

Treatment of Data in National Accounts*

Dylan G. Rassier
Robert J. Kornfeld
Erich H. Strassner

Paper prepared for the BEA Advisory Committee

May 2019

1. Introduction

The treatment of data stocks and flows in national accounts has been a topic of discourse among statistical agencies and international organizations since at least the last revision of the *System of National Accounts (SNA)*. Early conversations settled on the treatment of *databases* as an accompaniment to software in capital formation (Ahmad 2004, 2005). More recent conversations have a renewed focus on the treatment of the information content of databases – i.e., the embedded *data* – in response to the rapid increase in the collection and use of data among businesses, governments, non-profits, and households over the last 10-15 years (Ahmad and Van de Ven 2018, Nijmeijer 2018, Ribarsky 2019).

Anecdotal evidence on the value of data as a business asset or a commodity is abundant in popular media articles and other outlets (e.g., Chua 2013, Dance et al. 2018, Dezember 2018, Dwoskin and Timberg 2018, Giles and Ram 2019, Mandel 2017, Satariano and Isaac 2018, Valentino-DeVries 2018, World Economic Forum 2011). At the end of 2018, two of the largest global data firms – Google and Facebook – had a combined market capitalization of \$1.1 trillion and net income before tax of \$60.3 billion. The two firms amounted to 5.3 percent of the market capitalization of all S&P 500 firms and 3.6 percent of U.S. domestic corporate profits before tax. The value of data is also supported by literature on the economics of privacy (Acquisti et al. 2016). While the value of data may be apparent, the treatment of data in business accounting and economic accounting is not so apparent. Neither U.S. generally accepted accounting principles (GAAP) (nor International Financial Reporting Standards) nor *SNA 2008* provide adequate guidance on the treatment of data. Likewise, understanding and measuring the value of data presents challenges to both firms and economic statisticians (e.g., Akred and Samani 2018, Ballivian and Maret 2015, Hughes-Cromwick and Coronado 2019, Li et al. 2019, OECD 2013, Ribarsky 2019, Short and Todd 2017, U.S. Department of Commerce 2016, Waters 2018, Wilson et al. 2000, Wixom and Ross 2017).

* Corresponding author: Dylan Rassier at dylan.rassier@bea.gov or 301-278-9018. While the views (and any related errors) expressed in this paper are those of the authors, the topic addressed in the paper reflects an ongoing collaboration between the national accounts groups of the U.S. Bureau of Economic Analysis and Statistics Canada. We thank Michael Armah, Hussein Charara, Jen Lee, Wendy Li, Matt MacDonald, Emmanuel Manolikakis, Jessica Nicholson, Greg Prunchak, Marshall Reinsdorf, Jennifer Ribarsky, Carol Robbins, Rachel Soloveichik, Phil Smith, Jim Tebrake, Hal Varian, Peter van de Ven, Catherine Van Rompaey, Chander Velu, Dave Wasshausen, and participants at the November 2018 meetings of the OECD Informal Advisory Group on Measuring GDP in a Digitalized Economy and the UN's Advisory Expert Group on National Accounts for early discussions and suggestions.

This paper outlines preliminary thoughts and considerations for the inclusion of data stocks and flows in a national accounts framework. The next section of the paper briefly summarizes the current treatment of data in the *SNA* and in the U.S. national accounts. Section 3 presents considerations that serve as a foundation for a proposed treatment of data in national accounts, and section 4 discusses how data may fit into the *SNA* framework and presents some cursory estimates of data-related flows based on official statistical sources. However, we do not attempt at this point to measure the value of data stocks and flows for the U.S. economy, and our cursory measures cannot be used to compare against published measures in the U.S. national accounts. Section 5 concludes.

2. Current Treatment of Data in National Accounts

2.1. System of National Accounts

The 1993 version of the *SNA* includes only a brief paragraph on the inclusion of “large databases that the enterprise expects to use in production over a period of time of more than one year” as part of the computer software category of capital formation (*SNA 1993*, paragraph 10.93). There is no mention of embedded data in *SNA 1993*. Leading up to the 2008 version of the *SNA*, the Canberra II Group carefully considered the inclusion of embedded data in capital formation. To guide the Group’s discussions, Ahmad (2004) outlines two components of databases – supporting software and data stored in the database – and summarizes practical challenges that countries encounter while trying to implement the vague *SNA 1993* recommendation. In light of the challenges, Ahmad (2005) describes two definitions for databases considered by the Canberra II Group. One definition included the value of the information content to be stored in databases as long as the information had a useful life of more than one year, and one definition did not include the value of the information content. The Group recommended that the latter definition is preferable because the former definition would “open the door to the capitalization of knowledge” (Ahmad 2005, page 2). Based on the summary outlined in Ahmad (2005), the Canberra II Group primarily considered databases maintained by statistical agencies. The Group did not consider the use of data as intermediate consumption.

The recommendation that was ultimately written in *SNA 2008* includes databases combined with computer software as a separate category of intellectual property products in capital formation (*SNA 2008*, paragraphs 10.109-10.114). If a database is developed for own use, *SNA 2008* recommends a sum of costs approach to value the database. The sum of costs includes the cost of preparing data in a format that conforms to the database but excludes the cost of acquiring or producing the data. In addition, the sum of costs excludes the value of the database management system (DBMS), which is included instead with computer software. If a database is developed for sale or for license, the value should be determined by the market price, which includes the value of the information content. Thus, *SNA 2008* recommends an inconsistent treatment for data in capital formation depending whether a database is developed for own use or for sale or license. The *SNA 2008* does not include any explicit considerations for the use of data as intermediate consumption. However, the value of data acquired or produced for inclusion in databases is not to be treated as intermediate consumption in the sum of costs approach for own-account databases, which is inconsistent with the inclusion of intermediate consumption costs in the usual sum of costs measurement.

2.2. U.S. National Accounts

The U.S. national accounts are consistent with *SNA* recommendations on intellectual property products, including computer software (BEA 2017, Chapter 6). Similar to general practice in other countries, the U.S. accounts do not include a separation between software and databases (i.e., the software that houses data) in published capital stock and flow measures. The value of any data included in purchased software is included in measures of investment and capital stock. The value of any data in own-account software is excluded from measures of investment and capital stock.

BEA estimates three categories of software: 1) prepackaged, 2) custom, and 3) own-account. Benchmark estimates of prepacked and custom software are determined using a commodity flow method based on receipts reported in the U.S. Economic Census for NAICS 5112 (software publishers), 5182 (data processing, hosting, and related services), and 5415 (computer systems design and related services). For non-benchmark years, estimates are based on receipts reported in the U.S. Census Bureau's Service Annual Survey. Benchmark estimates of own-account software are determined using a sum of costs approach based on wage data in the Occupational Employment Statistics (OES) survey from the U.S. Bureau of Labor Statistics for Standard Occupational Classification (SOC) codes 15-1131 (computer programmers), 15-1132 (application software developers), 15-1133 (systems software developers), and 15-1121 (computer systems analysts) and on the Economic Census. For non-benchmark years, estimates are primarily based on the OES data. In addition to labor and intermediate input costs, own-account software includes a cost for capital services (Chute, McCulla, and Smith 2018).

3. Considerations for Data

This section first introduces a classification option for data and the idea of a data boundary and then describes a production process for data that national accountants have used from information science (e.g., Li et al. 2019). These considerations can be used as building blocks for a subsequent discussion in section 4 to understand how data may fit into a national accounts framework.

3.1. Classification and the Data Boundary

In order to understand how data may fit into a national accounts framework, a useful classification is necessary. The World Economic Forum (2011) and OECD (2013) identify two broad categories for data – personal data and institutional data – that may help guide classification. Table 1 summarizes the data categories and their content.

