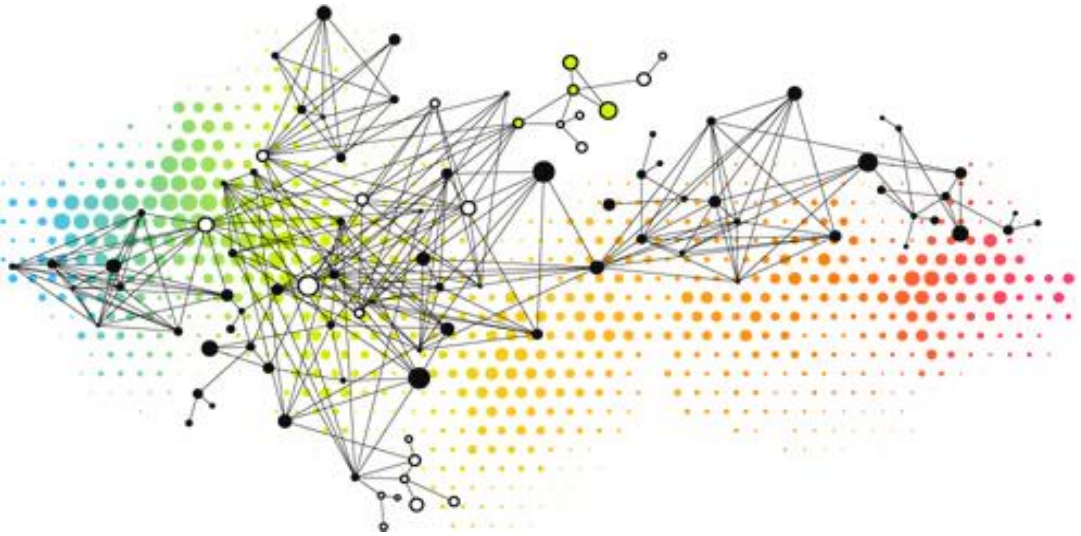


Measuring Digital Economy in Macroeconomic Statistics: The Role of Data

Jennifer Ribarsky, IMF Statistics Department

Group of Experts on National Accounts: Measuring Global Production
Geneva, Switzerland
April 10-12, 2019

Outline

- Motivation
- What's data, information, and knowledge?
- Data as a factor of production?
- Is digital data an asset? What's the perimeter? Input data or data-based information?
- What's in business accounting?
- How to value digital data-based information?
- What's recorded in national accounts?
- Summary and recommended further research

Motivation

Data flows through the modern economy

Cisco says annual global Internet Protocol (IP) traffic was 1.5 zettabytes* in 2017, projected to more than triple over the next 5 years.

- 75% of IP traffic was video in 2017.

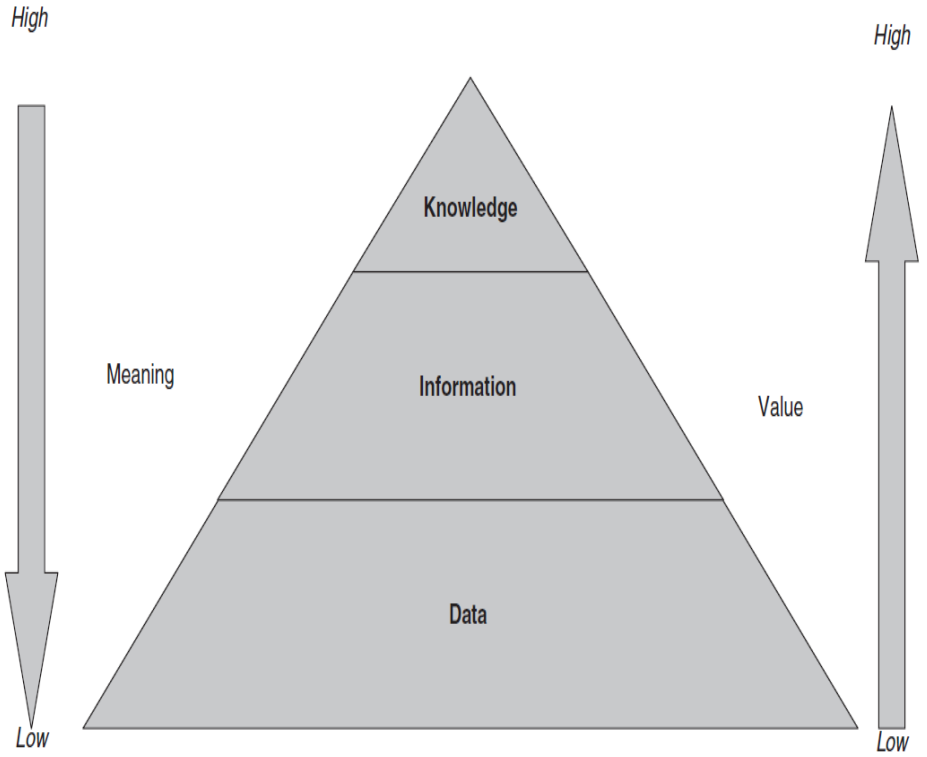
***Data flows can be a means of delivery of content*.**

