparable, feasible and meaningful; and certainly with respect to human capital, recording the activity as production could run the risk that it would swamp GDP and render it meaningless. It was this realisation that the value of data was intrinsically related to the underlying knowledge it embodied that led to it being recorded as de facto non-produced (i.e. goodwill) when a market transaction occurred. To do otherwise would open the door to the inclusion of all kinds of information or knowledge.

9. During the discussions leading up to the 2008 SNA, part of the thought process looked into electronic data as being in some way different to other forms of data, which could merit its inclusion as a produced and distinct type of data. But given the fact that the underlying value of the data does not reflect its electronic characteristics, but, rather, the inherent and intrinsic value of the knowledge or information, it was not possible to make a credible distinction between various types of data. Why would one limit the capitalisation of knowledge to information that is stored electronically? What makes the electronic data especially different to paper copies, that it should be treated differently? Should the storage media matter? Knowledge/information embodied in databases contains asset characteristics, but knowledge displays these characteristics whether it is embodied in a database or not.

10. Of relevance for the discussion below, and certainly for any discussions that reconsider the 2008 SNA decision not to treat data as a produced asset, is which costs related to the creation of a database should be included in the 2008 SNA definition of a database. As noted above, the 2008 SNA recognises only the costs in preparing data in the appropriate format, and any related staff costs and capital services of assets used in digitising data, together with any related intermediate costs, but not the acquisition or generation of data in and of itself.

11. In this respect, it is important to note that, in theory, these costs only pertain to cases where the data itself has a useful life of more than one year. Taking the examples of personnel or business accounting records within an enterprise, or statistical databases of statistical offices, one could argue that the records, which need updating on a quarterly or annual basis, have a service life of less than a year, where almost all of the value is embodied in the data for the very latest year (or quarter). But if one thinks of these databases as providing historical and recent data, where updates and revisions are to be looked upon as improving quality, the database has a service life of more than one year, and is to be considered as an asset.

12. There was also a discussion on limiting the scope of asset recognition of databases to data providers or commercial industries, thus excluding databases created by government which are publicly available and do not generate monetary returns. However, this idea was rejected, because it was considered conceptually incomplete. Excluding, for example, databases managed by statistical offices would be inconsistent with the recording of, for example, public infrastructure or public R&D.⁴

⁴ The issue of including/excluding freely available public knowledge, more specifically public R&D, in the asset boundary has re-emerged again; see e.g. Mark de Haan and Joseph Haynes, "R&D capitalisation: where did we go wrong?", EURONA, Issue 1, 2018; see <https://ec.europa.eu/eurostat/cros/system/files/euronaissue1-2018-article1.pdf>.

3. A taxonomy and classification of data-driven business models

13. Digitisation has led to a proliferation in the role of data in economic decision-making and society at large. As the blog by Jon Brock et al. (2013) on the Boston Consulting Group (BCG) notes: “From the dawn of time until 2003 – some 5 exabytes, according to Intel – is now created every two days”, while “Data processing and storage costs have decreased by a factor of more than 1 000 over the past decade”.⁵ In the five years since they wrote their blog, the pace of change has showed no signs of slowing. On the contrary, data are used at an ever-increasing speed, with new technologies for storage and analysis of data becoming available.

14. Business models increasingly rely on the use of data, to generate revenue directly, or indeed, as succinctly put by Aashish Pahwa, to form more effective and efficient business strategies:

*But it is important to remark that not every data-driven business model will involve monetising the data directly to earn revenue. Rather, most businesses not just collect, store, and sell data, but use the data internally to form more effective and efficient business strategies, make better decisions, and refine their business processes and models, which will eventually result in more profits.*⁶

15. Business has always tried to collect information to make their operations more efficient, to improve their sales performance, or directly generate revenues via provision of data and providing analytics. However, the sheer scale of data use has changed dramatically.

