Data, Intangible Capital, and Productivity*

Carol Corrado, The Conference Board and Georgetown University McDonough School Center on Business and Public Policy
Jonathan Haskel, Imperial College Business School and Bank of England
Massimiliano Iommi, ISTAT and LUISS University
Cecilia Jona-Lasinio, ISTAT and LUISS University
Filippo Bontadini, LUISS University

Keywords: intangible capital, data, innovation, productivity growth
JEL reference: O47, E22, E01

March 8, 2022
(Preliminary conference draft, not for quotation)

* Paper prepared for the NBER/CRIW Conference on Technology, Productivity, and Economic Growth, March 17-18, 2022. The authors are grateful to the OECD Economics Directorate and the European Commission's DG ECFIN for partial financial support of work reported in this paper.

1. Introduction

A defining aspect of the digital age is the use of data, specifically large stores of digitized information referred to as "bigdata." Much popular work on bigdata appears in the business strategy literature. By a very long way, the best-selling book on the subject is "Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Mayer-Schönberger and Cukier (2013). The book makes several statistical claims, suggesting that data will be used to disprove much casually held causal intuition and reduce many measurement problems—in line with some economists who describe the transformational promise of digital tools that adjust themselves to perform better as they are exposed to more and more data (e.g., Brynjolfsson and McAfee 2014) and assertions in the press that data is the new oil (e.g., The Economist Magazine 2017).

These statements—some from nearly a decade ago—would imply that data has had significant impacts on economic activity. But has it? Economic growth has slowed globally, and business productivity performance has been subpar. Though the ease with which business users can deploy modern digital tools needed to derive knowledge from data is frequently suggested as reason for this subpar performance, the very purpose of modern software and computing systems hosted in the cloud is to reduce technical barriers to user engagement in data analysis. As it seems unlikely that there has been a failure in the inherent productivity of AI and cloud-based technologies, this paper looks to the character of knowledge gained through data itself as a contributor to the slowdown in productivity growth.

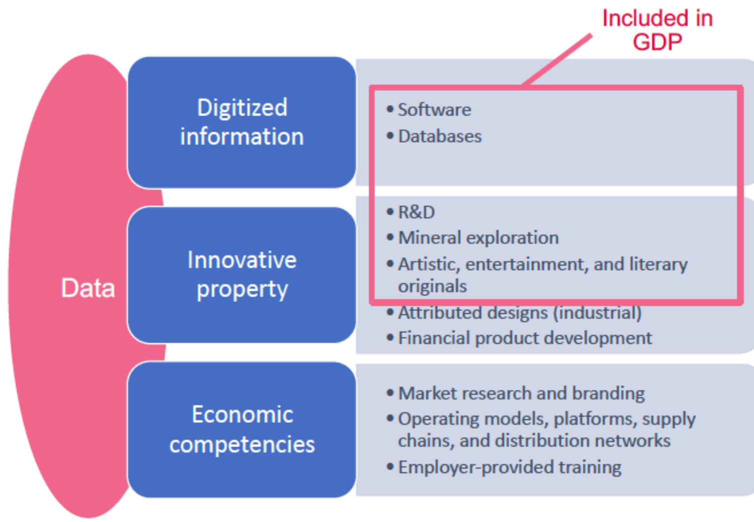
Data is conceptualized as an intangible asset in this paper: a storable, nonrival (yet excludable) factor input that is only partially captured in existing macroeconomic and financial statistics. A framework for capturing asset creation based on all digitized information that is processed and transformed into useable knowledge, or data capital, in an economy is set out in this paper, a framework amenable to measurement and quantitative analysis.

A contribution of this paper is the development of estimates of data capital coherent with both national accounts and widely used concepts in the intangible capital literature at the industry-level of analysis. We find that data assets, from stores of raw data to actionable intelligence derived via data analytic tools, are largely subsumed within the intangible capital framework attributable to Corrado, Hulten, and Sichel (2005, 2009), shown in figure 1. The analysis of data in this paper models the economic impacts of data capital using this framework, emphasizing how the relative importance of data capital within intangible capital lowers intangible asset prices and strengthens the (partial) appropriability of the asset class.

The goal of the paper is to model and estimate the impact of these changes due to the increased use of data in economies on (a) intangible capital and (b) productivity. We thus first address concepts associated with value creation due to data and assess how these concepts are covered in available measures of intangible capital.

Virtually all components of intangible capital are potentially driven by data….

Figure 1 Intangible Capital: Broad groups and investment types



After modeling how data capital, innovation, and productivity are related, the consequences for labor productivity of the increased use of data capital according to its breadth of use and relative efficiency in production are calibrated. We conclude by using our findings from the conceptual and empirical analysis of the relationship between data assets and intangibles to interpret the recent slowdown in total factor productivity. The analysis makes use of the recently issued EUKLEMS & INTANProd database, which includes estimates of intangible investment per figure 1 for EU countries, Japan, United Kingdom, and the United States.

Relation to recent literature

Approaches to the measurement of data are summarized in the measurement section. Here we attempt to place our empirical macro findings in the context of models that focus on economic mechanisms affected by data.

The innovative potential of data as an intangible asset rests in its ability to yield competitive returns to owners and "spillovers" elsewhere in an economy. Spillovers are produced when a technology or business idea is adopted relatively costlessly (or copied) by multiple firms in an economy, e.g., a blueprint or original software tool (Romer 1990, Jones 2005). Owing to this nonrival property, intangible assets are only partially appropriable by their owners/creators, creating a situation in which the asset class has increasing returns at the macro level.

At the micro level, data is usually assumed to have diminishing returns, e.g., Varian (2019) points out that there are diminishing returns to more and more training data fed to AI algorithms. Jones and Tonetti (2020) formulate an aggregate model of data in an economy in which data is a productive intermediate input with diminishing returns, not a "technology" that leads to increasing returns. In the intangible capital model set out in section 3, data are productive long-lived assets whose value stems in part from the application of data technologies. There are obvious differences between these approaches (e.g., data as an intermediate vs data as capital). The stylized Jones and Tonetti model is designed to highlight the aggregate welfare impacts of data sharing, while the intangibles framework applies to analyzing data value creation via business investment. The data/intangible capital approach of section 2 combined with the existence of productivity spillovers is a close representation of the welfare-enhancing processes theorized by Jones and Tonetti in that (a) data assets have diminishing returns in production but (b) returns to data asset ownership may spill over to other firms to the extent they are shared within an industry or economy.

Many models attribute rising market power and/or industry concentration to scale economies of intangible assets at the firm level (e.g., Crouzet and Eberly 2019, De Ridder 2019), suggesting that firm-level studies assessing changes in competitive conditions that ignore intangibles likely obtain biased results (due to the usual omitted variable bias argument). Closer to our findings are studies that attribute declining business dynamism to a slowdown in knowledge diffusion and offer that proprietary data play a larger role in modern production processes as a plausible story for slowdown (e.g., Akcigit and Ates 2021). Such a breakdown might occur if, in innovative data-intensive firms, diminishing returns to data assets can co-exist with rising market power or cost advantages due to scope economies and local scale effects.[1] Firms may also amass market power due to agglomeration effects that weaken the law of diminishing returns, e.g. by recombining data for different uses.[2] And sometimes an industry's scope of operations expands due to outside developments that create external economies of scale, e.g. network externalities enjoyed by social media and other digital sharing platforms.

The contribution of this paper is to offer a framework and macroeconomic empirics in line with the primary development behind the foregoing concerns—the increased use of data—and we hope, in so doing, that we can sharpen our understanding of the divergent perceptions its impact on economic activity.

2. Data Value Creation

[1] Unlike economies of scale, where unit costs fall as the volume of production rises, economies of scope are efficiencies that arise from variety, not volume, creating a situation where a company's average cost of production falls with product diversification. Economies of scope are often characterized by local cost complementarities among factors of production as well as the existence of fixed costs, especially in large enterprises (e.g., marketing, supply-chains, distribution systems, etc.).

[2] As used here, agglomeration effects refer to the fact that proprietary data assets of one type may be combined with another type to generate whole new uses or solutions, and to the extent this occurs within a single firm, it weakens the effect of diminishing returns to data.

Rise of Proprietary Data

Consider first examples of data use in modern economies. Table 1 lists examples of data uses, grouped according to whether the use is rival or nonrival. Though data is inherently nonrival, the classifications in the table are designed to reflect the degree to which data are openly shared with the public or other organizations in an industry (or the economy).

As may be seen, the uses listed on lines 1–5 mainly reflect applications of new digital technologies by firms, i.e., digital platform-based businesses and/or applications of machine learning and other AI-based algorithms to massive data. Product-led growth strategies (line 6) refers to marketing innovations based on user feedback data (also enabled by new technologies). Line 7, customer lists and after-sales customer feedback, long have been inputs to brand development, marketing, and customer retention strategies.

Examples of "nonrival" data use range from "new technology" marketers of personal data for B2C companies (line 8), to examples of longer-standing industry-level data sharing, e.g., financial records held by credit bureaus and shared across financial institutions (line 9), vehicle accident and major repair records shared by buyers and sellers in used car markets (line 10), and personal medical records shared across providers of medical care services (line 11), to cross-platform and cross-purpose uses (lines 12 and 13). Finally, the table lists two examples of government open data.

The examples in the table suggest that data has much potential for wide use and industry benefit when shared, though many of the "new" applications involve exclusive, proprietary use (mainly in marketing but also digital manufacturing operations).

Consumer privacy protection also engenders near exclusive use of business-held data—and excludability via policies that prohibit re-using of data (e.g., lifestyle data collected by marketers used for precision medicine solutions) potentially affects the pace of digital innovation. Conversely, policy intervention may be needed to

Table 1. Examples of Data Use

Rival	
1.	Product-level forecasting (e.g., Amazon)
2.	A/B Internet testing and marketing (e.g., Google)
3.	IoT factory systems (e.g., Siemens)
4.	Targeted advertising on consumer content platforms
5.	Fintech (e.g., algorithmic trading, digital lending, etc.)
6.	Product-led growth strategies (e.g., Slack)
7.	Customer lists/after sales services design
Nonrival	
8.	DaaS (Data as a Service) platforms (e.g., BDEX)
9.	Financial records (FICO scores)
10.	Vehicle records (CARFAX reports)
11.	Personal medical records (across service providers)
12.	Open-source data generated by web users (map data)
13.	Private by-product data put to alternative uses (e.g., Zillow data used for economic research)
14.	Genomic and other public biomedical research data
15.	Official statistics (economic, demographic, social)

Note: Data is inherently nonrival, and classifications reflect the degree to which owners share data with other organizations or the public.

ensure both thriving market competition in data-intensive markets (e.g., digital platform businesses) and protection of consumers.

A conceptual framework for measuring and analyzing data and its role in competition is necessary for understanding the factors affecting data markets and economic growth as data assets play an increased role in production. A framework needs to account for the following "special" characteristics of data: (a) data is nonrival and, like other intangible assets, capable of improving economic welfare via sharing, whether within an industry or general-purpose commons; (b) data, though nonrival, is frequently used exclusively, i.e., to business owners, data assets are trade secrets; and (c) data is different from other intangible assets in that it has a consumer privacy dimension. Data privacy laws often mandate exclusivity.

The Data Stack

Many economic models of data focus on data as a "free" by-product of economic activity, and many observers focus on certain special features of data, such as how rapidly it accumulates. In contrast our approach is based on the following observations:

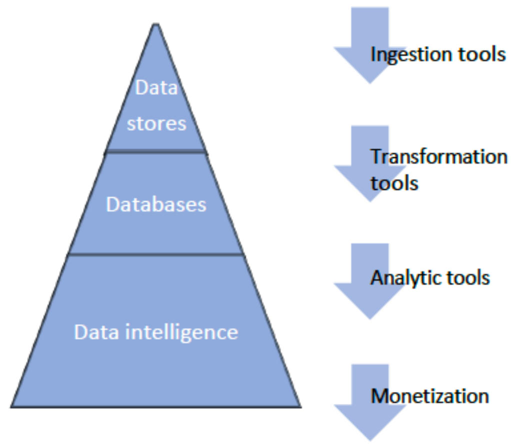
- Data, in the sense of raw digitized records, may accumulate at an astonishing pace and be stored at little to no cost. But that does not automatically provide a flow of services to production.
- The accumulation of data has the potential to boost real output only when the sector also invests in transforming such records, possibly along with other available economic or social information, into analytical insights and actionable business intelligence.
- Data stores and knowledge gleaned from data stores via application of data technologies are, in fact, long-lived intangible assets that contribute to final production in an economy. The long-lived appropriability of accumulated stores of digitized information implies that business spending on data and data transformation are intangible capital investments.

Our specific approach to data value creation embraces widely used approaches in the technology and management literatures. Technologists characterize data according to a "data stack" that describes the transformation of raw data into usable data structures and intelligence. Business management strategists use a value chain construct that adds monetization, or market implementation, as a capability (or tool) required for data value creation.[3]

[3] See again Mayer-Schönberger and Cukier (2013), also PriceWaterhouseCoopers LLP (2019).

Our framework for data value creation is illustrated in figure 2. Though it embraces both characterizations, note first that technologists usually stack a sequence of data forms and digital tools in a single pyramid.[4] Figure 2 separates these by identifying three major forms of data inherent in their characterizations. These forms, depicted on the left, reflect the business strategists' notion of an information value chain, where greater value is produced as data is processed into usable intelligence. The digital tools that enable value creation from raw, digitized information are depicted on the right. The sequencing of data assets with tools used in their formation is implied, i.e., ingestion tools are used to create data stores, etc.

Figure 2. The Data Asset Value Chain



Note. The stack to the left depicts the stages of data asset value creation based on applying layers of tools shown on the right.

The data asset stack has then three layers of value—data stores, databases, and data intelligence—each corresponding to an asset type amenable to measurement and analysis.[5] The asset types are defined more precisely as follows:

- *Data stores* are raw records that have been stored but not yet cleaned, formatted, or transformed for analysis, e.g., data scraped from the web or sensor and economic data captured from production or transactions activity. Raw records also cover the raw data collected from experiments, statistical surveys, or administrative records.
- *Databases* consist of transformed raw data, records that have been cleaned, formatted, and structured such that they are suitable for some form of data analytics or visualization.
- *Data intelligence* reflects the further integration of data with advanced analytic tools (e.g., machine learning training algorithms); data intelligence is a set of quantitative inputs that provide actionable guidance for decision-makers, including solutions to scientific problems.