The OECD (2013) outlines six types of personal data: 1) *user-generated content* such as photos, videos, and blogs; 2) *activity or behavioral data* such as internet search and online purchases; 3) *social data* such as contacts and friends; 4) *locational data* such as IP addresses, residential addresses, and geolocation; 5) *demographic data* such as age, gender, race, income, and political affiliation; and 6) *identifying data of an official nature* such as name, financial information, health information, and police records. Personal data receive a lot of attention recently because of data-dependent business models such as social media, internet search, and other online platform firms that rely on access to personal data for revenue and profits. Facebook and Google, for example, each cite security breaches of personal data and personal privacy regulations as risk factors in their

annual reports filed with the U.S. Securities and Exchange Commission. In July 2018, Facebook's stock price fell 19 percent in a single day as a result of missed revenue targets and lower revenue projections generated by personal data security and privacy concerns (Otani and Seetharaman 2018). Likewise, the economics of personal privacy has re-emerged as an area of keen interest in economic research as a result of the internet and business models driven by personal data (Acquisti et al. 2016). Thus, personal data are an important category of data that will require consideration for treatment in national accounts.

In addition to personal data, institutional data are important to the functions and profitability of businesses and to the functions of governments and non-profits. For businesses, institutional data include customer lists, personnel files, accounting records, and legal and financial documents. In addition to those traditional types of business data, we include internet of things (IoT) sensor technology under business. Sensors are increasingly embedded in manufactured products such as household appliances, manufacturing equipment, farm implements and other machinery, automobiles, and airplanes. Among other uses, sensors collect data on the performance of products and their parts, which can be transmitted to manufacturers and used by them as a signal to provide maintenance services and parts before a breakdown occurs.

For governments, institutional data include intelligence records, diplomatic cables, defense files, and statistical surveys as well as regulatory and administrative data such as social security profiles, tax returns, passports, environmental compliance records, financial accounting reports, and public health and safety records. Governments also collect data through advanced monitoring technologies such as traffic cameras, public transportation scanners, and satellites. For non-profits, institutional data include records on social and public policy programs that they administer. Similar to businesses, both governments and non-profits also have personnel files, accounting records, and other data for their own operations. Unlike businesses, governments and non-profits maintain confidential data and data for public consumption. They generally do not use data for commercial purposes, which does not mean that businesses do not benefit from government and non-profit data.

The categories in table 1 seem useful for national accounts because they can be matched with the institutional sectors of *SNA 2008*. However, there may be other classifications that should also be considered. Once a classification for data is constructed, decisions can be made about categories of data that fit into the scope of the national accounts framework and categories that are outside the scope. Figure 1 illustrates the extent to which data may fit into the scope of the national accounts framework. The green and orange circles represent the *SNA* asset boundary and production boundary, respectively. Non-produced assets – i.e., natural resources, contracts, leases, licenses, purchased goodwill, and marketing assets – fit within the asset boundary but not within the production boundary. Produced assets – i.e., fixed assets, inventories, and valuables – reflect capital formation that fits within both the asset boundary and the production boundary. The production boundary also includes non-assets that are used for final products, intermediate inputs, or exports.

The blue circle in figure 1 represents the data boundary. Areas B, C, and D overlap the *SNA* boundaries and represent data categories (e.g., personal and institutional) that fit within the scope of the *SNA* framework as either non-produced assets, produced assets, or non-assets, respectively.

Area A of the circle represents data that lie outside the scope of the *SNA* boundaries. Regardless of how data are classified, we expect that each of the areas in the blue circle will include some data categories. In other words, not all data in the data boundary are eligible for inclusion in the national accounts framework and not all data that fit within the framework should be treated as capital formation or as assets at all. Determining the actual content of each of the areas A, B, C, and D is a matter for future work.

3.2. *Production of Data*

Varian (2018) describes a data pyramid that is a variation of the data-information-knowledge-wisdom (DIKW) hierarchy introduced by Ackoff (1989) and subsequently used in information science (e.g., Rowley 2007). The data pyramid is used to illustrate the relationships among data, information, and knowledge. Figure 2 presents the pyramid. In the foundation of the pyramid, data is collected and stored as bits. In the next rung, processing and analysis generate information that is stored in documents. The third rung consists of learning that leads to knowledge, which is stored in humans. Action can be taken once the bottom three rungs are realized. Varian (2018) asserts that markets are well developed for information (e.g., economic statistics) and knowledge (e.g., economic statisticians) but not so well developed for data in its unorganized raw form. A point that is illustrated by the pyramid is that knowledge is embodied by humans, but data cannot be embodied by humans. Thus, data cannot be a knowledge asset but may become an information asset. Knowledge is generally outside the scope of the *SNA* asset boundary.

The concepts illustrated in the data pyramid are related to a data value chain presented in OECD (2013) and subsequently expanded in Moro Visconti et al. (2017). The data value chain presented in figure 3 demonstrates a five-stage production process for data from an unstructured form that has very little value to a structured form that can be leveraged in a business model or other usage. When high volumes of data are collected and accessed in the first stage, they may be unstructured such as data collected via electronic payment systems (e.g., credit card purchases and Venmo), internet-connected machines and devices (e.g., smart phones and IoT), or other methods. In this case, their accuracy has not been validated and they are not ready for use, so value is low. Likewise, data may be collected and accessed in the first stage from sources such as regulatory filings (e.g., tax returns and annual financial reports) and other methods that require less validation and are much closer to usage in the last stage. Thus, the data value chains presented in OECD (2013) and Moro Visconti et al. (2017) are designed with “big data” in mind but can also be used to understand the value of “small data” that may require less storage space and be subject to more traditional processing and distribution in steps 2 to 4.

In the second stage in figure 3, data are stored and aggregated in servers, individual data centers, cloud service facilities, and other locations until they are processed in the third stage. The value of data increases in the third stage with fusion and analytic techniques. Fusion enables a more comprehensive profile to be developed on a household or business—the larger the profile the more valuable the data record. Scientific methods for analyzing data include techniques such as statistical modeling, data mining, machine learning, artificial neural networks, and social network analysis. Distribution in the fourth stage is made possible by tracking companies, data exchanges, ad networks, data brokers, and other outlets, which makes data available to a wider audience.

The Moro Visconti et al. (2017) version of the data value chain focuses on business users in the fifth stage with the monetization of data via a business model. The OECD (2013) version includes business users and also includes government, non-profit, and household users. Businesses may be data-dependent or data-neutral. Data-dependent business models, such as advertising-supported media firms, rely exclusively on data for sources of revenue and profits. Data-neutral firms do not depend on data but can still realize benefits from data if it helps improve existing products or offer new products. Governments and non-profits use data to develop, administer, and improve regulatory and social programs. Households may use some data directly – such as tracking exercise progress on a Fitbit – and the OECD (2013) argues that households benefit from data-driven models through cheaper products, better product variety, improved services, and social and professional networking opportunities. While some household uses of data may be within scope of the *SNA*, the latter advantages are welfare effects and externalities that are outside the scope of national accounts.

3.3. Features of Data

Data share features with some *SNA* intellectual property products and with some *SNA* capital goods. Similar to R&D and software, data are non-rival but excludable, which poses a risk of multiple counting. Likewise, data do not necessarily experience wear and tear but can become obsolete. Like cultivated assets early in their service lives (e.g., breeding livestock and orchards), data may experience appreciation until they reach their peak point of production, so their depreciation profiles may not follow a typical pattern assumed by national accountants. For example, data may become more valuable through aggregation (Li et al. 2019). Alternatively, aggregation may actually create an entirely new product (i.e., data record) that just happens to be more valuable than the old product. In other words, there may be no appreciation at all. In any of these cases, the challenges imposed on national accountants are not new.

There are also features of data that are unique to conventions in the current *SNA* framework. First, data have characteristics of both goods and services—data are storable like goods and intangible like services (Mandel 2012). Second, the provision of personal data by households to other sectors at the beginning of the data value chain may yield transactions that are not customary in national accounts. Nevertheless, the transactions do fit into the current design of the *SNA* framework if we are willing to accept households as producers of data or personal data as “currency” for some products. We return to these two points in section 4.3. Third, the value of the same data may vary across users because the data may be used differently across users—i.e., the value of data depends on the context in which it is used (OECD 2013). Related to this, information asymmetries on the value of data may exist between the providers of data and the users of data. Solutions to these challenges are outside the scope of this paper.

