https://doi.org/10.56093/JISAS.V80I1.3
Available online at https://isas.live/newsite
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 80(1) 2026 21–37

Capitalization of Data in The Indian National Accounts

Gulab Singh[1] and A.C. Kulshreshtha[2]

[1]United Nations Statistics Division, New York, USA
[2]Ministry of Statistics and Programme Implementation, Govt. of India, New Delhi

Received 06 March 2026; Revised 21 March 2026; Accepted 24 March 2026

SUMMARY

The 2025 System of National Accounts (SNA) introduces a major conceptual advance by recommending the capitalization of organized, reusable data and databases as produced fixed assets within Intellectual Property Products. This paper examines the conceptual foundations, measurement framework, and practical implementation of this recommendation in the context of India's national accounts. It traces the evolution of the SNA asset boundary, clarifies the data–information chain, and establishes consistent terminology for data-related concepts.

The paper develops an operational methodology based on the sum-of-costs approach for estimating own-account data assets, supported by an explicit accounting identity linking labour inputs, involvement rates, and adjustment factors. Drawing on India's existing statistical infrastructure—Periodic Labour Force Survey (PLFS), Occupational Wage Survey (OWS), Annual Survey of Industries (ASI), Annual Survey of Unincorporated Sector Enterprises (ASUSE), Goods and Services Tax (GST) data, and Balance of Payments (BoP) statistics—the study demonstrates how occupation-specific labour costs and non-labour inputs can be combined to estimate gross fixed capital formation (GFCF) in data assets.

A numerical illustration using hypothetical data is provided to demonstrate the full estimation sequence from nominal investment to real GFCF and net capital stock using the perpetual inventory method (PIM). In addition, a sensitivity analysis evaluates the robustness of estimates under alternative assumptions regarding involvement rates, adjustment factors, and service lives.

Evidence from international studies and ADB pilot estimates suggests that data assets account for approximately 0.8 to 2.7 per cent of GDP, with India exhibiting rapid growth in data-related investment. While the short-term impact on GDP growth is modest, capitalization significantly enhances the measurement of capital stock and productivity. The paper concludes with practical recommendations for the National Statistical Office (NSO) to implement data capitalization in India's national accounts, emphasizing methodological transparency, use of existing data systems, and the need for continued empirical refinement.

*Keywords: C*apitalization of data, Data as asset; Digital economy; Gross Fixed Capital Formation; GDP; National Accounts Statistics; Measurement framework; Supply-Use Tables; Sum-of-costs valuation; Depreciation.

JEL classification codes: E01, E22, C82, L86, O33

0. **Dr. Prem Narain: A Tribute to an All-time Great Genetical Statistician**

I, A.C. Kulshreshtha, am grateful to the Indian Society of Agricultural Statistics (ISAS) for inviting me to contribute a paper for the special issue of the Journal of ISAS being brought out in honour of Late Professor Prem Narain. I happened to be at I.A.R.S. over sixty years ago during 1964-69being in the first batch of M. Sc (Agric Statist) and later a research student. Prem Narain did not teach us but being the Junior Faculty at that time, was responsible to organise our major field related courses. He used to assist senior Faculty (Drs G.R. Seth, D. Singh, V. N. Amble and B.V. Sukhatme) at that time. In 1969, I leftI.A.R.S. to join as Faculty at Meerut University and in the meanwhile, Prem Narain left for Edinburgh to take up research work in Genetical Statistics and returned as Dr Prem Narain, a totally revamped personality as visible expert in Genetical Statistics and senior Faculty at I.A.R.S. In 1972, I joined I.A.R.S. as a Junior Faculty member and worked under Dr. Prem Narain on a research project on Cost of Milk Study. Subsequently I joined the ISS and watched Dr Prem Narain growing to the highest level as Director of IASRI from 1981 to 1992. He was also the Secretary of the ISAS from 1982 to 1994. He undoubtedly made a significant contribution in the field of Agricultural Statistics. During 2001-04 I happened to be Additional Director General, CSO, MoSPI, inter-alia in charge of the then created National Awards. Therein Dr. Prem Narain was a contestant for the Prof. P V Sukhatme National Award in Statistics. Owing to his extra-ordinary unparallel contributions to Agricultural Statistics, Prof Prem Narain became the first recipient of Prof P V Sukhatme National Award in Statistics, and it was my pious privilege to facilitate that.

Corresponding author: A.C. Kulshreshtha
E-mail address: ackulshreshtha@yahoo.com

I, Gulab Singh, feel deeply honoured and privileged to contribute a paper to this special commemorative volume of the Journal of the Indian Society of Agricultural Statistics brought out in memory of the eminent statistician Prof. Prem Narain. My association with him dates back to 1976, when I joined the Institute of Agricultural Research Statistics (IARS) to pursue the Professional Statistician Certificate Course. At that time, Prof. Prem Narain served as the warden of the institute's hostels, and it was during this period that I had my first opportunity to observe closely his distinctive qualities as an administrator and mentor. Later, during my tenure as a Ph.D. student, I was entrusted with the responsibility of Prefect of the hostels in 1978, succeeding Dr. K.K. Tyagi. This was also a landmark year in the institute's history, when the IARS was rechristened as the ASRI, reflecting its expanding mandate in advancing research, training, and application of statistical methods in agricultural sciences. Serving as Prefect provided me with valuable opportunities to interact closely with Prof. Prem Narain on matters concerning the welfare and well-being of hostel residents, and through these interactions I came to appreciate his deep concern for the student community. He took a keen interest in sports, cultural activities, and the overall development of students, encouraging participation and fostering a vibrant academic environment within the institute. Although I did not have the privilege of attending any formal course taught by him, I retain very fond memories of my interactions with him. Within the institute he carried the aura of a firm and disciplined administrator, reflecting his commitment to maintaining high standards of institutional governance. Yet those who interacted with him closely knew that behind this dignified exterior was a kind-hearted and compassionate individual, always attentive to the welfare and aspirations of students. Looking back, I feel fortunate to have witnessed his leadership and humanity during my years at the institute, and his legacy continues to inspire generations of statisticians and researchers.

1. INTRODUCTION

The 2008 System of National Accounts (SNA), developed under the aegis of the Inter Secretariat Working Group on National Accounts (ISWGNA) comprising the United Nations, the IMF, the World Bank, the OECD and the Eurostat and published by the ISWGNA in 2009, is the current global Standard that National Statistical Offices use to measure economies in a consistent and comparable way. It produces the familiar headline figures such as Gross Domestic Product (GDP) and many supporting tables on investment, consumption, productivity, and growth. Because economies keep evolving—with new technologies, business models, and sources of values—the SNA is revised periodically roughly every fifteen years. The latest version 2025 SNA includes a significant recommendation: to treat ***data*** as a capital asset rather than simply as a day-to-day expense.

Capitalizing an item in the national accounts means changing how spending on it is recorded. Instead of treating the full cost as an immediate input that is used up in the process of production in the current period (called intermediate consumption), it is recognized that the expenditure creates something lasting that will support production for multiple future years. In everyday terms, this spending becomes "Capital Formation"—the building of productive capacity, just like buying a machine or constructing a factory.