[4] See, e.g., Roca (2021a), for a recent depiction. The data stack has its roots in information science, which uses the concept of a "data pyramid" to depict the relationship between data, information, and knowledge (Varian 2019).
[5] A multiple asset type conceptual approach has been used in previous work on defining and measuring data, including McKinsey Global Institute (2016), Statistics Canada (2019a, 2019b), Nguyen and Paczos (2020), and Goodridge, Haskel and Edquist (2021).

What separates the "modern" data stack from legacy systems is that modern systems are hosted in the cloud, requiring little technical configuration by users. According to technologists (e.g., Roca 2021b), "the modern data stack lowers the technical barrier to entry for data integration." And "components of the modern data stack are built with analysts and business users in mind, meaning that users of all backgrounds can not only easily use these tools, but also administer them without in-depth technical knowledge."

Implications of the Data Stack

The key implication of this framework is that data value creation reflects investment in the application of layers of data technologies and market implementation to create assets that generate productive value in an economy. New investment streams typically accompany the emergence of new technologies, e.g., the invention of the modern internal combustion (IC) engine was followed by a surge of spending on motorized equipment for transport. The seemingly sudden appearance of transport equipment stemmed from its many uses in consumption and production, e.g., personal travel, farming, goods delivery. The arrival of new data technologies such as AI might be likewise expected to cause a shift in the composition of business spending towards "all things data"—data analytic tools, data stores, structured dataset development, data-derived business strategies—i.e., the appearance of data assets capable of further use in production or for sale.

The data value chain framework, in which greater value added is created as raw data is processed and developed into insights and solutions, applies to data-driven development of engineering designs, customer platforms, and organizational practices, as well as to data-driven R&D processes.[6] This suggests that data assets are largely subsumed—though not explicitly identified—in available measures of intangible capital but not fully covered by investment in GDP/national accounts (see again figure 1).

From this perspective, i.e., a knowledge-based or intangible capital perspective, the increased use of data assets derived from modern digital technologies is an "innovation in the method of innovation." Modern data use fosters faster, more efficient experimentation and feedback in R&D processes, industrial production processes, marketing research, and business strategy and operating model development. This implies that, with increased use of data and application of

[6] Although the three asset types shown in figure 2 generally align with categories Statistics Canada set out in a conceptual framework for measuring data, Statistics Canada calls the third category "data science" and views it as unmeasured R&D, e.g., spending to develop new AI algorithms. Though data and data tools (AI) are inextricably bound via feedback and training data used to develop and refine AI tools, the data stack/data value chain notion of how value is created from data does not end with the development of algorithms.

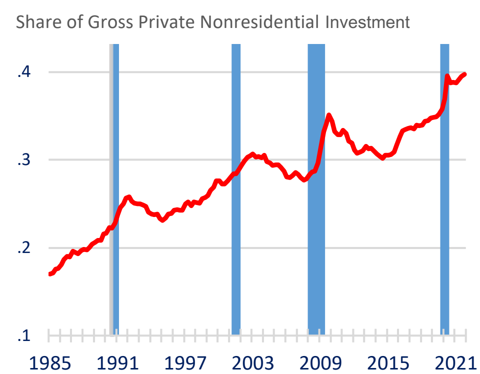
digital technologies, the "productivity" of these activities improves, i.e., that their resource cost per unit of final output falls, an implication discussed more detail in section 3 of this paper.

The depiction of monetization as a capability required data value creation in figure 2 refers to an organization's capability for implementing actions guided by data intelligence. Though these actions are relevant for understanding the macroeconomic impacts of the increased use of data in economies, they are played out via adjustments to existing primary factor inputs, i.e., labor and capital (tangible or intangible), in the short and long run.

Though the primary focus of this paper is on how data capital as intangible capital affects productivity growth, the rise of data capital as a strategic factor input also has the potential for altering cyclical patterns in macroeconomic data—patterns of investment and factor input demands, and perhaps the responsiveness of inflation to economic conditions in the short run. Though subjects for future research with more complete data, the partial incorporation of intangibles in quarterly GDP hints that there is indeed something different about the workings of the intangible macroeconomy.

Research on the formulation of investment demand argues that intangibles are less sensitive to changes in interest rates than tangibles due to their higher user cost and tendency to be less reliant on secured debt financing.[7] Figure 3 displays fluctuations in the intellectual property products share of private nonresidential investment using quarterly data from the U.S. national accounts, which also suggests that these investments are the last category of capital spending cut during downturns. Businesses may view the acquisition of software (and other intangibles) as moves to increase efficiency that dampen the impact of workforce layoffs and cutbacks in customer demand, i.e., that intangible capital (or some forms of it) may allow firms to adjust production relatively rapidly to changes in economic conditions, with possible implications for inflation dynamics and monetary nonneutrality.

Figure 3. U.S. Intellectual Property Products Investment, 1985Q1- 2021Q4



Source: Authors' elaboration of quarterly NIPA data.
Notes: Intellectual property products include software, R&D, and entertainment originals. Shaded areas are periods of business recession as defined by the NBER.

[7] See, e.g., Crouzet and Eberly (2019), Haskel and Westlake (2019, chapter 8), and Döttling and Ratnovski (2020) for further elaboration.

Furthermore, the most recent observations in figure 3 show that IPP investments remained relatively strong in the *recovery* from the economic downturn caused by the pandemic.[8] The fact that intangible capital increasingly reflects knowledge built from the analysis of data likely explains this persistence of relative strength. Half of the respondents in survey of companies administered by McKinsey & Company reported that the pandemic-induced economic downturn had *no effect* on their investments in AI, while 27 percent reported *increasing* them (AI Index Report 2021, page 103).

Data capital as Intangible capital

Intangible investment covers a wide class of investments, from databases to business processes, engineering design, and market research, that would appear to be relevant for analyzing the consequences of the increased use of data in economies. Let us then consider the definitional/conceptual overlap between the data assets in the data stack and activities covered by existing measures of intangible assets.

Identified intangible investment asset types are set out in table 2. Column 1 of the table shows that there are three major categories of intangible assets: digitized information, innovative property, and economic competencies. Column 2 reports specific assets used to populate each major category, and column 3 reports whether the asset is covered in national accounts. As may be seen, only lines 1 through 5 are included.

Table 2. Intangible Investment: Major Categories and Asset Types

Categories (1)	**Investment by Asset Type** (2)	**NA** (3)	**Examples of Assets and Property** (4)
Digitized Information	1. Software	Yes	Digital capabilities, tools
	2. Databases	Yes	Trade secrets (data)
Innovative Property	3. Research and development (R&D)	Yes	Patents, licenses
	4. Mineral exploration	Yes	Mineral rights
	5. Artistic, entertainment, and literary originals	Yes	Copyrights, licenses
	6. Attributed designs (industrial)	No	Patents, trademarks
	7. Financial product development	No	Trademarks, software patents
Economic Competencies	8. Brand and market research	No	Brand equity, customer lists, market insights
	9. Business process and organizational practices	No	

[8] IPP investments refers to the national accounts investment category, intellectual property products, which in the United States includes three components of intangible capital: software and databases, R&D, and artistic, literary and entertainment originals. International standards (e.g., OECD 2010) include mineral exploration in IPP but this is not done in the U.S. data.

	10. Employer-provided training	No	Operating models and platforms, supply chains and distribution networks, and management and employee practices Firm-specific human capital

Note. Column 3 indicates whether the asset type is currently included as investment in national accounts (NA).
Source: Updated version Corrado, Hulten and Sichel (2005) as set out in Corrado (2021).

At first blush one might infer from column 1 of table 2 that the digitized information grouping of intangible assets includes the data stack's individual asset types, but as may be seen in the itemized list in column 2 of table 2, only databases appear. This implies that national accounts' estimates of the value of investment in databases exclude the cost of acquiring or ingesting the data stores they contain; furthermore, as a matter of practice, outright purchases of data stores and databases are only included to the extent they are embedded or sold as software products.[9]

Consider now data intelligence, the most valuable, final stage of the data value chain per figure 2 and where the utility of the intangible capital framework becomes especially apparent. The knowledge created from data encompasses all modern, data-driven business, financial, marketing, engineering, and scientific intelligence. The inclusion of investments in business operations, financial products, and general marketing intelligence in intangible capital is readily seen via lines 7, 8 and 9 of table 2.

An increase in the use of data capital in R&D activities (line 3), will cover novel forms of data-derived scientific intelligence (including the development of new AI techniques and certain bio-engineered substances/formulas). It will exclude, however, many uses of modern data-driven engineering design that yield improved industrial production systems—such solutions typically are regarded as not sufficiently novel to be included in R&D. Investments in modern engineering design are covered in the intangible framework via line 6, and also line 9, which includes investments that re-engineer in-house computer systems and computer network platforms to make use of cloud infrastructure services, data analytic services, and data.

The intangibles framework thus covers most, if not all, forms of data intelligence as virtually all assets in the intangibles framework are potentially data driven. The perspective offered by the framework also informs the development of empirical estimates of data intelligence. Other approaches that adopt a value chain approach to measure data assets have missed some key application areas of modern data science. For example, the Statistics Canada (2019a, b)

[9] National accounts of most countries do not publish databases as a unique asset category. The combined "software and databases" measure, however, covers investments in digital tools used to create data assets.

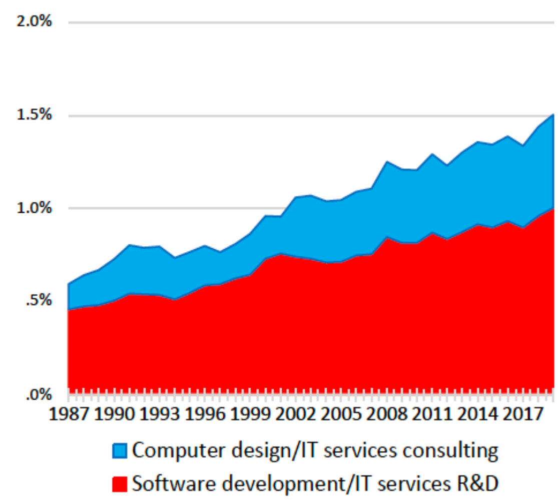
implementation covered financial and marketing forms of data-derived intelligence but did not include data-driven industrial and computing engineering design.

A meta-analysis of the joint evolution of engineering design (ED) and data science reports that, although ED is recognized as a key element of the innovation process at-large, only in recent years has data-driven engineering design become more prominent due to developments in AI (Chiarello, Belingheri, and Fantoni 2021). The emergence of digital platforms that use bigdata to design cost-efficient routes/processes for manufacturing parts production is a related development (Mandel 2019). Underlying factors affecting the increased accompaniment of ED with data include increased competition and digitization of manufacturing, coupled with new methodologies to collect data on product characteristics, product performance and customer requirements.

That modern data systems are AI-based and hosted in the cloud suggests looking at elements of intangible capital that may be picking up the increased use of data. This is provided in figure 4, which shows two series that arguably capture the data-driven demand for cloud services (Byrne and Corrado 2017). The two series are business R&D in IT services and software development and purchases of computer and network design consulting services; these are underlying components of the R&D and business process investment intangible investment categories listed on table 1 (lines 3 and 9, respectively).

Figure 4 Intangible Investment Components Related to Data, Cloud Services, and AI

Investment as percent of GVA, private industries, 1987 to 2019



Source: Authors' update of Byrne and Corrado (2017).

As may be seen, these data driven components of intangible investment have grown enormously, nearly tripling relative to private sector GDP over the period shown. This share relative to total GDP is 1.2 percent in 2018, which would *not* include public funding for AI research, suggesting that the true contribution of AI software research to total GDP is higher.

In summary, beyond the main message of this section that data capital is largely covered in intangible capital, key findings regarding the measurement of data capital are as follows:

- Data value creation involves the generation of data assets--data stores, databases, and data intelligence. This is in addition to the design and production of the digital tools used to create them.

- Data intelligence is the most valuable, and final, stage of the data value chain as it pertains to investments in modern digital business practices and engineering design.
- Data intelligence has many forms—operations, marketing, engineering, and scientific—and not all forms have been included in measurement schemes of previous works.
- Data stores, purchased databases, and most forms of data intelligence are not captured in official statistics.

3. Data Capital and Innovation

If firms are to use data assets, the capital must be produced, and its owners rewarded. GDP is designed to capture market production in a society, so how does data value creation fit into GDP?

An economic model[10]

Activity in the economy consists of (a) an "upstream" "innovation" or "commercialization" sector and (b) a "downstream" or "production" sector that uses the knowledge generated by the upstream sector to produce final output.[11] The upstream sector produces new ideas that can be monetized, e.g., a new system for organizing production or a software program adapted to the needs of the organization; upstream sector output each period is denoted N and its value $P^N N$.

The outstanding stock of commercial knowledge R reflects the accumulation of the upstream sector's supply, after adjusting for losses due to aging, and the income earned by the owners R is $P^R R$. The stock R consists of multiple assets, each type denoted a, and the filtering of time series for the real supply of new knowledge of type a at time t, $N_{a,t}$, into stocks is assumed to follow a perpetual inventory relationship, $R_{a,t} = N_{a,t} + (1 - \delta_a^R)R_{a,t-1}$, a calculation that assumes depreciation of each asset is geometric and constant across all vintages of the asset.[12] As explained in a series of works by the authors, the depreciation rate for intangibles reflects the "service life" of their income generating capacity.

The commercial knowledge stock is non-rival and appropriable, but appropriability is partial, i.e., lasting only for the time during which the producer/innovator can sell or rent it at a

[10] Based on Corrado, Hulten, Sichel (2005, 2009) as adapted and termed "upstream/downstream" in Corrado, Haskel, and Goodridge (2011). The upstream/downstream model was initially set out for the analysis of innovation and intangible capital; its applicability to the analysis of data capital does not require further adaption.

[11] "Final" output is output for sale to consumers or for investment, i.e., for simplicity we ignore intermediate inputs and assume a closed economy.

[12] It also assumes that once each $R_{a,t}$ for an industry is obtained, the usual procedures for aggregating over assets and industries apply. For a discussion of the determinants of the longevity of individual assets and the applicability of the perpetual inventory method to sums of spending streams by firms that compete against one another, see Corrado, Haskel, Jona-Lasinio, and Iommi (2022).

monopoly price to the downstream sector, which is a price-taker for knowledge. That innovators hold temporary product market power is a common feature of economic models of innovation. We write the production and income flows in this economy as:

(1a) $$N = A^N F^N(L^N, K^N, R^{Basic}); \quad P^N N = P^L L^N + P^K K^N + \pi^N$$

(1b) $$Y = A^Y F^Y(L^Y, K^Y, R); \quad P^Y Y = P^L L^Y + P^K K^Y + P^R R$$

where π^N is the upstream sectors' pure rents from innovation, rents that are then embedded in P^N and P^R.