4. Data in the *SNA* Framework

The data pyramid and data value chain imply a production process exists for data that should be considered further for treatment in national accounts. In this section, we relate concepts from the *SNA* framework to the data value chain and discuss how data may fit into the *SNA* framework. The *SNA* concepts we focus on include institutional sectors and industries, economic ownership, the supply-use identity of the goods and services account, and asset categories (produced and non-

produced). A stylized relationship between the data value chain and the *SNA* concepts is depicted in figure 4.

The scope of figure 4 is limited to data within the *SNA* production and asset boundaries—i.e., areas B, C, and D of the data boundary in figure 1. The top part of the diagram – panel I – loosely shows how the data value chain may fit into the production and asset boundaries of the *SNA*. The boundaries are shown with dashed ovals to reflect the fluidity with which data may or may not be included in each. For example, a photo of a new puppy that is posted to social media and used by advertisers to target the household for dog food ads may fall within the *SNA* production boundary because it is being used as data for production purposes. Alternatively, if the photo is simply viewed by other households, it is probably outside the scope of the *SNA* production boundary because the activity all takes place among households. The production boundary is likely to include stages 2-5 of the data value chain but may not include stage 1 depending how data are collected and accessed. For example, transactions data collected by credit card firms and location data collected by mobile phone carriers and mobile app providers may not be part of the production boundary if we consider them “naturally” occurring (i.e., collection is possible at very low cost as a by-product of other activity), but regulatory data collected by tax and financial authorities may be part of the production boundary because their collection requires more inputs. In either case, the data that are collected are subject to consideration for inclusion in the asset boundary—produced or non-produced.

4.1. Institutional Sectors and Industries

Panel II of figure 4 reflects institutional sectors and industries involved in each stage of the data value chain. All sectors are likely to be involved in the collection and usage of data in the first and last stages of the value chain, respectively, but households are not as likely to be involved in the intervening stages. While some industries in the business sector are engaged in the production of data as a primary product, data as a secondary product are likely to span many industries, especially with the increase of IoT sensor technology in manufactured goods. Likewise, an increasing number of industries are likely to use data as intermediate consumption or capital formation.

One option for providing more granularity on data in the *SNA* institutional sectors framework would be to include subsectors for data-dependent and data-neutral non-financial corporations. In the supply-use tables, data industries already exist to some extent under NAICS and ISIC – e.g., NAICS 518 and ISIC 631 (data processing, hosting, and related services) – and an additional industry could be considered for personal data provided by households, which would be similar to including owner-occupied housing in the real estate industry. In addition, a data commodity could be considered for inclusion in supply-use tables.

4.2. Economic Ownership

Transactions across institutional sectors and industries are recognized in the *SNA* when economic ownership changes. Economic ownership of data reflects an economic claim on any benefits that can be reaped from the data, which may be influenced by institutional factors across countries. In the U.S., for example, collection and control of large quantities of data are privatized to the extent that once households knowingly or unknowingly submit personal data such as mobile phone

location or internet search, economic benefits attached to the data also shift to the business sector (Valentino-DeVries et al. 2018). In the European Union, the General Data Protection Regulation (GDPR) gives households control over the collection and use of personal data, and GDPR and other proposed regulations are introducing uncertainty to the future profitability of data-dependent firms.

Economic ownership is reflected with the gold border in panel II of figure 4. The tapered shape of the border reflects the changing economic ownership of data as it moves through the data value chain. In the first stage, data are owned by the sector in which the data originate. As data move through the value chain, ownership may become more concentrated in specific sectors such as the business, government, and non-profit sectors.

Access to and control of large amounts of data in the U.S. business sector is currently concentrated among a small group of large firms, but that may change in the future with blockchain technology (Popper 2018a). Blockchain is usually associated with cryptocurrencies because it was introduced with Bitcoin, but the technology has uses in other areas such as smart contracts and supply chain management. Blockchains are an idiosyncratic form of database in which blocks of data are chained together by sophisticated mathematics. Appealing characteristics of blockchains generally include decentralization, anonymity, immutability, and security. The technology may be used between organizations or within an organization. An example of the potential usefulness of blockchain is in applications of artificial intelligence (AI) (Popper 2018b). Since data are an important input into AI, small AI firms are attempting to decentralize access to and control of data by developing block chain marketplaces that will serve as forums for buying and selling data. AI experts are hoping that blockchain can allow AI networks to access large stores of data without any big company in control of the data or algorithms. Thus, blockchain could change the landscape for economic ownership under *SNA* guidelines.

4.3. Supply and Use of Data

The supply-use identity of the goods and services account is depicted in panel III of figure 4. Recall two of the unique features of data from section 3.3—characteristics of both goods and services and provision of personal data by households at the beginning of the data value chain. As suggested by Mandel (2012) and proposed by Ribarsky and Ahmad (2018), panel III includes a goods, services, and *data* account to reflect the unique product characteristics of data. To reflect the role of households at the beginning of the data value chain, the account in panel III also includes non-monetary flows and consumer to business (i.e., C2B) intermediate consumption.

The scope of the goods, services, and data account is limited to the production boundary, which is illustrated in figure 4 with the orange box. To understand the supply of data as a product, we focus on the ways in which data enter the data value chain in stage 1. Likewise, to understand the uses of data as a product, we focus on stage 5. While the goods, services, and data account in orange spans all stages of the chain in which production takes place, much of the activity in intervening stages is traditional market activity that is likely captured one way or another in the supply-use identity—even if the current classification of some flows is subject to question.

Since data that are controlled by governments and non-profits are generally confidential or available for public use rather than commercial use, we do not consider them here. We also do not consider cross-border flows of data beyond including imports and exports in the goods, services, and data account. We instead focus on the provision of personal data by households to business and IoT sensor data collected by business.

Stage 1: Supply of Personal Data and IoT Sensor Data

Domestic production of goods and services generally comes from three sources: monetary transactions, non-monetary flows, or own-account flows. The supply of personal data in stage 1 fits into monetary transactions or non-monetary flows, and the supply of IoT sensor data fits into non-monetary flows or own-account flows.

Personal data may be purchased from households in traditional monetary transactions where the household receives money in exchange for data. Under non-monetary flows, households barter personal data for “free” services such as internet search, social networking, video and audio streaming, and mobile apps. In addition, large amounts of personal data are made available to business in stage 1 through sources that do not fit into monetary transactions or barter transactions, such as electronic payments and geolocation.

Data are collected on households via electronic payment systems such as credit cards for online and offline purchases. In addition to the payment record, a second record of the transaction is generated by the vendor. For example, an online purchase from Amazon using a Chase credit card yields data for both Amazon and Chase to build a profile on the customer. Likewise, an offline purchase at Whole Foods (owned by Amazon) using the same Chase credit card and a Prime membership discount yields data for Amazon. In the latter case, the transaction could be considered monetary since the household receives a discount.

Geolocation data are collected on households via smart phones and other mobile technology by mobile service providers and by location firms that embed code in apps (Valentino-DeVries et al. 2018). The carriers and location firms track patterns in movements of mobile users, and the data are then sold to advertisers, retailers, and investment firms who use the data to understand and target customers, predict demand, and make investment decisions based on patterns of workers showing up to factories or shoppers to retail outlets.

IoT data are collected by sensors embedded in an increasing number of internet-connected machines and devices used in homes, factories, farms, and other places. The sensors collect data that are then used to monitor and improve performance among other uses. For example, a sensor embedded on a home appliance or a farm implement can transmit data to the original equipment manufacturer (OEM) who then arranges parts and maintenance service on a just-in-time basis using distributed ledgers and additive manufacturing. Velu (2019) describes an example of a consumer appliance manufacturer that licenses the necessary intellectual property to a third-party contractor close to the customer and the contractor then uses 3-D printing to replace parts on demand without the need for long lags that result from low inventories or transportation. The intellectual property transfers and payments are all managed through a smart contract made possible with a distributed ledger.

Stage 5: Uses of Personal Data and IoT Sensor Data

Domestic uses of goods and services generally include final consumption, intermediate consumption, or capital formation. Uses of personal data in stage 5 fit into any of the three uses. Uses of IoT sensor data fit into either intermediate consumption or capital formation.