16. When looking at the different models, one could make the following basic distinction:⁷

- Providing services for free or at very low prices to gather data of users, which are subsequently used to detect behavioural patterns to provide other producers with targeted advertising services (Google Ads, Facebook, etc.), or to offer other services (e.g. using information from payment systems, to provide other services like specialised accounting services, support in taxation and pension arrangements, etc.).
- Using data generated as part of the primary production process, to improve the efficiency of the internal operations and/or to detect behavioural pattern to support own sales. Amazon using e.g. dynamically generated recommendations is an obvious example, but also Walmart using analytics to optimise supply chain and pricing models. VISA has applied integrated analytics to detect fraudulent transactions, saving 2 billion US Dollars between August 2011 and March 2013. Another example is a car manufacturer using data automatically created and stored by the car’s electronics. More generally, McAfee and Brynjolfsson (2012) note that

⁵ Jon Brock, Ralf Dreischmeier, James Platt and Rob Souza, “Big Data’s Five Routes to Value”, 26 September 2013. See www.bcg.com/publications/2013/information-technology-strategy-digital-economy-opportunity-unlocked-big-data-five-routes-value.aspx.

⁶ Aashish Pahwa, “The Data Monetization: Big Data Business Models”, 2 April 2018. See www.feedough.com/the-data-monetization-big-data-business-models/.

⁷ For this classification, we have combined the information from various websites on value creation using big data.

companies relying on data-driven decision-making are performing better in terms of productivity and profitability.

- Creation of new types of services by using and analysing big data. Examples are Swarmly, a smartphone application where users can share their current location and provide details of their sentiments, after which data are aggregated to provide a real-time map of popular venues; GoSquared, Mixpanel and Spinnakr, which use tracking codes embedded in their customers' websites, and provide analytics through a Web-based dashboard; and Automatic, a selling device that can be plugged into a car's data port and submits data via Bluetooth to the driver's smartphone, which are collected and analysed to provide feedback on driving style.
- Provision of data-related services, by collecting data from a vast number of different, mostly free, available data sources, normalising formats and providing access, with revenues from subscription or usage fees. An example is Gnip, which is a premium reseller of Twitter data. Instead of only providing "raw data", the services can also be extended to conducting analytics on freely available data and/or data provided by the customers. Here, one can think of consultancy type of services, such as fraud detection services, analysis to improve marketing activities, to improve customer service and relationships, or to "simply" increase sales.

17. A somewhat different type of service producers in the context of data concern the data facilitators, providers of data tools, such as providing storage media, servers and workstations, data collection, analysis and visualisation software, database management software, encryption technology and software, data protection technology, etc.

18. Finally, one should mention here the creation of freely available information or knowledge by communities of people, providing their contributions for free. Wikipedia is the best known example, but there are various other types of information platform such as ResearchGate. This type of knowledge creation differs significantly from the above types of business models in that it does not provide monetary revenues, neither directly via e.g. advertising revenues, sales or subscription fees, nor indirectly by enhancing the revenue generating potential of an enterprise. In the remainder of this paper, these types of knowledge creation are not further elaborated, as they are less relevant in a discussion on the role and recording of data as such.

19. All in all, the proliferation of the above four types of business model – and it is important to note that this is not an exhaustive list, further work will need to be done as the discussions on data unfold – shows that data, or access to data, can have a clear value added for enterprises; which was central to the discussions in the 2008 SNA. Perhaps an added complication since the 2008 SNA discussions however is the fact that the data are increasingly provided for 'free', i.e. without any monetary transaction, and, so, an approach that only recognises data when a monetary transaction occurs may no longer be tenable.

20. A consequence of any decision to review this decision is to add another complication, as imputations for data flows where no monetary exchange occurs necessarily implies some form of barter exchange. Doing so would explicitly lead to valuations of the currently 'free' benefits that consumers receive when making their data available, for example services related to social media, search engines, etc. That said, in considering the appropriate treatment, the debate also needs to consider analogies. For example TV, radio, newspapers, etc. have long used advertising based business models to provide free or discounted services to consumers, which in turn have generated longstanding and well understood value to consumers (but without widespread concern that

GDP was deficient). Indeed, this question of generating value to consumers is broader than that, for example it also relevant in the case of e.g. VIP-boxes at sports events.

