

## **On the Measurement of “Data” in Economic Analysis and National Statistics**

Michael Wolfson (University of Ottawa/ l'Université d'Ottawa)

There is growing interest in extending the SNA to include measures of the stocks and flows of data, as “data”, especially in digital form, has become a major factor in modern economies. From an SNA perspective, one key challenge is how to value data in monetary terms.

Digital data (actually a plural noun) conceptually consist of both data bases, stored on computers, and data flows, generated from myriad sources including but by no means limited to commercial transactions, continually updated satellite images, and blogs and videos posted or uploaded to various social media. There is also a spectrum from raw to curated data, and then to uses including analytical insights. Thus, “data” in modern economies are highly heterogeneous, and are changing rapidly due to continuing dramatic technological improvements.

The economic and social utility of data depend on a diversity of factors, including content, standardization, coverage, quality, modes / extent / forms of access, capacity to link to other data, timeliness, interoperability, available analytical methods, and the organizational infrastructures of statisticians, programmers and data scientists who transform data into useful information.

For SNA purposes, much of the discussion on how best to value “data” is premised on conventional official statistical data sources, such as labour force, firm, and trade surveys, in turn to observe costs of producing data and market values of transactions involving data. But prior to discussing how best to value data in monetary terms, it is critical to articulate (1) the ways that primary data on data bases and data flows (DBDFs) could be collected, and (2) the most important uses for national (and international) collections of statistics on “data” = DBDFs.

The main objective of this paper is to explore possible new kinds of statistical activities, including “thinking outside the box” of the SNA, for the development of national statistics on the roles and importance of DBDFs in modern economies.

We start by sketching a taxonomy of DBDFs to frame how best a national statistical office (NSO) might collect relevant information. For example, there are wide differences in DBDFs across social media (e.g. Facebook, Amazon) and private (e.g. banking, retail sales) versus public sector (e.g. tax returns, street maps) administrative data. Since the full gamut of DBFS is so diverse, we focus on two areas: health data (n.b. mostly non-market), and digital music – even though they are themselves both very heterogeneous.

A starting premise is that any NSO data on DBDFs should be microanalytic – as originally emphasized for the SNA by Richard and Nancy Ruggles in the 1960s. Next, we explore how such data could be used, not only for economic analyses but also for social statistics.

For health data, there are continuing efforts to deploy electronic medical records (EMRs) to improve the myriad existing but siloed computerized health records (e.g. hospital visits, drug prescriptions, covid infections). These administrative DBs are continually being updated with DFs, including new health care encounters. Uses range from patient care to health human resource planning, assessments of cost effectiveness of new (and existing) health technologies, pandemic preparedness, real-time projections of health resource bottlenecks, and health care quality evaluations. New initiatives in Canada to improve the quality, interoperability, and authorized accessibility of health data have the potential of major benefits in the efficiency of health care and for population health.

For digitally encoded music, recent decades have seen dramatic changes – from the near disappearance of vinyl, to music videos on YouTube, to streaming services like Spotify and emerging Twitch. Ticketmaster and Live Nation have essentially become monopolies. As a result, cultural policies such as copyright and related competition policies are seriously lagging these dramatic transformations.

The SNA and its main summary index, GDP, are very widely used for assessing countries' longer run economic growth, and in the shorter-term fiscal and monetary stabilization policies. Partially disaggregated SNA data are also considered important for analysis of productivity trends. Even though the welfare economics of the late 1940s and early 1950s showed that GDP cannot measure social welfare, GDP and GDP per capita continue in wide use for summary comparisons of countries' living standards.

For economic growth applications, the challenge with health DBDFs is that the “output” of the health sector is not properly measured by the SNA. One option would be to develop data on health sector DBDFs related to their impacts on the cost-effectiveness of health care and public health interventions, as well to measure trends in health sector productivity in natural units of health status. From a social welfare perspective, new kinds of data on health DBDFs could indicate the quality of measures of the levels, distributions, and determinants of population health.

The monetary “value added” of the music industry is already included in the SNA. However, conceptualizing let alone measuring its “productivity” in this era of dramatic technological and organizational change is more problematic. The evolution of digital music is a quintessential example of the “new goods” problem. The “price” of listening to a dozen songs has declined dramatically from the \$15 range to purchase an album or CD, to close to zero on

YouTube or Spotify. As a result, price indices and deflators can never be properly constructed (Wolfson, CJE, 1999).

From a social welfare perspective, perhaps the best indicators of the benefits derived from the huge increases in digital music availability could come from time-use surveys – though information on the secondary uses of time are needed as music is often consumed while engaged in other activities.

In sum, then, this paper will explore a range of new kinds of statistical data collections on DBDFs in the health and digital music domains that have the potential of major benefits for assessing their economic and social welfare impacts, aside from possible augmentations in the SNA.