A few examples of general uses of data by business include advertising, digital transformation, and artificial intelligence. Advertisers use data to target consumers based on revealed preferences from activities such as internet search and location patterns. Digital transformation includes the use of data to improve internal processes and decisions, wrap information around existing core products, or sell information offerings to new and existing markets (Wixom and Ross 2017). Artificial intelligence requires cloud computing equipment (i.e., physical capital), programming talent (i.e., labor), and lots of data to run through algorithms.

If personal data are used by households for final consumption, we consider the flows out of scope because all activity takes place within the household sector. Personal data that have been provided to business may be used for intermediate consumption, and we include a separate component of intermediate consumption for consumer to business (C2B) transactions to reflect the direction of flows from households to business. This type of transaction is not customary in national accounts but also does not require any modifications to the current design of the *SNA* framework. IoT sensor data that are used immediately in production with no future uses should be treated as intermediate consumption under business to business (B2B) transactions.

A logical starting point for the uses of personal data and IoT sensor data as capital formation is the expanded conceptual framework proposed in Corrado, Hulten, and Sichel (2005) for the Solow (1957) and Jorgenson and Griliches (1967) growth accounting model. Corrado, Hulten, and Sichel (2005) argue that, regardless of measurement challenges, economic theory supports treating investments in intangible capital symmetrically with investments in tangible capital. The framework they present includes three categories of intangible capital: computerized information, innovative property, and economic competencies. The category *computerized information* reflects “knowledge” embedded in computer programs and computerized databases. While Corrado, Hulten, and Sichel (2005, 2009) do not explicitly address the treatment of data in their framework, personal data and IoT sensor data fit well into the *computerized information* category of the framework and should be treated as capital formation if they are used repeatedly in future production.

4.4. Asset Categories

The *SNA* framework includes financial assets and non-financial assets. There are two categories of non-financial assets: produced and non-produced. According to *SNA* 2008 (paragraph 10.9), “Produced assets are non-financial assets that have come into existence as outputs from production processes that fall within the production boundary of the *SNA*.” In contrast, “Non-produced assets are non-financial assets that have come into existence in ways other than through processes of production.”

Produced assets include fixed assets, inventories, and valuables (*SNA 2008*, paragraphs 10.10-10.13). Fixed assets are used repeatedly in production for more than one year, and inventories are goods and services held for sale or use at a later date. Valuables are generally used as a store of value rather than for production or consumption.

Non-produced assets include natural resources, purchased goodwill, marketing assets, and contracts, leases and licenses (*SNA 2008*, paragraphs 10.14-10.17). Natural resources are naturally occurring resources that have economic value, such as land, water, and minerals. Contracts, leases, and licenses are assets only when their terms specify a price that differs from the price that would prevail in their absence and when on party is legally and practically able to realize the price difference. Purchased goodwill is only recorded when an entire institutional unit is purchased, and marketing assets are only recorded when an identifiable marketing asset is purchased.

Panel IV of figure 4 reflects asset categories. Given the production process for data that is implied by the data value chain, we generally consider data assets to be produced. However, we also acknowledge that some data may be “naturally” occurring before they enter the data value chain.

4.5. Valuing Data

The data value chain illustrates how the value of data evolves. Consistent with the data value chain, Li et al. (2019) present empirical evidence on the increase in the value of data across the chain using case studies for eight online platform companies. Likewise, Nakamura, Samuels, and Soloveichik (2018) offer empirical evidence on the relatively modest value of “viewership services” provided by U.S. households in exchange for advertising- and marketing-supported media, which we interpret as a proxy for an upper bound value of personal data in stage 1.

Valuation Methods

The preferred method of valuation in *SNA 2008* is a market-based approach using observed transactions. For some types of products – such as intellectual property products – *SNA 2008* recognizes active markets may not exist and, thus, recommends a cost-based approach using a sum of compensation, intermediate consumption, consumption of fixed capital, net return on fixed capital, and taxes less subsidies on production (other than taxes less subsidies on products). A third option available for hard-to-value products but not favored in *SNA 2008* is an income-based approach that models the discounted present value of a future profit stream.

OECD (2013) outlines five options for valuing personal data—not all of which are useful to national accountants. The most direct option is market prices paid and received in actual transactions—either legal markets or illegal markets. Another option is market capitalization or revenue (net income) per user or data record for data-dependent firms such as Facebook or Experian. A third option is measuring the economic costs to individuals who have their identities stolen or to firms that are victims of data breaches. A fourth option is households’ willingness to pay to protect their data by observing premiums for identity theft insurance or other types of protection. Finally, economic experiments and surveys can be used to determine how much compensation households would require to give up some of their personal data to firms.

Valuing Data for the U.S. Economy

We do not attempt to measure the value of data stocks and flows for the U.S. economy. To establish some context to understand the extent of data-related activity for the U.S. economy, we consider official statistical sources on data-related products for the business sector. This approach provides a starting point for assessing currently available sources and methods and provides insights into classification gaps. For simplicity, we do not consider data controlled by governments or non-profits because those organizations generally do not use data for commercial purposes. We look at potential sources to measure two categories of data-related products: purchases and own-account. We estimate cursory measures that are incomplete and cannot be used to compare against published measures in the U.S. national accounts, such as GDP and investment. Our approach is conservative and we do not assert that the resulting measures are reflective of the actual value of data stocks and flows in the U.S. economy. Finally, we do not attempt to relate the cursory measures estimated here to personal data or IoT sensor data that we focused on for the stylized relationship in figure 4.

Purchases

To estimate a cursory measure of purchased data-related products, we use published receipts from the 2012 Economic Census on the following four products for NAICS 518 (data processing, hosting, and related services) that directly reflect data-related services: data storage services, data management services, other data processing or IT infrastructure provisioning, and information and document transformation services. Table 2 summarizes the four products. All four products are primary for NAICS 518. Total receipts for NAICS 518 on the four data-related products in 2012 were \$21.6 billion, which is treated as intermediate consumption in BEA's accounts.

The four data-related products from the 2012 Economic Census provide the most direct measure possible with current U.S. Census Bureau sources. A less direct measure that demonstrates growth since 2012 is receipts on a group of products referred to as "data processing, IT infrastructure provisioning, and hosting services" (product code 6364) published for NAICS 518 in the Service Annual Survey. This group includes twelve products from the Economic Census that are summarized in table 3. BEA uses total receipts for the product "application service provisioning" (product code 34930) when measuring investment in prepackaged software, and the other eleven products are generally treated as final consumption or intermediate consumption in BEA's accounts. Thus, we exclude product receipts for application service provisioning from the group of twelve in the 2012 Economic Census and extrapolate total receipts for the remaining eleven products using the percent change in receipts for the entire group published in the Service Annual Survey for 2013-2017. Figure 5 demonstrates the growth in the indirect data-related product receipts for 2012-2017. The percent change is as low as 1.5 percent in 2014 and as high as 12.7 percent in 2017. Total receipts for 2017 are \$78.1 billion.

Own-Account

To estimate a cursory measure of own-account data-related products under a sum of costs method, we use published 3-digit industry-level wage data from the Bureau of Labor Statistic's OES Survey for 2012-2017 on the following 2010 SOC categories for all private industries except NAICS 518

(since we used 518 for purchased data-related products): information security analysts (15-1122), database administrators (15-1141), actuaries (15-2011), mathematicians (15-2021), operations research analysts (15-2031), statisticians (15-2041), and mathematical technicians (15-2091).

We choose SOC categories for computer and mathematical occupations that score 85 or higher based on data-intensive work activities in Hawk et al. (2015). Hawk et al. (2015) identify 91 data-intensive occupations from computer, mathematics, science, engineering, education, and healthcare occupational categories, but we limit our choices to computer and mathematics categories because the forthcoming 2018 SOC system includes newly evolved data-intensive occupations such as data scientists in computer and mathematics categories. We exclude the SOC categories that BEA uses to estimate own-account software.

Since the sum of costs includes all production costs and not just wages, we multiple total OES wages for each year by a blow-up factor. We calculate the blow-up factor by averaging the ratio of total operating expenses to payroll for NAICS 518 from the Service Annual Survey for 2013-2017. We do not attempt to adjust for any costs that may already be counted in capital measures of R&D. We also do not attempt to adjust for time spent on activities other than data-related activities. Figure 6 demonstrates the growth in the data-related production costs for 2012-2017. The percent change is as low as 4.7 percent in 2013 and as high as 12.1 percent in 2016. Total production costs for 2017 are \$74.9 billion.