4. Options for recording and measuring data in the SNA

21. From the above there appears to be little contention that data can have asset characteristics. Currently the SNA treats data as non-produced assets only when a market transaction materialises, and discounts their recording as produced assets on the grounds that to do so would conceptually open the door to the capitalisation of other forms of knowledge, that could distort meaningful analyses of GDP estimates. This ‘distortion’ matters less of course when data are treated as non-produced assets but the requirement that data ‘materialises’ in the accounts only when a market transaction occurs is being tested by the many innovations in data driven business models.

22. The starting point for the discussion below is that the concerns relating to a back-door capitalisation of knowledge through recommendations on data remain as valid today as they did in the run up to the 2008 SNA. That is not to say that of course that all is well, and further guidance is not needed. The explosion of data generated business models that collect data in return for ‘free’ services requires new thinking here, as the transactions have strong characteristics of bartering mechanisms. This issue, including the potential for valuing bartered data, is discussed in more detail below. At the end of this Section we also take another look at the SNA decision to treat data as non-produced, and consider the additional consequences of treatment as a produced.

23. However, setting aside the challenges raised by implicit barter transactions in data, there remain a number of practical considerations that are raised by the explosion in the acquisition and use of data. A pressing issue concerns the treatment of small value data transactions when there is a monetary exchange. Whilst the 2008 SNA recognises monetary transactions as goodwill when they occur, it is less clear what official national statistics do in practice, particularly for large volumes of small scale transactions, as is increasingly the case in some business models, as noted above. Following the 2008 SNA, in principle, for paid acquisitions of data, the acquiring firm would record an increase in non-produced assets with a corresponding counterpart sale (typically in the household sector). Notwithstanding reflections on whether the underlying data are produced or non-produced, there is not a great deal of contention, at least conceptually, around this, but in practice it is not inconceivable that official statistics present a different story, and further investigations will be needed.⁸ At present the likelihood is that official statistics record payments here as intermediate consumption.

⁸ One peculiarity here concerns the acquisition of data that have a service life of less than one year, which will, in theory, add to the stock of non-produced assets when they are acquired, but the related costs of integrating these data into a database will not be treated as gross fixed capital investment. Non-produced non-financial assets, in this respect, like financial assets have no one-year threshold. Furthermore, the 2008 SNA also includes an implicit threshold for distinguishing between intermediate consumption and gross fixed capital formation. Such a provision for small tools would again not apply to non-produced assets. However, also in the case of treating data as produced assets, one could argue that the multitude of small purchases still qualifies as investments.

4.1. Free data with and without a corresponding barter transaction

24. One of the most important factors that has brought the treatment of data back onto the conceptual agenda concerns the increasing acquisition of data via non-monetary transactions, akin in many respects to bartering. Typically, this occurs when consumers gain access to a platform providing media or social network services ostensibly for free but implicitly in exchange for provision of their data.

25. Whilst it is clear (under the 2008 SNA) that a monetary exchange in data should materialise as a goodwill transaction – and so de facto as non-produced asset –, no corresponding recommendation is made for acquisitions where there is no monetary exchange. As noted, this, to a large extent, reflects the relationship between data and knowledge and, so, a desire to avoid wholesale imputations of non-produced assets in the accounts, which would be nigh on impossible to achieve today. However, that desire to avoid imputations to realise the value of the underlying stock of non-produced assets where no monetary exchange takes place does not necessarily mean that they are not barter transactions.