***National accounts : Interested in how data is used in production*.**

* One zettabyte is 1 followed by 21 zeroes

What is data, information, and knowledge?

- Data are
  - Products of facts or observations.
  - Quantities, characters on which operations are performed by a computer (e.g. bits & bytes).
- Data is not knowledge.
- Information is inferred from data.
- Information is then used to create knowledge.



Source: Rowley (2017). Data, information, and knowledge according to Chaffey and Wood

Data transformation chain

- Organizing and processing data lends the data relevance for a *specific purpose* or *context*, and thereby makes it meaningful, valuable, and useful
- Potential value depends on where in the chain it lies.
- Input dataset has much less value than the value of the information and know-how once analysis is done.

Source: Mawer (2015), https://www.svds.com/valuing-data-is-hard/

How is input data obtained?

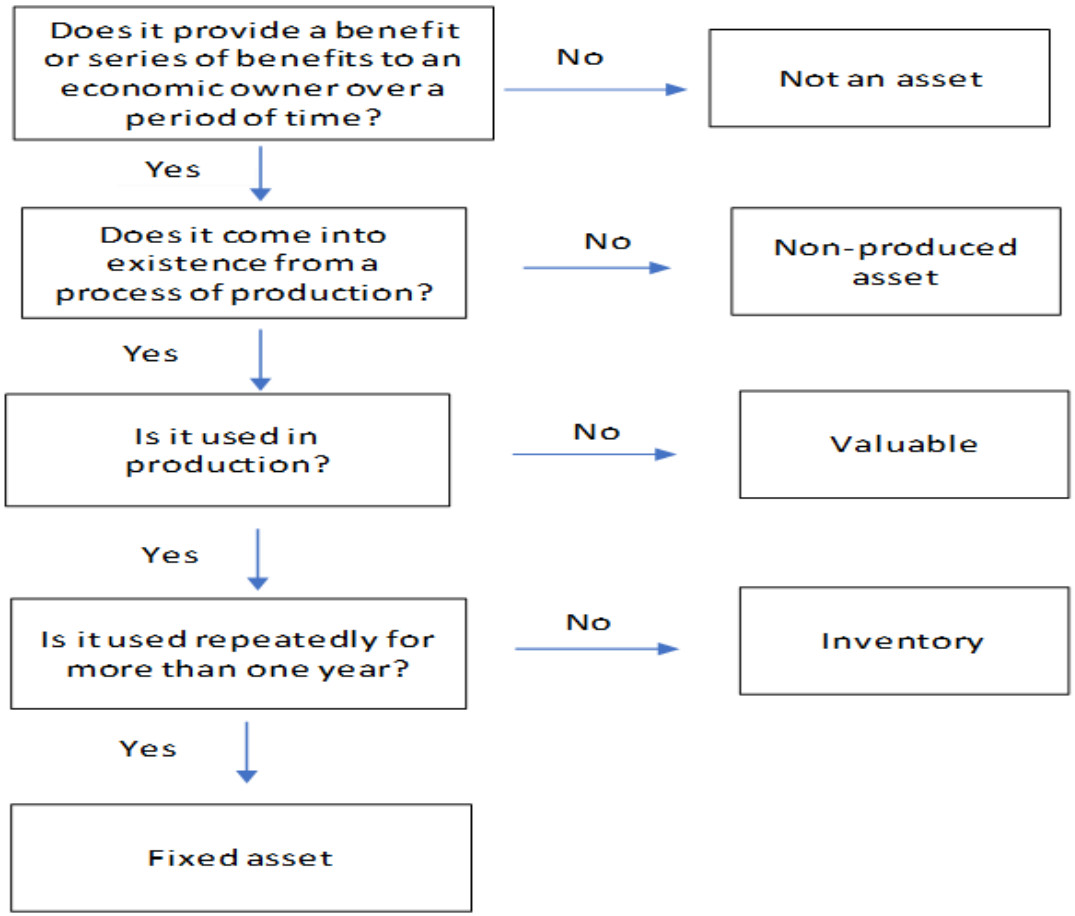
- **First-party data-** collected by the business itself about its users or customers (e.g., cookie-based data on browsing activity or data on past purchases)
- **Second-party data**- essentially someone else's first-party data. Businesses work out arrangements with trusted partners who are willing to share their customer data with them (and vice versa).
- **Third-party data**- any data collected by an entity that does not have a direct relationship with the user the data is being collected on.
- **Public data-** open or freely available without payment, e.g. data produced by the government.