5. Conclusions

This paper outlines preliminary thoughts and considerations for the inclusion of data stocks and flows in a national accounts framework. Based on work to date, we draw three conclusions. First, a classification for data is necessary to determine what data are in and out of scope of the *SNA* boundaries. Not all data in the data boundary are eligible for inclusion in the *SNA* framework and not all data that fit within the framework should be treated as capital formation or as assets at all. A data classification should provide clarity about specific uses of data in a goods, services, and data account. Second, the treatment of data in national accounts faces some of the same challenges as capital measures of R&D, computer software, and other assets – e.g., prices and depreciation – but a unique challenge imposed by data is the valuation and treatment of personal data collected by firms from households at the beginning of the data value chain and the treatment of IoT sensor data. Even though we focus on the use of personal data and IoT sensor data by the business sector for commercial purposes, data under the control of governments and non-profits also have value that should be considered for national accounts. Third, the cursory estimates we present for purchased and own-account data-related products demonstrate that measures of data-related products such as data mining and management are possible with existing industry and product classifications, statistical sources, and methodologies, which can be expanded and improved to include more explicit references to data in order to also measure stocks and flows of data that are deemed to be within scope of national accounts.

References

- Ackoff, Russel L. 1989. "From Data to Wisdom" *Journal of Applied Systems Analysis*, 16, pp. 3-9.
- Ahmad, Nadim. 2004. "The Measurement of Databases in the National Accounts" Issue paper prepared for the December 2004 Meeting of the Advisory Expert Group on National Accounts.
- Ahmad, Nadim. 2005. "Follow-Up to the Measurement of Databases in the National Accounts" Issue paper prepared for the July 2005 SNA Update Issue 12.
- Ahmad, Nadim and Peter van de Ven. 2018. "Recording and Measuring Data in the System of National Accounts" Paper prepared for the Meeting of the OECD Informal Advisory Group on Measuring GDP in a Digitalized Economy, November 9.
- Akred, John and Anjali Samani. 2018. "Your Data is Worth More than You Think" *MITSloan Management Review*, January 18.
- Acquisti, Alessandro, Curtis Taylor, and Liad Wagman. 2016. "The Economics of Privacy" *Journal of Economic Literature*, 54(2), pp. 442-492.
- Ballivian, Amparo and Fenohasina Rakotondrazaka Maret. 2015. "Measuring the Value of Data" Paper prepared for the World Bank DECDG Seminar, November 18.
- Binns, Reuben, Ulrik Lyngs, Max Van Kleek, Jun Zhao, Timothy Libert, and Nigel Shadbolt. 2018. "Third Party Tracking in the Mobile Ecosystem" *WebSci '18 Proceedings of the 10th ACM Conference on Web Science*, pp. 23-31.
- Bureau of Economic Analysis. 2017. *Concepts and Methods of the U.S. National Income and Product Accounts*.
- Chua, Faye. 2013. "Big Data: its power and perils" Accountancy Futures Academy.
- Chute, Jason W., Stephanie H. McCulla, and Shelly Smith. 2018. "Preview of the 2018 Comprehensive Update of the National Income and Product Accounts" *Survey of Current Business*, 98(4).
- Corrado, Carol, Charles Hulten, and Daniel Sichel. 2005. "Measuring Capital and Technology: An Expanded Framework" in Carol Corrado, John Haltiwanger, and Dan Sichel, eds., *Measuring Capital in the New Economy*, Vol. 65, University of Chicago Press.
- Corrado, Carol, Charles Hulten, and Daniel Sichel. 2009. "Intangible Capital and U.S. Economic Growth" *Review of Income and Wealth*, 55(3), pp. 661-685.
- Dance, Gabriel J.X., Michael LaForgia, and Nicholas Confessore. 2018. "As Facebook Raised a Privacy Wall, it Carved an Opening for Tech Giants" *The New York Times*, December 18.

Dezember, Ryan. 2018. "Your Smartphone's Location Data is Worth Big Money to Wall Street" *The Wall Street Journal*, November 2.

Dvoskin, Elizabeth and Craig Timberg. 2018. "Facebook Discussed Using People's Data as a Bargaining Chip" *The Washington Post*, November 30.

European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank. 2009. *System of National Accounts 2008*, New York, NY: United Nations.

Giles, Chris and Aliya Ram. 2019. "Tech Giants Should 'Open Up Their Customer Data to Others'" *Financial Times*, March 12.

Goldfarb, Avi and Catherine Tucker. 2019. "Digital Economics" *Journal of Economic Literature*, 57(1), pp. 3-43.

Hawk, William, Regina Powers, and Robert Rubinovitz. 2015. "The Importance of Data Occupations in the U.S. Economy" Economics and Statistics Administration Issue Brief #01-15.

Hughes-Cromwick, Ellen and Julia Coronado. 2019. "The Value of US Government Data to US Business Decisions" *Journal of Economic Perspectives*, 33(1), pp. 131-146.

Jorgenson, Dale W. and Zvi Griliches. 1967. "The Explanation of Productivity Change" *Review of Economic Studies*, 34, pp. 249-283.

Li, Wendy C.Y., Makoto Nirei, and Kazufumi Yamana. 2019. "Value of Data: There's No Such Thing as a Free Lunch in the Digital Economy" BEA Working Paper.

Mandel, Michael. 2012. "Beyond Goods and Services: The (Unmeasured) Rise of the Data-Driven Economy" Progressive Policy Institute.

Mandel, Michael. 2017. "The Economic Impact of Data: Why Data is not Like Oil" Progressive Policy Institute.

Moro Visconti, Roberto, Alberto Larocca, and Michelle Marconi. 2017. "Big Data-Driven Value Chains and Digital Platforms: From value co-creation to monetization" in *Big Data Analytics*, Arun K. Somani and Ganesh Chandra Deka, eds., Chapter 16.

Nakamura, Leonard, Jon D. Samuels, and Rachel Soloveichik. 2018. "'Free' Internet Content: Web 1.0, Web 2.0 and the Sources of Economic Growth" Paper prepared for the IARIW 35th General Conference.

Nijmeijer, Henk. 2018. "Issue Paper on Databases" Paper prepared for the Joint Eurostat-OECD Task Force on Land and Other Non-Financial Assets.

Otani, Akane and Deepa Seetharaman. 2018. “Facebook Suffers Worst-Ever Drop in Market Value” *The Wall Street Journal*, July 26.

Organization for Economic Cooperation and Development. 2013. “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value” OECD Digital Economy Papers No. 220, OECD Publishing, Paris.

Popper, Nathaniel. 2018a. “Confused about Blockchains? Here’s What you Need to Know” *The New York Times*, June 27.

Popper, Nathaniel. 2018b. “How the Blockchain Could Break Big Tech’s Hold on A.I.” *The New York Times*, October 20.

Ribarsky, Jennifer. 2019. “Measuring the Digital Economy in Macroeconomic Statistics: The Role of Data” Paper prepared for the April 2019 Meeting of the UNECE Group of Experts on National Accounts: Measuring Global Production.

Ribarsky, Jennifer and Nadim Ahmad. 2018. “Towards a Framework for Measuring the Digital Economy” Paper prepared for the 35th IARIW General Conference.

Rowley, Jennifer. 2007. “The Wisdom Hierarchy: Representations of the DIKW Hierarchy” *Journal of Information Science*, 33(2), pp. 163-180.

Satariano, Adam and Mike Isaac. 2018. “Facebook Used People’s Data to Favor Certain Partners and Punish Rivals” *The New York Times*, December 5.

Short, James E. and Steve Todd. 2017. “What’s Your Data Worth?” *MITSloan Management Review*, March 3.

Solow, Robert M. 1957. “Technical Change and the Aggregate Production Function” *The Review of Economics and Statistics*, 39(3), pp. 312-320.

U.S. Department of Commerce. 2016. “Measuring the Value of Cross-Border Data Flows” Paper prepared by the Economics and Statistics Administration and the National Telecommunications and Information Administration.

