Data, Intangible Capital, and Productivity*

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

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

June 12, 2023

Abstract
This paper analyzes how the increased use of data in economies affects productivity. We introduce a framework for measuring data and find that data assets are conceptually encompassed in the Corrado, Hulten, and Sichel (2005, 2009) intangible capital framework. Data assets are not explicitly identified in that framework, however, and to remedy this, the paper develops measures of industry-level investments in data for nine European countries. Analysis of the new measures concludes that about 50 percent of intangible capital is, in effect, data capital.

Next, a simple model of an economy with data/intangible capital is used to assess the impact of the increased use of data, especially proprietary bigdata. We find that there are two primary macroeconomic impacts. First, the greater relative efficiency of data capital boosts its contribution to labor productivity. Second, the increased data intensity of intangibles weakens knowledge diffusion and diminishes TFP growth. To frame the net magnitude of these offsetting effects, the paper uses the recently developed EUKLEMS & INTANProd database (LLEE 2023). The boost to labor productivity stemming from the estimated relative efficiency of data capital is offset by the appropriability effect, which shaved .3 and .4 percentage points off 2010-2019 TFP growth in Europe and the United States, respectively.

* Paper originally 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 those who believe in 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 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 it is frequently suggested that business users face high adjustment costs in deploying the modern digital tools needed to derive knowledge from data, 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 artificial intelligence (AI) and cloud-based technologies, this paper looks to the character of knowledge gained through data 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. The paper introduces a framework for capturing asset creation based on all digitized information that is processed and transformed into useable knowledge in an economy. This knowledge is referred to as "data capital." The framework is amenable to measurement and quantitative analysis.

The paper is in two main parts. Part I looks at how the increased use of data affects intangible capital conceptually and empirically. New estimates of investment in data are developed and analyzed. Part II models how data capital, innovation, and productivity are related, including an analysis of how data capital affects intangible asset prices, intangible asset returns, and potential labor productivity growth. The recent slowdown in total factor productivity is analyzed considering these results and making use of the recently issued EUKLEMS & INTANProd database (LLEE 2023). A feature of this database relative to its predecessors is that it includes estimates of intangible investment consistent with the intangible capital framework attributable to Corrado, Hulten, and Sichel (2005, 2009) for all European Union countries, Japan, the United Kingdom, and the United States.

This paper makes, we believe, three contributions. The first is its development and analysis of industry-level estimates of data capital consistent with concepts used by management strategists and technologists. We find that data assets, from stores of raw data to actionable intelligence derived via data analytic tools, are largely subsumed within intangible capital. A second contribution is the paper's modeling of the economic impacts of data capital, which finds that the efficiencies associated with modern data capital lowers intangible asset prices and strengthens the (partial) appropriability of the asset class. Third, the paper frames the likely macroeconomic impacts of these developments, finding that that the increased data intensity of intangible capital boosts its contribution to labor productivity growth by about ¼ percentage point. At the same time, the increased role of proprietary bigdata in production processes, especially in the production of commercially valuable knowledge (i.e., intangible capital), weakens the diffusion of innovations and slows growth in total factor productivity (TFP). All told we estimate that the appropriability effect has, so far, more than offsets the efficiency effect and that the rise of modern data capital has been a rather consequential contributor to the recent slowdown in productivity growth.

Relation to recent literature

Recent literature has advanced many models that focus on economic mechanisms affected by data (or intangibles). How does our approach relate to these models?

At the micro level, data assets are 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 artificial intelligence (AI) algorithms. Jones and Tonetti (2020) formulate a model of an economy in which bigdata are an intermediate input with diminishing returns and analyze the welfare consequences of data and data sharing. Despite treating data as an intermediate (versus as an asset as we do in this paper), the Jones and Tonetti (2020) analysis of data sharing transcends this distinction. Their analysis, like ours, is based on the observation that firms attempt to hoard proprietary data for their own use in fear of potential "creative destruction" and that this behavior stanches the reuse of data as a nonrival good and is inefficient. The analysis of Arrieta-Ibarra et al. (2018) also focuses on the inefficiency of current data use in economies, highlighting the monopsony position of firms collecting consumer data and arguing that this depresses the value of personal data and productivity gains from its use. While the concerns of these works are directly relevant for optimally allocating property rights of consumer data in economies, the concerns also impact knowledge diffusion and productivity growth, as elaborated in this paper.

Most descriptive models of data in the economics literature assume diminishing returns to data assets but show how they can co-exist with firm-level cost advantages due to the scale effects in data-dependent production processes. This is the gist of *The Economist* (2017) article, which emphasizes how data enable retail and social media giants to expand their customer base by exploiting network effects. *The Economist* also speculated that data hampers "creative destruction," noting that data enables giant firms to utilize surveillance systems to "see" when a new product or service gains traction and take action to avoid being blindsided by competition from new entrants (e.g., Facebook's acquisition of WhatsApp).

Business and management scholars frequently emphasize scale and scope economies created by data in business activities such as marketing, distribution, and after-market services. It should be noted, however, that there is little industry- or economy-level evidence for these effects. In fact, an econometric study of the global advertising and marketing services business found that scale efficiencies did not arise during the 1990's era of globalization (Silk and Berndt 2003).

The combined impact of these effects, whether data-induced scale or scope efficiencies in business functions, production processes, or network externalities induced on the demand side, will be seen as a rise in market power if data assets (and intangibles in general) are not accounted for in competition indicators—indicators such as rates of returns, price markups, and productivity dispersion. Virtually all studies finding a rise in market power in recent decades (e.g., as reviewed in the IMF's April 2019 *World Economic Outlook,* IMF 2019) rely on firm-level datasets that exclude or improperly account for intangibles. Evidence based on official macrodata—which currently accounts for about 40 percent of total intangible investment—is less dispositive on whether markups are rising (Basu 2019). In fact, aggregate data adjusted to include all Corrado, Hulten, and Sichel (2005, 2009) intangibles suggests that abnormal rents did not, in fact, accrue to U.S. private industries from 1995 to 2019 (see Corrado, Haskel, Jona-Lasinio, and Iommi 2022a, figure 4 and its discussion).

But industry concentration measures have risen and many attribute this to scale economies of intangible assets at the firm level (e.g., Haskel and Westlake 2018, Crouzet and Eberly 2019, De Ridder 2019). Economies due to data-driven intangible capital also could be driving the widening of within-industry productivity dispersion documented in Andrews, Criscuolo, and Gal (2017), and a study addressing whether the increase in within-industry productivity dispersion since 2000 can be attributed to intangibles found that it could (Corrado, Criscuolo, Haskel, Himbert, and Jona-Lasinio 2021). While this evidence was stronger at large firms, strictly speaking, economies of scale were not detected. Dispersion was rather explained by financial barriers to the accumulation of intangible assets—which affect some firms more than others—and by the complementarity of intangibles with expenditures on digitization.

Where does this leave us? First, the intangible capital framework is consistent with the welfare-enhancing processes of data sharing as theorized by Jones and Tonetti in that (a) data assets have diminishing returns in production but (b) returns to data assets will spill over to other firms to the extent data can be copied or shared within an industry or economy. Second, increased productivity dispersion between "leaders" and "laggards hints at how *average* productivity growth in an economy can slow at the same time as innovative, competitive firms amass bigdata: a breakdown in the diffusion of data-derived knowledge from the leaders to the laggards. In this paper we suggest that productivity spillovers to intangible capital have in fact weakened substantially as production processes—notably, production processes for intangible assets—have become dependent on proprietary bigdata.

In related work, Akcigit and Ates (2021) attribute slow productivity growth and declining business dynamism to a breakdown in knowledge diffusion. They give four reasons for this breakdown, three of which are supported by their R&D-inspired models of innovation and analysis of developments in patent protection by large firms. A fourth, and we believe the most compelling, reason for a breakdown in knowledge diffusion owes to the rise of proprietary bigdata used by productivity "leaders." Akcigit and Ates do not offer direct empirical support this proposition, but we believe the that the evidence advanced in this paper, combined with our previously mentioned microdata-based work on the determinants of productivity dispersion, provide a solid base of evidence for this view.

Rise of Proprietary Bigdata

As background for the arguments that we develop in this paper, consider first some examples of exclusive (or rival) versus nonrival use of data in modern economies as listed in table 1. Though data is inherently nonrival, the degree to which owners share data with the public or other organizations in an industry (or the economy) depends upon both context and competitive factors.

The examples listed on lines 1–5 of the table mainly reflect applications of bigdata using 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, which long have been inputs to brand development, marketing, and customer retention strategies, are fertile ground for application of data technologies.

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 (traffic patterns)
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. The grouping of examples in the table reflects the degree to which data owners share their data assets with other organizations or the public.

Examples of "nonrival" data use range from marketers of personal data for companies (line 8), to longer-standing examples of 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), personal medical records shared by medical care services providers (line 11), to newer cross-platform and cross-purpose uses (lines 12 and 13).

Finally, the table lists some examples of government open data. Governments generate rather vast stores of information and are working to make the data they collect more "open", i.e., freely available for anyone to download, modify, and distribute without legal or financial restriction. This suggests that are government statistics that are a public good externality. The UK Open Data Institute (ODI) estimates that the use of "core" public open data alone—data such as addresses, maps, weather, and land and property ownership records—currently contributes an additional ½ percent of the country's GDP in economic value every year (ODI, 2016). A review of estimates and surveys of the value of U.S.

government data in business decision-making concluded that (a) the value of government data is increasing and that (b) while companies are generating ever-increasing amounts of big data from their own operations, the combination of proprietary data with comprehensive government data often creates more strategic benefit (Hughes-Cromwick and Coronado 2019).

These examples suggest that while data has much potential for use and economic benefit when shared, many applications of bigdata involve proprietary use. Data-dependent business models are on the rise (Nguyen and Paczos 2020), as are regulations to protect consumer privacy, e.g., the General Data Protection Regulation (GDPR) in the European Union and the US equivalent, California Consumer Privacy Act (CCPA). These regulations limit third-party sales, even though certain cross-purpose uses of data (e.g., lifestyle data collected by marketers used in precision medicine solutions) have the potential to affect the pace of innovation. Conversely, policy interventions can facilitate data sharing and competitive entry, e.g., the data sharing environments facilitated by open banking policies in the United Kingdom and other countries.[1]

All told, the foregoing suggests that a conceptual framework for measuring and analyzing data needs to account for the fact that: (a) data is nonrival and capable of improving economic welfare when shared or replicated at low cost; but that (b) data, though nonrival, is frequently used exclusively.

Part I: Data and Intangible Capital

2. A Framework for Data Value Creation

The Data Stack

Many economic models of data focus on data as a "free" by-product of economic activity, and observers focus on certain special features of data, such as how rapidly it accumulates. Data, in the sense of raw digitized records, may of course accumulate at an astonishing pace and be stored at little to no cost. But accumulation of raw bits and bytes does not imply that a flow of services is being provided to production processes in an economy.

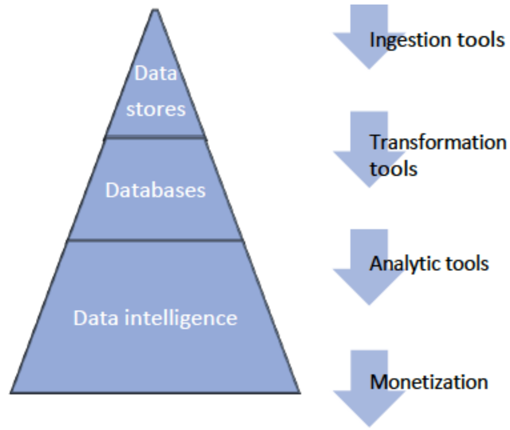
Our approach to developing a framework for the analysis of data is thus based on the following observations/assumptions: (1) The accumulation of data has the potential to boost real output only when producers also invest in transforming such records into analytical insights and actionable business intelligence. (2) Knowledge-based assets gleaned from the application of data technologies to data (raw or transformed) are productive assets in an economy. (3) The appropriability of returns to these assets implies that business spending on data accumulation and transformation, and data analytics are intangible capital investments.

Our specific approach to data value creation is illustrated in figure 1, which depicts a framework that embraces widely used approaches in both the technology and management literatures. Technologists

[1] Open banking refers to a data sharing environment in which financial intermediaries—both incumbents and fintech entrants—can compete for customers. For an analysis of the impact of open banking regulations on financial innovation globally, see Babina, Buchak and Gornall (2022).

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 includes monetization, or market implementation, as a capability required for creating value from data assets[2]. Technologists usually depict a sequence of data forms and digital tools in a single pyramid (e.g., Roca 2021a). The data pyramid has deep roots in information science; it was used by Varian (2019) to depict the relationship between data, information, and knowledge in his analysis of AI.

Figure 1. 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.

Figure 1 separates data forms, i.e., types of data, from the tools used to create them. Three major types of data appear in the business-oriented data value creation literature. These types are depicted on the left in figure 1 and reflect the business strategists' notion of an information value chain, where greater value is produced as data is processed into usable intelligence. The digital technology tools that enable value creation from nondigitized records are depicted on the right. The sequence of data assets is horizontally aligned with the tools used in their formation on the left, i.e., ingestion tools are used to create data stores, etc.

The three layers of value in figure 1's data stack—data stores, databases, and data intelligence—correspond to an asset type amenable to measurement and analysis. The data asset types are defined 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 provides actionable guidance for decision-makers and includes solutions to scientific and engineering problems.

What separates the "modern" data stack from legacy, on-premises data management 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

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

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."

Data capital as Intangible capital

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, and data-driven R&D processes. This suggests that data assets are largely subsumed—though not explicitly identified—in measures of intangible capital covering the full range of assets. 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. For this we refer to the latest EUKLEMS & INTANProd database (LLEE 2023), which as previously mentioned incorporates all Corrado et al. (2005) 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 in official measures of capital formation and GDP.

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

Note. Column 3 indicates whether the asset type is currently included as investment in national accounts (NA). In national accounts, only databases generated by firms for their own internal use or embedded and sold as software products are included.
Source: Updated version Corrado, Hulten and Sichel (2005) as set out in Corrado (2023).

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; in addition, as a matter of practice, outright purchases of data stores and databases are included only if they are embedded and sold as software products.

Consider now data intelligence, the most valuable, and final, stage of the data value chain per figure 1. This is where the utility of the intangible capital framework becomes especially apparent. The data stack pertains to all knowledge created from data, i.e., it encompasses all modern, data-driven services product development, marketing and business operations intelligence. Compared with national accounts, which may contain the software tools and databases used to produce this intelligence, intangible capital inherently contains the production of this intelligence from "soup to nuts" via, e.g., its inclusion of investments listed on lines 7, 8 and 9 of table 2—investments in new financial product development, marketing, and organization practices.

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 and business operations—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. Investments in business operations are a component of line 9, which includes outlays for the re-design of computer systems and network platforms to facilitate the ingestion and transformation of raw data.