The downstream sector is assumed to be competitive, i.e., product prices P^Y are competitive prices (given payments for the use of new commercial knowledge $P^R R$) and P^L and P^K are competitive factor prices for labor and capital unit inputs, respectively. Basic scientific knowledge, generated via public funds for basic research performed at universities, say, is represented as the input R^{Basic} in the upstream production function (1a). Though basic knowledge plays a role in the production of N, as seen to the left in (1a), there are no factor payments to R^{Basic} because its services are assumed to be freely available.

This model's depiction of the two sectors, though stylized, captures business innovation in modern economies in some important ways. The upstream sector may be considered as firms that are almost fully reliant on the production of innovations in the form of data capital, e.g., biotech startups using massive data experiments to produce new formulas for drugs, with the downstream sector comprising producers that acquire the use of the innovations via outright purchase ($P^N N$) or licensure agreements with annual payments ($P^R R$). Firms may also have their own "innovation labs" and "business strategy teams" that produce and commercialize new ideas for downstream production. These innovation labs and strategy teams are then upstream knowledge producers residing within larger organizations with $P^R R$ representing the contribution of these "factories within a factory" to total firm revenue. For example, many banks have teams of software writers developing software to run, for example, mobile banking apps. Such a team sits within the bank, where note R^{BASIC} would capture the fact that many banking apps use, in part, open-source software.

The asset price of commercial knowledge P^N and the price of its services for a year P^R are linked via the Jorgenson (1963) user cost expression in this model. The upstream/downstream model is then closed via arbitrage of after-tax returns to investments in innovation (that build data/intangible capital R) with returns to alternative long-term investments (that build tangible capital K). This arbitrage also operates as an intertemporal constraint, implying the existence of "abnormal" innovator (i.e., firm-level) profits for periods of time but zero profits (i.e., π^N = 0) in long-term equilibrium.

Data capital in GDP and growth accounting

Without the capitalization of intangibles, GDP consists solely of downstream sector output Y, but when investments in data capital/innovation are capitalized, aggregate value added Q reflects production in both sectors:

$$P^Q Q = P^Y Y + P^N N = P^C C + P^I I + P^N N \tag{2}$$
$$\equiv P^L L + P^K K + P^R R$$

As seen in the first line to the right, investment in final demand is expanded to include data value creation, i.e., spending on intangibles is no longer treated as intermediate expenditure and thus GDP is larger, a first order impact of capitalizing spending on intangibles. Factor income, the second line, accounts explicitly for returns to data and other intangible assets in total capital income. The term may contain monopolistic returns to innovation as discussed above.

When Solow's sources-of-growth decomposition is applied to GDP with investment expanded to cover data value creation, the usual log differentiation cum constant returns yields

$$dq = \sigma_Q^X dx + \sigma_Q^R dr + da \tag{3}$$

for output growth where σ_Q^X is the combined factor income share for conventional inputs L and K in total production and σ_Q^R is the factor income share attributed to owners of intangible capital.[13] This decomposition says that output growth consists of a contribution from conventional inputs $\sigma_Q^X dx$, a contribution from paid-for, commercially valuable knowledge $\sigma_Q^R dr$, and total factor productivity growth da. In practice, total factor productivity growth is calculated as residual from equations such as (3).

To return to the banking example, suppose a bank has a software writing team that ingests data, and a machine learning team that uses that data to develop improved marketing and credit-scoring processes. The use of open-source software is captured by *da*. The commercially written software and improved marketing and credit-scoring, if they are appropriable to a particular bank, is captured by $\sigma_Q^R dr$. When such knowledge diffuses then it becomes part of *da*. If there are economies of scale/scope in the joint intra-firm use of data, software, and analysis, this is in *da* because capital factor shares in (3) (and usual conventions) are based on constant returns.

[13] The notation used in equation (2) is as follows: "dz" the log change in "Z" where Z is any variable in the model. Conventional inputs K and L combined as X are weighted appropriately.

What is different in the model with intangibles then—and its major implication—is that the contribution of paid-for, commercially valuable knowledge $\sigma_Q^R dr$ becomes an accountable source of growth. And because this knowledge is nonrival and only partially appropriable, intangible capital is also determinant of measured da in (3) via the increasing returns mechanism featured in modern growth theory (Romer 1990, Jones 2005). This feature will be apparent in measured da to the extent that the innovations embodied in current and past vintages of R diffuse across firms and industries in an economy, e.g., via patent expiration, loss of first-mover advantage, etc., as R is filtered through an economy's innovation ecosystem (Nelson and Rosenberg 1993; Moore 1993).[14]

If the ecosystem's filtering process shuts down (for whatever reason), this would imply an enhanced earning potential of investments in R for investor/owners. The direct effects of this as they pertain to productivity statistics would be that (a) data capital would have a relatively longer service life compared with "traditional" intangible assets, and that (b) measured growth of total factor productivity would be lower due to fewer spillovers.

Interpretation of Innovation

When considering innovation, economists typically reach for TFP as a measure of underlying technical progress. It seems clear that TFP as a production function "shifter" is capturing innovation, being a residual after subtracting share-weighted paid-for inputs from output, but it is rather hard to talk meaningfully about a residual with others interested in innovation. Innovation analysts typically focus on how firms innovate (e.g., develop a new business model) and consider the resource cost necessary to bring about new products and change—aspects of innovation consistent with the intangible capital approach that sets out these activities as sources of growth.

Whether this connection helps depends on one's definition of innovation. In his evidence to the Gutierrez commission (Schramm et al. 2008), Dale Jorgenson explained growth by stressing innovation versus duplication. Consider this by asking, how might the firm Peloton make more sales? One way would be to employ more K and L to produce more bikes and treadmills, i.e., growth via *duplication*. The other path would be to get more sales from existing K and L: mixing new exercise music, developing new software, re-engineering the supply process. Jorgenson called this growth via *innovation*. The intangible capital framework gives this a natural interpretation: Innovation is output less the contribution of K and L, suggesting that innovation reflects the final two terms in (3) or that:

(4) $$Innovation = \sigma_Q^R dr + da$$

[14] For a recent review, articulation, and examples of innovation ecosystems, see Granstrand and Holgersson 2020).

This implies that when considering innovation, residually calculated TFP should not be the sole focal point of analysis. Rather an understanding of the intangible resources being deployed at firms, be they paid-for or "borrowed", enables links to literatures that analyze business practices, trends in entrepreneurship, and the like.

Data capital and intangible asset prices

As previously indicated, modern data use via the cloud fosters faster, more efficient experimentation and feedback in many business functions (R&D processes, industrial production processes, marketing research, etc.) and can be thought of an innovation in the method of innovation. How does this affect the workings of the upstream/downstream model?

Consider first the upstream innovation sector. Solving for the log change in intangible asset prices dp^N using the production and factor payment equations in (1a) yields

$$dp^N = \sigma_N^L dp^L + \sigma_N^K dp^K + \sigma_N^\pi d\pi^N - da^N \, , \tag{5}$$

which expresses dp^N as a weighted average of changes in input costs (the first two terms), plus changes in innovator profits, offset by changes in the efficiency of upstream production.

When input costs are expressed in "transactions units", changes in factor input quality are an offset to costs. Consider labor input. In the upstream sector L^N is an aggregate of services provided by a range of worker types, which implies that L^N differs from aggregate hourly input H^N (its "transactions unit" equivalent) by a composition effect, Θ^N, that accounts for differences in the marginal productivity of different worker types employed in production, i.e., we have, $L^N = \Theta^N H^N$. This implies that if we ignore capital inputs—attributing the impact of the ease with data can be processed and utilized via cloud services to (redefined) upstream TFP—equation (5) can be rewritten as:

$$dp^N = \sigma_N^L (dw - d\theta^N) + \sigma_N^\pi d\pi^N - da'^N \tag{5'}$$

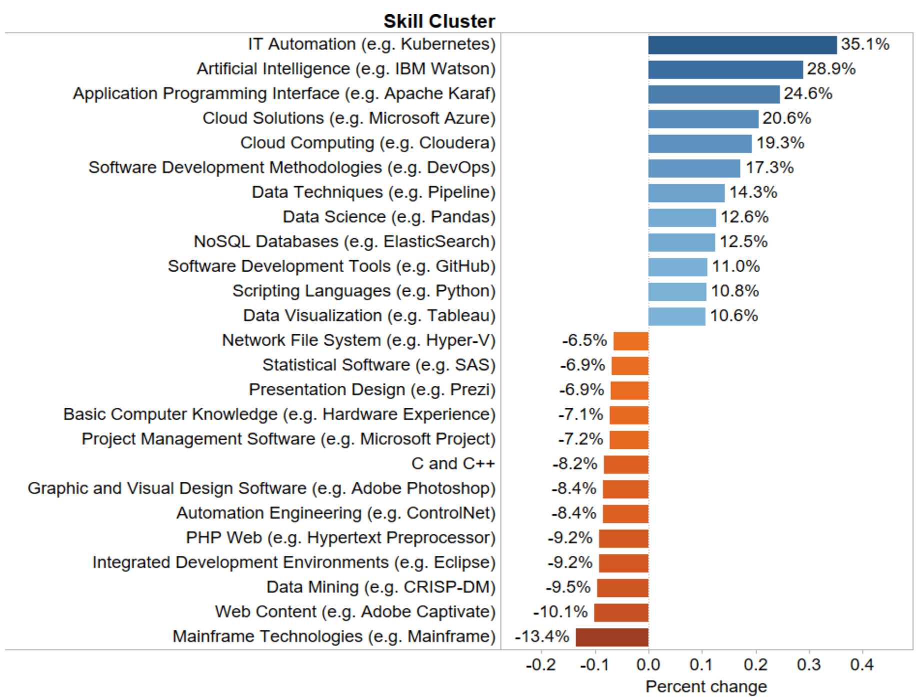
where the impact of *upstream* labor composition changes on data/intangible asset prices is explicit. Note that upstream labor composition effects are likely to reflect moves *within* the usual grouping of workers termed "high-skilled" and may not be accounted for in measures of labor composition used in growth accounting, which are developed using broad groupings of employment by worker type. All told then, equation (5) suggests that the data intensity of intangible capital will have, first, a restraining impact on price change for intangible assets through improvements in data asset production efficiency (da'^N), and second, that upgrades to the productivity/skill base of upstream workers ($d\theta^N$) are likely to offset associated requisite increases in wages paid.

Working against the restraint exerted by data technologies via increased data use on the full passthrough of costs to dp^N is the contribution of markups, expressed here as changes in innovator rents ($\sigma_N^{\pi} d\pi^N$). Measuring this term directly is highly problematic: markups may be transitory, and they are hidden via the typical exclusion of intangible assets from company financial accounts.

Direct evidence that employer demand for skills related to automation, AI, data connectivity, and cloud storage/computing is reshaping IT work—thus boosting $d\theta^N$—is suggested by figure 5, which shows that the demand for these skills accelerated the fastest among IT roles during the pandemic (figure 5), consistent with increased data use offsetting wage cost pressures on asset prices for data capital.

Figure 5. Emerging skill clusters including Artificial Intelligence and Cloud Solutions relative to other tech occupations

Percent change in the share of selected skill cluster mentions in job ads for tech occupations from 2019 to the last 12 months ending in February 2021



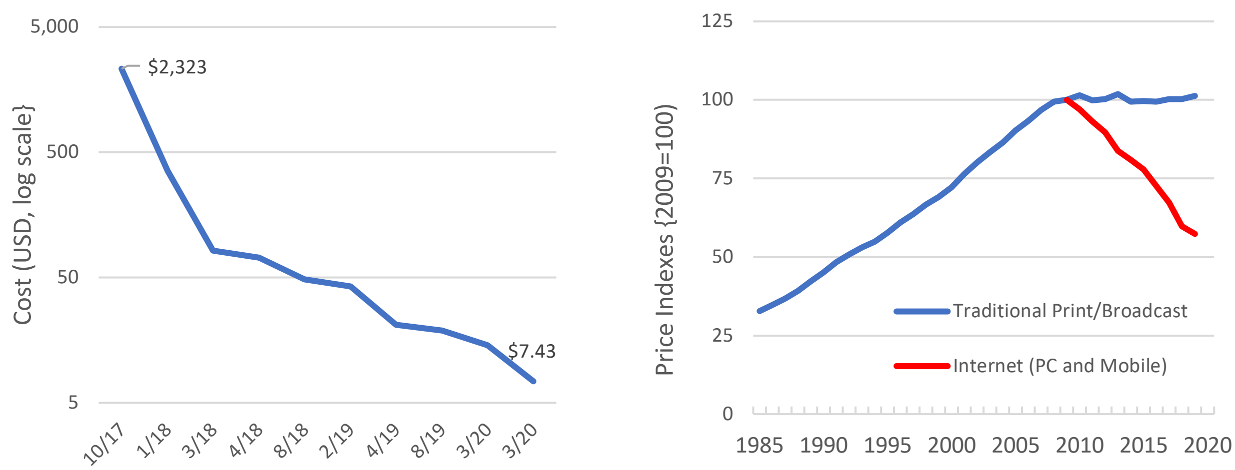
Source: The Conference Board®-Burning Glass® Help Wanted OnLine® (HWOL) data series

Many studies document improvements in the efficiency of modern data systems ability to ingest, store, process and analyze large quantities of data (e.g., Byrne, Corrado, and Sichel 2021, Coyle and Nguyen 2018). The findings are consistent with a strong impetus to productivity growth emanating from da'^N, an effect that will show through in productivity estimates only insofar as asset prices for data capital capture the impacts of these changes in data processing costs. Hard-to-measure services price research typically does not address intangible asset-producing activities—R&D labs, marketing teams, engineering design projects—nor are these activities viewed as hotbeds of rapid quality change missed by price collectors in assessments of productivity mismeasurement. But with the digital transformation of economies, rise of digitally enabled business models, and increased use of data in business more generally, the nature and efficiency of intangible asset-producing activities is arguably driving down costs and effective prices of these activities.