26. Not all data captured from households is necessarily part of a barter transaction. Data collected by supermarkets on customer preferences, for example, do not always provide consumers with something of value in return (although in cases where consumers collect points on expenditures and benefit from card related discounts, a strong case could be made for some form of exchange). From the consumer's perspective therefore the data would appear to have no value but not necessarily so, from the acquiring firm's perspective. This, in and of itself, is not necessarily unique nor problematic, as the adage one man's trash is another man's gold can well testify. In isolation the simplest approach would be for the accounts to treat these flows at the zero price that they appear to follow, albeit with some revaluation mechanism or other volume change – whether the asset was produced or non-produced – recorded by the data acquirer. Where this becomes more problematic however is in its differentiation of price for ostensibly the same data: data provided under a barter mechanism has value, but the same data provided in an unrequited manner has no value for the provider but value for the recipient.

27. Extending the principle that 'barter' transactions are in scope provides an entry point to imputing monetary flows for intra-firm transfers of data, which are currently invisible in the system (as indeed are flows on information more generally). Valuation in this respect would however be non-trivial. Moreover, if data were treated as produced, any imported data that does not satisfy the one-year criterion would reduce GDP. On the other hand for data treated as non-produced this would also have consequences for reinvested earnings and FDI statistics.

28. Finally, there has been a lot of discussion on the treatment of free services financed by pure advertising approaches directed at consumers as compared to business models based on data (even if these are used to generate advertising revenues). One concern from the discussions on the potential treatment of free services is that it could lead to an outcome that resulted in GDP increasing whenever consumers sat down and watched free TV financed via advertising. Whilst a conceptual case could be made for doing so,⁹ there is no escaping the fact that this would have serious implications for GDP's ability to act as an

⁹ See Nakamura, L.I. et al. (2016), "Valuing free media across countries in GDP", *Working Paper*, No. 16-24, Research Department, Federal Reserve Bank of Philadelphia, <https://ideas.repec.org/p/fip/fedpwp/16-24.html>.

indicator for macro-economic policy making. A similar argument could of course be made for imputations of free services financed by data. But there are important differences between advertising based business models and data based business models (including those that are ultimately used to generate advertising revenues) that could warrant different treatments, as the recipient of the services funded by the advertising provides nothing of value in return for the free services received (assuming of course, as the 2008 SNA currently does, that the viewing of advertising services is not a productive activity), whereas the 2008 SNA already recognises that data does have a value as a non-produced asset.

4.2. Potential valuation methodologies for barter transactions between firms and households

29. When it comes to barter transactions between firms and households,¹⁰ the challenge would be to estimate the value of the data provided (or indeed the value of the related services exchanged). Three main approaches avail in this respect. The first looks at market equivalent estimates of data (differentiating between digital footprint data and digital identity data).¹¹

- **Digital identity data** includes basic information on individuals that either remains unchanged or that gets modified only sporadically over the course of a lifetime. Identity includes, for example, name, gender, residence, language, education level, marital status, household composition. Digital identity also includes e-mail addresses. This type of information is provided on a voluntary basis in exchange for free digital services (e.g. Facebook or LinkedIn).
- **Digital footprint data** reflects data on behaviours, including observed data recording online activities of individuals such as location, websites visited (as in Clickstream Data), product or services purchased, travel. Users are not necessarily aware of the collection of such data.

30. Anecdotal evidence suggests that companies tend to assign a negligible value to digital identities and a higher value to information on consumers' preferences and digital footprints. According to an article in the Financial Times, information on age, gender, or residence was worth a mere USD 0.0005 per person in 2013. But information on what individuals are seeking to buy is more valuable. Data on potential auto buyers, for instance, are worth about USD 0.0021 a person. Knowledge that a woman is in her second trimester of pregnancy could add another USD 0.11 to the value of her data. Information on specific health conditions or on the use of prescription drugs is worth USD 0.26 per person. Maximum values of data on different aspects of digital identity and behaviour covered by the interactive calculator are shown in Table 1 below. Demographic and family information are grouped together under 'digital identity', and the other types of information are categorised under 'digital footprints'.