The intangibles framework thus covers most, if not all, forms of data intelligence. Indeed, while virtually all assets in table 1 are potentially data driven, the intangible framework can be used to inform the development of empirical estimates of data—and especially data intelligence. Approaches that lack this perspective, have missed key application areas of modern data science. For example, the Statistics Canada (2019a, b) implementation covered financial and marketing forms of data-derived intelligence but did not include engineering design and business operations.

Engineering design is generally recognized as a key factor in innovation. A meta-analysis of the joint evolution of engineering design and data science concluded that *data-driven* engineering design has become increasingly more prominent due to the 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).

In summary, the rise in use of data suggests that the composition of intangible investment is becoming more data intensive. This is especially true for investments in new financial products, industrial design, branding and marketing, and organizational processes, i.e., "data-intensive intangibles." Beyond this main message, 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 stores, purchased databases, and most forms of data intelligence are not currently captured in official statistics.
- 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.

3. Measuring Data: Approaches and methods

The economics literature has used diverse frameworks and approaches to measure the value of data. These frameworks and approaches are reviewed in Corrado, Haskel, Iommi and Jona-Lasinio (2022a, 2022b). The reviews cover how concepts used in company financial accounts differ from national accounts and how methods used to measure data vary according to the scope, context, and economic sector of origin of the data to be valued (i.e., personal or institutional). Takeaways from these reviews and an overview of the approach used to measure the value of data in this paper follow.

Takeaways from Prior Reviews

The economics literature has taken three main directions to develop estimates of the value of data. These include (a) approaches based on consumers' valuations, (b) approaches based on firm valuations and market transactions, and (c) approaches based on estimates of resource costs used to produce data assets.

The first takeaway from our prior reviews is that approaches aimed at valuing consumers' personal information will not encompass the full data value chain of figure 1, which includes data originating in business and government organizations. But because some of the largest and fastest growing tech companies are built mainly on the economics of transforming personal information into business and marketing intelligence, the valuation of personal data is viewed with keen interest. The World Economic Forum (WEF, 2011) and OECD (2013) identified two broad categories of data—personal data and institutional data—based on the economic sector of origin of the information. These categories are not very amenable to measurement.

A "personal data" value chain can be thought of as a construct that sits within the figure 1 data value chain in which public open data and business-specific information also reside and contribute to value creation. Because the overall value of data in economies derives, at least in part, from the combination of personal data with institutional and public open data, the value of personal data *as an economic resource* cannot be readily disentangled from the value of other data records in an economy.

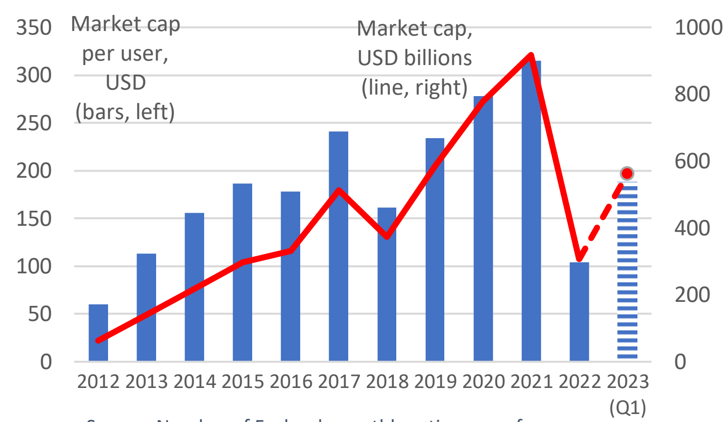
The second takeaway is then that it Is necessary to adopt a method that yields comprehensive coverage of data use in market activities. Ideally this implies using methods based on market prices and market valuations. Market prices paid and received in actual transactions are the best proxy for quantifying the value of data.

Adopting a market-based approach faces many obstacles, however. There is no well-defined market for most types of data, and when information on transactions is available, valuations are highly context dependent. Moreover, market transactions in unprocessed data will not capture the entire transformation chain necessary to generate values digitized information.[3] Studies of digital platform companies (e.g., Li, Norei, and Yamana 2019) underscore that the valuation of a company's data assets is highly dependent on the degree of vertical integration in a company's data value chain—consistent with the data stack in which only the monetization of data intelligence captures the value of the entire chain.

That said, an extensive line of research has examined the market capitalizations of companies that derive most (or all) of their income from advertising linked to personal data. 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.

Figure 2 illustrates that the volatility in market capitalizations of individual companies is a major pitfall of using financial market indicators as a measurement tool. The figure plots the value of an individual (active) record at Facebook/Meta (the bars in the figure), which has fluctuated with the company's market capitalization (the line in the figure). As a result, the value of a typical Facebook user's data, which averaged nearly 200 USD from 2013 to 2023 (Q1), stood both well above, and well below, that average at times (e.g., more than 300 USD in Q4 of 2021 but only 100 USD in Q4 of 2022).

Figure 2. Facebook/Meta Market Capitalization per Active User, 2012 to 2023 (Q1)



Source: Number of Facbook monthly active users from Statistica, Q4 of each year. Market capitalization is Nasdaq value for December of each year (except as noted for 2023).

A third takeaway is that, absent comprehensive arms-length transactions, a resource cost approach provides the best way forward for valuing data in an economy.[4] National accounts estimate investment

[3] Large-scale market transactions 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. Third-party data sit near the beginning of the value chain and their valuation will not reflect the entire chain.

[4] Though the approach differs substantially in context and application from the cost-based valuation method used in financial accounting, the concepts do overlap. See the previously cited reviews for further elaboration.

by asset type based on resource costs. They aim at consistently recording investment flows and capital stocks for each industry (or institutional) sector, which involves estimating values for all sources of supply for each asset and deriving the asset valuations and quantities using information on prices for newly produced assets and how fast they lose value as they age (economic depreciation).

If firms purchased all or most data assets from market transactions, as they do with tangible assets, measuring the cost of data would be conceptually like measuring expenditures for a construction firm's purchase of excavators and concrete mixers. Instead, most digitized information used by businesses is not transacted on markets but produced in-house. This is also the case for many of the components of intangible investment list in table 2. For intangible assets, compilers must come up with two components of nominal investment in each asset type: own-account investment (when assets are produced and used in-house) and purchased investment (when assets are bought from producers in arm's length transactions).

Information on data products usually is missing in surveys of production or capital spending, and the national accountant's total supply approach is difficult to implement for valuing data as an asset. Statistics Canada (2019a, 2019b) pioneered a practical approach to valuing data assets using the sum-of-costs approach national accountants employ for in-house investments, e.g., own-account software, to comprehensively measure data investments in an economy. This is the approach we use in this study.

Sum-of-costs Approach

The measurement of in-house investment using sum-of-costs methods is as follows: Imagine a firm having a "software factory" inside it, and your task 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). This task begins by identifying the workers in the factory-within-the-factory. Based on knowledge of the compensation paid to these workers, the total payments made to all factors involved in the in-house production can be estimated.

As a practical matter, workers are identified by occupation, and consideration is given to the likelihood they are not 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 50 percent of their time working in their firm's "software factory" to produce original code. In-house production of data assets can be estimated in a similar fashion.

Statistics Canada (2019a, 2019b) prepared experimental estimates for Canada's total economy and major institutional sectors—nonfinancial corporations, financial corporations, nonprofit institutions serving households, and governments. Occupational groups were selected from among those generally associated with converting raw data into digital formats suitable for knowledge creation and monetization.

The Statistics Canada schema included three asset types that generally align with those in the data stack of figure 1, though Statistics Canada called the third category "data science" and viewed it as unmeasured R&D, e.g., spending to develop new AI algorithms. Data and AI data tools are inextricably bound via feedback training data used to develop and/or enhance the performance of AI tools, but the data stack/data value chain notion of how value is created from data does not end with the development of *new* algorithms. Section 2's discussion of the data stack and its overlap with intangible investment suggests that value creation due to data intelligence also occurs when existing AI tools or analytics are fine-tuned with data to obtain firm-specific solutions for product design, services development, marketing campaigns, and business organization processes.

Statistics Canada estimated bands for the value of investment in the three data types that ranged from 1-3/4 to 2-1/4 percent of the country's GDP in 2018. They further found that about 47 percent of the total was accounted for by nonfinancial corporations, 31 percent by financial corporations, 20 percent by governments, and 2 percent by nonprofit institutions serving households. Statistics Canada produced estimates in volume terms (i.e., adjusted to consider price changes) and used them to develop capital stock measures. Price indexes were based on weighted input costs (without a productivity adjustment). Service lives were assumed to be 25 years for data stores, 5 years for databases (the same as software) and 6 years for their data science category.

Goodridge, Haskel, and Edquist (2021) took essentially the same approach as Statistics Canada to estimate the value of investments in data capital for 16 EU countries. 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 gross fixed capital formation by around 60 percent compared to own-account investment in software and databases measured in EU official national accounts. In GDP terms, their estimates are in line with the results for Canada.

Emerging work analyzes skill requirements and work activities (tasks) by detailed occupation to classify workers (Autor 2013), an approach thought to be well-suited for identifying workers engaged in own-account data asset production. In an early application, Squicciarini and Le Mouel (2012) used detailed task-level information on occupations from O*NET to measure own-account investment in organizational capital.[5] Babina, Fedyk, He and Hodson (2022) used skills information from Burning Glass job postings and resumes from Cognism to estimate industry-level AI workers and AI investments at U.S. publicly traded corporations.

Santiago-Calderón and Rassier (2022) also use skills information from Burning Glass but combine this with O*NET's detailed occupation descriptors, and employment and wage data by detailed occupation from the OEWS survey.[6] The authors employ a machine learning model to estimate the number of employees and employment time devoted to data asset production in the U.S. economy. They estimate that data investment was about 1 percent of U.S. business sector gross value added (GVA). Though less

[5] O*NET is a database covering nearly 1000 standardized occupations and their descriptors; descriptors include job requirements (knowledge, skills, and abilities) and work activities and tasks. Information is updated annually to capture changes in occupations as they evolve over time.

[6] OEWS refers to The Occupational Employment and Wage Statistics (OEWS) program of the U.S. Bureau of Labor Statistics, which produces employment and wage estimates annually for approximately 830 occupations.

than the implied estimate of about 1-1/2 percent for the business sector of Canada and market sector industries in European countries covered in Goodridge et al. (2021), the upshot of all three studies is that data investment is rather small.

4. Estimates of Data Investment in European Countries

The sum-of-costs method is used to develop estimates of data asset creation coherent with the value chain framework of section 2 and national accounts. Broadly speaking, the estimates are developed from occupation employment and wage share data from Labor Force Surveys conducted in nine European countries; these surveys use the International standard classification of occupations (ISCO) system to define occupations.

Features of the application

The identification of workers engaged in producing data includes workers engaged in data-driven engineering design, business operations, and marketing—in line with the discussion of data assets and intangible capital in section 2. This is a key feature of our approach; prior works excluded, or only partially included workers in these business functions in their estimates of the value of data.

Our occupation identification procedure exploits information from four primary sources (1) the results of the skills-based studies described above (Babina et al 2021; Santiago-Calderón and Rassier 2022), (2) the ISCO and detailed O*NET occupation descriptors, (3) detailed employment data by occupation from OEWS, and (4) a concordance of the ISCO system for classifying occupations in European economies and the SOC system used in North America. The skills-based studies and O*NET and OEWS data use the SOC system, for which more detailed information and estimates are openly available than for ISCO.

In addition to data assets, we generate sum-of-costs estimates of investment in software assets. National accounts combine software and databases in a single published asset category using a total supply approach, and these estimates are incorporated in the measures of intangible investment in EUKLEMS & INTANProd (and in this paper). The four sum-of-costs investment series` are generated for 14 NACE "letter" level nonagricultural market sector industries of nine European, mainly western, economies.[7] Estimates cover the years 2010 to 2019. An appendix to this paper provides further details on the procedures and data sources used to develop the data and software sum-of-costs estimates reported in this paper.

The estimates generated in this paper capture values produced in the market sector regardless of whether the produced output is intended for own final use or final sale. The produced value of data is a good proxy for data capital investment in the market sector if data transactions between the government and the market sector are rather small.

[7] Market sector industries exclude NACE sections L (real estate activities), O (public administration and defense; compulsory social security), P (education), Q (human health and social work activities), and T (activities of households as employers; undifferentiated goods- and services-producing activities of households for own use).

Strictly speaking, estimating market sector investment requires adjusting the estimates of in-house production for imports and exports of data flows as well. Information on transactions in data stores and databases, whether in domestic sales or international trade, is not readily apparent in official statistics (Ker and Mazzini 2020). This is mainly because official statistics are focused on industry activity, and industry classifications generate inexact identification of data production activities, i.e., Zillow sells its data on home real estate valuations, Nielsen sells it consumer survey data, as do credit agencies such as Experian, but these firms engage in widely different primary production activities and classified accordingly.

Empirical results

Our main results on the relative size and growth of market sector data asset production and intangible investment are shown in figures 3(a)–(d). Data asset production is shown according to the three segments in the data value chain in the upper panel of the figure. The total data value chain averages 7.8 percent relative to nonagricultural market sector gross value added (GVA) in the covered countries and years. The United Kingdom is the most data intensive of the countries included (9.1 percent), and Italy and Spain are the least (5.2 and 6.5 percent, respectively). Data intelligence is the largest segment in the data value chain, with databases the smallest.

The middle panel of figure 3 suggests that resources allocated to data asset production (the solid blue bars) were less than domestically produced intangible assets in 2019 (the solid gray bars). Averaged over countries (without regard to size of country), data asset production averaged 50 percent of intangible investment in 2019 and close to 50 percent of intangible asset domestic production. The panel also shows our estimates of the value of domestically produced software investment (the solid red bar stacked above the data asset production bar). The sum of data production and software production, which reflects in part the production of data tools, is 64 percent of intangible investment and 73 percent of domestic production of intangibles. Software production activity is found to be much smaller than production activity in the data value chain, reflecting the fact that a significant share (about ½) of net imports of intangibles is software.

The GVA shares of our sum-of-costs production estimates for software and databases are compared with those for official investment estimates for the combined category in the bottom left panel of the figure 3, in panel (c).[8] The correlation of our sum-of-costs estimates is not all that strong (R^2 of the simple regression line is .22), but there is a positive and significant intercept in the relationship shown on the figure, in line with the fact that that our sum-of-costs estimates miss net imports.[9]

---

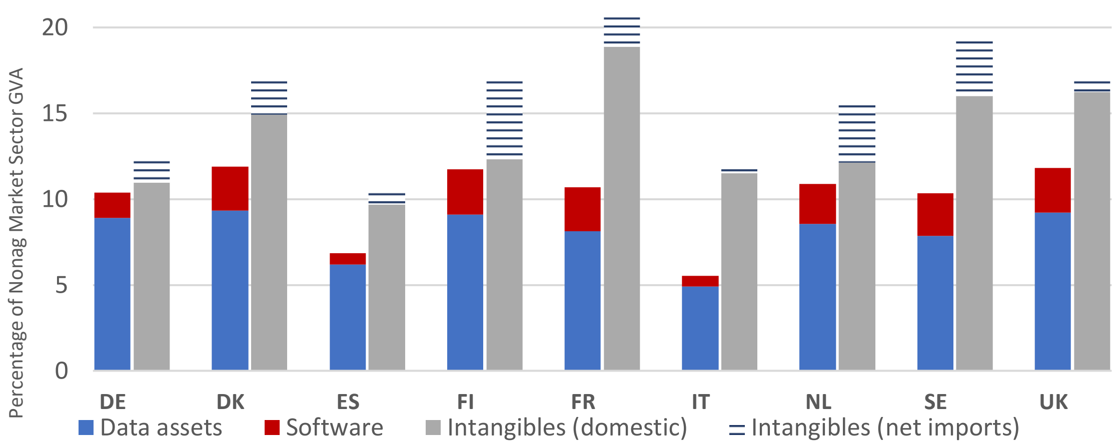
[8] Recall, national statistical agencies do not report software assets separate from database assets.