How much are these cost efficiencies? In terms of training a contemporary image recognition system, the answer, according to tests shown in the AI Index Report (2021, page 49), is "a few dollars in 2020, down by around 150 times from costs in 2017" (figure 6a). This representing progress in both algorithm design and drops in the costs of cloud-computing resources. Similar factors in conjunction with accumulating data on consumer buying patterns and tastes have affected (directly and indirectly) the advertising media costs of marketing (figure 6b).[15]

Figure 6 Data-driven cost efficiencies affecting intangible asset prices

(a) Training cost of image recognition *(b) Advertising media costs*



Data capital and "potential" labor productivity growth

Though we are still in the early stages of pinning down data capital in macroeconomic statistics, the foregoing suggests that the contribution of data capital to potential labor productivity growth can be calibrated using assumptions for data capital income shares and asset price change, σ_Q^R and dp^N. The utility of such a calibration sets the stage for the empirical sections of

[15] For further discussion of the how AI "fits" into the intangible framework, see Corrado, Haskel, and Jona-Lasinio (2021). The media cost price indexes are developed from detailed BLS input cost indexes aggregated using information from the Census Bureau and industry sources; see Corrado (2021) for further discussion.

this paper where we address how data capital can be measured and how much of its potential actually can be "seen" in the available statistics.

The long-term growth-promoting potential of a capital input depends on the extent to which its volume rises more rapidly than its relative price falls (i.e., that the input shares continue to rise). Though this is ultimately an empirical question about the degree of substitutability between data/AI and human efforts, the limits to which are discussed in Nordhaus (2021), section 2 suggests we can approach a calibration of data capital's potential impact on labor productivity using measured input shares for intangible capital as a guide. Ignoring spillovers, the impacts of data capital can then be calibrated using estimates of two effects, a "use" effect determined by input cost shares and relative prices, and a "production" effect determined by production shares and relative prices.

This approach has been a mainstay of productivity analysis with IT capital. As business services derived from IT equipment have shifted to the cloud, however, domestic production effects (via IT services production) have become more pronounced in calibrations of IT impacts on an economy (Byrne and Corrado 2017). When thinking about data capital, production effects are also likely to loom large because much production of data capital occurs within using firms.[16]

The results of the calibration exercise are reported in table 3, which shows alternative scenarios for the productivity-enhancing impact of data capital.[17] The scenarios vary according to the breadth of investments in data capital in an economy (broad or limited diffusion), the extent to which data assets are domestically produced, and the productivity advantage of data assets and data technologies (based on their relative price). The capital income share of data capital captures diffusion via use and is assumed to be less than the corresponding total intangible capital income; the ranges used in the table are based on actual shares in high vs low intangibles-intensive countries. The production share is assumed to be the capital income share +/- 10 percent, roughly the range for net exports of corresponding intangible investment services in high vs low intangibles-intensive countries.

---

[16] The available estimates for U.S. total intangibles suggest that about one-half are produced for use with the same organization (Corrado 2021).

[17] The calculations are based on the steady-state solution to the two-sector upstream/downstream model consisting of a data capital producing sector and all other goods and services producing sector. In this model, the contribution of the data sector to labor productivity equals the sum of the use effect, ${\bar{\sigma}^R}/{\bar{\sigma}^L}\ (-\dot{p})$ plus the production effect, $\bar{\omega}^D(-\dot{p})$, where $\bar{\sigma}^R$ and $\bar{\sigma}^L$ are steady state income shares of data capital and labor, $\bar{\omega}^D$ is the production share of data investments, and $(-\dot{p})$ is the relative productivity of data capital measured as the rate of decline in the *relative* price of data assets (sign reversed, i.e., - $(dp^N - dp^Y)$) using upstream/downstream notation. The calculations in the table assume labor's share of total income equals .7. For a derivation, see Oulton (2012) or Byrne and Corrado (2017).

Table 3. Productivity Scenarios: Contribution of data capital to potential labor productivity growth (percentage points)

	Productivity advantage (relative asset price growth differential) Narrow edge 1 percentage point	Large edge 5 percentage points
Broad use (and net exporter of data services) 10 percent capital income share 11 percent production share	0.25	1.26
Limited use (and net importer of data services) 5 percent capital income share 4.5 percent production share	0.12	0.58

Note: Contributions include the sum of the use and production effects of data capital.

The lower bound for the productivity advantage is drawn from recent evidence on the relative price differential implied by an intangible investment price index designed to capture the impacts of digitization on investments in brand and the IT consulting and marketing subcomponents of organizational capital, about 1 percentage point per year (Corrado 2021). This deflator takes many national accounts prices as given, notably asset prices for R&D, software and gross output deflators for industrial design and training, and thus is a lower bound in that these deflators do not incorporate efficiency gains due to increased application of AI methods or use of open-source content. The upper bound is guided by the long-term relative price decline of conventionally defined IT capital of about 15 percent per year (based on the estimates reported in Byrne and Corrado, 2017). It is conservatively set at one third of that, i.e., 5 percentage points per year.

All told, estimates of the contribution of data capital to labor productivity growth range by more than a factor of 10—from 0.12 percentage point per year to 1.26 percentage point per year. The range highlights the synergies among data capital efficiency and an economy's breadth of use and capability for digital transformation, which implies much scope for policies to affect outcomes. Promoting diffusion through the use effect (i.e., encouraging both data investments and data sharing) is very important, and a typical focus of traditional IT policies.

The table further implies that the course of data productivity is "doubly" important, operating as it does through both the use and production channels given that much data value creation occurs within firms and international trade in data assets and data asset services remains limited. Facilitating the creation of new scope economies within firms (more data-driven business functions) and new data-enabled firms will boost this potential, as would the development of well-functioning markets for data assets and data asset services.

4. Measuring Data

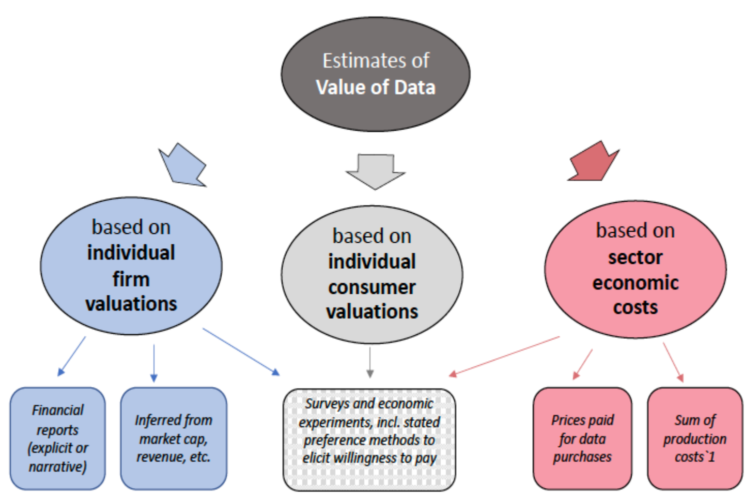
How much value do firms derive from data? And how is this related to the value of personal information or to costs sustained by firms to obtain the data that are used and/or created via the data stack? In this section we review methods and approaches that have been used to measure data and then present our own application.

Part 1: Methods and Approaches

In addressing these questions, one encounters different perspectives and different measurement approaches to the valuation of data. The economics literature has taken three main directions to develop estimates of the value of data. As depicted in the middle panel of figure 7, these include approaches based on individual firm valuations, approaches based on consumers' valuations, and approaches based on sector (and/or industry) economic costs. The bottom panel of the figure indicates methods used under each approach. Surveys and economic experiments (the middle box in the bottom panel) are of course methods that are not unique to a given approach, as the figure indicates.

Approaches aimed at valuing consumers' personal information will not encompass the full data value chain of figure 2, which covers the production of digitized information for all sectors of an economy. Our review of methods is targeted at methods that can yield comprehensive coverage of data use in market activities in economies, and we thus proceed as follows: We first discuss methods that have been to estimate the value of data for individual firms and/or based on individual firm-level data. That is followed a discussion of stated preference methods applied to measure the value of data used in business—a little appreciated niche in this literature, though long used by market researchers to study consumer preferences and recently employed by Brynjolfsson, Collis, and Eggars (2019) to estimate the value of free digital goods.

Figure 7. Approaches to the Valuation of Data



We then briefly summarize the gist of the sector cost approach as deployed by national accountants. This approach, also called the sum-of-costs approach, is used to develop the experimental estimates of data reported in part 2 of this section.

Methods Based on Firm Valuations

Below we review data valuation approaches used and/or emerging in financial reporting, followed by a review of methods used in key studies. These studies provide essential insights on measuring the value of data, even if their methods cannot be readily adapted to compile macroeconomic statistics sufficiently comprehensive to inform economic policy analysis.

- Business reporting

There is a growing consensus in the business literature that building a framework to discover and realize the potential of data is critical for increasing the value provided to shareholders (Deloitte 2020 and PWC 2019). The starting point for designing a data strategy is to assign a value to data as an asset, which requires i) completing an inventory of current data assets; ii) identifying how the organization is currently utilizing them and their possible alternative uses; iii) selecting a valuation method.

Most of the approaches adopted for valuing data in the business context consist of an implementation of the three traditional valuation methods used to value any asset type: income, market, and cost approaches. The income methodology measures the incremental cash flows (increased revenues and/or reduced costs) that the data are expected to generate in the future. The market approach captures the value of a given data asset using the information about the value of a comparable data asset whose value is observable in an active market or transaction. The cost approach estimates the value as the cost for recreating a replica of the data or replicating the data's utility.

The growing importance of intangibles in corporate activity and the evidence that they do not fit very well in the current financial reporting has generated a debate among the accounting community about the opportunity to deliver more information on intangibles promoting its disclosure of financial reporting (see, for instance, UK Financial Reporting Council 2019 and the assessment by CPA Ontario 2022) or by capitalizing intangibles as assets in balance sheets (ACCA 2019, Lev 2019). The UK Financial Reporting Council (2019) proposes two ways to get more information on intangibles in financial reporting. One is to revise the statement of profit or loss to provide information on expenditure on future-oriented intangibles, analyzed by nature. The other is the provision of more details on intangibles in the narrative sections of financial reporting.

The first option is more beneficial for compiling business statistics and for economic analysis based on firm-level data. First, it would facilitate gathering information via business surveys. Based on current financial reporting standards, respondents to business surveys would typically be unable to identify expenditures for data and several other intangibles separately. Second, improved and more comprehensive disclosure of spending on intangibles (in addition to the value of existing stocks) would be consistent with the needs of national accounts compilers of collecting information on outlays (not on the value of the assets). Finally, more precise information on expenditure for intangibles and data would be available to firm-level data users.

- Revenue-based approaches

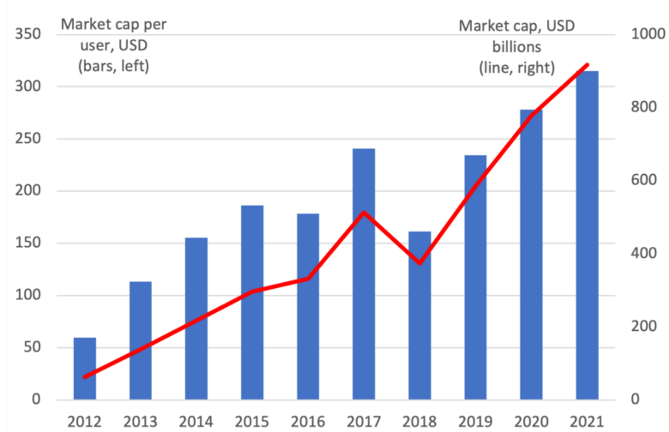
Another interesting approach suggested by Nguyen and Paczos (2020) aims at capturing the value of data based on the revenue shares driven by data monetization across different types of firms (e.g., manufacturers, utility providers, banks, or online platforms). Nguyen and Paczos (2020) adopt a stylized taxonomy of business models distinguishing two main categories: data-enhanced or data-enabled. Their assumption is that by looking at the business models adopted in different productive sectors it is possible to identify specific characteristics from which to infer a general measure of the value of data at the industry level. This approach can be easily implemented even if it requires additional efforts from national statistical institutes to conduct ad-hoc economic surveys and coordinate internationally to guarantee comparable results across countries.

- Depreciation-based approach

Coyle and Li (2021) develop a demand-side methodology for estimating the size of data markets using the recent finding that an online platform's entry can disrupt incumbent firms' organizational capital by affecting its depreciation rate (Li and Chi 2021). They calculate the stocks of organizational capital based on before-entry and after-entry depreciation rates. This difference captures the loss due to the failure of using data to cope with changes in competition due to the entry of an online platform. Thus, it can be used to measure the potential size of the demand for data by incumbent firms in the industry sectors disrupted by online platforms. In other words, they use the loss of the value incumbent firms' organizational capital to measure firms' maximum willingness to pay for the access to data. Coyle and Li (2021) apply their model to study the impact of the entry of Airbnb on existing firms in the hospitality industry. They find that the market size for data in the global hospitality sector is USD 43 billion in 2018 and that this data market has also grown rapidly at an average growth rate of 35%, meaning that its size has been doubling in under three years.

Consistent with the existing literature on measuring intangible capital from firm-level data, Coyle and Li (2021) use the selling, general, and administrative (SG&A) expenses as a proxy for a firm's investment in organizational capital. This includes expenditures for employee training costs, brand enhancement activities, consulting fees, and supply chains' installation and management costs, thus covering the economic competencies category in the list of intangibles as set out in table 2. On this basis, they estimate the value of data considering the extent to which online platform entry can disrupt incumbent firms' economic competencies assets.

Figure 7. Facebook/Meta Market Cap per Active User, 2012-2021



Sources: Number of FB monthly active users from Statistica, Q4 of each year (Q3 for 2021). Market capitalization for Facebook/Meta is Nasdaq value for December 11 of each year.

- Market prices

Market prices paid and received in actual transactions are the best proxy for quantifying the value of data. However, adopting this approach faces many obstacles. First, there is no well-defined market for many types of data, and, when available, transaction-based valuations may stem from obsolete information. Second, as the value of data is highly context-dependent, the same dataset might be valued differently across different data suppliers, users, and regulators (Nguyen and Paczos 2020). Finally, market transactions in unprocessed data would only capture the input data and not the entire transformation chain necessary to generate digitized information (Reinsdorf and Ribarsky 2019).

Large-scale market transactions typically exist primarily for third-party data produced by data brokerage or data aggregator companies. These companies usually collect information from publicly available personal records and then aggregate, store, and sell it to different customers through licensing subscriptions or contractual arrangements. As third-party data is widely accessible, they are less valued than first and, to a lesser extent, second-party data (Reinsdorf and Ribarsky 2019).

It is also illustrative to examine financial indicators per record from companies that derive most (or all) of their income from advertising linked to personal data, e.g., Facebook/Meta. Figure 7 shows that the value of an individual (active) record currently is more than 300 USD. The firm's valuation is approaching 1 trillion USD. Ahmad, Ribarsky, and Reinsdorf

(2017) calculate a value equivalent to around 0.02 percent of global GDP for the user data collected by five major digital services (Facebook, Twitter, Instagram, LinkedIn, and Gmail) based on the number of active users and assumed prices of a user profile.