Valentino-DeVries, Jennifer, Natasha Singer, Michael H. Keller, and Aaron Krolik. 2018. “Your Apps Know Where You Were Last Night, and They’re not Keeping it Secret” *The New York Times*, December 10.

Valentino-DeVries, Jennifer. 2018. “5 Ways Facebook Shared Your Data” *The New York Times*, December 19.

Varian, Hal. 2018. “Artificial Intelligence, Economics, and Industrial Organization” National Bureau of Economic Research Working Paper 24839.

Velu, Chander. 2019. “Management Information and Business Model Innovation: Unpacking the Productivity Paradox” in *Handbook on Digital Innovations*, eds., S. Nambisan, K. Lyytinen, and Y. Yoo, Edward Elgar Publishers.

Waters, Richard. 2018. “Turning Big Data into Money Proves Harder than Expected” *Financial Times*, October 4.

Wilson, Richard M.S., Joan Stenson, and Charles Oppenheim. 2000. “Valuation of Information Assets” Loughborough University Research Series Paper 2000:2.

Wixom, Barbara H. and Jeanne W. Ross. 2017. “How to Monetize Your Data” *MITSloan Management Review*, January 9.

World Economic Forum. 2011. “Personal Data: The Emergence of a New Asset Class”.

Table 1: Model Data Classification

<i>Personal Data</i>	<i>Institutional Data</i>		
	<i>Businesses</i>	<i>Governments</i>	<i>Non-Profits</i>
<i>User-Generated</i>	<i>Personnel Files</i>	<i>Personnel Files</i>	<i>Personnel Files</i>
Photos			
Videos	<i>Accounting Records</i>	<i>Accounting Records</i>	<i>Accounting Records</i>
Blogs			
	<i>Legal Docs</i>	<i>Legal Docs</i>	<i>Legal Docs</i>
<i>Behavior</i>			
Internet search	<i>Financial Docs</i>	<i>Financial Docs</i>	<i>Financial Docs</i>
Online purchases			
	<i>Customer Lists</i>	<i>Intelligence Records</i>	<i>Social Policy Programs</i>
<i>Social</i>			
Contacts	<i>IoT Sensors</i>	<i>Diplomatic Cables</i>	<i>Public Policy Programs</i>
Friends	Appliances		
	Equipment	<i>Defense Files</i>	
<i>Location</i>	Machinery		
IP address	Automobiles	<i>Statistical Surveys</i>	
Residential address	Airplanes		
Geolocation		<i>Regulatory Records</i>	
		Environmental	
<i>Demographic</i>		Financial	
Age		Safety	
Race		Public health	
Gender			
Income		<i>Admin Records</i>	
Political affiliation		Social security	
		Taxes	
<i>Official Identification</i>		Passports	
Name			
Financial info		<i>Monitoring Tech</i>	
Health info		Traffic	
Police records		Public transport	
		Satellites	

Source: Adapted from World Economic Forum (2011) and OECD (2013).

Table 2: Direct Data-Related Products for NAICS 518

<i>Product</i>	<i>Code</i>	<i>Description</i>
Data storage	36140	Managing or administering the storage and back-up of data, including data migration services
Data management	36150	Providing ongoing management and administration of data as an organizational resource, including modeling, mobilization, mapping, and mining of data
Other data processing	36170	Providing other IT hosting or infrastructure provisioning services, such as hosting applications, processing data, and computer time sharing
Transformation	36220	Information and document transformation services, including imaging and other data capture services and data conversion and migration services

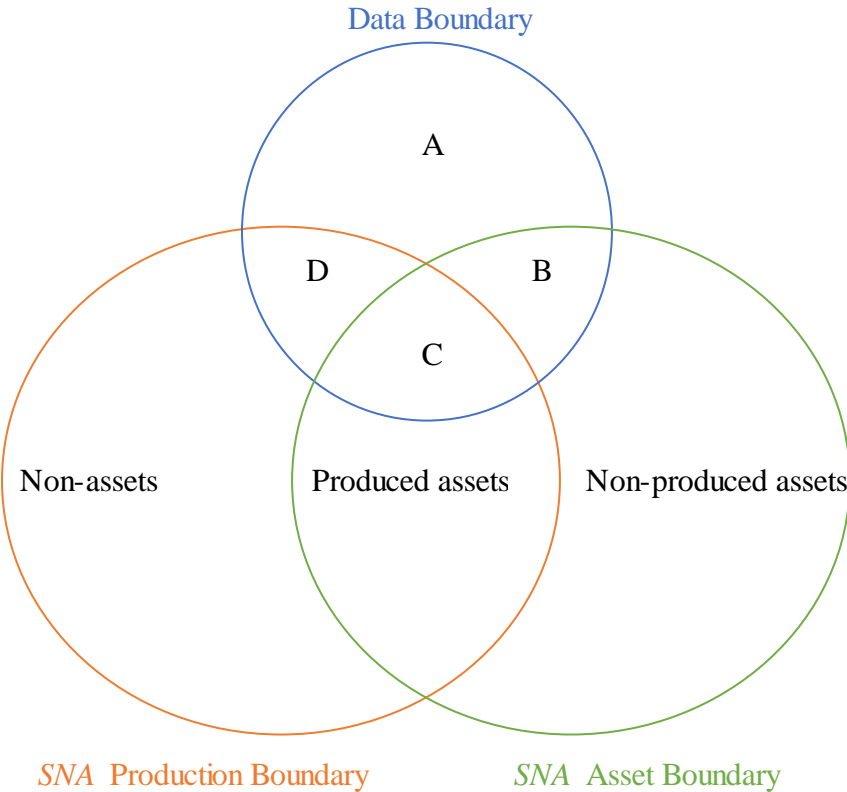
Source: 2012 Economic Census form for NAICS 518.

Table 3: Indirect Data-Related Products for NAICS 518

<i>Product</i>	<i>Code</i>	<i>Description</i>
Website hosting	36120	Hosting a website and related files in a location that provides a fast, reliable internet connection
Collocation	36130	Providing rack space within a secure facility for the placement of servers and enterprise platforms
Data storage	36140	Managing or administering the storage and back-up of data, including data migration services
Data management	36150	Providing ongoing management and administration of data as an organizational resource, including modeling, mobilization, mapping, and mining of data
Data processing	36160	Sending audio and video over the internet or providing services associated with the storage, production, and support of video and audio streaming over the internet
Other data processing	36170	Providing other IT hosting or infrastructure provisioning services, such as hosting applications, processing data, and computer time sharing
Internet access	36190	Broadband internet access services
	36200	Narrowband internet access services
Application service provisioning	34930	Providing software applications on a leased, fee, or subscription basis from a centralized, hosted, and managed computing environment
Business process management	34940	Providing a bundled service package that combined IT-intensive services with labor, machinery, and facilities to support and manage a business process
Network management	37511	Managing and monitoring communication networks and connected hardware to diagnose networking problems and gather capacity and usage statistics to administer network traffic
	37512	Providing data-to-day management and operation of a computer system