Data as a factor of production

- Data has always had a central role in business decision making.
  - Strive to gather data on customers, to improve products
  - Processes to enhance productivity, improve performance
- Data in electronic form allows it to be analyzed for insights and decision-making at an unprecedented scope and scale. Data has been transformed into **Digital Data**.
- **Digital data has allowed for new information/knowledge creation that could not have been done if the data were not in digital form.**

Is digital data an asset?

Is digital data-based information an asset?



Data provides economic benefits.

Data can be used repeatedly in production.

Over a period time?
Harder to determine: some data (use) is short-lived.

Digital data-based information provides economic benefits & comes into existence from a production process.

What's in business accounting?

Axciom's criteria for capitalizing data acquisition costs

Axciom is a data broker, amortizes some data acquisition costs: useful life is 2 – 7 years

Type of data	Whether data becomes stale	To what extent data will be replaced and updated
Whether stale data has value as historical data	Restrictions on use of data	Terms of license to use data

Nielsen's acquisition of Gracenote

Acquisition included a bundle of intangibles

- Nielsen is a leader in market research and ratings.
  - "an extensive foundation of proprietary data assets"
- In 2017, Nielsen's total revenue was USD 6.6 billion, balance sheet only includes a small amount of data assets (USD 168 million) that were recorded when Nielsen acquired Gracenote for USD 585 million. **Content database: 29% of acquisition value.**

(IN MILLIONS)

Description	Amount	Useful Life
Customer-related intangibles	$ 109	10 - 15 years
Content database	168	12 - 16 years
Trade names and trademarks	7	5 years
Computer software	57	7-8 years
Total	$ 341	

How to value digital data-based information?
Potential valuation methods

How to value digital data-based information?

