Data, Intangible Capital, and Productivity: Literature review, Theoretical framework and Empirical evidence on the UK

F.Bontadini,[1] C. Corrado,[2] M.Iommi,[3] C. Jona-Lasinio[4]

LUISS Guido Carli
LIBERA UNIVERSITA' INTERNAZIONALE DEGLI STUDI SOCIALI
LUISS LAB OF EUROPEAN ECONOMICS

[1] LUISS University and SPRU – University of Sussex.
[2] The Conference Board and Georgetown University McDonough School Center on Business and Public Policy.
[3] ISTAT and LLEE.
[4] LUISS Business School.

Table of Contents

Executive summary

This report sheds new light on the contribution of data to productivity, focusing on the UK economy. In doing so, the relationship between data and other kinds of intangible capital is disentangled from a conceptual point of view. We distinguish three key components of data capital: (i) Data store, in the sense of raw digitized records (ii) the accumulation and transformation of such records into usable databases and (iii) the knowledge gleaned from the data which produces long-lived assets that contribute to final production. Building from this conceptual framework, the report argues that data capital can largely be subsumed within the intangible capital.

As a result, the report discusses the methodological approach to estimate investment flows and stock of data capital, consistently with the *2008 System of National Accounts (SNA)* and the estimation of other intangible assets. A key contribution of this report is to augment the well-established sum of cost approach to measure intangible assets with the insights from a first of its kind survey across a representative sample of firms for the UK economy. The report puts forward and discusses novel measures of data investment and stocks at the industry level. The average value of investment in data assets in 2012-2020 was 10.4 percent of UK total economy value added and the share steadily grew from 2017.

The report also identified some informational needs that should be prioritized in future surveys on data usage in businesses. These include identifying the occupations mostly involved in producing data assets and the time spent in this activity, collecting information on the cost components of the data production process, and collecting quantitative information on the market for data.

The report also expands the intangible capital theoretical framework and highlights the importance of the notion of appropriability to understand the implications of the rise of data capital for productivity growth. As data capital changes the composition of the economic characteristics of other intangible assets, it also increases the appropriability of the economic benefits of such assets, reducing as a result the measured total factor productivity in the economy (Corrado and others, 2023).

Finally, the report discusses brand new econometric evidence confirming two key findings. First, growth in data capital is associated to growth in labour productivity for the UK economy over the period; second, intangible investments and the data stack assets strongly overlap, 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.

A prompt policy response to the new data driven challenges should be focused on the definition of clear rules favouring data sharing to foster competition and innovation considering the complementarities between data and other intangible assets.

1 Introduction and relation to recent literature

Perceptions that the fruits of rapid technological progress have been unevenly shared across businesses and households have raised questions about whether policy frameworks are suitably adapted to the digital age. One key aspect of these concerns pertains to the role of data, especially the data exchanged between firms and consumers.

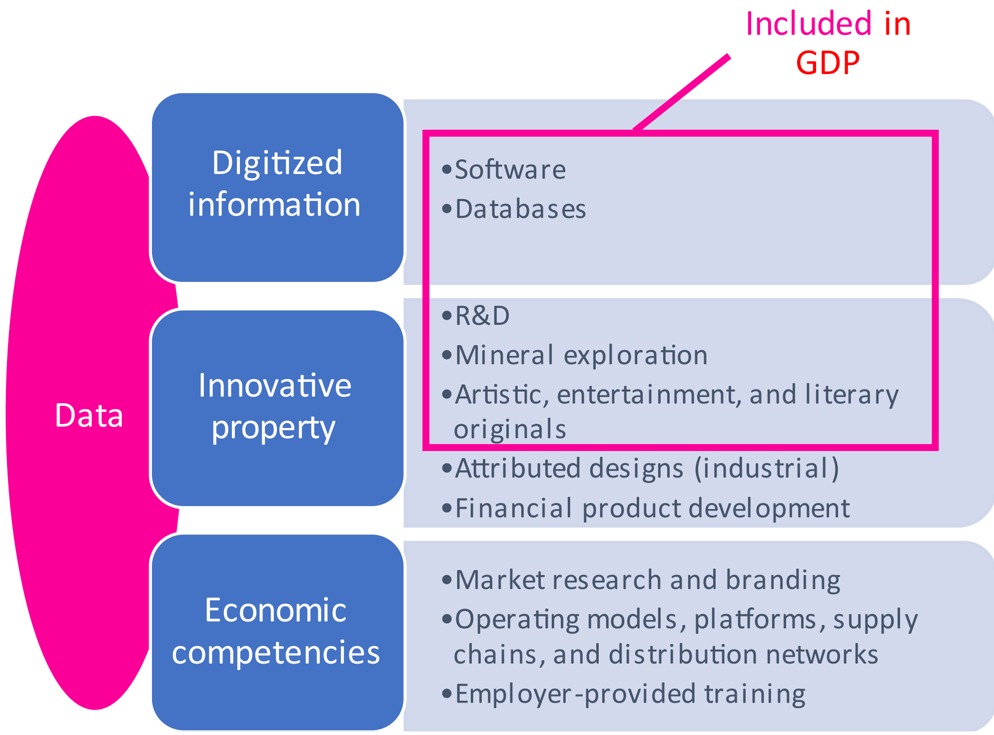
Data is an increasingly important input for businesses (for example, to train artificial intelligence algorithms) but is often neither accounted for in macroeconomic statistics nor part of business contracts for goods and services provided to customers. It is typical to associate economic growth with the emergence of new investment streams, the use of which leads to production efficiencies and proliferation of new goods. What, then, is new (or unique) about the increased use of data and data assets in economies? Did data have significant impacts on economic activity? What is the role of data in the UK economy?