Assets in the national accounts are grouped into financial assets (such as deposits or shares) and non-financial assets. Non-financial assets are further divided into produced assets (created through economic activity) and non-produced assets (such as natural resources, contracts, leases, licenses, and purchased goodwill). Produced assets include inventories, valuables held mainly for their store of value, and fixed assets. Fixed assets are long-lasting items used repeatedly in production — buildings, machinery, vehicles, and the group known as Intellectual Property-Products (IPPs), which already covers research and development, software and databases, and artistic originals. The 2025 SNA adds a new, clearly defined category inside IPPs: "data and databases."

For national-accounting purposes, data is defined as information content produced by accessing and observing real-world phenomena, then recording, organizing, and storing those observations in digital form so the resulting collection can be used repeatedly in productive activities and to generate ongoing economic benefit. Not every digital file qualifies—only organized, reusable datasets that act like tools or equipment for future production. This concept is easiest to grasp through the data-information chain, which is clearly illustrated above in Figure 1

Data-information chain from a SNA perspective

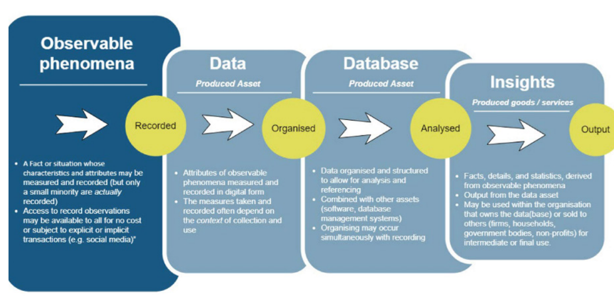

Fig. 1. Data-information chain from a SNA perspective

Source: Mitchell, Ker, and Lesher (2021)

This figure illustrates the transformation of observable phenomena into data, databases, insights, and final outputs, highlighting the stages at which data may qualify as a produced asset. The diagram in Fig. 1, presents a straightforward left-to-right flow with four main stages connected by arrows:

Observable

- ***Observable phenomena*** (starting point, shown in a blue box) Facts or situations in the real world whose characteristics and attributes may be measured and recorded. Only a small minority are actually captured. Access to record them may be free or may involve capital or implicit transactions (for example, social media activity).
- ***Data*** (light blue box, labeled "Produced Asset") A collection of observable phenomena that have been recorded and stored in digital form, including the necessary labels and context for their later use.
- ***Database*** (light bluebox, labeled "Produced Asset") The recorded data, now organized and structured in digital form, often combined with other assets such as software or database management systems. Organizing and structuring may happen at the same time.
- ***Insights*** (light bluebox, labeled "Produced goods / services") Facts, details, and derived information obtained by analyzing the database. These insights can be used inside the organization or sold to others (businesses, governments, non-profits, or for-profit entities).
- ***Output*** (final yellow box) The end result—goods or services produced using the insights.

This visual chain shows exactly how raw observations move through recording, organization, analysis, and eventual use. As illustrated in Fig. 1, the transformation from raw observations to structured databases forms the basis for identifying data as a capital asset. Depending on who uses the data, how often, and for what purpose, it can play three different economic roles: (1) a final consumable good or service, (2) an intermediate input that is fully used up in the process of production in one production cycle, or (3) a capital asset that keeps contributing to production again and again over multiple years. The distinction is not always obvious and depends on context, but the new guidance aims to make the classification clear and consistent across all countries.

By capitalizing data—especially the large volumes created inside companies and governments for e-commerce, fintech, social networks, scientific research, and public services—the national accounts will finally reflect an increasingly vital part of modern economies. Recording these expenditures as investment rather than current costs will give a more accurate picture of how economies grow, where capital is accumulating, and what resources are truly building future prosperity. The remainder of the paper explains in practical detail how countries can measure, value, and record data assets using the sum-of-costs approach and other recommended methods.

For conceptual clarity, the following distinctions are maintained throughout the paper:

- ***Data assets***: Organized, reusable digital datasets recognized as produced fixed assets under the SNA.
- ***Data capital***: The accumulated stock of data assets after accounting for depreciation.
- ***Databases***: Structured systems used to organize and manage data; may include both data and software components.
- ***Data-related investment (or data science investment)***: Broader expenditures related to data creation, processing, and analytics, not all of which qualify as capital formation under the SNA.

This distinction ensures consistency between conceptual definitions and their treatment within the national accounting framework.

The paper is organized as follows. Following this introduction, Section 2 traces the historical evolution of the SNA asset boundary and the progressive recognition of intangible assets, culminating in the landmark decision of the 2025 SNA to capitalize organized, reusable digital data and databases as produced fixed assets under IPP. Section 3 sets out the conceptual framework and measurement methodology for India, contrasting the OECD Digital Supply and Use Tables approach with the preferred sum-of-costs valuation method for own-account production; it details the estimation of labour costs (via occupation-specific wages and involvement rates), non-labour inputs, markup factors, and the full accounting sequence from nominal GFCF to net capital stock.

The Section 3 briefly describes the evolution of data as asset in the System of National Accounts (SNA), while Section 4 discusses framework and methodology for capitalizing data in the Indian National Accounts. Section 5 talks about Gross Fixed Capital Formation. While Section 6 presents various details on Labour Cost, the Section 7 discusses Total production cost adjustment factor**.** Net Stock and Backcasting are discussed in Sections 8 and 9 respectively. Section 10 identifies and evaluates the key Indian statistical sources—PLFS, ASI, ASUSE, GST returns, and BoP statistics—that enable practical implementation of these methods. The paper concludes with empirical findings from prior pilot exercises, a discussion of implementation challenges, and concrete recommendations for the NSO, MoSPI to integrate data assets into the India's national accounts in the forthcoming revision exercises for implementing 2025 SNA recommendations.

2. EVOLUTION OF DATA AS ASSET IN SYSTEM OF NATIONAL ACCOUNTS (SNA)

The SNA is an internationally agreed-upon framework that provides a standardized set of recommendations for compiling comprehensive measures of economic activity within a country, ensuring coherence, consistency, and integration across macroeconomic accounts. It encompasses concepts, definitions, classifications, and accounting rules to track how production is generated and distributed among consumers, businesses, governments, and foreign entities, while illustrating the flow of income and its allocation toward consumption, saving, and investment. Originally developed in 1953, the SNA has evolved through multiple revisions to support economic analysis, policy formulation, and the integration of various statistical systems, making it adaptable for nations at different development stages. The current version, the 2025 SNA, developed under the aegis of the ISWGNA, as the earlier (1993SNA, and 2008 SNA) ones is the latest update, building on the 2008SNA to incorporate contemporary economic realities and enhance global comparability.

The SNA has incrementally broadened its asset boundary to encompass intangible assets, mirroring the transition from industrial to knowledge-driven economies. The 1993 SNA pioneered the inclusion of R&D and computer software as produced assets, capitalizing outlays that deliver protracted benefits. The 2008 SNA advanced this by treating databases as fixed assets while omitting raw data, deeming it non-produced or intermediate consumption (United Nations *et al.,* 2008).

The 2025 SNA represents a landmark revision, expressly incorporating the information content of electronic data as a produced fixed asset under AN11532: Data and Databases, when employed in production exceeding one year (United Nations *et al.,* 2025). *Only the digital data is included, as non-digital forms fall outside the production boundary.*

The 2025 SNA defines data as "information content that is produced by accessing and observing phenomena; and recording, organizing, and storing information elements from these phenomena in a digital format, which provides an economic benefit when used in productive activities." The valuation of data assets is recommended using a sum of cost approach which would include: (i) costs of planning, preparing, and developing a data production strategy, (ii) costs associated with accessing, recording, and storing information, including already produced data, (iii) costs associated with designing, organizing, testing, and analyzing the data in order to draw information and conclusions from it, (iv) costs associated with processing, cleaning, and organizing the data to allow for use in productive activities, and (v) consumption of the fixed capital used (obsolescence), including return to capital.