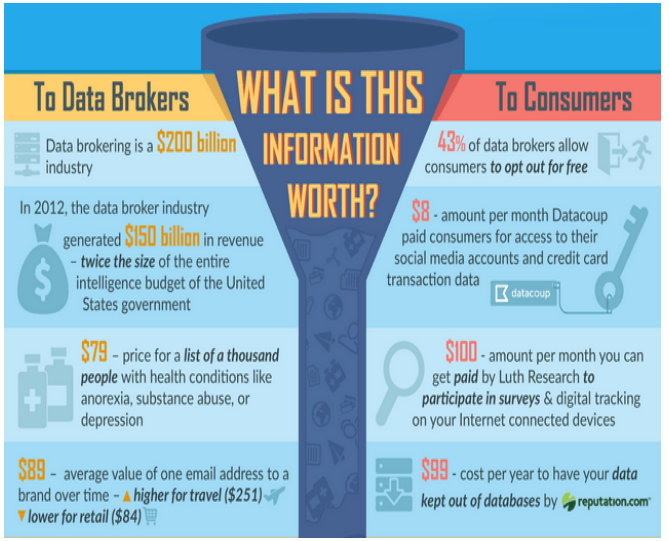
- **Market-based**: value is determined based on the market price of comparable products on the market.
- **Cost-based**: value is determined by how much it costs to produce the information/know-how derived from data.
- **Income-based**: value is determined by estimating the future cash flows that can be derived from the data.

Market-based approach

- In SNA, transactions should be valued at market prices or market-price-equivalents.
- On a *conceptual basis* the market-based approach is the preferred concept of the SNA.
- In most cases, a comparable product does not exist.
- 3rd party data can be bought and sold.
  - But user profile data has undergone processing (e.g. organizing, cleaning).
  - May require significant resources.

Data brokers sell consumer profiles in large batches.

A list of a thousand people with health conditions like anorexia, substance abuse, or depression was USD 0.079 per user profile.



Source:https://www.webfx.com/blog/general/what-are-data-brokers-and-what-is-your-data-worth-infographic/

Cost-based approach

- If no appropriate market price then SNA gives preference to valuation by costs.
- Own-account gross fixed capital formation (GFCF) in software and databases and research and development (R&D) are measured using the “sum-of-costs” approach.
  - For market producers, includes a mark-up that reflects the operating surplus or mixed income attributable to producer.

Direct costs for creating digital data-based information

1. Costs of acquiring data (survey, locate, capture, provide free services or discounts)
2. Costs of designing database management system (or purchasing database management services)
3. Costs on inputting and preparing data in appropriate format for storage (including organizing and metadata)
4. Costs of storing/warehousing data (including in the cloud)
5. Costs of tools used to analyze data (e.g. software, algorithms)
6. Costs of analyzing data (including data validation, cleaning, and contextualizing)

But what about data acquired in exchange for free services?