[9] Figure 3 (c) also suggests that the differences in reported software and databases investment rates across EU countries are as likely to be real rather than obscured by documented differences in measurement methods (see discussions in Colecchia and Schreyer 2001, Timmer, Inklaar, O'Mahony, and van Ark 2011, and Goodridge et al. 2021). Methods used to develop our data and software production estimates are harmonized (i.e., are the same; see appendix) across countries, yet notable differences remain.

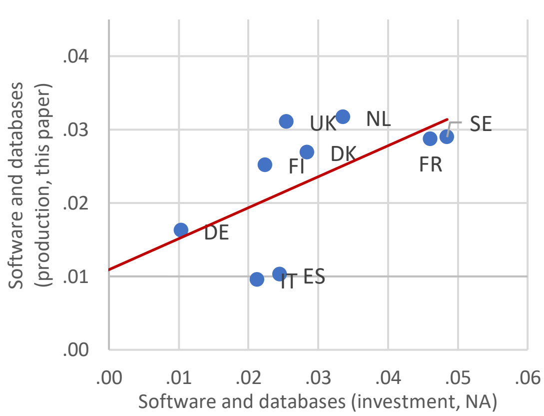
Figure 3. Estimates of the Data Value Chain

(a) Estimates for nine European countries, 2010-2019 average.

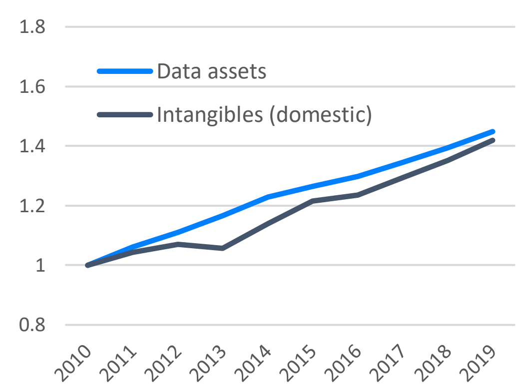
(b) Data and software production versus intangibles, 2019



(c) GVA shares, Software & databases, this paper versus national accounts (NA)



(d) Relative growth (2010=100), country-level estimates aggregated using PPPs)



European country abbreviations are as follows: Denmark (DK), Germany (DE), Finland (FI), France (FR), Italy (IT), Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK).

Table 3. Sectoral distribution of investment in data and total intangibles, percentages of sector gross value added in nine European countries, 2010 to 2019.

		Data Investment	***Intangible investment***
	Selected industry sectors	*(1)*	*(2)*
1.	**Professional, scientific & technical activities**	16.3	27.0
2.	**Information and Communication**	14.2	28.2
3.	**Financial and Insurance activities**	14.5	14.2
4.	**Manufacturing**	7.3	21.6
5.	**Administrative & support service activities**	4.8	11.1
Memo:			
6.	**Nonagricultural market sector**	7.8	16.9

Note: Each cell represents the unweighted average of investment as a percentage of sector gross value added over time and countries. Industries shown correspond to NACE letter sectors M (row 1), J (row 2), K (row 3), C (row 4), N (row 5) and B to K, M, N, R, and S (row 6). European countries include Denmark (DK), Germany (DE), Finland (FI), France (FR), Italy (IT), Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK).

The bottom right panel of figure 3, panel (d), illustrates that estimated nominal data production (aggregated over countries using PPPs) grew a 5 percent faster than intangibles through 2015, but not thereafter. Cost efficiencies enabled by data-driven forms of intangible investment imply that the *real* growth of data assets may have eclipsed that of overall real intangible investment, however; we discuss the rationale for this thinking and evidence for it in the next section of this paper.

Table 3 shows the sectoral distribution of data capital investment (column 1) compared with intangible investment (column 2) for 5 groups of NACE letter industries. Average rates for all nonagricultural market sector industries are shown in the memo item (line 6).

The most data intensive industry groups are shown on lines 1 to 3. These consist of the professional, scientific, and technical activities; information and communication services; and finance and insurance activities sectors. Along with manufacturing (line 4), these sectors also post high rates of intangible, knowledge-based investments (column 2 of the table). The manufacturing sector invests disproportionately in R&D compared with other intangibles, however, suggesting that R&D processes (in manufacturing) are less data intensive than business functions such as marketing, and supply logistics that are more predominant in services industries.

Data asset production by *type* of data also differs across industry sectors (not shown), but consistent with figure 3 (a), the most data-intensive sectors are those with high contributions from data intelligence. This underscores the reasoning from our analysis of the data stack, that data intelligence is the final and most valuable stage of the value chain.

Data capital and intangible capital: How much of an overlap?

Our section 2 discussion suggested that measured data capital and intangible capital overlap significantly and that, conceptually, data capital is subsumed within intangible capital, especially in its "data-intensive" components: new financial products, industrial design, branding and marketing, and organizational processes. This idea is supported by simple correlation tests indicating that data capital and intangibles are moderately linked (correlation coefficient is 0.5) and that correlation is stronger with data-intensive intangibles (correlation coefficient increases to 0.6).

To further explore this evidence, table 4 reports the results of a regression analysis of the relationship between labor productivity, data capital, and intangibles. The relationship is explored within a simple production function framework with controls for country, industry, and time fixed effects.

Table 4. Regression model estimates of the contribution of the growth in data and intangible capital deepening to labor productivity growth

	(1)	(2)	(3)	(4)	(5)
Intangible capital	.216***	.123***			
	(.027)	(.031)			
Tangible capital	.168***	.029	.166***	.034	.034
	(.029)	(.036)	(.029)	(.036)	(.036)
Data capital		.325***		.331***	
		(.05)		(.051)	
Data-intensive intangibles			.033**	.005	.007
			(.016)	(.016)	(.016)
National accounts intangibles			.073***	.05**	.049**
			(.02)	(.02)	(.02)
Training			.132***	.061**	.067**
			(.027)	(.029)	(.029)
Databases					-.006
					(.037)
Data stores					.005
					(.061)
Data intelligence					.316***
					(.065)
Observations	6614	6522	6680	6588	6588

Standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*
Note: Dependent variable is labor productivity computed as change in the natural log of value added per hour; value added is adjusted to include intangibles not currently capitalized in national accounts. All explanatory variables are in similar, i.e., per hour, delta natural log terms. Estimates are generalized least squares. All columns include time, industry and country fixed effects.

Column 1 is our benchmark specification showing that intangible and tangible capital are statistically significantly associated with labor productivity growth. Adding data capital in column 2 reduces the coefficient of intangibles by 50 percent and renders tangible capital statistically insignificant. We next run a Wald test to check if data and intangible capital coefficients are equal. The Wald test indicates that the null hypothesis of perfect equality can be rejected at 0.05 significance statistical level, implying that both capitals contribute to explaining labor productivity growth.

We than examine the overlap in terms of components of intangibles: data-intensive intangibles and other intangibles, with the latter distinguished between components included in national accounts and training (the only component not included in national accounts or the data-intensive intangibles grouping). These results are reported in columns 3 and 4, which shows that once we control for data assets, data intensive intangibles lose their statistical significance. This suggests that the data-intensive grouping of intangibles and data capital capture similar factors affecting productivity growth. Finally, column 5 looks at the individual categories of data capital, which reveals that the high statistical significance of the relationship of data capital with labor productivity growth is mainly driven by data intelligence.

Data and SNA guidance

International guidelines for compiling GDP given by the 2008 System of National Accounts (SNA) will be soon superseded by the planned 2025 SNA. This new SNA is expected to expand the types of intangible investment in GDP to include data. Though not officially endorsed as of this writing, the guidance of the United Nations digitalization task team (DZTT) is to create a new asset category called "data" and to report it separate from current category of computer software (United Nations DZTT 2023). The guidance defines data in a manner generally consistent with the data stack of figure 1 and calls for a sum-of-costs approach be used for measuring investment in data.

The analysis in the SNA guidance depicts a data value chain with three stages where the final stage is termed "insights" (rather than "data intelligence"). Studies investigating the measurement of data in preparation for the 2025 SNA (e.g., Statistics Canada 2019a, 2019b; Santiago-Calderón and Rassier 2022) conceptually include something like what we have set out as investment in data stores and databases but not the full scope of the data value chain including data intelligence.

The omission of data intelligence seems to be the primary reason why prior studies produce estimates of data investments that are much smaller than those developed in this paper. In a value chain the output of final stage reflects the further processing of the output of earlier stages. It is possible that national account practitioners interpret the final stage output of the data value chain as another good or service entirely, rather than as a data asset as done in this paper. If so, the distinction they are making is semantic because the output of the data value chain is still capital to the extent its use in production is long-lasting.[10] The distinction seems to be reinforced elsewhere in the 2025 SNA deliberations. Guidance from the Joint Globalization task team (GZTT) suggests that marketing assets may also be included in investment if measuring them is determined to be feasible (United Nations GZTT 2023). This of course is coherent only if marketing assets do not overlap with data assets, which goes against the logic of the data stack as well as the empirical analysis in this paper (e.g., table 4 and the correlations reported in the appendix).

[10] This logic is of course symmetric with that used for tangible capital, e.g., an engine block made by a machine tool is still a capital good (despite its difference in character).

The exclusion of data intelligence notwithstanding, the very small size of estimates generated by prior studies follows from the approach used to identify occupations for estimating the costs of data asset production—occupations with job and/or skill descriptors that include the word "data" in them (e.g., Customer Data Integration, Customer Service Database, Data Management, and Data Science to name just a few). Though we include these occupations in our analysis, we also allocate fractional amounts of the value of time of selected managers, professionals, and technicians to the production of data assets. To justify this approach, we appeal to both the data stack *and* the intangible capital framework—but note, the SNA guidance language also provides support.

The guidance by the United Nations Digitalisation Task Team (UN DZTT) for the 2025 SNA lists some specific costs that data asset investment should include, and these costs include more than the costs of digitizing and organizing records into databases for visualization and application of analytics. For example, the "costs of planning and developing a data production strategy" (UN DZTT, 2023, page 8) are to be included, an activity sure to involve a company's C-suite as well as its computer and data professionals. The costs of "analyzing data for the purpose of drawing conclusions from it" (United Nations DZTT, 2023, page 8) are also to be included, an ongoing activity that involves multiple types of managers and technical professionals that may, or more likely may not, have "Data" in their job descriptions.

Finally, as previously discussed, modern data use fosters faster, more efficient experimentation and feedback in business processes. Existing SNA guidance does not address how the relative efficiency of data assets is to be captured in price statistics and thereby reflected in real GDP and productivity calculations. This is a major omission and a topic discussed in some depth in the next part of this paper.

Part II: Data and Productivity

5. Data Investment, Growth Accounting and Innovation

The aggregate effects of the rise of data capital are analyzed using the upstream/downstream two-sector model summarized in Corrado et al (2022a). The model is based on Corrado, Hulten, Sichel (2005, 2009) as adapted and termed "upstream/downstream" in Corrado, Haskel, and Goodridge (2011).

As previously suggested, data affects innovation and productivity growth in divergent ways. On the one hand cutting-edge digital tools that exploit bigdata have the potential for making production and innovation processes more efficient. On the other hand, the data assets created by them may be inextricably bound with network externalities in customer demand that weaken competition and/or, due to the difficulty to replicate proprietary data assets, weaken knowledge diffusion in economies. This section investigates how these two forces—the "efficiency" promise of bigdata/AI versus the "appropriability effect" (that restrains TFP growth)—operate in a two-sector model with data/intangible capital.

Upstream/downstream model of an economy

A simplified model of an economy with data as an intangible asset divides production into two broad sectors: (1) an "upstream" sector that *produces new knowledge* that can be commercialized, e.g., a new or improved product design (or product formula), or a software program adapted to the needs of the organization; and (2) a "downstream" sector that *uses the knowledge* generated by the upstream sector to produce final output.

Box 1 sets out the upstream/downstream model in more detail, including sectoral inputs and their payments, sectoral outputs and their prices, and sectoral productivity. Intangible investment is the value of the upstream sector's output in this model—the investment stream corresponding to data asset creation in our prior discussion. The outstanding stock of data assets is then the accumulation of upstream output after adjusting for losses due to ageing (economic depreciation).

The downstream production sector uses the stock of data-derived intelligence to produce final goods, and the upstream sector is remanded a portion of the income earned from the sale of final goods in return. Because knowledge producers demand (and earn) returns on their investments, the value of the data knowledge stock must be included in calculations of the realized return to capital, which is arbitraged across sectors and asset types in competitive equilibrium. As mentioned previously, not recognizing data and other forms of intangible capital in calculations of the realized return to capital can cause observers to mischaracterize the competitive intensity of markets and industries.

To the extent there are pure rents from innovation in this model, they create a wedge between asset prices for data capital (P^N) and its production cost; by extension (see Box 1), they enter the per period remand paid by downstream producers for use of the data capital (P^R). The model thus allows for innovators/data capital owners to hold temporary market power, a common feature of many economic models of innovation, especially Schumpeterian-inspired models such as Aghion and Howitt (1992). In these models, innovation results from entrepreneurial investments motivated by prospects of monopoly rents.

The temporary nature of the market power is due to the inherent nonrival character of knowledge-based assets. As valuable commercial knowledge diffuses (is copied/replicated), innovator profits are competed away. This loss of revenue-generating capacity forms the conceptual basis for the relatively short service lives found for intangible capital in empirical studies (reviewed in De Rassenfosse and Jaffe 2017) and surveys (e.g., Awano et al. 2010).

Data capital in GDP and growth accounting

Without the capitalization of data assets, GDP consists solely of downstream sector output Y, but when upstream investments in building data stores, databases and developing data intelligence are capitalized, aggregate value added Q reflects production in both sectors:

(1a) $$P^Q Q = P^Y Y + P^N N = P^C C + P^I I + P^N N$$

(1b) $$\equiv P^L L + P^K K + P^R R$$

Box 1. A Model of an Economy with Intangibles

A simplified model of the economy divides production into two broad sectors: (1) an "upstream" sector that produces new knowledge that can be commercialized; and (2) a "downstream" sector that uses the knowledge generated by the upstream sector to produce final output. For simplicity, the model assumes that there are no exports or imports of intangible assets and no intermediate purchases of other goods and services.