Although it is likely that some own account data may be fully consumed within one year, due to practical limitations on delineating this data from data used repeatedly for more than a year, it is recommended to capitalize all own account production of data (United Nations *et al.* 2025).

Before recommending that expenditure on data be treated as investment in the 2025 SNA, the ISWGNA carefully examined whether countries would actually be able to implement such a change as it follows a transparent and consultative global process to ensure that proposed changes are conceptually sound, practically feasible, and internationally comparable. In the case of capitalizing data, the proposal was discussed in detail by expert groups under the global update programme for the 2025 SNA. Guidance notes were prepared by specialized task teams and reviewed by the Advisory Expert Group on National Accounts (AEG). Countries were invited to comment on these notes, and many statistical offices conducted experimental estimates to assess the availability of data sources, classification issues, valuation methods, and the likely impact on GDP and other aggregates. These practical exercises helped identify implementation challenges, especially for developing countries with limited data infrastructure. Global consultations and regional workshops were organized so that national statistical offices could share experiences and concerns. The aim was not simply to agree on a new concept, but to ensure that countries could compile the new estimates in a consistent and cost-effective way.

Thus, before incorporating the capitalization of data into the 2025 SNA framework, the ISWGNA ensured that the proposal was widely discussed, tested, and refined. The process, reflects the principle that changes to international statistical standards must balance conceptual improvement with operational feasibility and global comparability.

Prior to Country-level testing, the Asian Development Bank (ADB)-derived estimates for India, revealed data assets spanning 0.8% to 2.7% of GDP, with negligible GDP growth repercussions (0.00% for India in 2019) yet pronounced capital stock influences contingent on asset life, retirement profiles, and deflation methodologies (ISWGNA, 2023). For instance, Table 1 illustrates comparative values.

Table 1. Value of Data assets & theoretical impact on GDP growth (current prices, selected countries)

Country	Year	Value of Data Asset (% of Total GDP)	Percentage points Difference in Total GDP Growth for Year	Percentage points Difference in Total GCF Growth for the Year
(1)	(2)	(3)	(4)	(5)
Australia	2016	2.7%	0.016%	0.57%
Canada	2018	1.8%	-0.037%	-0.09%
Netherlands	2017	2.7%	-0.012%	-0.12%
India	2019	1.0%	0.000%	0.14%
USA	2020	0.8%	0.047%	0.26%

Source: ISWGNA (2023)

Academic studies have further explored this evolution, particularly in the context of productivity measurement. Goodridge *et al.* (2022) used Labor Force Survey data for European countries to estimate national investment in data assets, extending the asset boundary beyond software and databases as defined in the SNA. They found that in 2011–2018, 1.4% of EU-28 employment was engaged in the formation of software and data assets, growing at 5% per annum, highlighting the need for broader capitalization to capture data's role in productivity. One of the most important benefits of capitalizing data in the national accounts is that it can help explain the well-known "productivity paradox". This paradox refers to the puzzling fact that, even though data and digital technologies are now everywhere in our economy, official statistics continue to show very slow growth in productivity.

Goodridge *et al.* (2022) also argue that the main reason for this gap is that a large part of the money and effort spent on creating, cleaning, organizing, and analyzing data is not counted as investment under the current SNA rules. These activities are treated simply as day-to-day expenses and therefore do not appear in the capital stock. When the researchers expanded the asset boundary to include more of these data-related activities for 13 European countries during 2011–2016, the picture changed noticeably. Measured labour productivity growth rose from 0.79 % per year to 0.83 % per year. At the same time, the contribution of software and data capital to productivity growth increased more than three times — from 0.03 % to 0.10 % per year. They also found that these data-related activities that are currently left out of the accounts are growing faster than other parts of the economy, which means official productivity figures are probably being underestimated.

A similar situation is likely in India, which has experienced very rapid digital growth in recent years. According to the Asian Development Bank (ADB, 2021), investment in data assets in India grew at a compound annual rate of 7.8 % between 2011 and 2019 — faster than overall GFCF. Once India begins to capitalize these data assets, the measured productivity growth should therefore improve and give a more accurate picture of the economy's true performance.

India also faces some special challenges. A large share of data-related work takes place in the informal and unorganized sector, where measurement is difficult. These gaps could actually make the productivity paradox appear even worse in the official numbers. For this reason, the National Statistical Office, MoSPI will need to make suitable adjustments in its productivity calculations so that the important role of data in driving innovation, efficiency, and long-term growth is properly reflected.

3. FRAMEWORK AND METHODOLOGY FOR CAPITALIZING DATA IN INDIAN NATIONAL ACCOUNTS

As per 2025 SNA, data qualifies as a capital asset and can be capitalized whenever it is expected to deliver economic benefits for more than one year. The central test is simple: if the organized dataset will keep being used repeatedly or continuously in production — for example, a customer database that supports marketing campaigns year after year, a sensor dataset that guides factory operations for several years, or a health records database that informs ongoing medical research and policy — then the money spent creating or improving it counts as investment rather than a one-time expense. Short-lived or "ephemeral" data, such as raw observations collected for a single research project, a one-off marketing survey, or temporary data used only inside an R&D experiment, do not qualify. In those cases, the spending remains an ordinary current cost (intermediate consumption) or becomes part of another asset category like research and development.

The total value of a new data asset may be calculated using (a) the digital SUT framework or preferably (b) using the sum-of-costs approach, which adds up every real expense involved in its production. This includes labor costs — the salaries, benefits, and related expenses of data scientists, analysts, statisticians, programmers, and database managers who collect raw observations, clean them, organize them, label them, and store them in digital form. It also covers all intermediate inputs needed to do the job, such as cloud computing storage and processing fees, electricity, purchased software tools, data subscriptions, and any other goods and services consumed during production. For businesses operating in the market, statisticians add a markup for normal operating surplus (sometimes called a "profit" or "blow-up" factor) so the recorded value reflects the full economic cost that would be charged if the same dataset were bought from another company. Government and non-profit producers skip this markup because they do not aim for profit.

Finally, any spending that improves or expands an already existing data asset is treated as gross fixed capital formation (GFCF). Examples include adding new variables to a database, upgrading its structure for faster analysis, merging it with other datasets, extending its coverage to new year or regions, or improving its quality and security. These enhancements increase the asset's usefulness and productive life, exactly the same way renovating a building or releasing a new software version counts as investment. In the national accounts, such improvements are added to the stock of data capital and will be depreciated over time, giving a much clearer picture of how modern economies are building long-term productive capacity in the digital age.

(a) The OECD's Digital Supply and Use Tables (Digital SUTs) framework

It is a practical extension of the standard Supply and Use Tables (SUTs) used in the SNA to make digitalization more visible in economic statistics without creating a single, rigid definition of the "digital economy". Developed by the OECD's Informal Advisory Group on Measuring GDP in a Digitalized Economy (and detailed in the 2023 OECD Handbook), it focuses on three key dimensions:

- ***The "how" (nature of transaction)*** — the core innovation: splitting transactions by whether goods/services are digitally ordered (e.g., via website, app, or platform) and/or digitally delivered (remotely over computer networks, mainly services).
- ***The "what" (products)*** — ICT goods and services, cloud computing services (CCS), digital intermediation services (DIS), and digitally deliverable services.
- ***The "who" (industries)*** — seven new "digital industry" groupings (e.g., ICT-enabling firms, digital platforms charging fees, data/ad-driven platforms, e-tailers, and purely digital producers).