¹⁰ In the remainder of this section, reference is made to households, but it may refer to other consumers, including enterprises and government institutions, as well.

¹¹ The distinction between the two types of data matters for two reasons, firstly because the price necessarily differs, and secondly because behavioural data are generally short-lived compared to digital identity data. Therefore, any decision to treat data as being produced would, in all likelihood, result in the collection of behavioural data as intermediate consumption and digital data as investment.

Table 1. Maximum user data values based on Financial Times Simulator (USD), 2016

Information		Maximum value (USD)
Digital identity (DI)	Demographic	0.3699
	Family	0.1597
	Total	0.5296
Digital footprint (DF)	Consumer	0.0411
	Activities	0.3668
	Property	0.1958
	Health	4.0180
	Total	4.6217
Total (DI + DF)		5.1512

Source: Ahmad, Ribarsky and Reinsdorf (2017).

31. Many of the variables in digital footprint change rapidly enough so that repeated observations during the same year would all be valuable, while a few change slowly enough so that having data from prior years would reduce the value of data collected in the current year. Ahmad, Ribarsky and Reinsdorf (2017)¹² made the simplifying assumption that footprint data are all observed once per year and that their value is not diminished by the availability of similar information from prior years. In addition, they assumed that firms were able to observe all of the variables, and that privacy laws do not impede them from collecting the health variables, and that all consumer data (whether from low or high income economies had the same value, thus creating upper bound estimates, as the FT estimates are more likely to reflect data providers – i.e. consumers – from developed economies). By applying these values to numbers of users across a range of major providers of digital services they arrived at a value of user data equivalent to around 0.02% of global GDP (although this excludes other major providers of free digital services such as those in China (Tencent and Baidu); see Table 2.

Table 2. Relative share of user's data to the global economy, for selected providers, 2016

World		Facebook	Twitter	Instagram	LinkedIn	Total without Gmail	Gmail	Total
Average of the number of Monthly Active Users - MAUs (Millions)	Total	1 757	315	525	106	2 703	975	3 678
	New users	269	14	200	6	489	50	539
	Old users	1 488	301	325	100	2 214	925	3 139
Users data value (Millions of USD)	Total	8 263	1 461	2 532	493	12 749	4 533	17 282
	New users	1 386	72	1 030	31	2 519	258	2 777
	Old users	6 877	1 389	1 502	462	10 230	4 275	14 505
Users data value/GDP ratio (Percentage)		0.010%	0.002%	0.003%	0.001%	0.015%	0.005%	0.020%

Source: Ahmad, Ribarsky and Reinsdorf (2017).

32. A second approach, although assuming some steady state equilibrium, would be to assume a relationship between the underlying value of free services received by consumers and advertising revenues generated from the database. This could be used to generate upper-bound estimates but would require adjustments to exclude the additional value

¹² N. Ahmad, J. Ribarsky and M. Reinsdorf (2017), "Can potential mismeasurement of the digital economy explain the post-crisis slowdown in GDP and productivity growth?", *OECD Statistics Working Papers*, 2017/09, OECD Publishing, Paris, <http://dx.doi.org/10.1787/a8e751b7-en>.

generated by platforms in assimilating data – including related acquisition costs, capital services from the platforms used to collect data, and associated costs related to data analytics.

33. A third approach could look at user-based valuations of their data, or shadow price valuations provided by consumers on their willingness to pay for free services. However, some care is needed here. Where actual market based transactions are available, it cannot be assumed that those individuals represent a representative consumer. For example, some consumers pay USD 3 per month to use FastMail for an advertising-free, no-tracking email service, but the vast majority of consumers do not. Other anecdotal information paints a similar picture, for example a 2013 survey of Facebook users by Greenlight,¹³ revealed that only 15% of users would be prepared to pay Facebook to see no ads, with 8%, saying they would pay USD 10 per month. An alternative approach, by Brynjolfsson et al. (2018)¹⁴, is to estimate a shadow price for Facebook access. From their research, which is primarily using methods to estimate numbers on willingness to pay, they assess a value of around USD 500 per year for the median Facebook user.