To increase our understanding of the role of data assets in modern economies and to offer an answer to the above questions, this report provides an overview of the literature developed so far on the theoretical and methodological approaches to capitalize the value of data and an empirical investigation of the contribution of data asset to productivity growth in the UK. We propose a model for capturing asset creation based on all digitized information that is processed and transformed into useable knowledge in an economy. We define this knowledge as “data capital,” conceptualized as an intangible asset within a framework that is amenable to measurement and quantitative analysis.

The report shows that data assets – from stores of raw data to actionable intelligence derived via data analytic tools – are largely subsumed within the intangible capital framework attributable to Corrado, Hulten, and Sichel (2005, 2009), shown in figure 1 and discussed in more detail below. The report models the economic impacts of data capital using this framework, emphasizing how the relative importance of data capital within intangible capital lowers intangible asset prices and strengthens the (partial) appropriability[5] of the asset class.

[5] Appropriability is the capacity of a business to retain the added value it creates for its own benefit.

Figure 1 - Intangible Capital: Broad groups and investment types



Source: Adaptation of Corrado, Hulten, and Sichel (2005, 2009)

A key question is then how the proposed approach to estimate the value of data asset relates to existing literature. Recently, several research efforts, proposed many models that focus on the economic mechanisms affected by data (or intangibles).

At the micro level, data is usually assumed to have diminishing returns, for example, Varian (2019) points out that there are diminishing returns to more and more training data fed to AI algorithms. Jones and Tonetti (2020) formulate an aggregate model of data in an economy in which data are productive intermediate inputs with diminishing returns, not a “technology” that leads to increasing returns. In contrast, in the intangible capital model set out in Corrado and others (2023) data are productive long-lived assets whose value stems in part from the application of data technologies. There are obvious differences between these approaches, most notably while the former considers data as an intermediate input the latter conceptualizes data as capital.[6]

The intangible framework can thus be used to analyse data value creation and classify it as business investment. The identification of data as an asset is a significant improvement for productivity analysis as the role of data as an input of production combined with the existence of spillovers turns out to be a close representation of the welfare-enhancing processes theorized by

[6] Nonetheless, it is worth highlighting that the stylized Jones and Tonetti (2020) model is designed to highlight the aggregate welfare impacts of data sharing, and its implications transcend this conceptual distinction.

Jones and Tonetti (2020) in that (a) data assets have diminishing returns in production but (b) returns to data asset ownership may spill over to other firms to the extent they are shared within an industry or economy. Indeed, the innovative potential of data as intangible capital rests in its ability to yield competitive returns to owners and "spillovers" elsewhere in an economy. For further discussion of the correspondence between endogenous growth theory (Romer 1990, Jones 2005) and the intangible capital framework, see the discussion in Corrado, Haskel, Jona-Lasinio, and Iommi (2022).