How it works: National statisticians start with their existing SUTs and *reallocate* rows and columns to show these breakdowns. This keeps everything consistent with SNA while producing richer insights.

What it delivers (high-priority indicators)

- Output and value added of digital industries
- Household and business spending on digitally ordered goods/services (e-commerce)
- Use of cloud computing and digital platforms across the economy
- Digitally ordered and digitally delivered trade (exports/imports)
- Share of the economy transacted digitally

The framework is flexible — countries can compile experimental tables at different levels of detail — and serves as the foundation for digital economy satellite accounts. It links directly to the OECD/IMF/UNCTAD/WTO (2023) *Handbook on Measuring Digital Trade*.

Using this framework The ADB furnishes a supply-use table modality for India's digital economy, segmenting core sectors and approximating data assets through input-output interrelations. It quantifies data investments leveraging labor costs (from census and ILO data) and capital inputs, categorizing data within intellectual property products (ADB, 2021). Data is delineated as the information content produced by accessing phenomena, recording, organizing, and storing digitally for economic gain. This aligns with extensions proposed in academic literature, such as Goodridge *et al.* (2022), who advocate for including data transformation and analytics beyond current SNA definitions for databases.

This methodology assures India's national accounts encapsulate digital verities, potentially elevating GDP by 0.5-1% via data capitalization (Ministry of Electronics and Information Technology, 2025). Table 2 exemplifies ADB's India estimates:

Table 2. Investments in Data, Databases, and Data Science in India (₹ billion)

Category	2011 (Lower/Upper, ₹ billion)	2019 (Lower/Upper, ₹ billion)	CAGR (%)
(1)	(2)	(3)	(4)
Total Data-Related	914/1,102	1,663/2,004	7.8
Data	496/624	980/1,225	-
Databases	201/225	373/418	-
Data Science	217/253	310/361	-
Net Capital Stock	2,944/3,631	5,511/6,794	8.2

Source: ADB (2021).

(b) Sum of Costs Approach:

The 2025 SNA recommends to use sum of costs approach to measure data assets that companies create on their own. These are called own-account data assets. It is important to know what statistics we need to estimate and publish. A full accounting for any asset class gives information on flows and stock each period. Flows mean changes over time. For example, flows can show how much new data is added each year. Stock means the total amount at a certain time. An example of stock is the total value of all data a company has at the end of the year.

The main statistics are the annual real investment and the net stock. Annual real investment is also called gross capital formation. For instance, this could be the real value of data created in a year after adjusting for prices. The net stock shows how much of the asset can still provide capital services. An example is the remaining useful value of data after subtracting wear and tear.

To estimate both nominal and real investment in data, additional inputs are required beyond simple expenditure totals. These inputs help explain the level of investment and its growth over time. Price information, for example, is essential to convert nominal investment into real investment. Nominal investment is measured in current monetary terms—that is, the amount spent during the year. Real investment adjusts this amount for price changes, removing the effects of inflation so that changes reflect the actual increase in the volume of investment rather than higher prices alone.

Depreciation rates are also necessary. They indicate how an asset gradually loses its capacity to provide services over time. In the case of data, depreciation may occur because information becomes outdated, less accurate, or no longer relevant. For instance, customer records may lose value as contact details change or market conditions evolve. Estimating appropriate service lives and depreciation patterns is therefore important for calculating capital stock and consumption of fixed capital in a consistent manner.

The measurement of own-account data production is based on international experience with own-account research and development (R&D) and own-account software within the System of National Accounts. These activities typically do not involve observable market transactions, as they are produced internally for use within the same enterprise. Consequently, no direct transaction prices are available for valuation. The recommended method is to measure output at the sum of production costs. This involves adding together all relevant costs incurred in creating the asset.

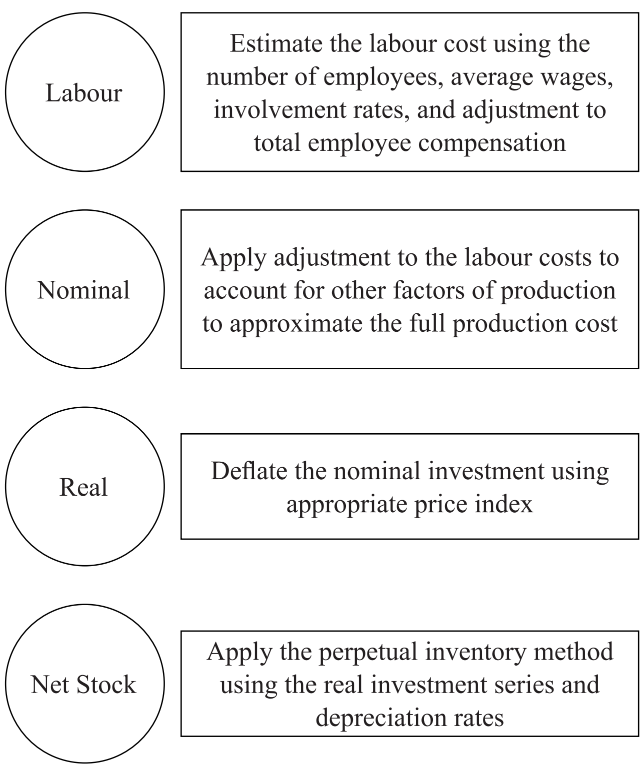

Fig. 2. Full accounting for asset category using the sum of costs approach

The figure presents the sequential steps from labour cost estimation to real investment and net capital stock using the perpetual inventory method.

In the case of R&D, this would include researchers' salaries, laboratory expenses, materials, and an appropriate share of overheads. Similarly, for own-account data, the costs may include compensation of data scientists and IT personnel, database infrastructure expenses, data collection costs, and related inputs. This cost-based approach ensures consistency with established national accounting principles while enabling practical and comparable estimation. Fig. 2 shows the sum-of-costs approach and all parts needed to fully account for an asset (Santiago Calderon, J.B. (2024)).

The first step is to estimate labour costs. These are the wages and salaries paid to workers who create and manage the data. This includes people such as data analysts, programmers, and statisticians who collect, organize, and process information. Since data production mainly depends on skilled workers, labour costs form the main part of the estimate.

The second step is to add other production costs. These are costs that are not related to wages but are still necessary to produce the data. For example, this may include spending on software, computers, servers, cloud storage, and other materials or services used in the process. Adding these to labour costs gives the total cost of producing the data asset.

The third step is to adjust the total investment for price changes. This is called deflating the nominal investment. Nominal investment shows the amount spent in current prices. After adjusting for inflation using a price index, the result is real GFCF. This helps show whether investment increased because more data was produced or simply because wages and other costs became higher.

The final step is to estimate the net capital stock. This means calculating the total value of data assets after considering both past investments and depreciation. Depreciation reflects the loss of value over time as data becomes outdated or less useful. For example, if a company invested 100 units last year and the data lost 20 units in value this year due to depreciation, the remaining net stock would be adjusted accordingly.

This measure shows the current value of data assets that can still provide benefits in the future.