Stated Preference Methods

Some studies have provided estimates of data value using stated preference methods (including contingent valuation, conjoint analysis, and discrete choice analysis). This approach asks surveys participants to directly report their willingness to pay (WTP) to obtain a specified good or willingness to accept (WTA) to give up a good. The value of a non-market good or service is the amount that users are “willing to pay” for it, or “willing to accept” in return for not having it. Contingent valuation methods are widely used to understand consumer valuations and preferences in contexts with no monetary prices, such as environmental or cultural goods (see, e.g., Carson, Flores, and Meade, 2001 and McFadden and Train, 2017 for surveys).

An excellent example to illustrate the use of stated preference methods for business valuation of data is the case of Landsat. The Landsat program consists of a series of Earth-observing satellite missions jointly managed by NASA and the US Geological Survey. Landsat data products are processed and made available for download to all users at no cost. Based on surveys of data users, Miller et al. (2013) estimated the economic benefit of Landsat data for the year 2011 to be $1.79 billion for US users and $400 million for international users. The annual benefit to US users is two times greater than the cost of building and launching Landsat-8, the Landsat satellite launched in 2013 and still operating.

From a different perspective, a growing literature relies on stated preferences methods to study the monetary valuation of privacy. Prince and Wallsten (2020) conducted a discrete choice survey across six countries: the United States, Mexico, Brazil, Colombia, Argentina, and Germany. They find that Germany places the highest value on privacy compared to the US and Latin American countries. Across countries, people place the highest value on keeping financial and biometric information private.

Stated preference methods are also used to assess the value of public information assets, e.g., official statistics. The United Nations Economic Commission for Europe (UNECE 2018) has called on national statistical agencies to develop approaches to calculate the monetary value of official statistics, which cannot be measured using market prices as many official statistics datasets are accessible under public license with no monetary price. UNECE (2018) recommends various possible valuation methods, including using the stated preference method and reports that it was used to explore the economic value of the UK Economic and Social Data Service (ESDS). ESDS is a distributed service that aims to promote the broader and more informed use of data for research and teaching in social sciences. In the study, respondents were asked to express their willingness to pay in terms of an annual (subscription) fee and on a pay-per-access basis. This resulted in an estimated willingness-to-pay of around £25 million per annum among the survey population.

Sum-of-costs approach

National accounts estimate investment by asset type based on a sum-of-costs approach. Though the approach differs substantially in context and application from the cost-based valuation method used in financial accounting, the concepts do overlap. National accounts aim at consistently recording investment flows and capital stocks every year and doing so involves estimating values for all sources of supply for each asset and deriving the asset valuations and quantities using information on price change in newly produced assets and information on the rate at which an asset's value declines as it ages.

If firms purchased all or most data from market transactions, as they do with tangible assets, measuring the cost of data would be like measuring expenditures for a construction firm's purchase of excavators and concrete mixers. Instead, most digitized information used by businesses (and other intangibles such as software and R&D as well) is not transacted on markets but produced in-house. Thus, national accounts compilers must come up with two components for intangible investments, own-account investment (when data are produced and used in-house) and purchased investment (when data are bought and sold in market transactions), for measuring nominal investment flows in data assets. Consider now how each component might be estimated.

- In-house production: the "factory within a factory"

  The approach to in-house production is as follows: Imagine a firm having a "software factory" or "R&D factory" inside it—and the task at hand is to estimate the gross output of this hypothetical factory based on the market value of the payments made to factors employed by it (labor, capital, and intermediates). In practice, the key to accomplishing this task is to identify the occupations of workers employed in the factory and to estimate their compensation. Based on knowledge of the compensation paid to these workers, the total payments made to all factors involved in the in-house production is then estimated (i.e., capital and intermediate costs are added to labor costs). The identified workers may not be involved in producing new assets their entire workday; for example, the conventional approach to measuring in-house software production in national accounts is to assume that software developers spend just 50 percent of their time working in their firm's "software factory" to produce original code. In-house production of data assets is estimated in a similar fashion.

The SNA explicitly recommends that national statistical offices use the sum-of-costs approach to estimate software and databases (unless produced for sale) and R&D (unless the market value of the R&D is observed directly) and the own-account component of any product for which it is not possible to find the price of a similar product. The INTAN-Invest database uses a sum-of-costs approach to estimate the own account component of non-national accounts intangibles.

- Purchased data assets

Purchased data should be valued at the transaction price. Although conceptually simple, measuring the purchased component of data investment is challenging because comprehensive data sources are scant. All told, information about the expenditures on data usually is missing in surveys of production or capital spending, and the national accountant's total supply approach is difficult to implement. Ker and Mazzini (2020) considered business statistics sources and looked at the revenues generated by firms that create explicit value from data (those collecting, compiling, and selling databases). But they found that focusing mainly on industry classifications is likely to generate an inexact identification of these activities. For example, Zillow sells its data on home real estate valuations, Nielsen sells it survey data, as do credit agencies such as Experian, but these firms are in widely different industries. Also, monetizing databases is not necessarily the primary line of business for many firms who charge for purchased databases or are in the business of producing data intelligence (e.g., Gartner, McKinsey).

This study vs prior studies using sum-of-costs approach

Statistics Canada (2019a, 2019b) prepared experimental estimates of in-house investments in data based on a sum-of-costs approach, counting in effect all production as in-house production. Occupational groups were selected from among those generally associated with converting observations into digital format (the process of digitization). Their estimated values for investments in all three data types ranged from 1-3/4 to 2-1/4 percent of the country's GDP in 2018.

Goodridge at al. (2021) took the same approach and estimated the combined value of software and databases (from national accounts) and other data capital investments for 16 EU countries using essentially the same implementation in terms of occupations covered. Their results suggest that including the Statistics Canada grouping of occupations engaged in producing data stores and data intelligence (which they refer to as data transformation and knowledge creation) raises own-account GFCF by around 60 percent compared to own-account investment in software and databases measured in EU official national accounts.

In the next section, we implement a sum-of-costs approach to estimate in-house production of three types of data assets. Our identification of workers engaged in producing data intelligence yields a broader list of occupations than used in previous works. In line with the intangible capital framework, our estimates of data intelligence include business and marketing strategy and data-driven engineering design. These forms of data intelligence were not included in prior works.

Part 2: Value of Data in Selected European Countries

We implement a cost-based approach based on the value chain framework illustrated in section 2 and report experimental estimates of (domestically-produced) investment in data stores, databases, and data intelligence for the market sector of 9 European, mainly western, economies for the years 2010 to 2018; these countries include Denmark (DK), Germany (DE), Finland (FI), France (FR), Italy (IT), Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK).[18] To the best of our knowledge, these are the first harmonized and internationally comparable measures of data investment produced so far for these countries. Although the estimates are experimental and preliminary, our conceptual approach emphasizes the use of a measurement framework consistent with national accounts for computing time series of investments in data assets, including estimates in volume terms (i.e., adjusted to consider price changes) and data capital stock measures.

Measures of data have been generated using the information on total in-house costs incurred for transforming raw digitized information into data assets by considering the occupation types engaged in producing the three data asset types set out in section 2. A similar exercise is conducted for software. The software estimates are valuable for comparative analysis with national accounts, which combines software and databases, and with intangible investment, which uses national accounts estimates for these assets.

Our measures of data assets and software capture the value produced in the market sector regardless of whether the produced output is intended for own final use or final sale. In what follows, we consider the produced value of data as a good proxy for data capital investment in

[18] We define market sector as all industries excluding NACE sections O (public administration and defense; compulsory social security), P (education), Q (human health and social work activities), T (activities of households as employers; undifferentiated goods - and services - producing activities of households for own use), and the imputed rents of owner-occupied dwellings component section L (real estate activities).

the market sector assuming that data transactions between the government and the market sector are rather small. Strictly speaking, estimating market sector investment would require adjusting the estimates of in-house produced output for data transactions between government and the market sector as well as for imports and exports data flows. As previously indicated, information on transactions in data stores and databases, whether domestic or for international trade, are not apparent in official statistics.

An appendix to this paper details the procedures and data sources used to develop our estimates of the value of data asset production in the 9 European countries listed above.

Results

Our main results regarding the relative size and growth of market sector data asset production and intangible investment are shown in figures 8a–d. Data asset production is shown according to the three segments in the data value chain in figure 8a. The total data value chain averages 5.3 percent relative to market sector GVA in the covered countries and years (2010–2018). The United Kingdom is the most data intensive of the countries included (6.5 percent), and Italy and Spain are the least (3.8 percent). Data intelligence is estimated to be the largest segment in the data value chain.

Figure 8b shows that resources allocated to data asset production in 2018 was rather less than intangible asset production (60 percent less). And figure 8c shows that nominal data production grew just a tad faster than intangibles, and at that, only until the final two years of the analysis. As stressed in the previous section, however, the cost efficiencies enabled by data-driven forms of intangible investment imply that the *real* growth of data assets likely eclipsed that of overall real intangible investment. When we examine results for the information and communications services producing sector (NACE industry sector J, not shown), we find that its data asset production share is a tad higher and grows slightly faster in relation to its intangible investment than comparable statistics for the overall market sector.

Owing to the complementarity between data and software tools, Figure 8b also shows a sum-of-costs estimate of domestic software production, which averages 25 percent of the data value chain plus software. To put this in perspective, figure 8d shows the relationship between the percent change over the sample period for the sum-of-cost estimate of software and databases with the national accounts software investment (which includes own-account production of databases and imports and exports, as previously discussed). As may be seen, production shares expand a bit less than investment shares, and production shares are, on average, less than investment shares (2.6 percent versus 3.1 percent).

Figure 8a Estimates of the Data Value Chain for 9 European Countries, 2010-2018 average

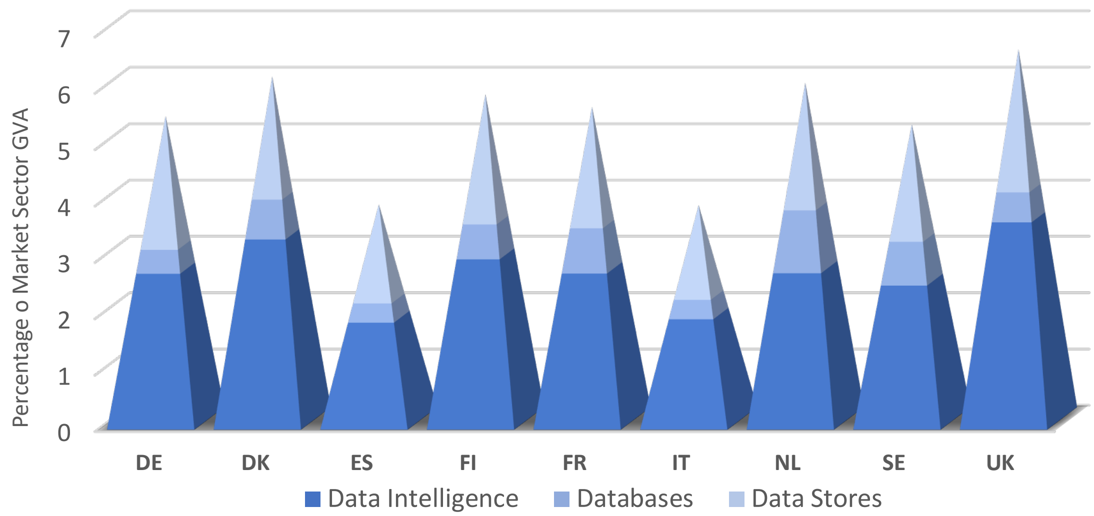


Figure 8b Data and software asset production versus intangibles, 2018

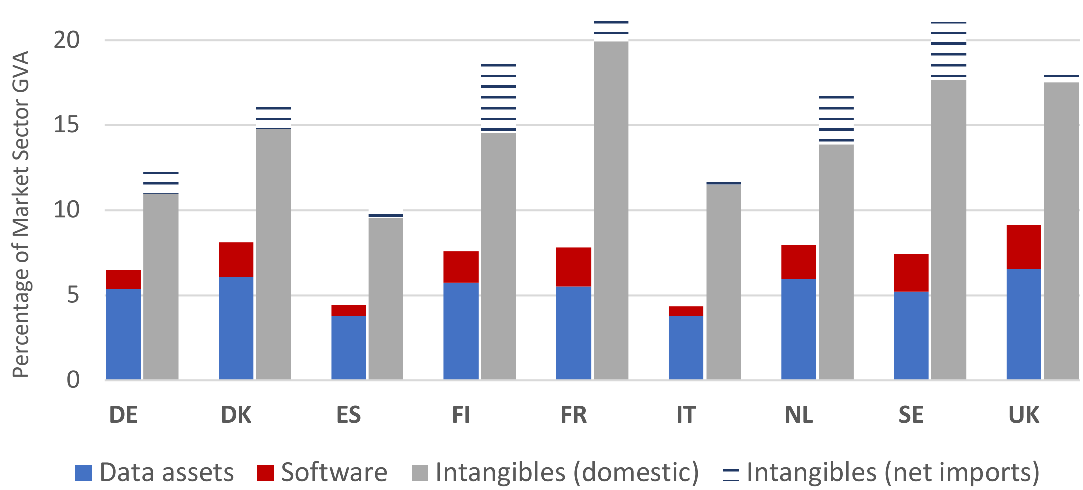


Figure 8c Relative growth (2010=100)

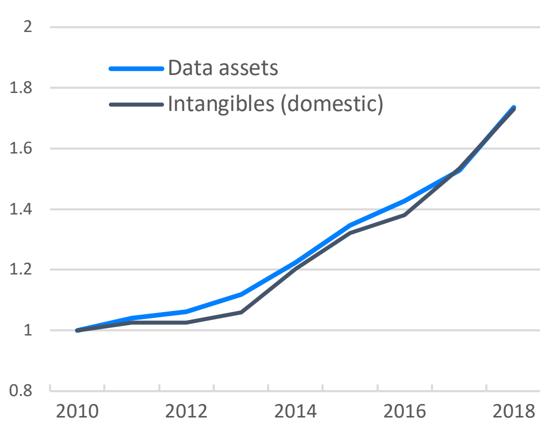


Figure 8d Software & databases, this paper versus national accounts (% chg. GVA share)