34. Although the Greenlight survey was more about advertising than data, extrapolating from the findings suggest that there are significant differences in the prices consumers would be willing to pay for free services, and, so, implicitly, significant differences in their valuations of the price of their data. A broad brush approach to valuation therefore may create particular challenges for comparisons of distributional income and inequalities, especially, as seems plausible, higher income groups are more willing to pay than lower income groups. Indeed this seems likely to be equally true for business valuations of consumer data.

4.3. Data: non-produced versus produced assets

35. Currently the 2008 SNA treatment implies that when data materialise there is no direct impact on GDP – the effects are only indirect as data generates value from its use in producing other products. Moving away from the idea that data only materialises when there is a monetary transaction, and recognising, at least conceptually, that data provided in exchange for access to services is akin to a barter transaction, opens the door to including data in the system as a non-produced asset whenever such an exchange occurs. The logical consequence of such a treatment (i.e. to impute a barter transaction), as the ‘free’ or ‘bartered’ services typically received are produced, would be to increase GDP; namely, through households providing their data (a non-produced asset) in exchange for services provided by the (typically a platform) firm accessing data.

36. As noted above, the decision not to treat data as produced was in large part a function of the fact that to do so would lead to an implicit recognition that all knowledge was produced, and as such should be valued as contributing to GDP; which would in effect significantly distort the ability of GDP to serve as a meaningful measure of macro-economic policy making. However, adopting a similar approach to that currently used for

¹³ www.prnewschannel.com/2013/07/25/greenlight-poll-shows-facebook-users-would-be-willing-to-pay-10-to-see-no-ads-at-all/.

¹⁴ Brynjolfsson, E. et al. (2018), “The digital economy, GDP and consumer welfare: Theory and evidence”, Paper prepared for the 6th IMF Statistical Forum on “Measuring Economic Welfare in the Digital Age: What and How?”, Washington DC, 19-20 November 2018.

data – i.e. to record flows only when a monetary transaction is revealed – could mitigate the distortionary impacts on GDP statistics.

37. In addition, following the arguments set out above for barter transactions, a similar line of reasoning could be used to record activity when data is provided in return for platform related services. However, it is important to note that recording flows for data that satisfies asset requirements in this way would have a greater impact on GDP than would be the case in treating them as non-produced activities, especially for digital identity data. In the first instance, the original ‘own-account’ production of data – akin to the creation of a knowledge asset - increases GDP. If those data - or copies of the data - are subsequently used to generate an asset (the database), GDP is again increased. And finally, the use of that asset, for example to generate advertising services, increases GDP a third time. One potential way to avoid this triple impact, would be to record the initial creation of the original as an ‘other changes in volume’ in the accounts of the entity producing the data, in which case the build-up of an asset by the database owner would correspond to negative investment on the part of the household generating the data.

38. In addition, whereas the treatment of non-monetary flows of non-produced assets exchanged without any bartering occurring have no impact on GDP, the same does not necessarily follow if data are treated as being produced, and have the same value as the same data provided in a monetary exchange or in an implicit barter transaction, unless the acquirer of data also records an other volume change for the ‘free’ data that acquire as they build up their produced asset.

5. Summary and conclusions

39. Presently the SNA only, explicitly, recognises transactions related to data when a monetary transaction occurs, and only then as goodwill, meaning that it is implicitly treated as a non-produced asset. However, such an approach means that the current accounting framework is not well equipped to reveal the current data revolution that is driving many new digitally related business models. It is clear that the accounts require innovations that better reveal the role of data, and its use in driving new forms of production.

40. The discussion above considers a variety of different approaches that could provide a better view of that role compared to the current SNA; some of which require changes to the conceptual treatment of data, with consequential impacts on GDP. In the simple vanilla case concerning the treatment of data as goodwill when it appears in the accounts, one simple elaboration, that would improve visibility, would be to recognise data as a separate and distinct category of non-produced assets.