Google's Traffic Acquisition Costs

Millions of U.S. Dollars

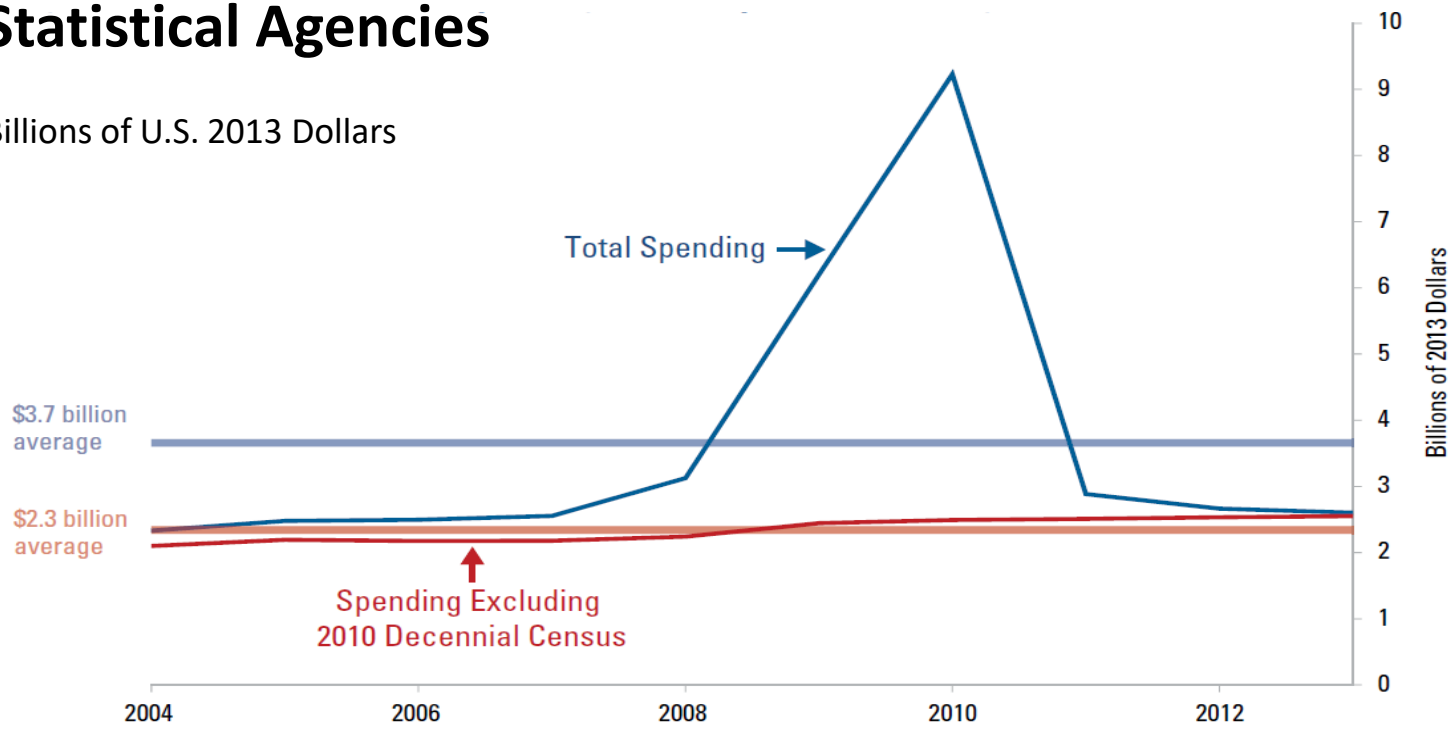
Traffic acquisition costs (TAC) to Google Network Members and distribution partners

	Three Months Ended December 31, 2017	Three Months Ended December 31, 2018
TAC to Google Network Members	$3,674	$3,930
TAC to Google Network Members as % of Google Network Members' properties revenues	74%	70%
TAC to distribution partners	$2,776	$3,506
TAC to distribution partners as % of Google properties revenues	12%	13%
Total TAC	$6,450	$7,436
Total TAC as % of Google advertising revenues	24%	23%

But what about government acquired data?

U.S. Federal Government Spending on the Principal Statistical Agencies

Billions of U.S. 2013 Dollars



Sources: Budget information compiled from *Analytical Perspectives, President's Budget*; *Statistical Programs of the U.S. Government Supplement to President's Budget*; actual agency budgets; *Principles and Practices for a Federal Statistical Agency*

Note: Budget amounts converted to real 2013 dollars using Government Consumption Expenditures deflator.

Income-based approach

- While the income-based valuation approach is an acceptable method, the 2008 SNA advises caution in its use.
- SNA gives preference to market-based and cost-based approaches… *because it may be difficult to determine the future earnings with the appropriate degree about the* ***asset's life length*** *and the* ***discount factor*** *applied.*
- One of the primary difficulties with this approach is **distinguishing the cash flows** (net of associated costs) **uniquely related to the asset from the cash flows related to the whole company**.
- Income-based approach is recommended in valuing musical, literary, and photographic works– industries where there is an established system of royalty flows.

Value of data-based information for U.S. Internet Publishing and Broadcasting and Web Search Portals industry, 2017

Billions of U.S. Dollars

	Discount factor: 8% Service life assumption		Discount factor: 5% Service life assumption		
Net stock, 2017	**7 years**	**3 years**	**7 years**	**3 years**	**NIPA stock**
Data-based information, portion of AD space	148.3	85.5	162.2	89.0	…
Data-based information, all AD space	285.2	164.4	311.8	171.1	…
Software, NIPA current-cost net stock of private fixed assets	…	…			644.4
Prepackaged software, NIPA current-cost net stock of private fixed assets	…	…			176.4
Custom software, NIPA current-cost net stock of private fixed assets	…	…			321.6
Own-account software, NIPA current-cost net stock of private fixed assets	…	…			146.3

What's recorded in National Accounts?