Many models in a parallel stream of literature have attributed rising market power and/or industry concentration to scale economies of intangible assets at the firm level (for example., Haskel and Westlake 2018, Crouzet and Eberly 2019, De Ridder 2019). It is noteworthy that diminishing returns to data assets can co-exist with market power/cost advantages due to scope economies and local scale effects.[7] Recombining data for different uses may also create agglomeration effects that weaken diminishing returns.[8] Recent work at the Organisation for Economic Co-operation and Development (OECD) addressed whether within-industry dispersion of productivity can be attributed to intangibles and found that it could, but not necessarily due to economies of scale at the firm level (Corrado, Criscuolo, Haskel, Himbert, and Jona-Lasinio 2021).

Closer to our findings are studies that attribute declining business dynamism to a breakdown in knowledge diffusion and suggest that the breakdown owes, at least in part, to the increased use of proprietary data in modern production processes. Akcigit and Ates (2021) make this suggestion, but their empirical analysis focuses on knowledge derived by R&D, and by extension, data capital created via the conduct of R&D. Here we look at the creation of commercial knowledge created via modern data use more broadly, for example, via marketing or logistics experiments as well as R&D.

This report contributes to this literature providing an evaluation of the economic impact of data assets for the UK economy. We generate industry level estimates of data assets coherently with both national accounts and widely used concepts in the intangible capital literature and also consistent with concepts used by management strategists and technologists.

The estimates of data assets for the UK are obtained using the results of the Business data use and productivity Study (hereinafter "the Survey"), which was conducted by IPSOS UK in collaboration with LUISS University and NIESR as part of this project.[9] To the best of our

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

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

[9] In particular, the questionnaire was developed by Ipsos, NIESR and LUISS working closely with DSIT.

knowledge, it is the first time that measures of investment in data capital are generated using a dedicated survey.[10] More precisely, the data investment estimates rely on information on the time that different occupational groups spend producing data assets which was collected in the Survey specifically to provide inputs to measure data investment.

Resorting to these estimates, we investigate the contribution of data asset to UK productivity growth and provide some policy insights.

Our results, consistent with those found by Corrado and others (2023) for a sample of European economies, suggest that the increased data intensity of intangible capital boosts its contribution to labour productivity growth by about 0.25 of a percentage point over the sample period.

The report is structured as follows. Section 2 provides an overview of the definitions and macroeconomic implications of considering data as an asset. Section 3 illustrates the methodology to generate estimates of data assets while section 4 shows the sources and main measurement issues for the UK. Then section 5 reviews the conceptual framework by Corrado and others (2023) and section 6 offers an econometric analysis of the contribution of data to UK productivity growth. Section 7 concludes and proposes some policy insights.

2 Data Value Creation: definitions and macroeconomic implications

2.1 Use of data

A key element for the definition of a conceptual framework for measuring and analysing data needs to consider in the first place some of its characteristics: (a) data is nonrival[11] and capable of improving economic welfare when shared or replicated at low cost; but (b) data, though nonrival, is frequently used exclusively. Table 1 shows some examples of exclusive (or rival) versus nonrival use of data in modern economies. Also, although 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.

[10] The target population of the survey was all UK business. It was a telephone and online survey, with a total of 1,962 business decision makers taking part.

[11] In economics, a good is said to be rival if its consumption by one consumer prevents simultaneous consumption by other consumers, or if consumption by one party reduces the ability of another party to consume it. A good is considered non-rival if, for any level of production, the cost of providing it to a marginal (additional) individual is zero. See Apesteguia, J; Maier-Rigaud, F (2006). "The Role of Rivalry: Public Goods Versus Common-Pool Resources". *Journal of Conflict Resolution*. **50**: 647. doi:10.1177/0022002706290433. S2CID 6738663 – via SAGE journals.

Table 1 - Examples of Data Use

Rival:

1. Product-level forecasting (for example, Amazon)
2. A/B Internet testing and marketing (for example, Google)
3. IoT factory systems (for example, Siemens)
4. Targeted advertising on consumer content platforms
5. Fintech (for example, algorithmic trading, digital lending)
6. Product-led growth strategies
7. Customer lists/after sales services design

Non-rival:

8. DaaS (Data as a Service) platforms (for example, 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
    (for example, Zillow data used for economic research)
14. Genomic and other public biomedical research data
15. Official statistics (economic, demographic, social

Notes: Data is inherently non-rival, and classifications reflect the degree to which owners share data with other organizations or the public

The examples listed on lines 1–5 of the table mainly reflect applications of big data using new digital technologies by businesses, that is, 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.