41. An extension of this, requiring a change to the current SNA, as at present exchanges in data are only explicitly recognised when a monetary transaction occurs, would be to identify flows whenever a barter transaction can be identified, which would increase GDP. The challenge here is in developing meaningful valuation mechanisms for the data which have been exchanged. This is particularly onerous when one considers the fact that some data (or rather individuals) are more valuable than others, at least when viewed from the firm’s perspective. On the other hand, there is a danger that an imputation of data based on a unitary value that does not discriminate on the basis of ‘whose data’ will create a potentially distortionary view of measures of inequality – inequality gaps of expenditure, income and disposable income will inevitably decline with such an approach, which has significant political ramifications. National accountants will almost certainly be accused of

politicising the account framework. On the other hand, viewed from the perspective of consumption, and because a barter mechanism implies an exchange of products of equal value, it is hard to justify having different imputed prices for the same products provided to high net worth individuals and lower income groups.

42. Extending the principle beyond barter mechanisms to include data that is provided in an unrequited manner can of course fit within the accounting framework but this too comes at a cost, requiring at a minimum, in the case of a treatment as non-produced assets, an appearance of a non-produced asset and a subsequent current transfer.¹⁵ In the case of recording data as produced assets – and certainly to avoid significantly increasing GDP, given the multiplicative effects that data have in producing downstream GDP through the provision of advertising services and the creation of an asset – at a minimum, it would be necessary to impute a similar transaction, for example ‘another change in volume’ transaction.

43. Overall, the arguments in favour of treating data as produced remain as contentious as they did when the debate on databases took place during the 2008 SNA, and, on balance, data should continue to be treated as non-produced, irrespective of whatever decisions are made concerning the treatment of unrequited or bartered data. But that is not to say that some modifications to the valuation of databases are not necessary. One clear improvement that could and perhaps should be made is to extend the sum of costs approach to valuing databases to include any resources that are used to do data analytics on the collected data, as ultimately the research into and the detection of patterns in behaviour is what generates the value inherent in the data. Further work will need to be undertaken to determine how this should be defined and indeed how it could be measured.

44. Other issues that will need to be explored going forward, concern the classification of activities (or firms) that provide bartered free services. Social media platforms are clearly in scope, as are many other digital activities that rely on the collection of data, such as search engines, and even in some cases e-mail accounts. It will also be necessary to identify platforms that collect data but in an unrequited way and indeed to note that not all digital free services are necessarily in scope. For example the provision of free services that are not dependent on data collection but rather act as an entry point and a vehicle for potential sales by the provider of the free service or indeed other firms that pay to access consumers through the platform (the Android system and Google Play may be cases in point).

45. Whatever the final outcome however, it is imperative that data flows wherever they arise should be recorded as a complement to the current accounting framework; indeed this is the approach being pursued by the OECD in its development of a satellite account.¹⁶ The satellite account should necessarily differentiate between imputations undertaken in a barter mechanism and data that have been provided with no exchange.

46. Certainly given the potential impact on GDP (and indeed measures of inclusiveness) from the inclusion of ‘free’ services in the account merit a prudent approach

¹⁵ One could also consider recording the appearance of the asset as another change in the volume of assets in the accounts of the relevant corporations, thus not needing an exchange of the assets between the households and the enterprises.

¹⁶ See Ahmad, N., and J. Ribarsky (2018), “Towards a framework for measuring the digital economy”, Paper presented at the 35th General Conference of the International Association for Research on Income and Wealth (IARIW), held on August 20-25, 2018 in Copenhagen, Denmark, www.iariw.org/copenhagen/ribarsky.pdf.

going forward, and, so, **the recommendation here, at least until robust measurement approaches have been developed, is to continue to define GDP as it is currently recommended, albeit with a new separate entry in the accounts for data, as a non-produced asset.**