Own-account data is similar to software because both involve creating and using information inside a company. They often use the same kinds of workers, such as programmers, IT specialists, and data experts. However, data assets can be different from software in how they are produced and used. Software may be written once and used for many years, with only small updates. In contrast, data usually needs regular updates to stay accurate and useful. Because of this, the costs and the loss of value over time may be different for data compared to software.

The method for measuring data investment can use different assumptions to see how the results change. For example, the estimates depend on how much time workers are assumed to spend creating and managing data. If it is assumed that workers spend more time on data-related tasks, the estimated investment will increase. The results also depend on how large the non-labour costs are. In industries that use a lot of technology, costs such as servers, software tools, and cloud services may be high. If these costs are higher, the total production cost and the investment estimate will also be higher.

Each country can adjust the method to match its own data sources. Countries use their national information on wages, employment, and business costs to prepare the estimates. For example, the United States may use information from the Bureau of Labor Statistics, while other countries may use labour surveys or census data. This makes the method practical for different national statistical systems.

At the same time, the approach follows the rules of the SNA, which provides a common framework for measuring the economy. Countries in the European Union, for example, work with Statistical Office of the European Union to ensure that their national accounts are prepared in a comparable way. This helps make economic data consistent and easier to compare across countries.

4. GROSS FIXED CAPITAL FORMATION (GFCF)

The sum-of-costs method used to estimate gross GFCF requires several connected decisions that influence the final estimates of real annual investment. The general approach to estimating nominal investment involves three main steps. First, the total yearly compensation of employees is calculated by occupation within each industry. Second, the portion of this compensation that relates to the production of data assets is identified and assigned to GFCF. Third, an adjustment factor is added to include non-labour costs, such as materials, services, and other expenses. Together, these steps provide an estimate of total nominal investment in data assets.

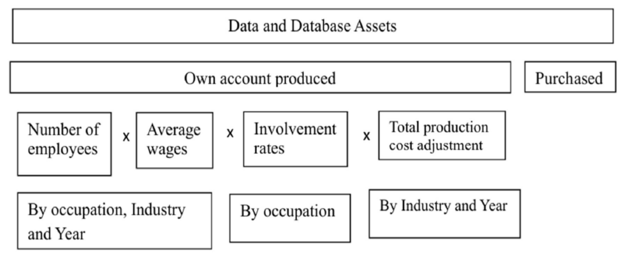

Fig. 3. Estimation of GFCF in Data Assets Using Labour Cost and Adjustment Factors

As depicted in Fig 3, above, the annual capital formation (C) series for data asset produced on own account can be estimated using an account identity decomposing total investment into labour cost components adjusted for non-labour inputs and other production factors, as indicated below:

$$C_{i,t} = \alpha_{i,t} \sum_{o} r_o N_{o,i,t} W_{o,i,t}$$

Where,

- $C_{i,t}$ = Gross Fixed Capital Formation (GFCF) in data assets for industry i at time t
- $\alpha_{i,t}$ = Adjustment (markup) factor for non-labour costsfor industry i at time t
- $\sum_o$ = Summation over all occupations o
- r_o = Involvement rate that captures the proportion of time and effort employees in a given occupation o dedicate to building or developing data-related capital formation
- $N_{o,i,t}$ = Number of workers in occupation o in industry i at time t
- $W_{o,i,t}$ = Average wage of occupation o in industry i at time t

To estimate nominal investment, three main steps are followed. First, the total yearly pay of employees is calculated for each occupation in every industry. Second, the share of this pay that relates to the creation of data assets is identified and treated as investment in data. Third, an adjustment is added to include non-labour costs, such as materials, services, and other operating expenses. By combining these steps, the total nominal investment in data assets can be estimated in a clear and consistent way.

The parameters are estimated using a combination of survey-based and administrative data sources. Employment and wage data are obtained from Periodic Labour Force Surveys (PLFS) and Occupational Wage Survey (OWS), while involvement rates are derived using expert judgement, time-use surveys, or international benchmarks. The adjustment factor is estimated from industry-level ratios of gross output to labour compensation using ASI and national accounts data.

These estimates are subject to uncertainty arising from measurement errors in survey data, assumptions regarding involvement rates, and heterogeneity across industries. Accordingly, results should be interpreted as approximations within a structured accounting framework rather than precise point estimates.

5. LABOUR COST

This section explains the process for identifying and using three key components needed to estimate labor costs involved in the GFCF of data and databases. It will guide the reader step by step on how each component contributes to calculating these estimates and how to combine them effectively for accurate results.

5.1 Pay-roll

The main goal of this part is to take the total labor costs shown in the national accounts for each industry and fairly divide them among all the different occupations in that industry. This provides a much clearer and more detailed picture of exactly how the money spent on labor is spread across various types of employees. At the same time, a standardized way needs to be created to measure these labor costs using one common unit—such as per employee, per full-time equivalent worker, or per actual hour worked. This step is essential because it allows the total amount of work being done (the quantity of labor) to be cleanly separated from the actual cost of each individual unit. Those two separate figures can then be used to build accurate and reliable prices for the specific asset category under consideration.

In India, clear and reliable data on the different kinds of jobs people do and the wages they earn in those jobs is readily available from two key government surveys. The Occupational Wage Survey (OWS), run regularly by the Labour Bureau, provides detailed information on average wages, hourly rates, and other earnings for specific occupations in important sectors like manufacturing, mining, plantations, and services, with breakdowns by industry and region. At the same time, the Periodic Labour Force Survey (PLFS), conducted by the National Statistical Office, offers wider and more frequent coverage across the entire country—both rural and urban areas—showing average earnings linked to various occupation categories for all types of workers. Together, the OWS and PLFS serve as a key source for understanding occupation- and activity-specific wages in India, helping everyone from policymakers to businesses track labour market trends accurately.

Using the information from these sources, the estimated shares of jobs and total payroll for each industry may be calculated and applied to the overall compensation of employees reported in the national accounts. Finally, to measure the change in price per unit of labor, the estimated compensation for employees in that industry is divided by the total hours actually worked within that industry.

5.2 Involvement rates

Involvement rates, also known as time-use factors, show how much of their working time and effort people in a particular job spend on specific tasks or activities. These rates can be estimated in several ways, including direct observation at the workplace, surveys of workers, or advice from experts. In the context of national accounts and the sum-of-costs approach, the main focus is on identifying which occupations actually carry out the work activities that create particular assets. For example, calculating the cost of developing software inside a company involves using a portion of the salaries paid to software developers, web developers, programmers, and testers.

A clear and precise definition of each asset class is therefore essential so that only the relevant work activities contributing to its production are included. Activities that merely involve using or analyzing data are not counted. However, the fact that data assets are being used provides strong evidence that the data and databases were produced earlier and are now being applied — often repeatedly in ongoing production — to generate real benefits. When the usage context shows that the data is applied continuously or repeatedly, this clearly indicates that the data itself should be treated as an asset.

6. TOTAL PRODUCTION COST ADJUSTMENT FACTOR

The total production cost adjustment factor, also known as a markup or blow-up factor, is used to raise the initial estimates that are based only on labor costs. This step ensures that other important elements of production costs are fully taken into account, giving a more complete picture of the true cost of creating an asset.