Examples of “non-rival” data use range from marketers of personal data for companies (line 8), to longer-standing examples of industry-level data sharing, for example, 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”, that is, freely available for anyone to download, modify, and distribute without legal or financial restriction. In fact, 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).

These examples suggest that while data has much potential for use and economic benefit when shared, many applications of big data involve proprietary use. Data-dependent business models are on the rise (Nguyen and Paczos 2020), as are regulations to protect consumer privacy, for example, 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 (for example, 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, for example, the data sharing environments facilitated by open banking policies in the United Kingdom and other countries.[12]

A conceptual framework for capitalizing data as an asset must consider all these data characteristics. In what follows, we illustrate the proposed approach and its main macroeconomic implications.

2.2 The Data Value Chain

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. By contrast our approach is based on the following observations:

- Data, in the sense of raw digitized records, may accumulate at a fast 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 the economy.
- The accumulation of data has the potential to boost real output when producers also invest in transforming such records, possibly along with other available economic or social information, into analytical insights and actionable business intelligence.[13]
- Data stores and knowledge gleaned from data stores via application of data technologies are, in fact, long-lived intangible assets that can contribute to final production in an economy. The appropriability of returns to the intelligence gleaned from accumulated stores of digitized information implies that business spending on data and data transformation are intangible capital investments.

Our specific approach to data value creation embraces widely used approaches in the technology and management literatures. Technologists characterize data according to a “data stack” that describes the transformation of raw data into usable data structures and intelligence. Business

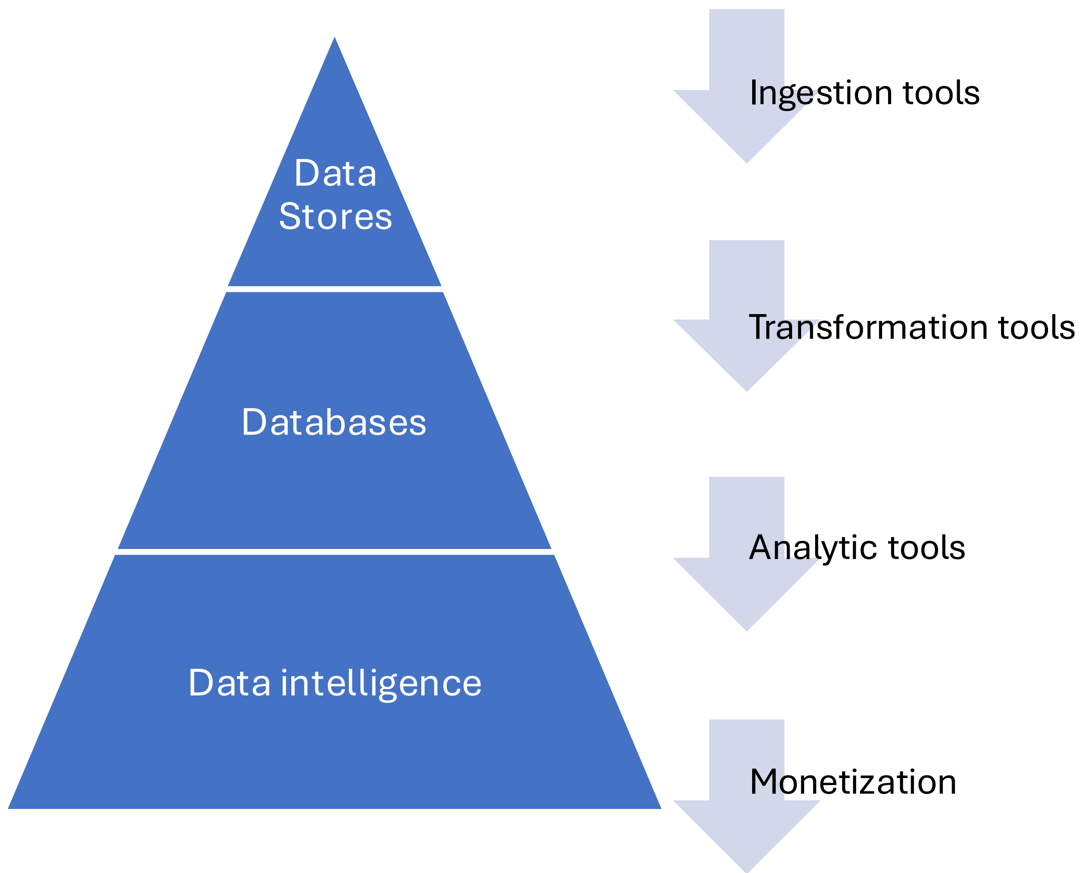
[12] Open banking refers to a data sharing environment in which financial intermediaries—both incumbents and fintech entrants—can compete for customers. For an analysis of the impact of open banking regulations on financial innovation globally, see Babina, Buchak and Gornall (2022).