Sectoral activity is described and denoted as follows:

- Upstream output reflects the production of new commercial knowledge. This is also intangible investment, which in volume terms intangible investment is N and in nominal terms is $P^N N$, where P^N is a price index for intangible assets.
- Downstream output reflects the production of (tangible) investment and consumer goods, $P^Y Y$, or $P^I I + P^C C$, in nominal terms.
- The outstanding stock of commercially valuable knowledge reflects the accumulation of the upstream sector output after adjusting for losses due to economic depreciation (i.e., ageing), i.e., the stock of intangible capital, R, is given by the perpetual inventory relationship, $R_t = N_t + \delta^R R_{t-1}$.
- Freely available basic knowledge, scientific or otherwise, is represented by R^{Basic}. It is an input to upstream production, e.g., open-source software, which we assume is produced outside the model. (This assumption can be relaxed with no major change in model implications.)
- The value of intangible capital, defined as its replacement cost, is given by $P^N R$. The payments made to the owners of R are denoted by $P^R R$, where P^R is the per period rental price equivalent of using intangible capital in production.
- The stock of tangible assets is denoted by K, its value by $P^I K$, and payments to owners by $P^K K$. Labor inputs and their price are L and P^L, respectively.
- Total factor productivity in the upstream and downstream production functions is given by A^N *and* A^Y

Regarding monopoly power:

- R is inherently nonrival and thus only partially appropriable. Appropriability lasts for the time the producer-innovator can sell or rent the knowledge to the downstream sector at a monopoly price.
- The downstream sector is assumed to be a price-taker for knowledge, i.e., monopoly power resides in the upstream sector. Final output prices for consumption and tangible investment are assumed to be competitive, as are factor input prices for labor and tangible capital.

The sectoral production and income flows in this economy are written as follows:

(B1-1) $$N = A^N F^N(L^N, K^N, R^{Basic}); \quad P^N N = P^L L^N + P^K K^N + \pi^N$$

(B1-2) $$Y = A^Y F^Y(L^Y, K^Y, R); \quad P^Y Y = P^L L^Y + P^K K^Y + P^R R$$

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

In this model, 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 $P^R = (r + \delta^R)P^N$ ignoring asset price inflation. The user cost of tangible capital is similarly linked to its asset price. The model is closed via arbitrage of returns (r) across sectors, i.e., returns to investments in innovation (that build intangible capital R) with returns to alternative long-term investments (that build tangible capital K).

The model allows for the existence of "abnormal" innovator profits for periods of time, but intertemporal arbitrage operates to constrain innovator profits to zero (i.e., $\pi^N = 0$) in long-term equilibrium. As a practical matter, with continuous entry of innovators (and waves of technological change), the model is consistent with varying degrees of market power continuously embedded in time series for intangible asset prices.

As seen in (1a) to the right, investment in final demand is expanded to include data value creation and thus GDP is larger. Factor income, the second line (1b), accounts explicitly for returns to intangible assets in total capital income. The term may contain monopolistic returns to innovation in the price element P^R 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:

(2) $$dq = \sigma_Q^X dx + \sigma_Q^R dr + da$$

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 data/intangible capital.[11] 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 $\sigma_Q^R dr$, plus total factor productivity (TFP) growth da. What is different in this model then is that the contribution of paid-for data capital has become a source of growth.

Data capital and knowledge diffusion

The intangibles framework also helps explain the origins of TFP growth, and this is no less true when the framework is applied to data capital. Unappropriated returns that the economy enjoys when knowledge-based assets are copied and used at low-cost in production elsewhere in an economy are a source of growth in measured TFP. The costless diffusion (or "spread") of innovators' knowledge from one organization to another—a phenomenon termed "knowledge spillovers" by Griliches (1992, 1994) in the context of R&D—drives the increasing returns on investments in knowledge that play a central role in modern growth theory (Romer 1990, Jones 2005).

From this perspective, whether data are proprietary or freely available (per the range of examples given in table 1) becomes crucial for assessing the productivity implications of data assets. Consider how Lyft was able to duplicate and compete against Uber's innovative, data-enabled ride-sharing business model. The idea of ridesharing as a business model was freely available once Uber became a fast-growing enterprise. So were the mapping and traffic data needed for ridesharing implementation because governments make this information freely available. But when data-enabled innovations are based on proprietary data (Amazon's very efficient delivery system, Google's targeted advertising systems, etc.), they operate more like trade secrets than patented technologies. After all, patented technologies are disclosed when filed and protected only for a time. Innovations stemming from trade secrets are not easily (or ever) duplicated, and first-mover advantage may be maintained.

So, when proprietary data-derived knowledge assets are a prevalent source of innovation, knowledge diffusion—and TFP growth—will weaken. This is the "appropriability" effect of data capital, i.e., unless

[11] The decomposition is obtained via the usual log differentiation of (1b) assuming constant returns to scale and that factors are paid their marginal revenue product. The notation "dz" is the log change in "Z" where Z is any variable in the model. Conventional inputs K and L are combined as X and weighted appropriately.

offset by moves to promote industry data sharing, an increase in the share of data-derived intangible capital in total intangible capital will lead to lower measured growth of total factor productivity due to fewer spillovers from a given stream of investment.[12]

The "appropriability effect" is not the whole story of the impact of data on innovation but it has much potential for being very significant. Aggregate productivity in the upstream/downstream model can be expressed as

$$da = s_Q^Y da^Y + s_Q^N da^N \tag{3}$$

I.e., the share-weighted sum of total factor productivity growth in each sector. To the extent proprietary data assets are like trade secrets and generate commercial knowledge that is not easily replicated at low cost, the appropriability effect operates largely via its impact on da^Y, the first term in (3). The production shares in figure 4b imply that the weight on downstream productivity s_Q^Y ranges anywhere from 80 to 90 percent, so even small changes in da^Y have significant impacts on aggregate (measured) productivity da.

Data capital and data technologies

The efficiencies of modern data technologies are an opposing force to the diminishment of productivity spillovers to investments in intangible capital. To the extent the latest wave of AI-driven digital technologies cum data assets produces innovations more efficiently, the second term in (3), upstream total factor productivity da^N, is boosted. Though the impact will become larger as the production share of data capital in intangibles increases, it is also possible that data and data technologies create innovations and efficiencies that are impactful enough to offset the heavily weighted, diminished pace of da^Y.

As the composition of intangible capital becomes, in effect, data capital, the relative efficiency of data capital will be "seen" as lower relative prices for intangible assets. The extent of this relative decline is the "efficiency effect" of data capital. This is analogous to the situation with ICT capital, whose relative efficiency is a familiar theme in the productivity literature. In the initial phases of ICT innovations in the 1990s through the rapid adoption of mobile by the early 2000s, ICT capital asset prices fell very rapidly—anywhere from 10 to 20 percent. These price drops were indicative of the relative productivity of the asset class.[13]

6. Intangible Asset Prices

[12] A secondary aspect of this effect is that proprietary bigdata will create longer-lasting positions of competitive advantage (all else equal), which implies that data-derived knowledge stocks have longer service lives (i.e., lower values for δ^R).

[13] See, for example, Byrne and Corrado (2017a, 2017b) for this analysis of ICT capital and the measurement of its prices.

Let us then look at prices of intangible assets to see what they may be signaling about the relative productivity of data capital.

Intangible asset price change in the upstream/downstream model

The change in intangible asset prices dp^N is obtained by log differentiation of the upstream factor payment equation (shown in the Box as equation B1-1), yielding:

(4) $$dp^N = \sigma_N^L dp^L + \sigma_N^K dp^K + \sigma_N^\pi d\pi^N - da^N \ .$$

This equation expresses dp^N as a weighted average of changes in constant quality unit input costs (the first two terms) and innovator profits (the third), both of which are offset by changes in the efficiency of upstream production (the last term).

Consider now how changes marginal productivities of factor inputs may offset changes in input costs. Let upstream input, L^N, be an aggregate of services provided by a range of worker types that differs from its "transactions unit" equivalent, hourly input H^N, by a composition effect, Θ. The composition effect accounts for differences in the marginal productivity of different worker types employed in production; thus, aggregate labor input may be expressed as $L^N = \Theta H^N$. A similar logic applies to capital inputs but, for simplicity, in what follows we combine the impact of its productivity with upstream TFP in a single term, denoted as db. In effect, we are simplifying the modeling of upstream production to be a function of multiple labor types and "free" technology. "Free technology" includes all aspects of data technologies, e.g., open AI engines that engender data production cost efficiencies; cloud technologies that enable computing and data processing efficiencies; and trends in the availability and use of open-source software (OSS) and open data.

This simplified data capital production model enables equation (4) to be rewritten as

(4') $$dp^N = \sigma_N^L(dw - d\theta) + \sigma_N^\pi d\pi^N - db$$

where the impact of *upstream* labor input composition changes on data/intangible asset prices is explicit in the term $d\theta$. Grouping the terms in (4') according to factors that "amplify" versus "restrain" price change, we have,

(5) $$dp^N = (\sigma_N^L dw + \sigma_N^\pi d\pi^N) - (\sigma_N^L d\theta + db)$$

where first group of terms includes increases in wages and innovator profits, factors that amplify price change. The second group reflects the combined effect of upstream capital deepening, improved marginal productivity of the sector's average worker, and upstream total factor productivity, factors that are restrain price change.

Relative efficiency of data capital: Some evidence

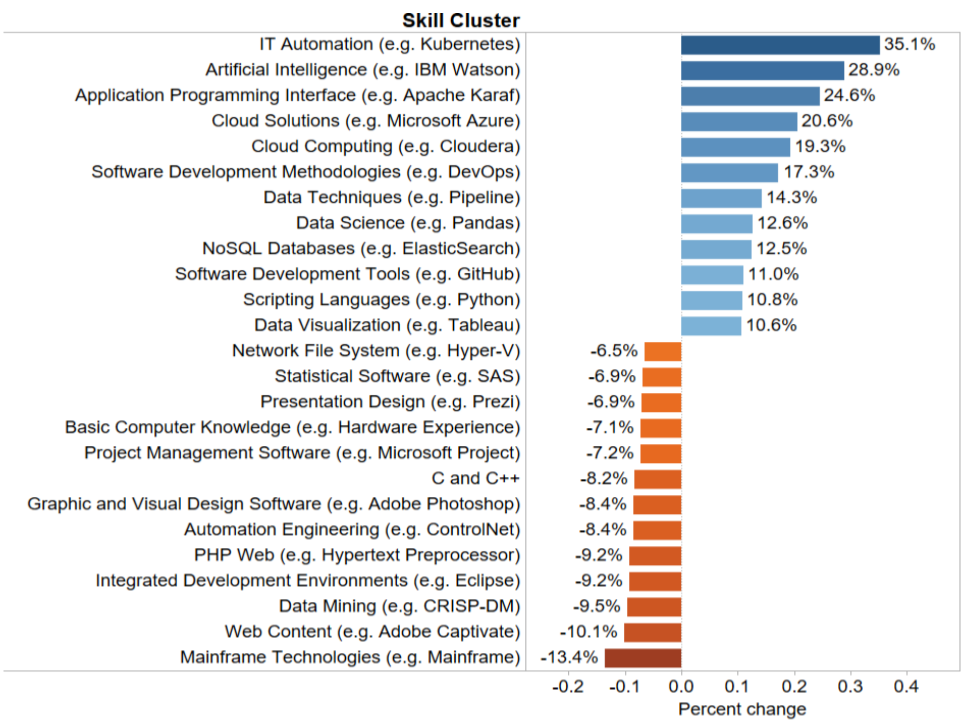
Equation (5) suggests how data and data technologies might affect intangible asset prices. Evidence that the impact of data technologies on intangible capital asset price change might be rather powerful include: (a) strong relative demand for skills used in the production of data capital, (b) stunning growth in the availability of open-source software based on data technologies, and (c) direct indicators of data production capital cost efficiencies, i.e., costs of algorithm design, cloud computing, and advertising/marketing media. This evidence is reviewed below:

<u>AI/Cloud systems skill demand</u>.

Skills related to automation, AI, data connectivity, and cloud storage/computing is reshaping IT work. Direct evidence of employer demand for these skills—is suggested by figure 4, which shows that the demand for AI and cloud systems skills accelerated the fastest among IT roles during the pandemic. To the extent this shifted the composition of the upstream workforce, it boosts $d\theta$ and suggests that workforce composition changes associated with increased data use may have significantly offset wage pressures on asset prices for data capital.

Figure 4. 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 (2021) based on The Conference Board®-Burning Glass® Help Wanted OnLine® (HWOL) data series.

The figure also underscores that upstream labor composition effects are moves *within* the usual grouping of workers termed "high-skilled" in measures of labor composition used in practical growth accounting. The latter are developed using broad groupings of employment by worker type, implying that the usual growth accounting understates the contribution of upstream labor composition to labor productivity growth and, consequently, thereby elevating da^N.

Open-source software.

Studies that quantify the resource cost of open-source software (OSS) activity suggest significant value creation, much of which is arguably correlated with the production of data capital.

Robbins et al. (2021) set out a sectoral framework for measuring investments in OSS in GitHub repositories, where much cutting edge open-source software is held. They use software engineering metrics (lines of code and project complexity) to estimate OSS resource cost in terms of global person-months of effort, enumerating results by country from 2009 to 2019. For the United States, their person-months estimates translated to 38 billion dollars in new OSS investment activity in 2019, having grown nearly 20 percent per year from 2014 to 2019—and likely boosting da^N (though to an unquantified degree).[14] Their estimates of person-months of effort in value creation for the nine European countries for which data capital production estimates were reported in Section 4 grew about 22 percent annually, moving the level of person months in these countries from 87 to 94 percent of US person-months from 2014 to 2019.

The very rapid growth in the value of OSS in GitHub repositories owes, at least in part, to the relative growth in AI applications in overall OSS software. AI application software ranges from general purpose algorithms to specific application-tuned systems, e.g., the software that runs industrial Internet-of-Things (IoT) installations and advanced robots. The OECD developed a classification algorithm to determine the fraction of AI GitHub repositories in all OSS in GitHub repositories, reporting very fast relative growth of AI open-source applications based on measures of “commits” (changes to code).[15] Despite their estimates of stunning growth of AI software in OSS repositories (a three-fold relative increase), the OECD also reports that the proportion of AI in the total is still rather small.

Cloud and other efficiencies.

Many studies document improvements in the efficiency of modern cloud systems 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 upstream productivity growth. But the effect will show through in productivity estimates only insofar as these changes in intangible/data capital production costs are captured in intangible asset prices.