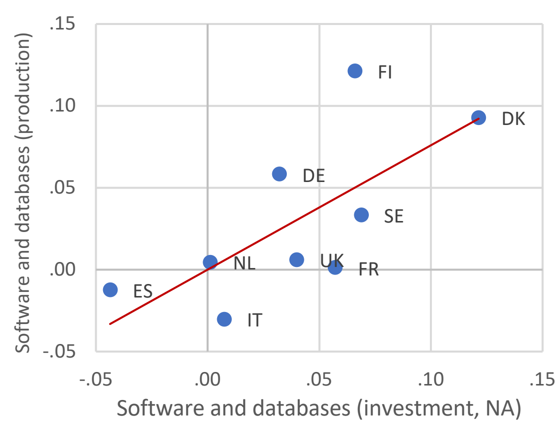
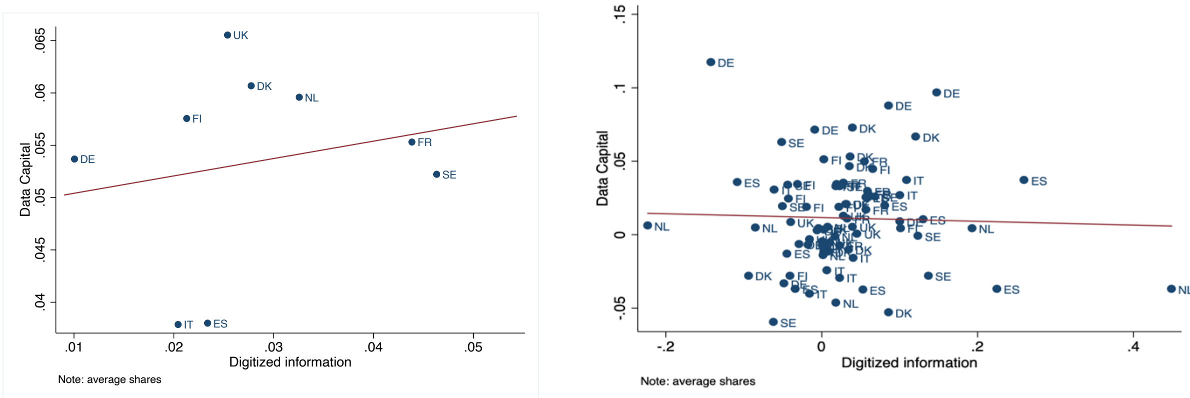
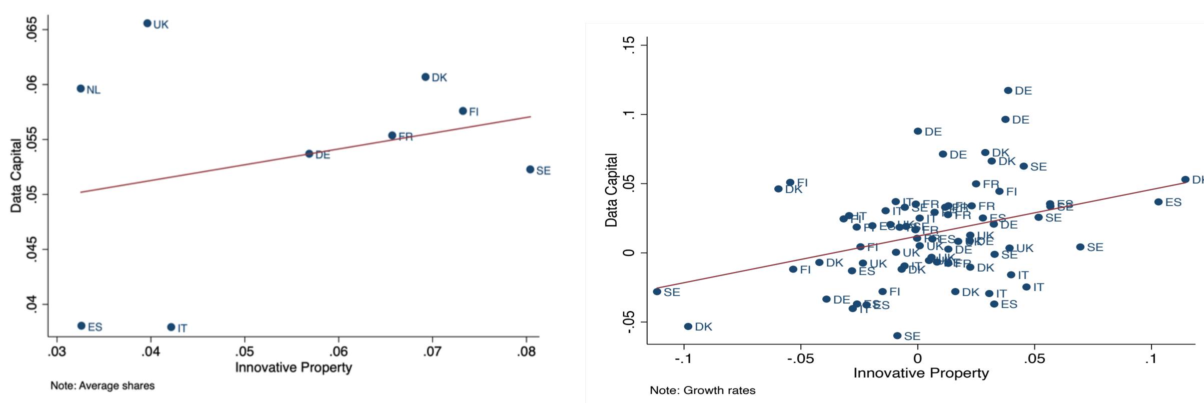


Figure 9 Data Capital Production and Intangible Investment by Broad Category 2010 to 2018 (average shares on the left, growth rate of shares on the right)

9a – Data and Digitized Information



9b – Data and Innovative Property



9c – Data and Economic Competencies

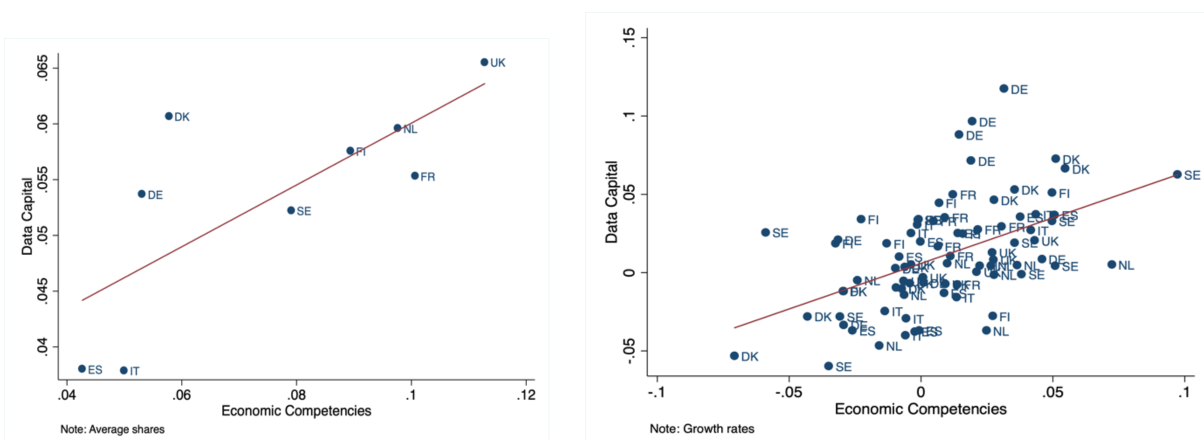


Figure 9 shows empirical results on cross-country relationships between the data value chain and intangible investment components. Data shares of market sector GVA are found to correlated with shares of the three broad groups of intangibles: digitized information, innovative property, and economic competencies (the right panels in the figure). And growth rates of data shares are strongly correlated with investments in the innovative property and economic competencies groups of assets. All told, the correlations are strongest for the economic competencies grouping, as suggested by the discussion of the relationship between the data value chain and intangible capital in section 2.

Tables in the paper's appendix 3 report detailed results of correlations among aggregates and components of the data value chain, our software production sum-of-costs estimate, and detailed components of intangible investment.

5. Data Capital and Total Factor Productivity

To calculate total factor productivity, we use the recently issued EUKLEMS & INTANProd database, which reports productivity data including investment streams for the intangible assets listed in table 2 for most of Europe, as well as for the United States and Japan.[19] The investment and capital estimates for assets not regularly capitalized in national accounts are developed using national accounts-consistent methods, i.e., they are not calibrations of a model or developed from data in company financial reports.[20]

Below we report and analyze estimates of total factor productivity that cover the nine European countries included in the empirical analysis of the data value chain and the United States from 1998 to 2018. The results for Europe are aggregated using production-side purchasing power parities (PPPs) to facilitate comparative analysis with the United States.[21] It should be noted that INTANProd includes estimates of intangible investment for all 27 EU countries (though histories are short for some); the limitation on countries included in this

---

[19] This update/expansion is funded by the European Commission's Directorate General for Economic and Financial affairs.

[20] Methods used to develop the harmonized estimates of intangible investment are documented in Bontadini et al (2022), available on the EUKLEMS & INTANProd portal at https://euklems-intanprod-llee.luiss.it. Compared with previous estimates issued via the INTANInvest database and website (www.intaninvest.net), current figures reflect significant improvements to the own-account components of intangible investment and to intangible asset price deflators for non-national accounts components. As in previous work that developed productivity estimates using INTANInvest, these estimates reflect price deflators for IT equipment and software whose product quality change component is harmonized across countries. These deflators are developed and supplied by the OECD.

[21] Productivity comparisons at the industry level should use PPPs that adjust for differences in industry product output and input prices across countries rather than overall prices derived from expenditure component of final demand. Methods for obtaining production side PPP estimates from unit value production statistics and adjusted expenditure PPPs are set out in van Ark and Timmer (2009) and Inklaar and Timmer (2008). Updated production-side PPPs will be produced as part of the EUKLEMS & INTANProd project and released this summer.

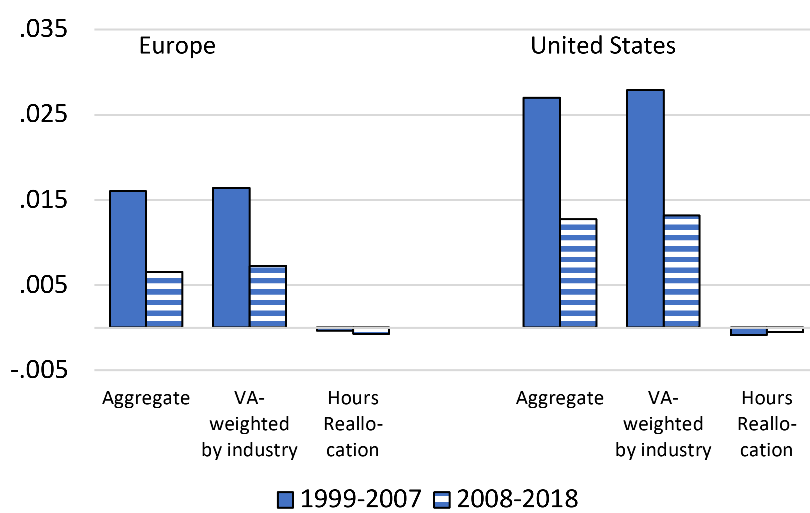
analysis rather mainly the availability of 20 years of requisite growth accounting data by industry. The EUKLEMS & INTANProd data used here are preliminary; statistics for recent years will be updated and the database enhanced this summer.

For international comparability, the intangible capital estimates reflect the incorporation of price deflators for brand and marketing that are harmonized to include the drop in advertising media marketing costs shown in figure 6b. (Similarly, the deflators for computer, and communications equipment and software are harmonized).

Growth decompositions

The growth accounting reported below is in per hour terms, i.e., it decomposes the growth in output per hour for both the European aggregate and the United States. The accounting for the European aggregate is developed at the country-industry level, where industries are aggregated to "market" sector aggregates for each country and then weighted accordingly to form the European aggregate. Market sector aggregates used here exclude the public sector and majority-public (or heavily subsidized) industries, resulting in coverage that is broadly similar, though not identical to the nonfarm business sector used for headline productivity statistics in the United States.[22]

Figure 10. Labor Productivity and Reallocation of Hours



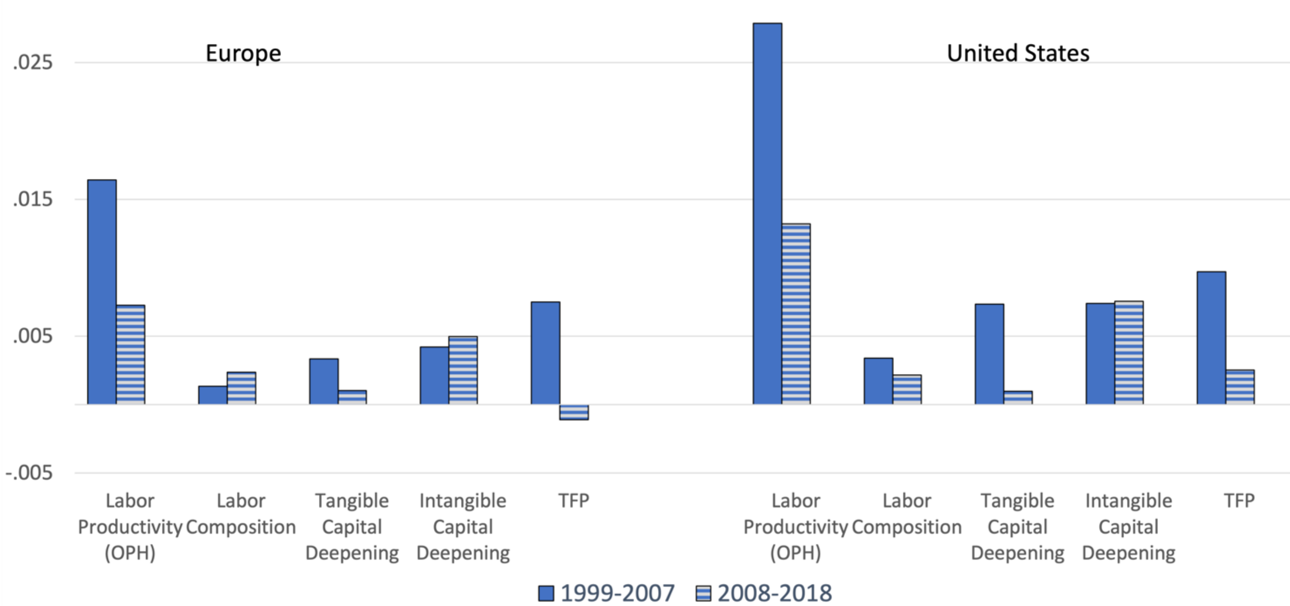
As commonly understood, country-level output per hour reflects both "within" and "between" industry sector effects, with the reallocation of labor across sectors (the "between" effect), e.g., out of agriculture to manufacturing in developing economies, an important factor driving change in low-income countries. Though reallocation may also affect modern economies as jobs move from manufacturing to services, figure 10 shows that the

[22] The market sector aggregates are formed using 25 individual industries that cover 10 NACE letter-level industry sectors: B (Mining), C (Manufacturing), F (Construction, G (Wholesale and retail Trade; repair of motor vehicles), H (Transportation and storage), I (Accommodation and food S=services), J (Information and Communication activities), K (Finance and insurance activities), M (Professional, scientific, and technical activities), N (Administration and support activities). NACE is an international system for industry classification used in Europe; for a concordance to the NAICS system used in North America, see the Bontadini et al. (2022) documentation on the EUKLEMS & INTANProd project portal.

reallocation of hours across market sector industries has had a negligible impact on broad changes in market sector output per hour in Europe and the United States in recent decades. The rate of change in labor productivity thus dropped precipitously in market-dominated industries of both regions with the onset of the global financial recession in 2008 due to its "within" effect—despite the likely boost of .1 to .2 percentage points from data capital based on estimates reported in table 3 and section 4 (and assuming the contribution of data capital was nil in the prior period).

Figure 11 sets out decompositions of the within-industry change in labor productivity. Comparing the first set of columns in figure 11 for each region with the last set, the drop in growth of output per hour (OPH) is seen to be mainly accounted for by a substantial slowdown in total factor productivity (TFP) growth, i.e., the da in equation (3) is 0.9 percentage point less per year in the period after 2007 compared with prior years in Europe and .7 percentage point less in the United States. The contribution of the second set of bars (labor composition) reflects the per hour contribution of increases in (employed) human capital, i.e., the contribution to the change in OPH of changes in the proportion of high-skilled/high wage jobs in an economy. Though this effect works in opposite directions in Europe vs the United States, its contribution to explaining developments in productivity growth in these regions during the past 20 years is relatively small.