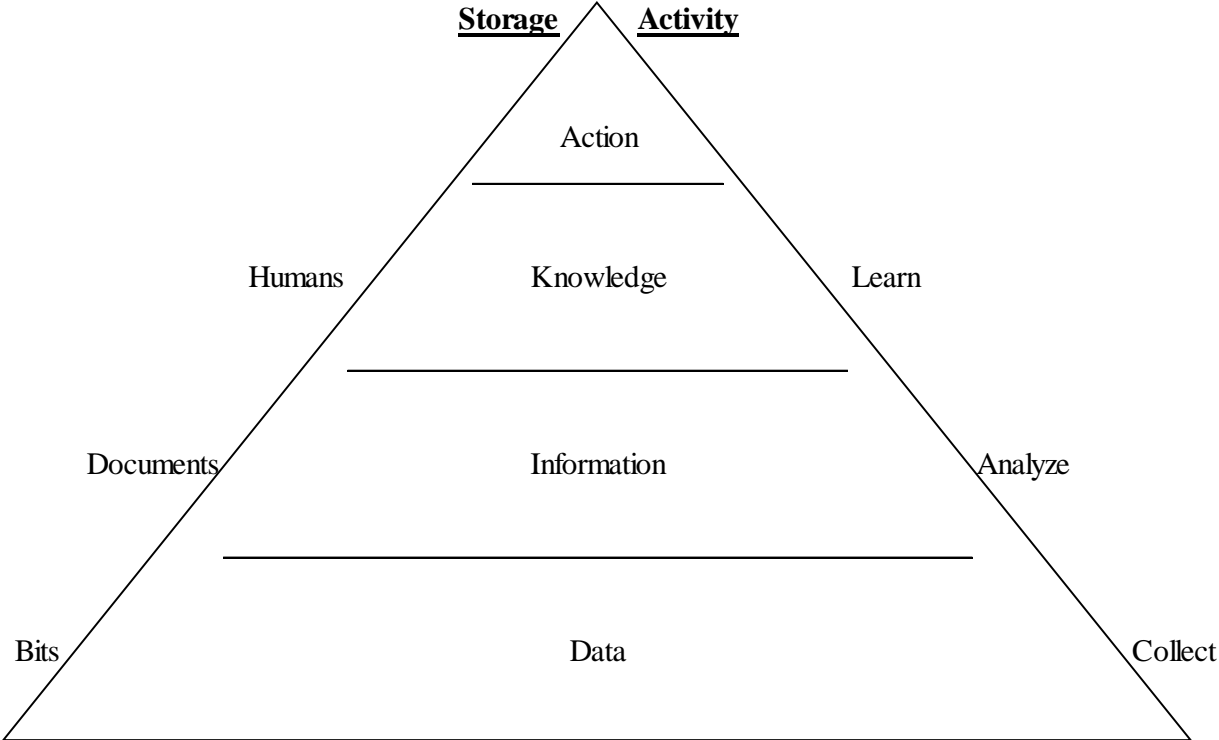
Source: 2012 Economic Census form for NAICS 518. These products are referred to as “data processing, IT infrastructure provisioning, and hosting services” (product code 6364) for NAICS 518 in the Service Annual Survey.

Figure 1: The Data Boundary



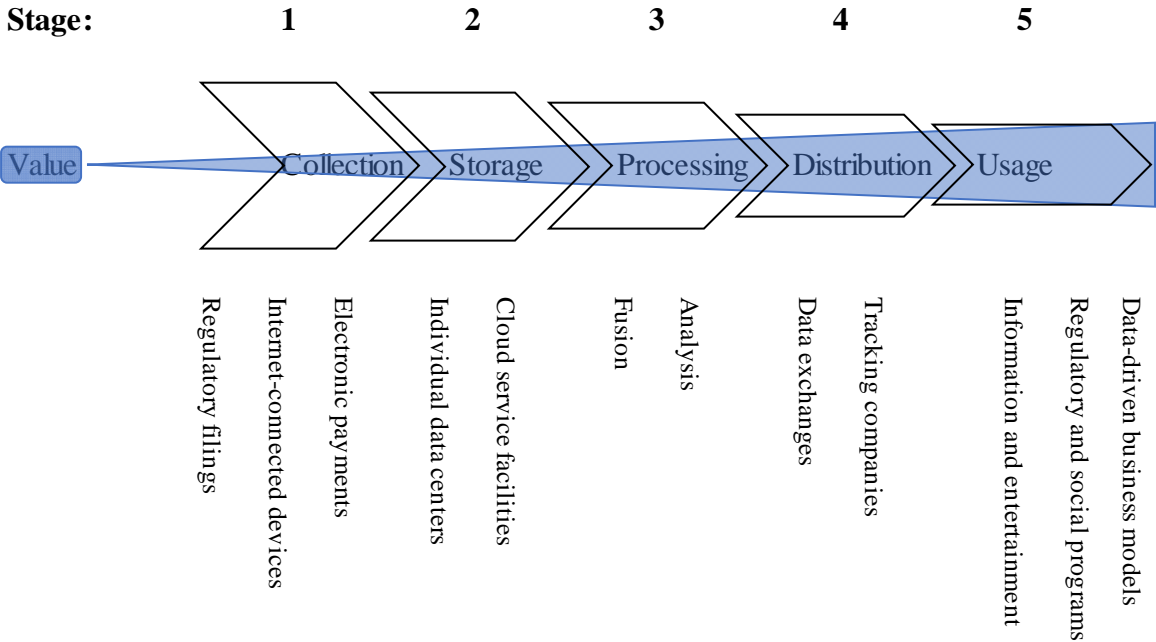
Source: Adapted from *SNA 2008*.

Figure 2: The Data Pyramid



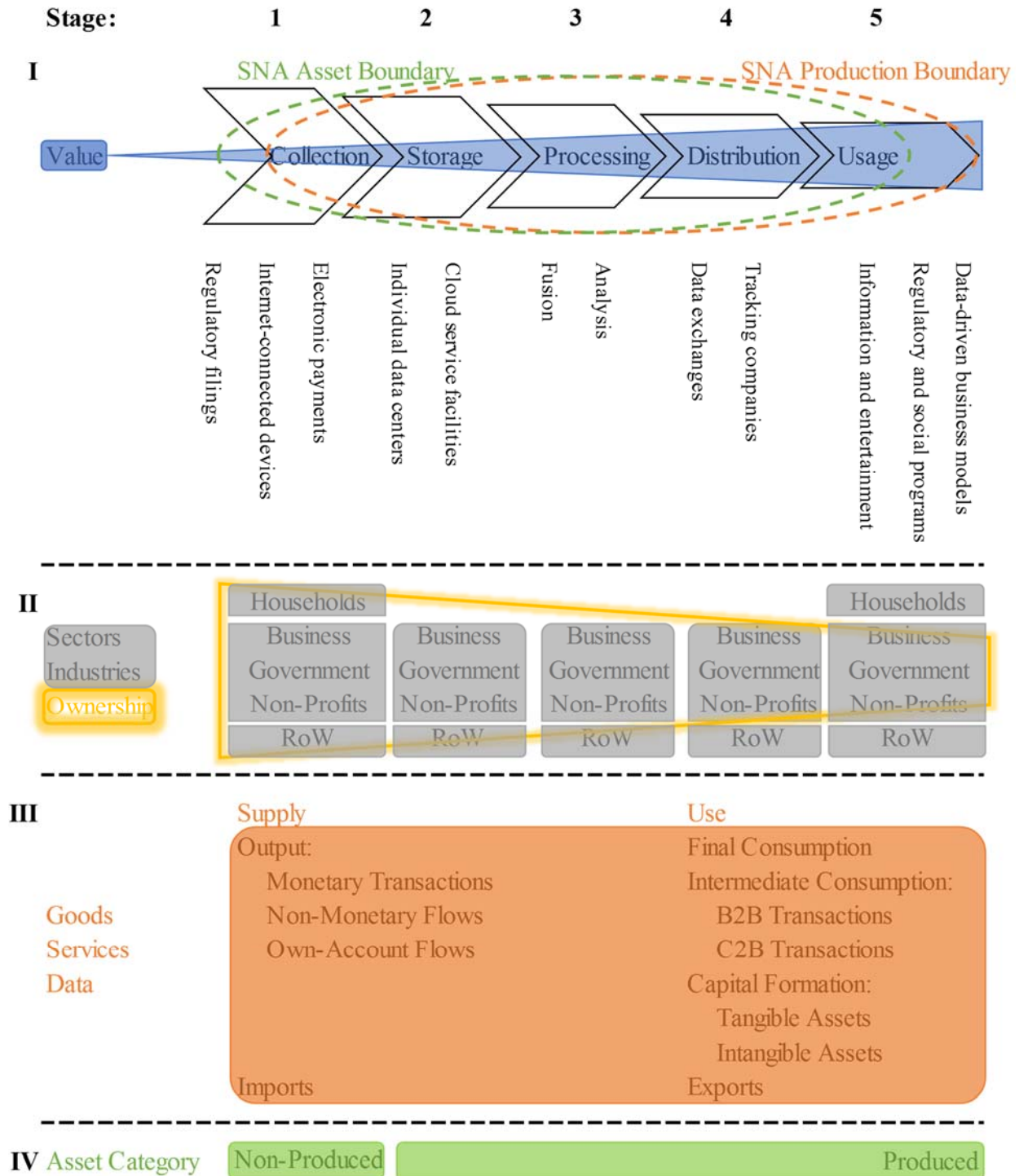
Source: Adapted from Varian (2018).