There are several ways to apply this adjustment factor. One of the simplest and most common methods is to calculate the ratio of an industry's gross output to its labor component — for example, total wages and salaries or overall employee compensation. These ratios differ widely across industries because each sector uses a different mix of labor, materials, and other inputs. In the professional, scientific, and technical services sector, for instance, production depends heavily on skilled workers, so labor makes up a large share of costs and the adjustment factor remains relatively small. By contrast, manufacturing relies much more on raw materials and intermediate inputs, which means a far larger portion of costs lies outside labor and the adjustment factor becomes significantly bigger.

According to the guidelines of the SNA, the adjustment factor for market producers must also include an allowance for "normal profits" — the reasonable return a business expects under normal conditions. However, estimating what counts as normal profit can be quite difficult in many real-world industries, especially those that do not operate under perfect competition where prices and profits are more stable and predictable.

6.1 Representative Industries

The most important first step is to choose a representative production function that closely matches the activity we want to measure. For data and databases in India's national accounts, the recommended representative industries are:

- Computer programming, consultancy and related activities (NIC Division 62), and
- Data processing, hosting and related activities; web portals (NIC Division 63 / Group 631).

The National Statistical Office (NSO) under MoSPI is encouraged to use these exact NIC codes or the closest available matching industries under India's National Industrial Classification system.

For example, just as the United States combines related NAICS codes (such as 518 and 519) into a single category for its adjustment calculations, Indian statisticians can group NIC 62 and NIC 631 together when preparing the total production cost adjustment factor (markup).

When using more than one industry, they apply a simple aggregation method: they calculate a weighted average of the adjustment factors, giving more importance (weight) to the industry that has a larger share of labour costs. This ensures the final markup accurately reflects how data and database production actually happens across these key sectors in India.

6.2 Non-market producers

In the Indian context, data and database assets are produced and used across all sectors of the economy — private businesses, non-profit institutions serving households (NPISH) such as educational trusts, hospitals run by NGOs, and charitable organizations, as well as the government sector. These sectors, especially educational services, healthcare, and public administration, are major contributors to India's Gross Fixed Capital Formation (GFCF) through their large-scale creation and use of data assets. When applying the sum-of-costs method to value data assets in NPISH and government, the National Statistical Office (NSO) must carefully adjust the total production cost adjustment factor (markup). This adjustment is essential to remove any implied normal profit component, because non-profit and government entities do not operate to earn profits like private market producers. This ensures the valuation remains accurate and consistent with the principles of India's national accounts.

7. NET STOCK

Most countries that have tried to calculate the net value of data assets using the perpetual inventory method (PIM) (OECD 2009b). The assumptions they use (such as how long the assets last) are the same as those they already use for other intellectual property products (IPPs) in their own country. However, these assumptions differ from country to country.

Unless there is very strong evidence to do something different, the recommendation is simple: each country should use the same PIM assumptions for data assets that it already uses for its other IPP assets.

For the international guideline, experts have mainly discussed the average service life of data assets. The current proposal suggests an average life of three to five years. This is shorter than for other IPPs such as software, which means data assets lose value (depreciate) faster. In future, if better information becomes available, countries can use different service lives for different types of data assets, depending on the industry that produces or uses them.

8. NUMERICAL ILLUSTRATION OF ESTIMATING DATA ASSET GFCF FOR INDIA

To demonstrate the practical application of the proposed methodology, this subsection presents a simplified numerical illustration using hypothetical values. The purpose is not to provide an official estimate, but to show clearly how the sum-of-costs method can be operationalized using the accounting identity introduced above.

Suppose that in a given year, one representative industry in India—say, Computer programming, consultancy and related activities (NIC 62)—employs the following categories of workers engaged partly in the creation of reusable data assets.

Table 3. Hypothetical labour inputs for estimating data asset investment

Occupation (o)	Number of workers $N_{o,i,t}$	Average annual wage $W_{o,i,t}$	Involvement rate r_o	Imputed labour cost for data asset creation (₹ lakh)
(1)	(2)	(3)	(4)	(5)
Data scientists	100	12.0	0.40	480.0
Data analysts	200	8.0	0.35	560.0
Database administrators	80	9.0	0.50	360.0
Software developers	150	10.0	0.20	300.0
Statisticians / research staff	50	11.0	0.30	165.0
Total	580	—	—	**1,865.0**

The imputed labour cost attributable to data asset creation is calculated as:$\sum r_{o,i,t} N_{o.i.t} W_{o,i,t}$

Thus, for this illustrative industry $\sum r_{o,i,t} N_{o.i.t} W_{o,i,t}$=1,865 lakh rupees

Suppose further that, based on ASI or industry accounts, the adjustment factor for non-labour costs and other production components is estimated as: $\alpha_{i,t}$=1.60

Then nominal gross fixed capital formation (GFCF) in data assets is:

$$C_{i,t} = \alpha_{i,t} \sum r_{o,i,t} N_{o.i.t} W_{o,i,t}$$

=1.60×1,865=2,984 lakh rupees

Conversion to real GFCF

To derive real GFCF, the nominal estimate must be deflated using an appropriate price index. Suppose the relevant price index for data-related production costs is 120 with base year = 100. Then real GFCF is obtained as:

Real GFCF = 2984/120 = 2,486.7 lakh rupees

Thus, the estimated real GFCF in data assets is ₹2,486.7 lakh, or about ₹24.87 **crore** at base-year prices.

Illustrative net stock estimation using PIM

Suppose the same industry recorded the following real GFCF in data assets over the previous four years:

Table 4. Real GFCF Time Series in data assets (hypothetical)

Year	Real GFCF (₹ lakh)
(1)	(2)
t-3	1,600
t-2	1,900
t-1	2,200
t	2,487

Assume further a straight-line depreciation pattern and an average service life of 4 years, implying an annual depreciation rate of 25 per cent. Under a simple perpetual inventory approach, the approximate net capital stock at time t can be expressed as:

$$K_t = I_t + (1-\delta)\, I_{t-1} + (1-\delta)^2\, I_{t-2} + (1-\delta)^3\, I_{t-3}$$

where δ=0.25

Substituting the values:

K_t =2,486.7 + 0.75(2,200) + 0.75^2 (1,900) + 0.75^3 (1,600)

= 2,486.7+1,650+1,068.75+675

= 5,880.45 lakh rupee

Interpretation

This stylized example demonstrates how existing Indian data sources can be used to generate an operational estimate of data-related GFCF. First, occupation-wise wage and employment data provide the labour cost component. Second, involvement rates identify the share of labour input attributable to the production of reusable data assets. Third, a markup factor extends labour cost to total production cost. Finally, the use of a price index and perpetual inventory method enables estimation of real investment and net capital stock. Although the numbers used here are hypothetical, the illustration shows that the proposed framework is transparent, replicable, and feasible for implementation in the Indian national accounts.

The above illustration is purely hypothetical and is included only to clarify the estimation steps. Official estimates would require empirically derived involvement rates, industry-specific adjustment factors, appropriate deflators, and validated assumptions regarding depreciation and service lives. This numerical illustration strengthens the empirical tractability of the proposed framework and demonstrates its immediate applicability within India's existing statistical system

9. BACKCASTING

The incorporation of this new asset class creates a challenge because its data series must be integrated into existing reports and publications. The new series will affect several important accounts, including intellectual property products (IPPs), non-residential fixed assets, industry growth rates, and GDP.

A recommended approach is to estimate the investment and net-stock values for this asset going back in time. This step prevents sudden jumps that could disrupt the smooth patterns desired in the data. Back casting of this kind has been applied successfully to other new assets in the past, and those earlier examples provide useful guidance for handling the inclusion of the new data.