2008 SNA intellectual property products linked to data (1)

- 2008 SNA currently recognizes several types of intellectual property products linked to data: **software and databases, research and development (R&D), and goodwill and marketing assets**.
- **Databases**
  - **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 based on 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." (SNA para 10.113)

2008 SNA intellectual property products linked to data (2)

- **R&D**
  - **Research and [experimental] development consists of the value of expenditures on creative work undertaken on a systematic basis in order to increase the stock of knowledge**, including knowledge of man, culture and society, and **use of this stock of knowledge to devise new applications**. This does not extend to including human capital as assets within the SNA. …Unless the market value of the R&D is observed directly, it may, by convention, be valued at the sum of costs, including the cost of unsuccessful R&D. (SNA para 10.103)

2008 SNA intellectual property products linked to data (3)

- **Goodwill**
  - **The value of goodwill and marketing assets is defined as the difference between the value paid for an enterprise as a going concern and the sum of its assets less the sum of its liabilities, each item of which has been separately identified and valued.** Although goodwill is likely to be present in most corporations, for reasons of reliability of measurement it is only recorded in the SNA when its value is evidenced by a market transaction, usually the sale of the whole corporation. Exceptionally, identified marketing assets may be sold individually and separately from the whole corporation in which case their sale should also be recorded under this item. (SNA para 10.199)

Intangible assets of data-driven businesses & potential overlap

- Data combined with software and databases (storage for the data) and R&D are all likely intangible assets of data-driven businesses.
  - Facebook's financial reports show significant R&D expenditures (USD 10.3 billion in 2018, primarily of compensation for software engineers and other technical employees)
- Valuing data-based information has considerable overlap with other intangible assets already capitalized within the 2008 SNA
  - May be hard in practice to distinguish between data and other types of intangible assets.
  - Own-account software, databases, and R&D production/GFCF are all estimated using the sum-of-costs approach.

What costs are missing to fully account for data-based information?

- Conceptually excluded:
  - Data acquisition costs
  - Costs incurred in analyzing the data
  - Use of open source software that is not produced in-house
- Potentially missing costs (conceptually included, but may be missed in practice):
  - Payments for cloud storage services.
- Unclear overlap with R&D:
  - Some of Facebook's R&D expenditure may be related to advancements in mining videos, pictures, and text for information which could also be considered as expenditure for creating a data-based information asset

Summary (1)

- Digital data is a key factor of production in the modern data-driven economy, but the value of data inputs and the information derived from the data is hard to determine.
- What's the perimeter? **Input data or data-based information**?
- Is it a (fixed) asset? **Input data and data-based information provides economic benefits and it can be used repeatedly in production**.
  - Further research needs to be done to determine long-lived (more than one year) versus short-lived data-based information.

Summary (2)

- Three approaches to valuing data-based information.
  - Market-based approach appears least feasible.
  - Income-based approach could be a viable method for data-based information that can easily be tied to a particular use, such as targeted advertising.
    - May be hard to use for all types of data-based information.
  - Cost-based approach appears to be a feasible approach for national statistical offices to implement.
    - A fair amount of data-based information already capitalized within the SNA (e.g., own-account software, databases, and R&D).
    - Using similar approach may help identify overlaps with what is already capitalized.

Areas for further research (1)

- Further research into business accounts of firms that provide free services in exchange for data to see if certain costs can be identified as being related to data acquisition.
  - Impact of including the acquisition costs of data in the calculation of databases (including government)
- Explore the possibility of surveying businesses for the costs associated with developing digital data-based information.
- Further research into potential overlap between R&D and digital data-based information.

Areas for further research (2)

- Explore the possibility of using an income-based approach to valuing data-based information, especially for advertising funded digital platforms.
- Refine the classification systems to better identify data-related activities, products, and occupations.
- Further explore link with digital trade: Cross-border data flows; Further research into recording in Balance of Payments and International Investment Position statistics.