---

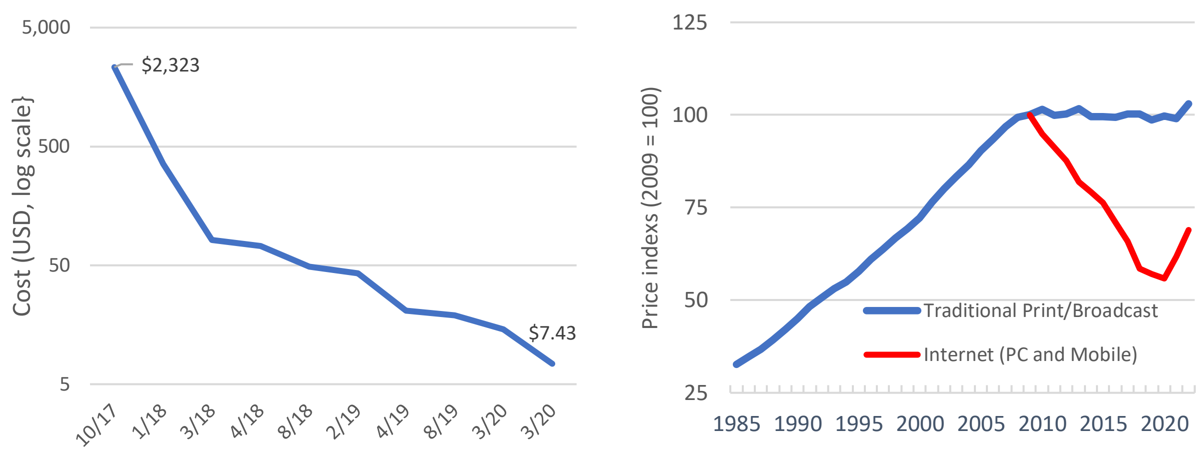
[14] By contrast the U.S. national accounts reports that gross fixed capital formation on software by private sector industries grew 7.5 percent per year from 2014 to 2019.

[15] Barruffaldi et al. (2020, page 32) report “…”in 2010, there were 50 active AI GitHub repositories that had gathered 1,350 commits from contributors, making up 0.26 percent of total commits on GitHub that month. In June 2017, AI software activity had increased to 26,275 commits on 1,533 projects, making up 0.74 percent of total commits on GitHub. …most of [the 2010 to 2017] growth [took] place since 2014. … [from 2014 to 2017], AI open-source software grew about three times as much as the rest of open-source software.”

Indicators of these cost efficiencies are shown in figure 5. According to tests shown in the AI Index Report (Zhang et al 2021, page 49), the costs of training a contemporary image recognition system was "a few dollars in 2020, down by around 150 times from costs in 2017" (figure 5, left panel). This dramatic reduction represents progress in both algorithm design and drops in the costs of cloud-computing resources. Similar factors have affected the accumulation of data on consumer buying patterns and tastes that have lowered (directly and indirectly) advertising media costs (figure 5, right panel) and marketing, though internet advertising media costs reversed course and began to rise sharply in the aftermath of the pandemic (2021 and 2022).[16]

Figure 5. Data-driven Cost Efficiencies

(a) Training cost of image recognition *(b) Advertising media costs, 1985 to 2022*



Sources. Panel (a): AI Index Report (2021, page 49). Panel (b): Corrado (2023), figure 6.

Hard-to-measure services price research typically does not address intangible asset-producing activities—R&D labs, marketing teams, engineering design projects—nor do assessments of productivity mismeasurement view these activities as hotbeds of rapid quality change missed by price collectors. That said, the digital transformation of economies, rise of digitally enabled business models, and increased use of data in business more generally is arguably driving down the production costs of intangible assets.

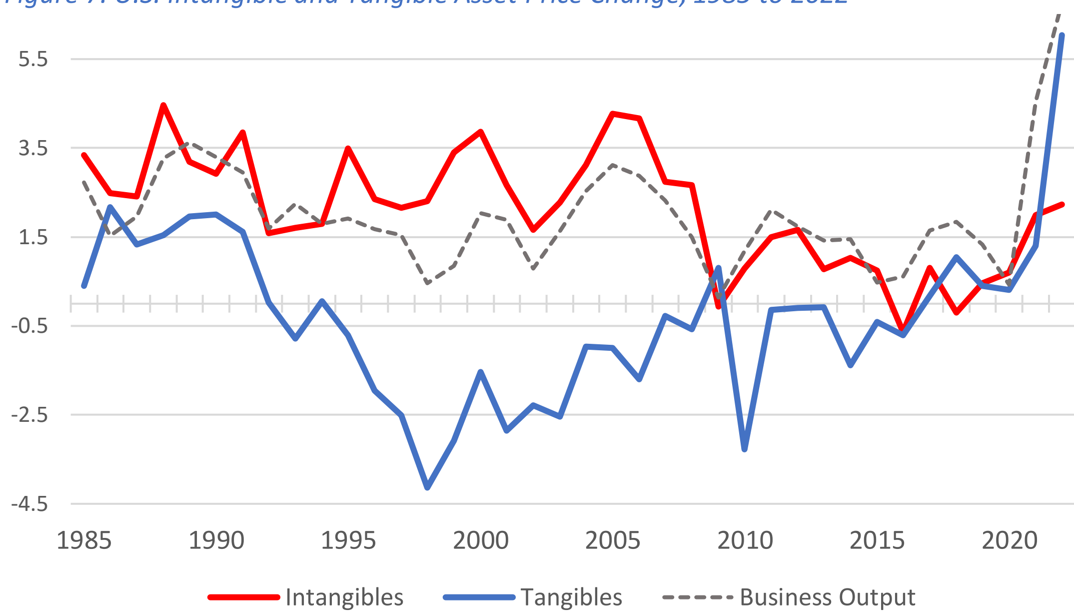
[16] The media cost price indexes are developed from detailed BLS input cost indexes aggregated using information from the Census Bureau and industry sources. The appendix in Corrado (2023) provides additional details.

A new intangible asset price deflator

A price deflator for U.S. intangible investment, reported in Corrado (2023), has been constructed using a brand and marketing investment price deflator calculated using the media input cost price indexes shown in figure 5 (b) and a production/content creation cost price index based on the gross output price index for the advertising and public relations industry (NAICS 5418). Changes in the resulting price index for intangible assets are shown in figure 7 and table 5.

Figure 6 shows that prices for intangible assets exhibit a disinflationary trend beginning in 2009, in line with the prediction that increased data intensity of intangible capital improves its production efficiency and slows its price change. The changed pace of price change mainly reflects the net effects of sharpy slower price change for investments in brand and marketing and in organizational capital (table 5, lines 5 and 6).

Figure 7. U.S. Intangible and Tangible Asset Price Change, 1985 to 2022



Note. Private nonresidential assets. Natural log changes, annualized.
Source. Intangibles price, Corrado (2023); Tangible price, constructed from NIPA table 5.3.4, Business output price, NIPA table 1.3.4.

Table 5. U.S. Intangible and Tangible Asset Price Change, selected periods

		1995 to 2009	2009 to 2019	2019 to 2022
	Asset group	*(1)*	*(2)*	*(3)*
1.	Intangibles	2.7	.7	1.6
2.	Tangibles	-1.7	-.4	2.5
Intangibles, selected components:				

3.	Software	-1.9	-1.6	-1.0
4.	R&D	2.2	1.8	3.4
5.	Brand and marketing investment	3.4	-.8	1.0
6.	Organization process investment	3.0	-.7	.8
Relative price change (asset price/business output price):				
7.	Intangibles, total	1.2	-.8	-2.5
8.	Software	-3.4	-3.1	-5.1
9.	Brand and marketing investment	1.9	-2.3	-3.1
10.	Organization process investment	1.4	-2.2	-3.3

Note: Private nonresidential assets. Natural log changes, annualized.
Sources. Lines 1 and 2 and denominator of lines 7 to 9, see Figure 7; lines 3 and 4, NIPA table 5.3.4; lines 5 and 6, Corrado (2023).

From 2009 to 2019, the relative price of total intangible assets fell 0.8 percent per year in the United States (table 5, line 7) and the relative price of the data-intensive components, investments in branding, marketing and organizational process change (lines 9 and 10), fell 2-1/4 percent per year. During the pandemic and subsequent global inflation (2019 to 2022), the decline in the relative prices of these assets and total intangibles fell even faster, primarily reflecting the sharp rise in overall business output prices (4.1 percent from 2019 to 2022, not shown on the table).

Advances in data technologies have not slowed (e.g., see the 2023 AI Index Report, Maslej et al. 2023), and the composition of intangibles will likely continue to shift toward data assets. This implies that declines in the relative price of intangible assets in the 1 to 2 percent range seem likely to persist for a time, which in turn implies a range for declines in the relative price of data-intensive intangible assets, i.e., data capital, of about 2 to 4 percent per year.

Relative asset prices and "potential" growth in labor productivity

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). In the context of data/AI, this is typically viewed as a question about the degree of substitutability between AI/data capital and human efforts, the limits to which are discussed in Nordhaus (2021).

We have argued in sections 3 and 4 that the rise of modern data capital is mainly a shift in the composition of intangible capital. This suggests data capital may then be viewed as improving the productivity of capital, i.e., it is an *efficiency* effect resulting from the substitution of data capital for other capitals (tangible or intangible). As a first step then, we can estimate the impact of data capital on labor productivity by making assumptions about data capital's relative productivity and income share, assuming labor's share is fixed.

The steady-state solution to the two-sector upstream/downstream model provides a starting point for calibrating estimates of the growth-promoting potential of data capital. To obtain a simple, closed-form steady state solution for this model, simplifying assumptions must be made, mainly, that the sectoral production functions (Box 1 equations B1-1 and B1-2) exhibit constant returns and differ only by their "A" terms and that there is faster TFP growth in the data capital-producing sector, i.e., $da^N > da^Y$. For further details on this solution in similar models, see Oulton (2012) and Byrne and Corrado (2017a).

The contribution of data capital to the growth in labor productivity in this solution is the sum of a "use" or "investment" effect plus a "production" effect that may be expressed as follows:

$$(6) \qquad \underbrace{\bar{\sigma}^R/_{\bar{\sigma}^L}\ (productivity\ advantage)}_{\text{investment or use effect}} + \underbrace{\bar{\omega}^D(productivity\ advantage)}_{\text{production effect}}.$$

The "overbar" notation in (6) denotes steady-state solution values. Thus, $\bar{\sigma}^R$ and $\bar{\sigma}^L$ represent steady state income shares of data capital and labor, respectively, and $\bar{\omega}^D$ is the steady state domestic production share of data investments.

"Productivity advantage" is the steady-state solution for the relative productivity of data capital. By assumption there is faster TFP growth in the data capital-producing sector, i.e., $da^N > da^Y$, and the solution for this productivity difference is (the negative of) data capital's relative price change. Thus, the relative productivity of data capital in steady growth is given by,

$$(7) \qquad productivity\ advantage \ = -(d\ln p^N - d\ln p^Y),$$

i.e., the rate of decline in the *relative* price of data assets (sign reversed).

Table 6 presents alternative scenarios for potential labor productivity growth using equations (6) and (7). The scenarios vary according to assumptions regarding the breadth of data capital use and production share (the rows of the table) and its productivity advantage (the columns). The cells represent simple scenarios that vary according to assumptions for $\bar{\sigma}^R$, $\bar{\omega}^D$ (limited or broad) use and production of data capital) and the productivity advantage of data capital, where the assumptions are drawn from measures developed and reviewed in this paper.

The capital input shares of data capital are assumed to range from 5 to 10 percent, i.e., a bit above the approximate band about the estimates for the penetration of data capital in intangible capital as discussed in section 4. Production shares are assumed to be the capital income share +/- 10 percent, a rough estimate of the range for net exports of intangibles (excluding R&D and software, not shown but embedded in the share of intangible investment attributed to net imports in figure 4b).

The upper and lower bounds for the productivity advantage are drawn from the relative price differential implied by the data-intensive components of intangibles investment shown in table 5. They are set at 1 and 5 percent, respectively. This lower bound is a bit below the US historical experience, whereas the upper bound is higher. Deflators for data-intensive components of intangibles rely on national accounts prices, e.g., gross output deflators for industrial design and management consulting that are unlikely to incorporate efficiency gains due to increased application of AI or use of open-source content, and it seems prudent to consider these measurement realities.

For the upper bound, consider first the long-term price decline of conventionally defined IT capital, about 15 percent per year (based on the estimates reported in Byrne and Corrado, 2017). Our best estimates of price declines for two data-intensive intangibles are extremely modest by comparison, and an upper bound for the relative productivity of data capital at 5 percentage points per year is likewise very prudent. All calculations assume labor's share of total income $\bar{\sigma}^L$equals .7.

Table 6. Productivity Scenarios: Contribution of data capital to potential labor productivity growth (percentage points)

	Productivity advantage (Relative asset price growth differential)	
Income and production share	Narrow edge 2 percentage point	Large edge 5 percentage points
Broad use (and net exporter of data services) 10 percent capital income share 11 percent production share	0.51	1.26
Limited use (and net importer of data services) 5 percent capital income share 4.5 percent production share	0.23	0.58

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

All told, estimates of the contribution of data capital to labor productivity growth range by more than a factor of 5—from 0.23 percentage point per year to 1.26 percentage point per year. The range highlights the synergies among data capital efficiency and an economy's capability for digital transformation of its production processes.

Having established that data capital has considerable potential for impacting labor productivity growth, let us now address how data capital affects measured total factor productivity da.

7. 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 Europe, the United States, and Japan.[17] 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.[18]

[17] This update/expansion of EUKLEMS was funded by the European Commission's Directorate General for Economic and Financial affairs.

[18] Methods used to develop the harmonized estimates of intangible investment are documented in Bontadini et al (2023), available on the EUKLEMS & INTANProd portal at https://euklems-intanprod-llee.luiss.it. Compared with previous estimates issued via the INTANInvest database (at www.intaninvest.net), the figures in EUKLEMS &

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 1995 to 2019. The results for Europe are aggregated using purchasing power parities (PPPs) to facilitate comparative analysis with the United States. 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 analysis is the availability of sufficient years of consistent input and output data for industry-level growth accounting.

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 5 (b) (Bontadini et al. 2023a, pages 31-32). The quality change component of price deflators for computer, and communications equipment and software also are harmonized across countries, and net stocks of capital (intangible and tangible) are estimated using common rates of economic depreciation. The EUKLEMS & INTANProd data used here are from the project's analytical module, accessed late February 2023.[19]

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.[20]

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, is an important factor driving productivity change in some countries. Figure 7 shows that the 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, however.[21]

INTANProd reflect significant improvements to the own-account components of intangible investment and to intangible asset price deflators for non-national accounts components.

[19] Compared with previous editions of EUKLEMS, the LUISS update presents productivity in two "modules": one with capital stock estimate issued by national statistical offices ("statistical" module) and another harmonized for international comparability ("analytical" module).

[20] 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.