Strategies for back casting involve adjusting the methodology according to the availability and quality of data sources. For example, it may include using different primary data sources and calculation methods for own-account software investment switching approaches whenever better information became available. Other methods include using decreasing involvement rates for certain occupations as a proxy to extend the series backward, or blending additional indicator series to produce the most realistic estimates until the target starting period is reached.

The resulting historical series smooths the impact on related accounts and serves as an essential input for applying the Perpetual Inventory Method (PIM) to generate accurate net-stock estimates. For these reasons, constructing a reliable historical series is a critical part of the process.

10. ADJUSTMENTS

This section explains both the theory and the real-world practice behind the production boundary in national accounts. The production boundary is the line that decides which activities count as economic production (and therefore belong in measures like GDP) and which ones do not. For data and databases to be treated as a new type of fixed asset, any spending on them must clearly qualify as gross fixed capital formation.

To make sure this happens, the section sets out the specific conditions that the production process must meet. These conditions act like a checklist: the activity must create something that lasts over time, can be used repeatedly in future production, and meets all the other rules for capital formation. Only when these criteria are satisfied can the money spent on creating or improving data and databases be counted properly as investment in this new asset class.

10.1 Multiple counting

To avoid counting the same spending more than once in India's national accounts, we must clearly understand the production boundary for new asset classes like Intellectual Property Products (IPPs). India's economy — with its booming IT sector, oil & gas exploration, scientific research, and massive entertainment industry — follows the United Nations System of National Accounts (SNA 2008) guidelines. This helps the Ministry of Statistics and Programme Implementation (MoSPI) correctly measure GDP and capital formation.

Here are simple explanations of the main IPPs used in Indian national accounts:

- *Research and experimental development (R&D):* This covers creative and planned work done to grow our overall knowledge — about people, culture, society, health, or technology — and then turn that knowledge into new products, medicines, or solutions. In India, this includes work at ISRO, DRDO, IITs, CSIR labs, and private pharma companies.
- *Mineral exploration and evaluation*: This means collecting data and information while searching for petroleum, natural gas, coal, or other minerals, and later checking how valuable the discoveries are. In India, this applies to activities by ONGC, Oil India, Geological Survey of India, and private mining firms.
- *Computer software*: This includes actual computer programs, their user manuals, design documents, and all supporting materials for both operating systems (like Windows or Linux) and application software (like banking apps or mobile games). India's huge software industry in Bengaluru, Hyderabad, and Pune makes this a very important asset class here.
- *Entertainment, literary and artistic originals (ELAO)*: These are the original master creations — such as films, music recordings, books, manuscripts, scripts, tapes, or digital models — on which movies, songs, TV serials, radio shows, dramas, sports broadcasts (like IPL), dance performances, or literature are recorded or stored. In India, this covers Bollywood and regional cinema, music from T-Series or Saregama, books by Indian authors, and digital content on platforms like Netflix or Hotstar.

By clearly defining these boundaries using the UN guidelines, India can correctly record spending on these items as investment (gross fixed capital formation) instead of routine expenses. This gives a more accurate picture of the country's growing knowledge-based and creative economy.

11. INDIAN DATA SOURCES FOR CAPITALIZING DATA AS AN ASSET IN NATIONAL ACCOUNTS

In India, the NSO under MoSPI uses a sum-of-costs method to estimate the value of data assets for own-account production. This method adds up employee compensation, intermediate consumption (such as ICT spending), consumption of fixed capital, and a markup for operating surplus where applicable. Although full capitalization of data as a produced asset (similar to software and databases under intellectual property products) is still emerging and aligned with international SNA guidelines, the key Indian data sources are already in use for related estimates like software and R&D (MoSPI, 2026).

(a) Periodic Labour Force Survey (PLFS)

The PLFS, conducted annually by the NSO, is the primary source for employee compensation — the largest part of data asset costs. It provides detailed data on wages, salaries, and earnings broken down by occupation (for example, data analysts, data scientists, software developers, and database administrators) and industry. This allows statisticians to identify how much time and pay go toward producing data assets and to apply time-use factors (such as the share of working hours spent on data-related tasks). PLFS covers both rural and urban areas and all types of workers, making it essential for building labour costs in the sum-of-costs approach (NSO, various years).

(b) Annual Survey of Industries (ASI)

The ASI is an annual census-like survey of registered manufacturing and some service enterprises conducted by MoSPI. It supplies detailed information on intermediate consumption (spending on cloud services, software tools, data storage, and ICT equipment), consumption of fixed capital (depreciation of servers and computers), and overall production costs. For data assets created within companies, ASI helps calculate the non-labour parts of the cost and provides benchmarks for organized-sector enterprises producing or using large amounts of data (MoSPI, various years).

(c) Annual Survey of Unincorporated Sector Enterprises (ASUSE)

ASUSE (previously part of NSS enterprise surveys) covers small and unorganized businesses, which form a big part of India's economy. It gives data on wages, intermediate inputs (ICT spending), and fixed capital use in informal sectors. This source is crucial for capturing own-account data production in smaller firms and startups that may not appear in ASI or corporate filings. In the new 2022-23 national accounts series, ASUSE is now used annually for direct estimation of the household and quasi-corporate sector (MoSPI, 2026).

(d) Goods and Services Tax (GST) Data

GST returns from the Goods and Services Tax Network provide rich administrative data on spending on digital services, cloud computing, data processing, software subscriptions, and other intermediate inputs. Statisticians use product-wise (HSN/SAC codes) GST data to measure how much businesses spend on items that help produce data assets. This source has become increasingly important in the new 2022-23 GDP series for cross-checking estimates and capturing modern digital transactions (MoSPI, 2026).

(e) Balance of Payments (BoP) Statistics

Prepared by the Reserve Bank of India, BoP tracks international flows of digital services, data processing, and software-related trade. It helps value cross-border data assets and imported components (such as foreign cloud services or data tools) used in domestic production. This source ensures that transnational data flows are properly included when estimating the total value of data assets in India's national accounts (RBI, various years).

Together, these sources allow NSO to build reliable estimates of data assets using detailed occupation-based wages, enterprise costs, digital spending, and international flows. The approach closely follows international best practices, such as those tested in ADB reports for India, which showed data-related gross fixed capital formation growing at 7.8 per cent per year between 2011 and 2019 (Asian Development Bank, 2021). As India moves toward full capitalization of data assets in line with the updated SNA, these sources will be refined further with granular GST, MCA corporate filings, and PLFS occupation data. This will give a more accurate picture of data as a valuable produced asset in the Indian economy (MoSPI, 2026).

12. SENSITIVITY ANALYSIS OF KEY PARAMETERS

The estimation of data assets relies on several key assumptions, including involvement rates, markup factors, depreciation rates, and asset service lives. To assess robustness, sensitivity analysis can be conducted by varying these parameters within plausible ranges.