Figure 11. Decompositions of Labor Productivity Growth



The terms in capital deepening are part of the slowdown story, directly and indirectly. A drop in tangible capital deepening directly explains 22 percent of the drop in OPH in Europe and whopping 44 percent of the drop in the United States. The rate at which workers in both regions were equipped with intangible capital was maintained, or edged up a tad, over the entire period, however. That resources continued to be invested in innovation in both regions during the period of the slowdown in productivity suggests that the slowdown story must be about, at least in part, changes in the costless diffusion of innovations across firms and industries in these economies. Before turning to discuss this, consider first that GDP measurement may also be a contributor to the productivity slowdown depicted in figure 11.

Besides missing intangible investments that, note, cover AI R&D and most business applications of AI, many believe that official statistics miss major aspects of how consumers benefit from the digital economy. The falling cost of consumer digital content delivery, i.e., the value consumers obtain from their paid-for wireless data and video subscription services, is chief among them.[23] Available research quantifies both very fast drops in prices for consumer digital services (esp. mobile data and streaming) and increased shares of consumer spending allocated to subscriptions for these services—telltale signs that the missed price drops have an *increasing* deflationary impact on consumer price inflation. The missed price drops are in fact estimated to have overstated consumer price change by 0.3 percentage points per year from 2007 to 2018, which when translated to figure 4, potentially explains about 1/3 of the estimated drop in TFP growth in Europe and nearly half of that in the United States.[24]

Diffusion of commercial knowledge and increased productivity dispersion

The diffusion of commercially valuable knowledge is the primary determinant of total factor productivity growth (measured da) according to the upstream/downstream model set out in section 3. A relationship also is a regularity in past productivity data, insofar as cross-country and firm-level econometric work have estimated increasing returns (or knowledge spillovers) to intangible capital. In simple terms, these works imply that a proportional relationship,

[23] That consumers also benefit from free content delivered via their paid-for digital services (e.g., value derived from user-generated content in social media) is a related matter, but however significant, its impacts fall outside the market activity scope of the productivity analysis reported in figure 11.

[24] The aggregate estimate is from Byrne and Corrado (2020, 2021), which applies to the United States and covers mobile voice and data, internet access, cable TV, and video streaming; this estimate is consistent with Abdirahman, Coyle, Heys, and Stewart (2020), who find comparable, rapid rates of price drops for mobile voice and data in the United Kingdom, and Edquist, Goodridge, and Haskel (2021), who document very rapid drops in global prices for music streaming.

expressed as $da \approx .2\ dr$, could be used to represent the costless diffusion of commercially valuable knowledge in an economy.[25]

As dr (per worker) did not slow after 2007, the logical (endogenous) explanation for the slowdown in measured da is that factors driving these increased returns ceased to operate as strongly as they previously had. Despite their nonrival character, the potential for productivity spillovers to intangible investments is determined by an innovation ecosystem, e.g., competition intensity and regulation, intellectual property rights and their enforcement, privacy laws, broadband access, etc. It is very difficult, however, to see how the workings of this ecosystem could change so seriously and suddenly on both sides of the pond (though a possible worrisome decline in competitive intensity in the United States is under active debate).

On the other hand, the composition of knowledge assets directly affects the strength of the diffusion process. Data capital and software code tend to be regarded as trade secrets, intentionally undisclosed and thus difficult to replicate. Moreover, though a data-enabled business model may be apparent to competitors, and the model's training data are not. As the digital economy has boosted *real* investment in data-intensive forms of intangible capital, the mechanisms that generate increasing returns to the macroeconomy via the "free" diffusion of innovations arguably have weakened.

As intangible capital has become, in effect, data capital, there also has been an increase in dispersion of firm-level productivities *within* industry groups attributed, at least in part, to increased investments in economic competencies by market services industries.[26] The changed composition of intangible investment then likely has also led to scale economies within certain firms, e.g., data agglomeration effects in digitally enabled firms, that affected competition.

The data capital framework set out in this paper frames the maximum impact of these developments on market sector productivity as follows: With post-2007 growth of intangibles averaging 2-3/4 percent per year in the European countries and more than 3-1/2 percent per year in the United States, a *complete* cessation of the diffusion mechanism could shave as much as .5 to .6 percentage point per year off measured da for these regions. These are rather sizeable impacts.

---

[25] This refers to the aggregate implications of estimates for R&D spillovers reported by Griliches for manufacturing (e.g., Griliches 1992) and for nonR&D intangibles (especially, the industrial design, employer-provided training, and organizational capital components) by Corrado, Haskel, and Jona-Lasinio (2017).

[26] This was documented globally in Andrews, Criscuolo, and Gal (2016), who characterized the development as a worrisome decline in the global diffusion of new ideas and technologies since 2000. The growing relative importance of intangible assets was identified as a mechanism behind increased firm-level productivity dispersion in follow on work (Corrado, Criscuolo, Haskel, Himbert, and Jona-Lasinio 2021).

Productivity growth via the costless replication of commercial knowledge is highly unlikely to have ceased entirely, however, and other factors including structural or policy-induced factors may have contributed to its slowdown. But our analysis of the data intensity of intangible capital, combined with the tendency for data assets to be closely held, suggests that spillover effects that prevailed in the past likely have diminished with the increased use data and slowed the growth in total factor productivity.

6. Concluding Remarks

We have used an intangible assets approach to the question of how much data might affect productivity. We argue that data, or more accurately transformed raw digitized records, and the capital services derived therefrom fit neatly into both the management/technology literature on the data stack and the economics literature on intangible assets. We outlined a simple two-sector growth accounting framework to articulate how much the relative technology gains from the use of data might have affected overall productivity growth. And using a new data set, we documented and analyzed the slowdown in total factor productivity growth in major developed economies.

A full explanation for the recent productivity slowdown perhaps remains elusive, but we are hopeful that the methods outlined, along with improved measurement, have demonstrated that the increased importance of data assets in intangible capital is a factor in the explanation.

References

Abdirahman, Mohamed, Diane Coyle, Richard Heys, and Will Stewart. 2020. "A Comparison of Deflators for Telecommunications Services Output." *Economie & Statistique*; Paris Iss. 517-518-519, (Oct 2020): 103–122. DOI:10.24187/ecostat.2020.517t.2017

ACCA - Association of Chartered Certified Accountants (2019). ""Business Reporting of Intangibles: Realistic proposals – Comments from ACCA". Accessed on 15 September 2021 at: https://www.accaglobal.com/gb/en/technical-activities/technical-resources-search/2019/may/frc-intangibles-response.html.

Zhang Daniel, Saurabh Mishra, Erik Brynjolfsson, John Etchemendy, Deep Ganguli, Barbara Grosz, Terah Lyons, James Manyika, Juan Carlos Niebles, Michael Sellitto, Yoav Shoham, Jack Clark, and Raymond Perrault, "The AI Index 2021 Annual Report," AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, March 2021.

Akcigit, Ufuk, and Sina T. Ates, 2021. "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory." *American Economic Journal: Macroeconomics* 13 (1): 257– 98.

Ahmad Nadim & Jennifer Ribarsky & Marshall Reinsdorf, (2017). "Can potential mismeasurement of the digital economy explain the post-crisis slowdown in GDP and productivity growth?,"OECD Statistics Working Papers 2017/9, OECD Publishing.

Andrews Dan & Chiara Criscuolo & Peter N. Gal, (2016). "The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy,"OECD Productivity Working Papers 5, OECD Publishing.

Brynjolfsson, E., & McAfee, A. (2014).The second machine age: Work, progress, and prosperity in a time of brilliant technologies (First Edition.). W. W. Norton & Company.

Brynjolfsson, Erik, Avinash Collis and Felix Eggers (2019). “Using Massive Online Choice Experiments to Measure Changes in Well-being.” *Proceedings of the National Academy of Sciences* 116(15): 7250-7255.

Byrne, David and Carol Corrado (2017). “ICT Prices and ICT Services: What do they tell us about Productivity and Technology?” *International Productivity Monitor*, Number 33 (Fall), p. 150-181.

Byrne, David and Carol Corrado 2020. “The Increasing Deflationary Influence of Consumer Digital Access Services.” *Economic Letters* 196 (November), 109447.

Bryne, David and Carol Corrado 2021. “Accounting for Innovations in Consumer Digital Services: IT Still Matters.” In Corrado, C, Haskel, J., Miranda, J. and Sichel, D., Measuring and Accounting for Innovation in the Twenty-First Century, 471- 517.

Carson Richard & Nicholas Flores & Norman Meade, (2001)."Contingent Valuation: Controversies and Evidence," Environmental & Resource Economics, Springer; European Association of Environmental and Resource Economists, vol. 19(2), pages 173-210, June.

Chiarello, Filippo, Paola Belingheri, and Gualtiero Fantoni (2021). “Data Science for Engineering Design: State of the Art and Future Directions.” *Computers in Industry* 129 (103447).

Clarkson, Kenneth W. 1996. “The Effects of Research and Promotion on Rates of Return.” In Robert B. Helms, ed., *Competitive Strategies in the Pharmaceutical Industr*y, 238-268. Washington, D.C.: American Enterprise Institute for Public Policy Research.

Corrado, Carol (2021). “Intangible Capital.” Chapter in forthcoming Brookings-Hutchins Center volume (available upon request from author).

Corrado, Carol, Chiara Criscuolo, Jonathan Haskel, Alexander Himbert, and Cecilia Jona-Lasinio (2021). “New Evidence on Intangibles, Diffusion and Productivity.” OECD STI Working Paper 2021/10 (July) https://doi.org/10.1787/de0378f3-en

Corrado, Carol A. and Goodridge, Peter and Haskel, Jonathan E. 2011. “Constructing a Price Deflator for R&D: Calculating the Price of Knowledge Investments as a Residual.” Available at SSRN: https://ssrn.com/abstract=2117802 or http://dx.doi.org/10.2139/ssrn.2117802

Corrado, Carol, Jonathan Haskel, and Cecilia Jona-Lasinio 2017. “Knowledge Spillovers, ICT, and Productivity Growth,” *Oxford Bulletin of Economics and Statistics* 79:4 (August), 592-618.

Corrado, Carol, Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi (2022). “Intangible Capital and Modern Economies. “ *Journal of Economic Perspectives (forthcoming).*

Corrado, Carol, Charles R. Hulten, and Daniel Sichel (2005). “Measuring Capital and Technology: An Expanded Framework,” in C. Corrado, J. Haltiwanger, and D. Sichel (eds.), *Measuring Capital in the New Economy*, Volume 66, Studies in Income and Wealth, 11–46. Chicago: University of Chicago Press.

Corrado, Carol, Charles R. Hulten, and Daniel Sichel (2009). “Intangible Capital and U.S. Economic Growth,” *Review of Income and Wealth* 55 (3): 661–685.

Coyle, Diane and Wendy C.Y. Li (2021). “The Data Economy: Market Size and Global Trade”, ESCoE Discussion Paper No. 2021-09

Coyle, Diane and David Nguyen, 2018. "Cloud Computing and National Accounting," Economic Statistics Centre of Excellence (ESCoE) Discussion Paper DP-2018-19. London, UK: Economic Statistics Centre of Excellence (ESCoE).

Crouzet, Nicolas and Janice C. Eberly (2019). "Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles." NBER Working Paper Series No. 25869 (May).

Deloitte (2020). "Data Valuation: Understanding the Value of your Data Assets." Accessed on 15 March 2021 at:https://www2.deloitte.com/global/en/pages/finance/solutions/datavalue.html.

De Ridder, Maarten, (2019), Market Power and Innovation in the Intangible Economy, No 1907, Discussion Papers, Centre for Macroeconomics (CFM), https://EconPapers.repec.org/RePEc:cfm:wpaper:1907.

Döttling, Robin and Lev Ratnovski (2020). "Monetary Policy and Intangible Investment." IMF working paper WP/20/160 (August).

The Economist (2017). "Leader: The world's most valuable resource is no longer oil, but data." The Economist Group Limited (6 May).

Goodridge, P., J. Haskel, and H. Edquist (2021), "We See Data Everywhere Except in the Productivity Statistics." Review of Income and Wealth. https://doi.org/10.1111/roiw.12542

Granstrand, Ove and Marcus Holgersson (2020). "Innovation ecosystems: A conceptual review and a new definition." *Technovation*: 90–91, 102098.

Griliches, Zvi. 1992. "The Search for R&D Spillovers." *Scandinavian Journal of Economics* 94 (Supplement): S29-47.

Haskel, Jonathan and Stian Westlake. 2018. *Capitalism without Capital: The Rise of the Intangible Economy*. Princeton, N.J.: Princeton University Press.

Inklaar, Robert & Timmer, Marcel P., 2008. "GGDC Productivity Level Database: International Comparisons of Output, Inputs and Productivity at the Industry Level,"GGDC Research Memorandum GD-104, Groningen Growth and Development Centre, University of Groningen.

ISWGNA (2020). "Recording and Valuation of Data in National Accounts", issue note prepared by the ISWGNA subgroup on digitalization for the 14th Meeting of the Advisory Expert Group on National Accounts, October 2020.

Jones, Charles I., 2005."Growth and Ideas." In Philippe Aghion & Steven Durlauf (ed.), *Handbook of Economic Growth*, edition 1, volume 1, chapter 16, pages 1063-1111, Elsevier.

Jones, Charles I. and Christopher Tonetti (2020). "Nonrivalry and the Economics of Data." NBER working paper #26260 (April).

Jorgenson, Dale. "Capital Theory and Investment Behavior." American Economic Review 53, no. 2 (1963): 247-259.

Ker Daniel & Emanuele Mazzini, 2020."Perspectives on the value of data and data flows,"OECD Digital Economy Papers 299, OECD Publishing.

Lev, B. (2019), "Ending the Accounting-for-Intangibles Status Quo", European Accounting Review, 28:4, 713-736, DOI: 10.1080/09638180.2018.1521614

Li, Wendy C.Y. and P.J. Chi (2021), "Online Platforms' Creative "Disruption"" in Organizational Capital – The Accumulated Information of the Firm, Moon Economics Institute Discussion Paper, No. 1, August.

Mandel, Michael (2019). "Building the New Manufacturing Stack." Forbes Magazine (August 20). Available at https://www.forbes.com/sites/michaelmandel1/2019/08/20/building-the-new-manufacturing-stack/?sh=623f722b561f

Mayer-Schönberger, V., and K. Cukier (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

McKinsey Global Institute (2016), "The Age of Analytics: Competing in a Data-Driven World", MGI Research Report (December). Available here.