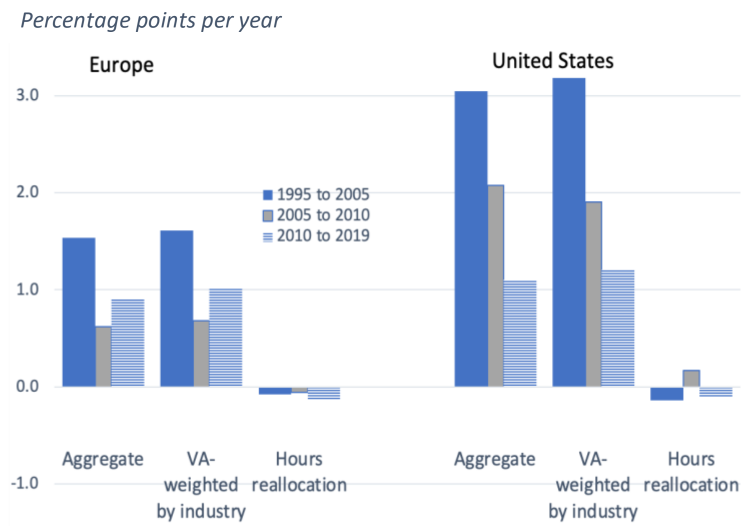
[21] Reflecting the global pandemic, reallocation of hours jumped in 2020 but was largely reversed in 2021.

The calculations reported in figure 7 follow Stiroh (2002). The first set of columns shows changes in aggregate labor productivity $d\ ln\ (V/H)$ where H denotes aggregate hours and V real value-added output, and second two sets of bars display its decomposition into two terms, the impact of productivity growth in each industry (a "pure productivity" or "within" effect)—and a reallocation effect, the impact of hours worked moving from low- to high-productivity industries, i.e.,

(8) $$d\ln(V/_H) = \sum_i \sigma_i^{PV} d\ ln\ \left({V_i}/_{H_i}\right) + \underbrace{\sum_i \sigma_i^{PV} d\ ln\ H_i + d\ln H}_{Reallocation\ effect}$$

where σ_i^{PV} denotes the value-added (Divisia) weight of industry i in (nominal) aggregate value added. As explained in Stiroh (2002, page 1572), aggregate hours growth, $d\ ln\ H$, approximately weights industries by their (lagged) share of aggregate hours, so aggregate productivity rises if industries with value-added shares above labor shares experience growth in hours.

Figure 7. Labor Productivity and Hours Reallocation



Source: EUKELMS & INTANProd. LUISS (2023).

To analyze the drop in "pure" labor productivity growth in market-sector dominated industries in the aftermath of the global financial recession (GFR), period averages from 2010 to 2019 are compared with those from 1995 to 2005 for two reasons. First, the comparisons exclude the onset and initial recovery years of the GFR during which changes in factor utilization created sharp, temporary fluctuations in productivity that obscure underlying trends.[22] Second, regime-shift detection methods estimate 2005 as a break date for TFP growth in both Europe and the United States (Fernald 2014, 2015; Fernald and Inklaar 2020).

Figure 8 sets out decompositions of the within-industry change in labor productivity that account for the full range of intangibles per table 2. These decompositions are calculated as follows:

(9) $$\sum_i \sigma_i^{PV} d\ ln\ \left({V_i}/_{H_i}\right) = \sum_i \sigma_i^{PL} d\ ln\ \left({L}/_{H_i}\right) + \sum_i \sigma_i^{PK} d\ ln\ \left({K_i}/_{H_i}\right)$$
$$+ \sum_i \sigma_i^{PR} d\ ln\ \left({R_i}/_{H_i}\right) + d\ln a$$

---

[22] This is not to imply that these temporary factors are similar for each region shown. A greater emphasis on job retention strategies and job stability in the Euro area relative to the United States creates rather disparate short-run impacts on job levels in response to a common shock. To the extent these labor market differences affect long-term growth trends, their ultimate impacts are included in our analysis.

where the first term is the contribution of labor composition, the second two are contributions of capital intensity (tangible and intangible), and the final term total factor productivity.[23] Comparing the first set of columns in figure 8 for each region with the last set (i.e., comparing average changes and contribution from 1995 to 2005 with 2010 to 2019), the drop in labor productivity growth is mainly accounted for by slowdowns in tangible capital deepening and total factor productivity (TFP) growth. In the United States, TFP growth is 0.9 percentage point less per year in the period after 2010. In Europe the drop is much less, about 0.2 percentage points, though it followed 5 years during which TFP posted a sharp negative average change (-0.5 percent per year). That said, the drops in labor productivity, contribution of tangible capital intensity and TFP growth are rather more dramatic for the United States.[24]

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 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, the contribution of changes in labor composition to explaining developments in productivity growth in these regions over the past 20 years is relatively modest.

The important takeaway from figure 9 is that *direct* contribution of intangible capital deepening does not contribute to the productivity slowdown. The rate at which workers in both Europe and the United States were equipped with intangible capital was well maintained for the period shown in the figure. And the edge down in intangible capital deepening shown in the post-GFR period for the United States reverses direction if we include figures for the pandemic period (2020 and 2021), which are available for the United States but not for Europe; U.S. labor productivity and TFP growth are a tad higher too when the pandemic period is added.[25] That resources continued to be invested in innovation while the growth of TFP slowed suggests that the slowdown story must be about, at least in part, changes in the diffusion of the fruits of those investments across firms and industries in these economies.

Before we turn to discussing innovation diffusion, consider further how we might interpret the fact that the contribution of intangible capital deepening did not slow as productivity growth slowed.

---

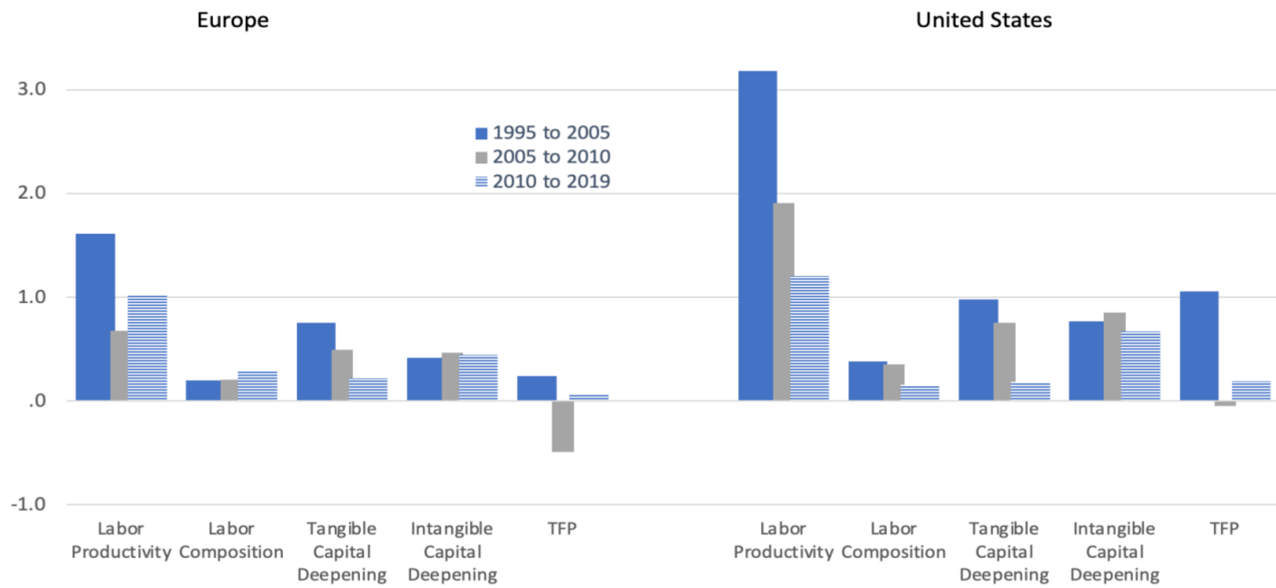
[23] For simplicity, the superscript notation for the income shares on the right side of equation (10) do not indicate that the nominal values for each factor's income have their own price element.

[24] These drops from the so-called "bubble" years may seem exaggerated compared to the 20 years of slow productivity growth that preceded them. But when the recent experience is compared with longer-term averages, i.e., from 1948 to 2005, the drops in the contribution of tangible capital intensity and TFP growth shown in figure 8 are in line with historical experience. According to official BLS historical statistics, the contribution of tangible intensity to labor productivity is the same from 1948 to 2005 as it is from 1995 to 2005, and TFP growth is just 0.1 percentage point less. Labor productivity growth was exceptionally high in the United States in the bubble years, however, averaging 0.6 percentage points per year higher compared with growth during the preceding 47 years.

[25] Based on information from the U.S. BLS; as of this writing, the EUKLEMS & INTANProd database reports data through 2020 but not 2021.

Figure 8. Decompositions of Growth in Labor Productivity

Percentage points, annual rate



Note: The bars labeled "Labor productivity" correspond to the value-added weighted industry aggregate shown in the middle set of bars in figure 8. European countries include Denmark (DK), Germany (DE), Finland (FI), France (FR), Italy (IT), Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK). Source: EUKELMS & INTANProd. LUISS (2023).

Interpretation of Innovation

When considering innovation, economists typically look to 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 practitioners interested in innovation. Innovation strategists typically focus on how individual firms innovate (e.g., develop a new business model, launch a new product) and recognize that business decision-makers need to weigh the costs involved in bringing about change through innovation against retaining profits. Intangible investment captures the cost of implementing and managing innovations.

Whether this connection between innovation and intangible investment helps depends on how innovation at the individual organization level is discerned in economists' productivity measures. 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 (Goodridge, Haskel and Wallis 2012). Innovation is output less the contribution of K and L, which suggests that innovation reflects the final two terms in (9), or

$$\text{(10)} \qquad Innovation = \sigma_Q^R dr + da$$

using the terminology of the upstream/downstream model of section 5. This implies that when considering innovation, residually calculated TFP growth (da) should not be the sole focal point of analysis but rather a fuller picture includes the appropriated returns intangible investment.

The table below summarizes the fraction of labor productivity growth accounted for by the combined contributions of intangible capital deepening and TFP growth, which is, in effect, an innovation account (Corrado and Hulten 2014). Seen from this perspective, innovation (a) became *more relevant* as a factor accounting for labor productivity growth in both regions since 2010 and (b) has been unequivocally the dominant factor in the United States since 1995.

Table 7. Innovation Accounting: Innovation Share of Labor Productivity Growth

		Europe (1)	United States (2)
1	1995 to 2005	.41	.57
2	2010 to 2019	.50	.69

Note: Innovation share is the combined contributions of intangible capital deepening and TFP growth to labor productivity growth as a fraction of labor productivity growth.
Source: Equation (10) and Figure 9.

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) in the intangible capital framework. Cross-country and firm-level econometric work has repeatedly estimated increasing returns (or knowledge spillovers) to intangible capital.[26] In simple terms, these works imply that the proportional relationship, $da \approx .2\ dr$, could be used to represent the costless diffusion of commercially valuable knowledge in market economies. As dr (per worker) did not materially slow after 2010, the logical (endogenous) explanation for the slowdown in

[26] This refers to the aggregate implications of estimates for R&D spillovers reported by Griliches for manufacturing (e.g., Griliches 1992) and for intangible capital spillovers (excluding R&D and software) reported by Corrado, Haskel, and Jona-Lasinio (2017). The latter study used a dataset covering the late 1990s to the onset of the GFR.

measured da is that factors driving these increased returns ceased to operate as strongly as they previously had.

Productivity spillovers to intangible capital and to data capital are under investigation in forthcoming cross-country econometric work using the new EUKLEMS & INTANProd database (Bontadini et al. 2023b). While many avenues remain to be investigated, the study preliminarily finds that, while the previous proportional relationship $da \approx .2\ dr$ still holds for intangibles in the pre-GFR period, the same or similar econometric specifications fail to detect similar-sized productivity spillovers to intangible capital in the GFR's aftermath, from 2010 on. This coincides with the rise of modern data use.

How should we interpret these results? The traditional view of knowledge diffusion is that its potential for boosting market sector productivity growth rests in an economy's innovation ecosystem, consisting mainly of (a) the commercialization of new knowledge created outside of industry for use in industry, (b) competition intensity within industries, and (c) intellectual property rights and their enforcement. It has also been suggested that the ability to utilize spillovers from other firms depends on their geographical proximity, and while firms in both regions have increased their use of offshore production, with possible negative consequences for knowledge diffusion, the practice was in full force prior to the GFR when productivity growth is estimated to have been significantly boosted by knowledge spillovers.

The larger role of (untraded/untradable) big proprietary data used by global productivity "leaders" in the post-GFR period is consistent with a breakdown in knowledge diffusion. It is also consistent with increased industry concentration and changes in regulatory frameworks regarding data. If large firms with strong internal cash flows disproportionately invest in, and benefit from, untraded bigdata, industry concentrations may rise. If regulatory policies prevent third-party trading of data, digital productivity leaders (whether large or small) will be able to distance themselves from competitors via their use of data capital. Apart from efficiency effects discussed in section 6, data capital benefits the larger economy through its (costless) replication. All told, as intangible capital has become partly data capital, this change in the character of intangible investment seems to have directly affected the average rate of productivity growth. It also likely boosted the growth of intangible capital to the extent innovations embedded in data capital is mainly copied by additional capital spending.

These observations on the impact of proprietary bigdata deserve further study and econometric analysis. But the empirics developed in this paper are sufficient to frame their likely macroeconomic impact. With post-2010 growth of intangible capital services in nonagricultural market sector industries averaging 3.3 percent per year in Europe and 4.4 percent in the United States and if data capital accounts for 50 percent of that, a complete cessation of data capital diffusion mechanism could shave .33 to .44 percentage point per year off measured total factor productivity growth in these regions, respectively. This is a sizeable impact, which note, overexplains the 2010 to 2019 drop in European TFP growth and approaches one-half of the drop in TFP growth in the United States (recall from figure 8 these drops were .2 and .9 percentage points, respectively).

Productivity growth via the costless replication of commercial knowledge is unlikely to have staunched completely, and other structural or policy-induced factors or mismeasurement may have contributed to the productivity slowdown. But our analysis of the increased data intensity of intangible capital, combined with the tendency for data assets to be closely held and the econometric results from Bontadini et al (2023b), suggest that the spillover effects that drove TFP growth in the past likely have diminished with the increased use or proprietary bigdata.

8. Digital Paradoxes Unpacked and Concluding Remarks

The changing nature of the global economy has put a spotlight on intangible capital as a source of economic growth and driver of innovation. But intangible capital itself has been evolving with the digitization of modern economies and growing use of proprietary bigdata in production processes. At the same time productivity growth—labor productivity and total factor productivity—has slowed, created what some have called a digital productivity paradox.

We have used an intangible assets approach to address the question of how the increased use of data in economies affects productivity. We argue that data, or more accurately transformed raw digitized records, and the capital services derived from them fit neatly into both the management/technology literature on the data stack and the economics literature on intangible capital. We developed new measures of industry-level data investments, new price measures for intangible assets most affected by data and outlined a simple two-sector growth accounting framework to articulate how much the relative technology gains from the use of data affect labor productivity growth. Then, using a new multi-country total factor productivity dataset, we documented and analyzed how the slowdown in total factor productivity growth in major developed economies may have been affected by the increased use of proprietary bigdata.