Two alternative scenarios are constructed:

Low Estimate Scenario (Conservative Case)

- Lower involvement rates (reflecting stricter definition of data-producing activities)
- Lower adjustment factor (labour-intensive production)
- Longer service life (slower depreciation)

Table 5. Assumptions Across Scenarios

Parameter	Low Scenario	Baseline (from illustration)	High Scenario
(1)	(2)	(3)	(4)
Involvement rates	Reduced by 25%	As in Table 4.X.1	Increased by 25%
Adjustment factor α	1.40	1.60	1.80
Price index	120	120	120
Service life (years)	5	4	3
Depreciation rate	20%	25%	33.3%

High Estimate Scenario (Expansive Case)

- Higher involvement rates (broader inclusion of data-related work)
- Higher adjustment factor (greater non-labour input intensity)
- Shorter service life (faster obsolescence of data)

Step 1: Labour Cost Under Alternative Scenarios

Baseline labour cost (from earlier):=1,865 lakh

Adjusted Labour Cost

- Low scenario:1,865×0.75=1,398.75
- High scenario:1,865×1.25=2,331.25

Step 2: Nominal GFCF Estimates

$C_{i,t}$=α×Labour Cost

Table 6. Nominal GFCF Across Scenarios

Scenario	Labour Cost (₹ lakh)	Adjustment Factor	Nominal GFCF (₹ lakh)
(1)	(2)	(3)	(4)
Low	1,398.75	1.40	1,958.25
Baseline	1,865.00	1.60	2,984.00
High	2,331.25	1.80	4,196.25

Step 3: Real GFCF Estimates

Real GFCF=Nominal GFCF/1.20

Table 7. Real GFCF Across Scenarios

Scenario	Nominal GFCF (₹ lakh)	Real GFCF (₹ lakh)
(1)	(2)	(3)
Low	1,958.25	1,631.88
Baseline	2,984.00	2,486.70
High	4,196.25	3,496.88

Step 4: Net Capital Stock (PIM)

Using the same historical series scaled proportionally and scenario-specific depreciation rates:

Table 8. Net Capital Stock Across Scenarios

Scenario	Depreciation Rate)	Net Capital Stock (₹ lakh)
(1)	(2)	(3)
Low	20%	6,950
Baseline	25%	5,880.45
High	33.3%	4,9504,950

Key Insights

The sensitivity analysis highlights several important findings:

- High sensitivity to involvement rates: A ±25% change leads to substantial variation in estimated GFCF
- Adjustment factor is a major multiplier: Small changes significantly affect total investment
- Depreciation assumptions matter for stock estimates:
  - Longer service life → higher capital stock
  - Shorter service life → faster decline in value
- Range of estimates:
  - Nominal GFCF varies from ₹1,958 lakh to ₹4,196 lakh
  - Real GFCF varies from ₹1,632 lakh to ₹3,497 lakh

The results indicate that estimates of data asset investment are sensitive to key modelling assumptions, particularly involvement rates and markup factors. This underscores the importance of developing empirically grounded parameter estimates using Indian data sources. At the same time, the relatively bounded range of estimates suggests that the proposed framework provides a stable and informative basis for integrating data assets into national accounts.

13. CONCLUDING REMARKS

The recommendation of the 2025 SNA to recognize organized, reusable digital data and databases as produced fixed assets represents a significant step toward aligning national accounts with the realities of a data-driven economy. This paper has demonstrated that the conceptual framework is well grounded, and that the preferred valuation method—the sum-of-costs approach—can be operationalized in the Indian context using existing statistical systems.

By explicitly formulating the estimation process through an accounting identity and supporting it with a numerical illustration, the paper shows that data asset measurement is both transparent and implementable. The incorporation of sensitivity analysis further highlights the importance of key parameters—such as involvement rates, adjustment factors, and depreciation assumptions—while confirming that the resulting estimates remain within a plausible and policy-relevant range.

India is particularly well positioned to adopt this framework, given the availability of rich and complementary data sources, including PLFS and OWS for occupation-specific labour costs, ASI and ASUSE for enterprise-level inputs, GST data for digital intermediate consumption, and BoP statistics for cross-border flows. Together, these sources provide a strong empirical foundation for estimating data-related gross fixed capital formation without the need for entirely new data collection systems.

Empirical evidence from international experience suggests that while the capitalization of data may have limited immediate impact on GDP growth, it significantly improves the measurement of capital stock, productivity, and the contribution of intangible assets. For India, where digitalization is expanding rapidly across both formal and informal sectors, incorporating data assets into national accounts will provide a more accurate representation of economic structure and growth dynamics.

The successful implementation of this framework will require careful attention to practical issues, including the selection of representative industries, estimation of adjustment factors, development of appropriate price deflators, and refinement of service life assumptions. Equally important is maintaining a clear distinction between capitalized data and intermediate consumption to avoid misclassification and double counting.

Looking ahead, continued experimental estimation, firm-level studies, and the development of digital economy satellite accounts will be essential to further refine measurement practices. By integrating data assets into the national accounts, India can enhance the relevance of its macroeconomic statistics, strengthen evidence-based policymaking, and position itself at the forefront of global efforts to measure the digital economy in a comprehensive and consistent manner.

ACKNOWLEDGEMENTS

The authors are grateful to the reviewer for providing insightful and constructive comments. The suggestions have been extremely helpful in improving the clarity, methodological rigor, and overall presentation of the paper.

REFERENCES

Asian Development Bank. (2021):*Capturing the Digital Economy: A Proposed Measurement Framework and Its Applications*. Manila: ADB. Available at https://www.adb.org/sites/default/files/publication/722366/capturing-digital-economy-measurement-framework.pdf

Goodridge, P., Haskel, J., & Edquist, H. (2022): We See Data Everywhere Except in the Productivity Statistics. *Review of Income and Wealth*, **68**(4), 862-894.

Inter-secretariat Working Group on National Accounts (ISWGNA): (2023). *Recording of Data in the National Accounts*. United Nations. Available at https://unstats.un.org/unsd/nationalaccount/RADOCS/ENDORSED_DZ6_Recording_of_Data_in_NA.pdf

Ministry of Electronics and Information Technology: (2025). *Estimation and Measurement of India's Digital Economy*. New Delhi: Government of India.

Ministry of Statistics and Programme Implementation. (2025):*National Accounts Statistics 2025*. New Delhi: Government of India.

Ministry of Statistics and Programme Implementation (MoSPI). (2026). Report of the Sub-Committee for Incorporation of New Data Sources, Rates and Ratios. Government of India. Available at: https://www.mospi.gov.in/

Mitchell, J., D. Ker and M. Lesher (2021), "Measuring the economic value of data", OECD Going Digital Toolkit Notes, No. 20, OECD Publishing, Paris, https://doi.org/10.1787/f46b3691-en

National Statistical Office (NSO). (various years). Periodic Labour Force Survey (PLFS) Annual Reports. Ministry of Statistics and Programme Implementation.

MoSPI. (various years). *Annual Survey of Industries (ASI)*. Government of India.

Reserve Bank of India (RBI). (various years). *Balance of Payments Statistics*. RBI.

OECD/IMF/UNCTAD/WTO (2023), Handbook on Measuring Digital Trade, Second Edition, OECD Publishing, Paris. https://doi.org/10.1787/ac99e6d3-en

OECD (2009). *Measuring Capital - OECD Manual 2009*: Second edition. OECD Publishing, Paris. DOI: https://doi.org/10.1787/9789264068476-en

Santiago Calderon, J.B. (2024): Recording Data as an Asset in National Accounts. Paper presented at the at the 34th AIRIW Conference, 22 August 2024.

United Nations, European Commission, IMF, OECD, & World Bank. (2008): *System of National Accounts 2008*. New York: United Nations

United Nations, European Commission, IMF, OECD, & World Bank. (2025): *System of National Accounts 2025*. New York: United Nations.