Figure 3: The Data Value Chain



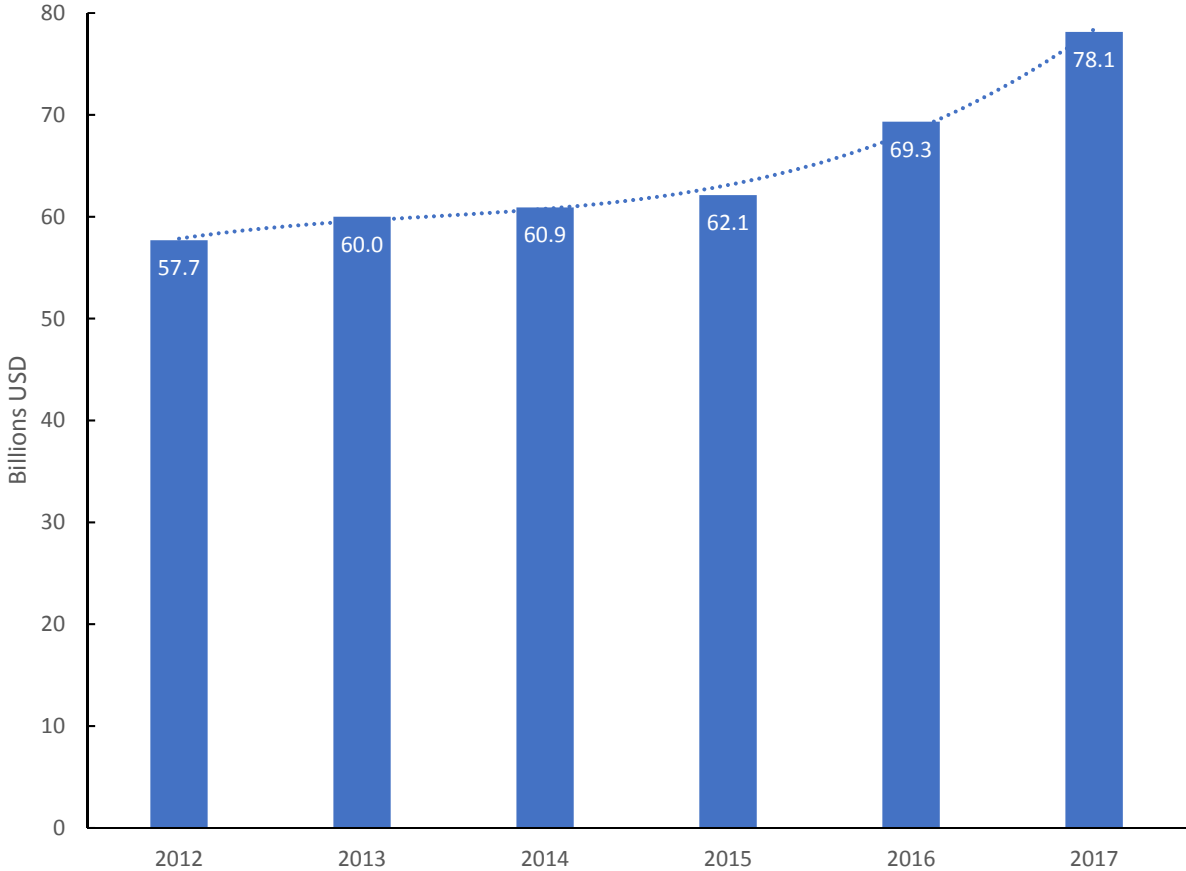
Source: Adapted from OECD (2013) and Moro Visconti et al. (2017).

Figure 4: Data in the System of National Accounts Framework



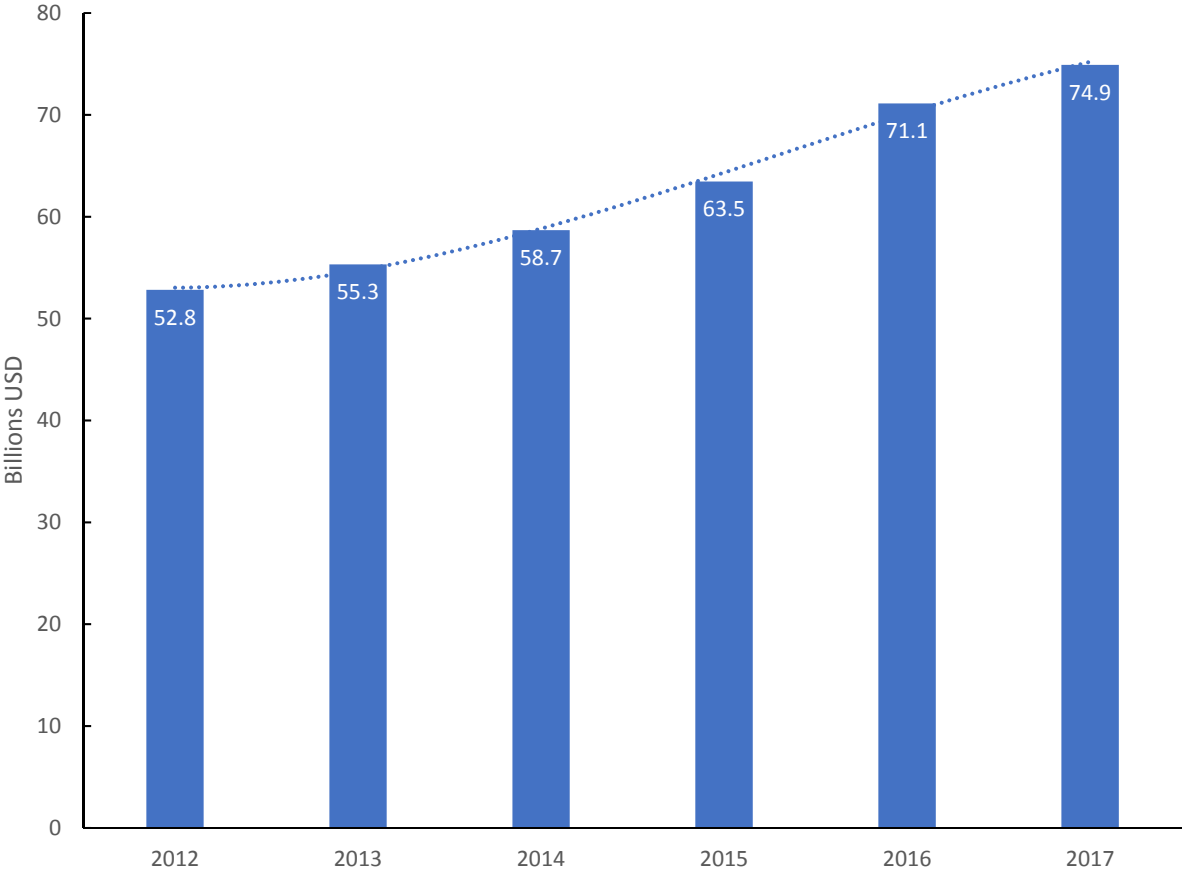
Source: Adapted from OECD (2013) and Moro Visconti et al. (2017).

Figure 5: Indirect Data-Related Product Receipts for NAICS 518



Source: Calculated from the 2012 Economic Census and the Service Annual Survey for 2013-2017. Products include website hosting services, business process management services, collocation services, data storage services, data management services, video and audio streaming services, other data processing services, IT infrastructure and network management services, and internet access services.

Figure 6: Data-Related Production Costs for Private Industries (except NAICS 518)



Source: Calculated from the Occupational Employment Statistics Survey for 2012-2017 and the Service Annual Survey for 2013-2017. Standard Occupational Classification categories include information security analysts, database administrators, actuaries, mathematicians, operations research analysts, statisticians, and mathematical technicians.