McFadden and Train, 2017. *"*Contingent Valuation of Environmental Goods. A Comprehensive Critique" *Elgar online* ISBN: *978 1 78643 468 5*

Miller, Holly H., Leslie A. Richardson, Stephen R. Koontz, John Loomis, and Lynne Koontz. 2013. "Users, Uses, and Value of Landsat Satellite Imagery: Results from the 2012 Survey of Users." https://doi.org/10.3133/ofr20131269.

Nelson, R.R. and Rosenberg, N. (1993) Technical Innovation and National Systems. In Nelson, R. R. (ed) 1993 National Innovation Systems: A Comparative Analysis. Oxford: Oxford University Press.

Nguyen, David and Marta Paczos (2020). "Measuring the Economic Value of Data and Cross-border Data flows: A business Perspective." OECD Digital Economy Papers No. 297. OECD Publishing, Paris.

Oulton Nicholas, 2012. "How To Measure Living Standards and Productivity," *Review of Income and Wealth*, vol. 58(3), pages 424-456, September.

Prince J. and S. Wallsten (2020), "How Much is Privacy Worth Around the World and Across Platforms?", TPRC48: The 48th Research Conference on Communication, Information and Internet Policy. Accessed o 0 october 2021 at https://techpolicyinstitute.org/publications/privacy-and-security/how-much-is-privacy-worth-around-the-world-and-across-platforms-2/

PriceWaterhouseCoopers LLP (2019). "Putting a value on data." Available at: https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf

Reinsdorf, M. and J. Ribarsky (2019). "Measuring the Digital Economy in Macroeconomic Statistics: The Role of Data", paper presented at the ASSA 2020 Annual Meeting

Reinsdorf, M. and P. Schreyer (2019), "Measuring consumer inflation in a digital economy", OECD Statistics Working Papers, No. 2019/01, OECD Publishing, Paris, https://doi.org/10.1787/1d002364-en.

Roca, Ciara (2021a). "How to build a modern data stock from the ground up". Available at: https://fivetran.com/blog/how-to-build-a-modern-data-stack-from-the-ground-up

Roca, Ciara (2021b). "What is the modern data stack?" https://fivetran.com/blog/what-is-the-modern-data-stack (last accessed August 25, 2021)

Romer, Paul M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 95(5): S71–S102.

Statistics Canada (2019a). "Measuring Investment in Data, Databases and Data Science: Conceptual Framework." (June)

Statistics Canada (2019b). "The Value of Data in Canada: Experimental Estimates." (July).

UNECE (2018), "Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics"

United Nations (2008). "System of National Accounts 2008"

UK Financial Reporting Council (2019), "Business Reporting of Intangibles: Realistic proposals". Accessed on 15 September 2021 at https://www.frc.org.uk/consultation-list/2019/discussion-paper-business-reporting-of-intangibl.

Varian, Hal (2019). "Artificial Intelligence, Economics, and Industrial Organization." In Arjay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda.* Chicago: University of Chicago Press.

Zéghal D. and A. Maaloul (2011). "The accounting treatment of intangibles – A critical review of the literature." *Accounting Forum*, 35 262– 274.

APPENDIX 1– Data Sources and Estimation Method for Data Assets and Computer Software

In this appendix, we describe the cost-based approach and the data sources we have used to estimate market sector investment in data stores and data intelligence (the components of data not currently included in official national accounts), databases and computer software. The same method is used to estimate the own-account brand, design, organizational capital, and new financial products in the EUKLEMS & INTANProd database.

In broad terms, cost-based estimates of output for any given asset can be derived as follows:

Estimated value of output at basic prices
equals
Labour costs of relevant personnel (compensation of employees)
plus
Intermediate costs used in these activities
plus
Cost of capital services used in these activities (gross operating surplus)
plus
Net taxes on production related to these activities

The standard way to implement the above calculation is to first estimate the labor cost component as:

Labour costs of relevant personnel
equals
Total number of employees working on producing the relevant asset
times
Average remuneration
times
Proportion of time spent on these activities

Then, gross output at basic prices is obtained as:
Gross output at basic prices
equals
Labor costs of relevant personnel

times
Blow-up factor

For each asset, the calculation requires i) a detailed list of occupations; ii) occupation-specific (and industry-specific, if relevant) assumptions on the share of time spent in producing the asset; iii) data on the number of employees for the relevant occupations and their compensations; iv) blow-up factors to account for other cost components (intermediate consumption and gross operating surplus) to derive an output measure consistent with national accounts definitions.

Occupations identified from the ISCO-08 as engaged in data assets and computer software capital formation are presented in Table A1, along with time-use assumptions.

The selection of relevant occupations is constrained by the level of detail of the available data sources. For this paper, we use micro-data of the EU Structure of Earnings Survey (SES) for 2010, 2014, and 2018 and the EU Labour Force Survey (LFS) for 2008-2019. The SES provides information on the number of employees by occupation (at the three-digit level of the 2008 International Standard Classification of Occupations, ISCO) and economic activity and their annual earnings. The LFS, instead, only provides data on employment with no information on wages. In the LFS, occupations are available at the three-digit level of ISCO for all countries, while SES data for 11 countries are available at two-digit. We have disaggregated two-digit ISCO into three digits for these countries based on the share of each relevant three-digit occupation from the LFS.

In the LFS, occupations are available at the three-digit level of ISCO for all the six countries, while SES data for Finland, Germany, Netherland, Spain, and Sweden are available at two-digit. We have disaggregated two-digit ISCO into three digits for these five countries based on the share of each relevant three-digit occupation from the LFS.

Some occupational groups relevant for data-related asset production are only identifiable at the four-digit level. Thus, we have tweaked our assumptions on the time-use factors accordingly to consider that the occupational groups include workers not engaged in data assets production.

For each asset, the calculation for 2010, 2014 and 2018 (the years for which we have the SES) is as follows:

1. Calculate total employment for each relevant occupational group involved in producing the asset (identified at three-digit ISCO)

2. Apply occupation-specific time-use assumptions to each occupation's employment

3. Calculate total wages for each (time-use adjusted) relevant occupation.

4. Calculate the total share of all occupations involved in producing the asset in total wages from the SES

5. Calculate labor cost component consistent with national accounts by applying the share calculated at step 4 to national accounts' compensation of employees.

6. Calculate gross output by applying blow-up factors to the labor cost component derived at step 5. We have used blow-up factors equal to 1.6 for organizational capital and 1.8 for the other assets.

We have derived the wage shares of the intervening years (when the SES is not available) based on information from the LFS. For each country, we have calculated the share of (time-use adjusted) relevant occupations in total employment for 2010-2018 from LFS. We have then used the employment share as an indicator to extrapolate/retropolate the wage shares obtained from the SES.

We have made the calculations by industry, at the level of Nace sections, and then aggregated the result to the market sector, defined as all industries excluding Nace sections O (public administration and defense; compulsory social security), P (education), Q (human health and social work activities), T (activities of households as employers; undifferentiated goods - and services-producing activities of households for own use), and the imputed rents of owner-occupied dwellings component section L (real estate activities).

Table A1. Relevant Occupations in Measurement of Investment in Data Assets and Computer Software and Time-use Assumptions

ISCO-08 sub-major group	**ISCO-08 minor group**	**Occupation description**	**Time-use (%)**			
			Data Stores	**Data Intelligence**	**Databases**	**Software**
21 - Science and	211	Physical and earth science professionals	10	25	0	0

engineering professionals	212	Mathematicians, actuaries and statisticians	10	25	0	0
	213	Life science professionals	10	25	0	0
	214	Engineering professionals (excluding electrotechnology)	10	25	0	0
	215	Electrotechnology engineers	10	25	0	0
	216	Architects, planners, surveyors and designers	10	10	0	0
24 - Business and administration professionals	241	Finance professionals	10	25	0	0
	242	Administration professionals	0	10	0	0
	243	Sales, marketing and public relations professionals	10	10	0	0
25 - Information and communications technology professionals	251	Software and applications developers and analysts	10	0	10	50
	252	Database and network professionals	10	0	50	10
33 - Business and administration associate professionals	331	Financial and mathematical associate professionals	10	25	0	0

35 - Information and communications technicians	351	Information and communications technology operations and user support technicians	10	0	0	0
41 - General and keyboard clerks	413	Keyboard Operators	5	0	0	0
42 - Customer services clerks	422	Client information workers	5	0	0	0
43 - Numerical and material recording clerks	431	Numerical clerks	5	0	0	0
	432	Material-recording and transport clerks	5	0	0	0

Note: 413 includes data entry clerks (4132); 422 includes survey and market research interviewers (4227); 431 includes statistical, finance and insurance clerks (4312).

APPENDIX 2– Data Sources and Estimation Method for the Domestic Component of Intangible Investment

In this appendix, we describe how we have estimated the domestically sourced component of intangible investment.

We have estimated domestically produced investment in R&D and computer software and databases based on data from national supply and use tables. We have calculated the share of gross output in total resources for domestic use, SGOD, as follows:

$$SGOD_i = (GO_i - EX_i) / (GO_i - EX_i + IM_i)$$

where GO_i, EX_i and IM_i are gross output, exports and import of product i (i= CPA_M72 for R&D, and CPA_J62-63 for computer software and databases).

Then, we have estimated the domestic component for each of the two assets, multiplying national accounts investment by the corresponding share of gross output in total resources for domestic use. The calculation assumes that the share of the domestic component is the same across different uses (intermediate consumption, final consumption, and investment).

For non-national accounts intangibles, we have calculated the domestic component of brand, design, and organizational capital as the sum of own-account investment and an estimate of the domestically sourced purchased component. New financial products are only domestically produced. No data sources for estimating imported training are available, but we deem that the imported component is very small and can be ignored.

Domestically sourced purchased component of brand, design and organizational capital has been calculated based on information from the world input-output tables available from World Input-Output Database (WIOD, available at https://www.rug.nl/ggdc/valuechain/wiod/?lang=en). For each industry in a country, world input-output tables report intermediate use of domestic output and intermediate use of imports from other countries disaggregated by product. Based on these tables, we have calculated the share of domestic output in market sector intermediate consumption of the following products: advertising and market research services (CPA M73), architectural and engineering services, technical testing, and analysis services (CPA M71) and legal and accounting services, services of head offices and management consulting services (CPA_M69_70). Finally, we have calculated the domestically sourced purchased component of brand, design and organizational capital multiplying purchased investment by the share of domestic output in total intermediate consumption for the relevant products (CPA M73 for brand, CPA_M71 for design, and CPA_M69_70 for organizational capital).

The 2016 WIOD release provides an annual time-series of world input-output tables from 2000 to 2014. For the most recent years, we have extrapolated the shares regressing 2000-2014 shares on a linear time trend.

Appendix 3 Correlation Tables

Appendix Table A2 **Shares of value added**

	Databases_P	Data_Stores_P	Data_Intelligence_P	Data_capital_P	Intangibles_I	Intangibles_NonNatAcc_I	Intangibles_NatAcc_I	Software_P	Software_DB_I	R&D_I	Org Cap_I	NFP_I	Design_I	Brand_I
Databases_P	1													
Data_Stores_P	0.393***	1												
Data_Intelligence_P	0.382***	0.861***	1											
Data_capital_P	0.602***	0.917***	0.954***	1										
Intangibles_I	0.612***	0.482***	0.549***	0.629***	1									
Intangibles_NonNatAcc_I	0.579***	0.624***	0.676***	0.740***	0.914***	1								
Intangibles_NatAcc_I	0.484***	0.149	0.219*	0.299**	0.833***	0.536***	1							
Software_P	0.647***	0.726***	0.805***	0.865***	0.872***	0.898***	0.589***	1						
Software_DB_I	0.587***	-0.0000921	0.0745	0.194	0.762***	0.564***	0.815***	0.593***	1					
R&D_I	0.268*	0.202	0.186	0.241*	0.599***	0.304**	0.833***	0.337**	0.389***	1				
Org Cap_I	0.649***	0.528***	0.491***	0.619***	0.863***	0.935***	0.520***	0.832***	0.666***	0.227*	1			
NFP_I	0.141	-0.168	0.2	0.102	-0.0468	-0.104	0.0445	0.0661	0.0885	-0.05	-0.221*	1		
Design_I	0.0759	0.132	0.216	0.188	0.571***	0.348**	0.713***	0.341**	0.358**	0.721*	0.18	-0.011	1	
Brand_I	0.211	0.572***	0.608***	0.589***	0.416***	0.687***	-0.0728	0.565***	0.0398	-0.18	0.568***	-0.128	-0.189	1

="* p<0.05 ** p<0.01 *** p<0.001"

Appendix Table A3 **Shares of value added (percentage changes)**

	Databases_P	Data_Stores_P	Data_Intelligence_P	Data_capital_P	Intangibles_I	Intangibles_NonNatAcc_I	Intangibles_NatAcc_I	Software_P	Software_DB_I	R&D_I	Org Cap_I	NFP_I	Design_I	Brand_I
Databases_P	1													
Data_Stores_P	0.365***	1												
Data_Intelligence_P	0.697***	0.659***	1											
Data_capital_P	0.728***	0.857***	0.921***	1										
Intangibles_I	0.0770	0.144	0.150	0.150	1									
Intangibles_NonNatAcc_I	0.255*	0.557***	0.608***	0.589***	0.374***	1								
Intangibles_NatAcc_I	-0.0275	-0.0631	-0.0784	-0.0724	0.931***	0.0562	1							
Software_P	0.707***	0.610***	0.751***	0.766***	0.158	0.452***	-0.0158	1						
Software_DB_I	0.00481	-0.0491	-0.0299	-0.0333	0.759***	0.0256	0.796***	0.0173	1					
R&D_I	-0.0684	-0.0789	-0.104	-0.105	0.859***	0.0586	0.957***	-0.0361	0.645***	1				
Org Cap_I	0.213	0.297**	0.381***	0.365***	0.353**	0.754***	0.101	0.367***	0.147	0.0511	1			
NFP_I	0.559***	0.432***	0.730***	0.661***	0.0450	0.275*	-0.0671	0.526***	0.0333	-0.107	0.0741	1		
Design_I	0.0818	0.417***	0.319**	0.361***	0.0874	0.611***	-0.117	0.192	-0.192	-0.067	0.342**	0.133	1	
Brand_I	0.158	0.487***	0.497***	0.474***	0.299**	0.741***	0.0827	0.330**	0.0621	0.103	0.365***	0.230*	0.125	1