Our empirical analysis of the relationship between the intangible investment estimates in EUKLEMS & INTANProd and the "data stack"-inspired estimates of data investment found the strongest overlap of data capital with intangible capital was in components hypothesized to be most likely driven by modern data use: investments in brand and marketing, marketing research, industrial design, and organization processes and structure. This does not imply that all intangible capital in these categories is data capital, but that each of these components overlaps strongly with data capital, with the degree of overlap expected to vary across industry sectors and with time.

Our modeling of the economic impacts of data capital produced two results and associated predictions not featured in prior works. Both results stem from the fact that modern data capital, though inherently nonrival, reflects the rise of proprietary bigdata and its use in production primarily as an "innovation in the method of innovation." Our analysis implied that the efficiencies associated with modern data capital lowers intangible asset prices and that the proprietary nature of bigdata strengthens the (partial) appropriability of the intangible asset class. The first-order impacts of these results on productivity are that the use of data capital boosts labor productivity growth (the efficiency effect) but that the increased data intensity of intangibles weakens commercial knowledge diffusion and diminishes TFP growth (the appropriability effect).

In stepping through all of this—from how we should conceptualize and measure bigdata as an asset to framing its first-order impacts on macro productivity growth—we shed light on several frequently mentioned paradoxes of modern digital economies. The first is that the much-touted efficiency gains promised for businesses as they incorporate bigdata into decision-making likely *are* present in economies, but they are not so readily "seen." Our estimates of data investments generally exceed measured IT equipment investment in national accounts, i.e., they are large and associated with components of intangible capital not included in national accounts investment. Their relative efficiency, which this paper calibrated in terms of investment prices, does not come close to the relative efficiency of IT equipment investment, but it is estimated to be two-thirds of the relative efficiency of software as currently measured in national accounts. These areas of price research (software and data-intensive intangibles) deserve more attention as their impact on economies will be larger than reported in this paper as data technologies and data assets gain additional traction in use.

This paper's productivity estimates reflect, to the extent possible, all intangibles and the increased relative efficiency of data assets, but the boost to labor productivity appears to be offset by the appropriability effect that diminishes TFP. One might ask whether the downside nature of data capital's appropriability effect is essentially what others have considered the paradox of the digital transformation of modern economies, i.e., the unleashing of anti-competitive and winner-take-all forces. There are several points to make here, the first of which is that in the context of the intangible capital framework, the downside to the rise of proprietary data capital is no paradox. Second, we emphasized the generally well maintained rate if resources devoted innovation in the modern digital age—the average share of market sector labor productivity growth accounted for by innovation in fact increased dramatically in 2010 to 2019 relative to 1995 to 2005—because intangible capital deepening was maintained.

Third, the winner-take-all view that recent growth has been slowed by diminished competition is often framed as a correlate of rising industry concentration, but measures of industry concentration have been rising for a very long time (Kwon, Ma, and Zimmerman 2023), suggesting the modern digital age and use of bigdata is unlikely a *fundamental* driver of increasing industry concentration. The view is also typically supported by empirics that do not account for all intangibles (or neglect them entirely), leading observers to misstate the degree of excess profits or market power in an economy. Here we account for all intangibles after which measures such as realized after-tax rates of return suggest there has been no material increase in market power of late (Corrado et al 2022a). And while the rise of proprietary bigdata has apparently staunched knowledge diffusion and diminished TFP growth, these are outcomes with much scope for reversal by policies that promote industry data sharing (Jones and Tonetti 2020).

A full explanation for the recent productivity slowdown perhaps remains elusive, but we are hopeful that the findings reported in this paper, along with its new measurements, have demonstrated that the increased importance of data assets in intangible capital is a factor in the explanation.

REFERENCES

Awano, Gaganan, Mark Franklin, Jonathan Haskel, and Zafeira Kastrinaki. 2010. "Investing in Innovation: Findings from the UK Investment in Intangible Assets Survey." NESTA Index Report.

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.

Andrews, Dan, and Chiara Criscuolo and 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.

Autor, David H. 2013. "The 'Task Approach' to labor Markets: An Overview." *Journal of Labour Market Research* 46(3). DOI 10.1007/s12651-013-0128-z

Babina, Tania, and Greg Buchak and Will Gornall. 2022. "Customer Data Access and Fintech Entry: Early Evidence from Open Banking." Stanford University Graduate School of Business Research Paper (March 31, 2022). Available at SSRN: https://ssrn.com/abstract=4071214 or http://dx.doi.org/10.2139/ssrn.4071214

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2022. "Artificial Intelligence, Firm Growth, and Product Innovation." Available at SSRN: https://ssrn.com/abstract=3651052 or http://dx.doi.org/10.2139/ssrn.3651052

Bontadini, Fillipo, Carol A. Corrado, Jonathan E. Haskel, Massimiliano Iommi, and Cecilia Jona-Lasinio. 2023a. "EUKLEMS & INTANProd: Industry Productivity Accounts with Intangibles." LUISS Lab of European Economics. https://euklems-intanprod-llee.luiss.it/wp-content/uploads/2023/02/EUKLEMS_INTANProd_D2.3.1.pdf.

Bontadini, Fillipo, Carol A. Corrado, Jonathan E. Haskel, Massimiliano Iommi, and Cecilia Jona-Lasinio. 2023b. "Productivity, Inflation and Intangible Capital: Recent evidence from the EUKLEMS & INTANProd database." Paper presented at ESCoE Annual Conference on Economic Measurement, Kings College, London (May).

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

Byrne, David and Carol Corrado. 2017a. "ICT services and their prices: What do they tell us about productivity and technology?" *International Productivity Monitor*, Number 33 (Fall), p. 150-181.

Byrne, David and Carol Corrado. 2017b. "ICT Asset Prices: Marshalling Evidence into New Measures." Finance and Economics Discussion Series 2017-016. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2017.016r1

Byrne, David, Carol Corrado, and Dan Sichel. 2021. "The Rise of Cloud Computing: Minding your P's, Q's and K's." In Corrado, C, Haskel, J., Miranda, J. and Sichel, D., *Measuring and Accounting for Innovation in the Twenty-First Century*, 519-551. Studies in Income and Wealth, Volume 78. Chicago: Chicago University Press.

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).

Colecchia, Alessandra and Paul Schreyer. 2001. "ICT Investment and Economic Growth in the 1990s: Is the United States a unique case? A Comparative Study of Nine OECD Countries." OECD STI Working Paper 2001/7, OECD Publishing, Paris.

Corrado, Carol. 2023. "Intangible Capital: What it is and why it matters." In Marshall Reinsdorf and Louise Sheiner, eds., *Measuring the Hard-to-Measure in a Changing Economy*. Chicago: University of Chicago Press (forthcoming).

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 (Juy). Paris: OECD Publishing. 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.

Corrado, Carol, Jonathan Haskel, Massimiliano Iommi and Cecilia Jona-Lasinio. 2022a, "The Value of Data in Digital-based business models: Measurement and economic policy implications", *OECD Economics Department Working Papers*, No. 1723 (August). Paris: OECD Publishing, https://doi.org/10.1787/d960a10c-en.

Corrado, Carol, Jonathan Haskel, Massimiliano Iommi and Cecilia Jona-Lasinio. 2022b. “Measuring Data as an Asset: Framework, methods and preliminary estimates.” OECD Economics Department Working Papers, No. 1731 (November). Paris: OECD Publishing. https://doi.org/10.1787/b840fb01-en.

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* (Summer): 36 (3): 3-28.

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.

Corrado, Carol A. and Charles R. Hulten. 2014. “Innovation Accounting.” In Dale W. Jorgenson, J. Steven Landefeld, and Paul Schreyer (eds.), *Measuring Economic Sustainability and Progress*, Volume 72, Studies in Income and Wealth, 595-628. Chicago: University of Chicago Press.

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).

De Rassenfosse, Gaétan, and Adam B. Jaffe. 2017. Econometric Evidence on the R&D Depreciation Rate. NBER Working Paper Series No. w23072 (January).

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.

Economist. 2017. “Leader: The world’s most valuable resource is no longer oil, but data.” *Economist Magazine*. The Economist Group Limited (6 May).

Fernald, John G. 2014. “A Quarterly, Utilization-Adjusted Series on Total Factor Productivity.” Working Paper no. 2012-19. Federal Reserve Bank of San Francisco.

Fernald, John G. 2015. “Productivity and Potential Output Before, During, and After the Great Recession.” NBER *Macroeconomics Annua*l 29: 1–51.

Fernald, John, Robert Inklaar. 2020. “Does Disappointing European Productivity Growth Reflect a Slowing Trend? Weighing the Evidence and Assessing the Future,” Federal Reserve Bank of San Francisco Working Paper 2020-22. https://doi.org/10.24148/wp2020-22

Goodridge, Peter, Jonathan Haskel, and Harold Edquist. 2021. “We See Data Everywhere Except in the Productivity Statistics.” Review of Income and Wealth. https://doi.org/10.1111/roiw.12542

Goodridge, Peter, Jonathan Haskel, and Gavin Wallis. 2012. "UK Innovation Index: Productivity and Growth in UK Industries", NESTA Working Paper 12/09.

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.

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

International Monetary Fund (IMF). 2019. "The Rise of Corporate Market Power and Its Macroeconomic Effects." In *World Economic Outlook* (April), Chapter 2, 55-76. Available here.

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. 1963. "Capital Theory and Investment Behavior." *American Economic Review* 53 (2): 247-259.

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

Kwon, Spencer Yongwook, Yueran Ma, and Kaspar Zimmermann. 2023. "100 Years of Rising Corporate Concentration." University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2023-20 (February). Available at SSRN: https://ssrn.com/abstract=4362319 or http://dx.doi.org/10.2139/ssrn.4362319

Lall, Subir and Li Zeng. 2020. "Intangible Investment and Low Inflation: A Framework and Some Evidence." IMF Working Paper WP/20/190 (September).

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.

LUISS Lab of European Economics (LLEE). 2023. "EU KLEMS & INTANProd database." LUISS University [producer]. Accessed at https://euklems-intanprod-llee.luiss.it/ February 25, 2023.

Mandel, Michael (2019). "Building the New Manufacturing Stack." Forbes Magazine (August 20). Available here.

Maslej, Nestor, Loredana Fattorini, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Vanessa Parli, Yoav Shoham, Russell Wald, Jack Clark, and Raymond Perrault. 2023. "The AI Index 2023 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA (April).

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

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.

Open Data Institute (2016), "Research: The economic value of open versus paid data." Accessed March 2021 at https://theodi.org/article/research-the-economicvalue-of-open-versus-paid-data

Oulton, Nicholas. 2012. "Long Term Implications of the ICT Revolution: Applying the Lessons of Growth Accounting and Growth Theory." *Economic Modelling,* 29 (5): 1722–1736.

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

Robbins, Carol, Gizam Korkmaz, Ledia Guci, José B. Santiago Calderón, Brandon L. Kramer. 2021. “A First Look at Open-Source Software Investment in the United States and in Other Countries, 2009-2019.” Paper presented the IARIW-ESCoE Conference, Measuring Intangible Capitals and their Contribution to Growth (November, RSA House, London).

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 (last accessed August 25, 2021).

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.

Santiago-Calderón, Jose B. and Dylan Rassier. 2022. “Valuing the US Data Economy Using Machine Learning and Online Job Postings.” Paper prepared for the NBER/CRIW Conference on Technology, Productivity, and Economic Growth, March 17-18, 2022.

Silk, Alvin J. and Ernst R. Berndt. 2003. “Scale and Scope Economies in the Global Advertising and Marketing Services Business.” NBER working paper series no. 9965 (September).

Squicciarini, Mariagrazia, and Marie Le Mouel. 2012. “Defining and Measuring Investment in Organisational Capital: Using US Microdata to Develop a Task-based Approach.” OECD Science, Technology and Industry Working Papers 2012/5. OECD Publishing, Paris.

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).

The Conference Board (2021). “Skills Growing in Demand for Tech Jobs.” (March 11). Available at https://www.conference-board.org/topics/help-wanted-online/skills-growing-in-demand-for-tech-jobs

Timmer, Marcel, Robert Inklaar, Mary O’Mahony, and Bark van Ark. 2011. “Productivity and Economic Growth in Europe: A Comparative Industry Perspective.” *International Productivity Monitor* 21 (Spring): 3–23.

United Nations Digitalization Task Team for the SNA 2025 (DZTT). 2023. “DZ.6 Recording of Data in National Accounts.” Last accessed March 15.

United Nations Globalization Task Team for the SNA 2025 (GZTT). 2023. “G.9 Payments for Nonproduced Knowledge-Based Capital (Marketing Assets)”. Last accessed March 15.

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.

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. 2021. “The AI Index 2021 Annual Report,” AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA.

APPENDIX

This appendix provides a description of the cost-based approach and data sources used to estimate market sector investment and capital stocks in data stores and data intelligence (the components of data not currently included in official national accounts), databases and computer software introduced in this paper. The methodological approach is the same adopted to estimate own-account brand, design, organizational capital, and new financial products in EUKLEMS & INTANProd database (Bontadini et al

2023). The data sources and main steps of the calculation are described below. A final section provides an illustration of the approach followed to estimating domestically sourced component of intangible investment.

Data assets measurement

The method adopted to generate estimates of data assets follows a cost-based approach assuming that the value of an asset can be obtained as the sum of the costs sustained for producing it. The basic approach can be summarized as follows:

$$Y^i_{bc} = COMP^i + IC^i + CK^i + T^i \qquad (A1)$$

where i = asset type, Y is the value of the produced asset at basic prices, $COMP^i$ is the labor cost of the relevant personnel measured as compensation of employees, IC^i are intermediate costs related to the activity, CK^i refers to the costs of capital services and T^i to net taxes on production related to these activities. But notice that besides $COMP^i$ the remaining components in equation (A1) are not directly measurable, thus the sum of these unmeasurable components, set equal to α **,** is a factor that must be approximated**.** So equation (A1) can be re-written as:

$$Y^i_{bc} = COMP^i + \alpha^i \qquad (A2)$$

where, notice that Y^i_{bc} can be measured directly by computing the compensation of employees (COMP) and finding a proxy for α. Thus COMP can be obtained as:

$$COMP^i = EMP^i_{tot} * W^i_{avg} * t^i \qquad (A3)$$

where EMP^i_{tot} is the total number of employees employed for producing the relevant asset*,* W^i_{avg} is the average remuneration (average wage) and t^i refers to the time spent on these activities (table A1 below shows the time assumptions (t) underlying the calculations developed in this paper). Using equation (A3) and substituting it in equation (A2) where it is assumed that $\alpha^i = COMP^i * bp^i - COMP^i$, the value of the produced asset is determined as:

$$Y^i_{bc} = COMP^i * bp^i \qquad (A4)$$

where *bp* is a blow-up factor that accounts for other cost components besides the compensation of employees and essential to develop a measure of output consistent with national accounts. The blow-up factors for each asset are measured using the ratio of gross output (GO) over the compensation (COMP) of all persons engaged where GO is adjusted to exclude national accounts own-account intangibles and intermediate purchases of intangibles that are capitalized in our framework. COMP is compensation of employees plus an estimate of compensation of self-employed. This adjustment is relevant, especially for those industries producing a sizable amount of intangibles whose production structure is assumed rather similar to the internal intangible factory described in the main text. The blow-up factors are estimated at the detailed industry level using US supply and use tables. The bp of the relevant industries averaged over 1997-2020 are then applied to each data asset.[27] In this paper, the blow-up factors are set equal to 1.7 for data intelligence and 1.8 for the other assets.

Main sources

The estimates of data assets illustrated in this paper have been produced applying equation (A4) across industries and countries. The main information needed to implement the calculation for each individual data asset is the following: i) a detailed list of occupations engaged in producing data assets; ii) occupation-specific (and industry-specific, if relevant) assumptions on the share of time spent in producing each data 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.

More specifically, table A1 shows the list of occupations that are assumed to be engaged in producing data assets and computer software capital formation based on ISCO-08 codes and of the time-use assumptions by asset as detailed below.

Notice that the selection of relevant occupations is constrained by the level of detail of available data sources. For this paper, we use micro-data from the EU Structure of Earnings Survey (SES) for the years 2010, 2014, and 2018 and the EU Labour Force Survey (LFS) for 2008 and 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, provides data on employment with no information on wages. In the LFS, occupations are available at three-digit level of ISCO classification for all countries. On the other hand, the SES provides data for 11 countries at two-digit level, that have to be further expanded to three-digit level to get a coherent information set. The higher level of disaggregation for the SES wages from two to three digits has been obtained computing the share of relevant three-digit occupation from the LFS then applied to the SES variables.

Estimating the time-use factors

In this paper, we make a further step in developing the estimates of the time use factors compared to previous literature. In particular, we generate our measures resorting to a very high level of disaggregation of the occupation classification, four-digit level, for a group of European economies[28].

[27] For data stores, we apply the average of blow-up factors calculated for Miscellaneous professional, scientific, and technical services, Computer systems design and related services, and Publishing industries, except internet (includes software); for databases and software, the average of Computer systems design and related services, and Publishing industries, except internet (includes software); for data intelligence, the average of Management of companies and Miscellaneous professional, scientific, and technical services.

[28] Ideally, the selection of the occupations and the definition of the corresponding time-use assumptions would have to be done at four digits level in the ISCO classification. But the available sources, the LFS and SES provide data at three digits level thus requiring a careful analysis of each code to identify the best time use assumption. As

First we have identified the relevant occupations and the corresponding time-use assumptions at four-digit level of the US Standard occupations Classification (SOC 2010) and then we have used the data from the US Occupational Employment and Wage Statistics 2019 (OEWS) to compute a set of weights to generate anestimate of the time-use factor aggregated at 3-digit ISCO level.

More precisely, the approach for measuring the time-use factor for each data asset can be summarized as follows: i) Identification of the ISCO unit groups included in each ISCO minor group[29]; ii) for each ISCO unit group, identification of the corresponding occupation in the 2010 SOC (based on a crosswalk between the ISCO-08 and the 2010 SOC available from the US Bureau of Labor Statistics); iii) assignment of a time-use factor to each four-digits SOC occupation.

Finally, the time use factors at three digit level have been computed as the weighted average of the time-use factor of the corresponding four-digits SOC where the weights have been generated from employment data gathered from the 2019 OEWS[30].

Capital stock and price deflators

The estimates of capital stock in real terms used in the econometric analysis are generated applying the perpetual inventory method (PIM) based on the aggregation of real investment over time allowing for declines in efficiency and value until the assets reach the end of their service lives and are retired[31].

We assume economic depreciation is geometric, in which case the real stock of data asset j in industry i at the end of year t ($Kq_{i,tj}$) is defined as:

$$Kq_{i,tj} = Kq_{i,t-1j} * (1-\delta_j) + Iq_{i,tj} \qquad \text{(A5)}$$

---

a matter of fact, not all the occupations within each ISCO code devote their working time to data production to the same extent or to produce the same type of asset.

So far, approaches to measure data assets dealt with aggregate occupation groups tweaking the assumption on the time-use factors to consider that each minor group includes workers who differ in how much time they spend producing a given data asset (or who do not create data assets at all). However, this approach adds a further layer of imprecision to estimates which are in themself based on many assumptions.

[29] In the ISCO classification, minor groups are occupations defined at the level of three digits and unit groups at the level of four digits.

[30] However, this approach has some drawbacks. It is implicitly assumed that the employment shares of four-digit occupations within each three-digit group are the same in all European countries and equal to those in the US. In addition, the aggregate percentages are used to develop industry-level calculations.

[31] See Hulten Charles R. "The Measurement of Capital" in Ernst R. Berndt and Jack E. Triplett, eds., ifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth, Chicago, University of Chicago Press, 1991

TABLE A1. RELEVANT OCCUPATIONS FOR DEFINING TIME-USE ASSUMPTIONS TO ESTIMATE DATA ASSETS AND COMPUTER SOFTWARE

ISCO-08 sub-major group	ISCO-08 minor group	Occupation description	Time-use (%)			
			Software	Databases	Data Stores	Data Intelligence
1	**1**	**Managers**				
11	**11**	**Chief executives, senior officials and legislators**				
111	11	Legislators and senior officials	0.00	0.00	0.00	0.00
112	11	Managing directors and chief executives	0.00	0.00	0.00	0.17
12	**12**	**Administrative and commercial managers**				
121	12	Business services and administration managers	0.00	0.00	0.00	0.00
122	12	Sales, marketing and development managers	0.00	0.00	0.00	0.10
13	**13**	**Production and specialised services managers**				
131	13	Production managers in agriculture, forestry and fisheries	0.00	0.00	0.00	0.00
132	13	Manufacturing, mining, construction, and distribution managers	0.00	0.00	0.00	0.00
133	13	Information and communications technology service managers	0.00	0.10	0.10	0.10
134	13	Professional services managers	0.00	0.00	0.00	0.00
14	**14**	**Hospitality, retail and other services managers**				
141	14	Hotel and restaurant managers	0.00	0.00	0.00	0.00
142	14	Retail and wholesale trade managers	0.00	0.00	0.00	0.00
143	14	Other Services Managers	0.00	0.00	0.00	0.00
21	**21**	**Science and engineering professionals**				
211	21	Physical and earth science professionals	0.00	0.00	0.10	0.40
212	21	Mathematicians, actuaries and statisticians	0.05	0.06	0.13	0.32
213	21	Life science professionals	0.00	0.02	0.11	0.28
214	21	Engineering professionals (excluding electrotechnology)	0.00	0.00	0.10	0.25
215	21	Electrotechnology engineers	0.00	0.00	0.10	0.36
216	21	Architects, planners, surveyors and designers	0.00	0.00	0.18	0.09
24	**24**	**Business and administration professionals**				
241	24	Finance professionals	0.00	0.00	0.10	0.26
242	24	Administration professionals	0.00	0.07	0.07	0.11
243	24	Sales, marketing and public relations professionals	0.00	0.00	0.15	0.21
25	**25**	**Information and communications technology professionals**				
251	25	Software and applications developers and analysts	0.48	0.10	0.03	0.01
252	25	Database and network professionals	0.10	0.47	0.10	0.03
263	26	Social and Religious Professionals	0.00	0.00	0.00	0.03
33	**33**	**Business and administration associate professionals**				
331	33	Financial and mathematical associate professionals	0.00	0.00	0.10	0.25
332	33	Sales and purchasing agents and brokers	0.00	0.00	0.00	0.00
333	33	Business services agents	0.00	0.02	0.02	0.04
334	33	Administrative and specialised secretaries	0.00	0.00	0.00	0.00
335	33	Regulatory government associate professionals	0.00	0.00	0.00	0.02
35	**35**	**Information and communications technicians**				
351	35	Information and communications technology operations and user support tec	0.01	0.02	0.03	0.04
352	35	Telecommunications and broadcasting technicians	0.00	0.00	0.10	0.05
41	**41**	**General and keyboard clerks**	0.00	0.00	0.07	0.00
42	**42**	**Customer services clerks**				
422	42	Client information workers	0.00	0.00	0.05	0.00
43	**43**	**Numerical and material recording clerks**				
431	43	Numerical clerks	0.00	0.00	0.16	0.00
432	43	Material-recording and transport clerks	0.00	0.00	0.05	0.00

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

where $Kq_{i,t-1j}$ is the real stock of asset j in industry i at the end of year t-1, δ_j is the annual depreciation rate for asset j and $Iq_{i,tj}$ is real investment for asset j in industry i during year t. Note that depreciation rates are asset-specific and are assumed to not vary across industries and over time.

Real investment in each type of data asset is obtained by dividing its nominal investment flow by an appropriate price index. The real value for data intelligence has been computed applying the deflator of non-national accounts intangibles while for data stores, computer software and databases exploiting the harmonized software deflator developed for the analytical module of the EUKLEMS & INTANProd databases (see Bontadini et al 2023).

For what concern the depreciation rates, for data intelligence it is the average of the depreciation rates of non-national accounts intangible assets (0.35), while for the other data assets it is the same depreciation rate used for software in EUKLEMS & INTANProd (0.315).

Estimates of data assets have been developed by industry, at the level of Nace sections, and then aggregated at the market sector level, defined as all industries excluding Nace sections L (real estate activities), 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).

Domestic Component of Intangible Investment: Data Sources and Estimation Method

In what follows we illustrate the main steps for estimating domestically sourced component of intangible investment.

Measures of domestically produced investment in R&D and computer software and databases are obtained using data from national supply and use tables. First, it is necessary to compute the share of gross output in total resources for domestic use, SGOD, as follows:

$$SGOD_i = (GO_i – EX_i) / (GO_i – EX_i + IM_i) \qquad (A6)$$

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, the domestic component of each asset is generated multiplying national accounts investment by the corresponding share of gross output in total resources for domestic use. A main assumption underlying this approach is that the share of the domestic component is the same across different uses (intermediate consumption, final consumption, and investment).

For those intangible assets not included in national accounts (brand, design and organizational capital), the domestic component is computed as the sum of the own-account investment and an estimate of the domestically sourced purchased component. That is, for organizational capital:

$$I_{dom}^{OrgCap} = (I^{OrgCap(OA)} + I_{dom}^{OrgCap(Purchased)}) \qquad (A7)$$

New financial products are only domestically produced, and that there are no available data sources for estimating imported training. The imported component for training is deemed very small and is ignored.

To estimate equation (A7) it is essential to measure the domestically sourced purchased component of brand, design, and organizational capital. Estimates of these components are generated from the information gathered from the world input-output tables reporting the intermediate use of domestic output and intermediate use of imports from other countries disaggregated by product, for each industry in each country (World Input-Output Database (WIOD, available at https://www.rug.nl/ggdc/valuechain/wiod/?lang=en).

Resorting to these data, it is possible to compute 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). Then the domestically sourced purchased component of brand, design and organizational capital can be generated by multiplying the purchased investment component by the share of domestic output in total intermediate consumption for the relevant products listed before (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, shares have been extrapolated regressing the 2000-2014 shares on a linear time trend.

Correlation tables

Tables A2 A) and B) below shows the correlations between the estimated 2010 to 2019 time series for the value-added shares of data assets and intangibles and for the rate of growth of the shares.

Table A2. Correlation of estimates

A) Investment shares of adjusted value added

	Data capital	Data stores	Databases	Data intelligence	**Software**	**Intangible capital**	Digitized information	Innovative property	Economic competencies	**Data-intensive intangibles**	Organization	Design	Brand	R&D
Data capital	1													
Data stores	0.746***	1												
Databases	0.663***	0.594***	1											
Data intelligence	0.905***	0.442***	0.369***	1										
Software	0.795***	0.284**	0.526***	0.859***	1									

Intangible capital	0.553***	0.0476	0.461***	0.634***	0.866***	1								
Digitized information	0.111	-0.263*	0.377***	0.156	0.577***	0.763***	1							
Innovative property	0.247*	0.0102	0.222*	0.283**	0.389***	0.632***	0.368***	1						
Economic competencies	0.678***	0.244*	0.422***	0.740***	0.836***	0.791***	0.478***	0.101	1					
Data-intensive intangibles	0.643***	0.140	0.410***	0.741***	0.900***	0.921***	0.602***	0.367***	0.943***	1				
Organization	0.576***	0.140	0.456***	0.627***	0.838***	0.864***	0.673***	0.210*	0.946***	0.937***	1			
Design	0.142	-0.121	0.0827	0.235*	0.369***	0.577***	0.369***	0.904***	0.0825	0.386***	0.198	1		
Brand	0.610***	0.366***	0.242*	0.646***	0.554***	0.399***	0.0233	-0.176	0.782***	0.667***	0.557***	-0.188	1	
Memo: R&D	0.244*	0.0824	0.317**	0.214*	0.328**	0.601***	0.364***	0.943***	0.0998	0.309**	0.219*	0.726***	-0.193	1

B) Growth rate of Investment shares of adjusted value added

	Data capital	Data stores	Databases	Data intelligence	**Software**	**Intangible capital**	Digitized information	Innovative property	Economic competencies	**Data-intensive intangibles**	Organization	Design	Brand	R&D
Data capital	1													
Data stores	0.834***	1												
Databases	0.746***	0.419***	1											
Data intelligence	0.919***	0.577***	0.729***	1										
Software	0.770***	0.528***	0.710***	0.778***	1									

Intangible capital	0.0963	0.0784	0.0609	0.0910	0.102	1								
Digitized information	-0.0361	-0.0755	-0.00605	-0.0114	0.0291	0.750***	1							
Innovative property	-0.0917	-0.0748	-0.0873	-0.0803	-0.0350	0.918***	0.622***	1						
Economic competencies	0.552***	0.518***	0.373***	0.472***	0.368***	0.364***	0.135	0.122	1					
Data-intensive intangibles	0.622***	0.581***	0.378***	0.546***	0.418***	0.320**	0.0391	0.116	0.904***	1				
Organization	0.409***	0.331**	0.340**	0.351***	0.370***	0.330**	0.180	0.106	0.785***	0.766***	1			
Design	0.350***	0.403***	0.148	0.262*	0.161	0.0432	-0.204	0.0136	0.247*	0.598***	0.301**	1		
Brand	0.494***	0.492***	0.233*	0.440***	0.268*	0.267*	0.0769	0.0996	0.844***	0.732***	0.379***	0.0905	1	
Memo: R&D	-0.109	-0.0913	-0.105	-0.0918	-0.0349	0.894***	0.631***	0.988***	0.114	0.0769	0.0743	-0.0721	0.120	1

*Notes: ** p<0.01, *** p<0.001 . Countries include Denmark (DK), Germany (DE), Finland (FI), France (FR), Italy (IT), Netherlands (NL), Spain (ES), Sweden (SE) and the United Kingdom (UK). Gross value added is adjusted to include all intangibles, as reported in the EUKLEMS & INTANProd database (analytical module).*