[13] Farboody and Veldkamp (2021) show that increasing data availability affect business’s expectations on future economic trends thus potentially impacting on output growth.

management strategists use a value chain construct that adds monetization, or market implementation, as a capability (or tool) required for data value creation.[14]

Our framework for data value creation is illustrated in figure 2 below. Although it embraces both characterizations, note first that technologists usually stack a sequence of data forms and digital tools in a single pyramid.[15] Figure 2 separates these by identifying three major forms of data inherent in their characterizations. These forms, depicted on the left, reflect the business strategists' notion of an information value chain, where greater value is produced as data is processed into usable intelligence. The digital tools that enable value creation from raw, digitized information are depicted on the right.

Figure 2 - The Data Asset Value Chain



Notes: The stack to the left depicts the stages of the data value creation, which are created by applying the tools illustrated on the right.

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

[15] See, for example, Roca (2021a), for a recent depiction. The data stack has its roots in information science, which uses the concept of a "data pyramid" to depict the relationship between data, information, and knowledge (Varian 2019).

The data asset stack has then three layers of value—data stores, databases, and data intelligence—each corresponding to an asset type amenable to measurement and analysis.[16] The asset types are defined more precisely as follows:

- *Data stores* are raw records that have been stored but not yet cleaned, formatted, or transformed for analysis, for example, 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 (for example, machine learning training algorithms); data intelligence is a set of quantitative inputs that provide actionable guidance for decision-makers, including solutions to scientific problems.

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

Estimates of data assets guided by figure 2 will capture the resource cost value of all data processed, transformed, and used in an economy. It does not, however, isolate *personal* information, the valuation (and protection) of personal data is viewed with keen interest. This is in no small part because some of the largest and fastest growing tech companies (such as Alphabet, Amazon) are built mainly on the economics of transforming personal information into business and marketing intelligence,

The World Economic Forum (WEF, 2011) and OECD (2013) identify 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 but help clarify conceptual issues regarding the relative importance of personal information. Figure 3 sets out examples of "raw" data by economic sector origin based on the WEF classification.

[16] A multiple asset type conceptual approach has been used in previous work on defining and measuring data, including McKinsey Global Institute (2016), Statistics Canada (2019a, 2019b), Nguyen and Paczos (2020), and Goodridge, Haskel and Edquist (2021).

Figure 3 - Classification of data by origin of raw information

	Institutional data		
Personal data	*Businesses*	*Governments*	*Non-Profits*
User-Generated	Personnel Files	Personnel Files	Personnel Files
Behaviour	Accounting Records	Accounting Records	Accounting Records
Social	Legal Documents	Legal Documents	Legal Documents
Location	Financial Documents	Financial Documents	Financial Documents
Demographic	Customer Lists	Intelligence Records	Social Policy Programs
Official Identification	IoT Sensors	Diplomatic Cables	Public Policy Programs
		Defence Files	
		Statistical Surveys	
		Regulatory Records	
		Admin Records	
		Monitoring Technology	

Source: Adaptation of Kornfeld, Robert (2019), slide 8, drawn from WEF (2011) and OECD (2013).

The collection and use of personal data (circled data items in figure 3), that is data on persons collected online and data on production processes collected via Internet of Things (IoT) sensors, are generally considered sources of information made possible by the advent of modern digital systems. Operational data and customer lists have long been part of legacy systems exploited for competitive advantage, but the information in these systems are now targets for modern digitization, including integration with the newer "big data" on consumers and business processes.

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

2.3 Macroeconomic Implications of the Data Stack

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

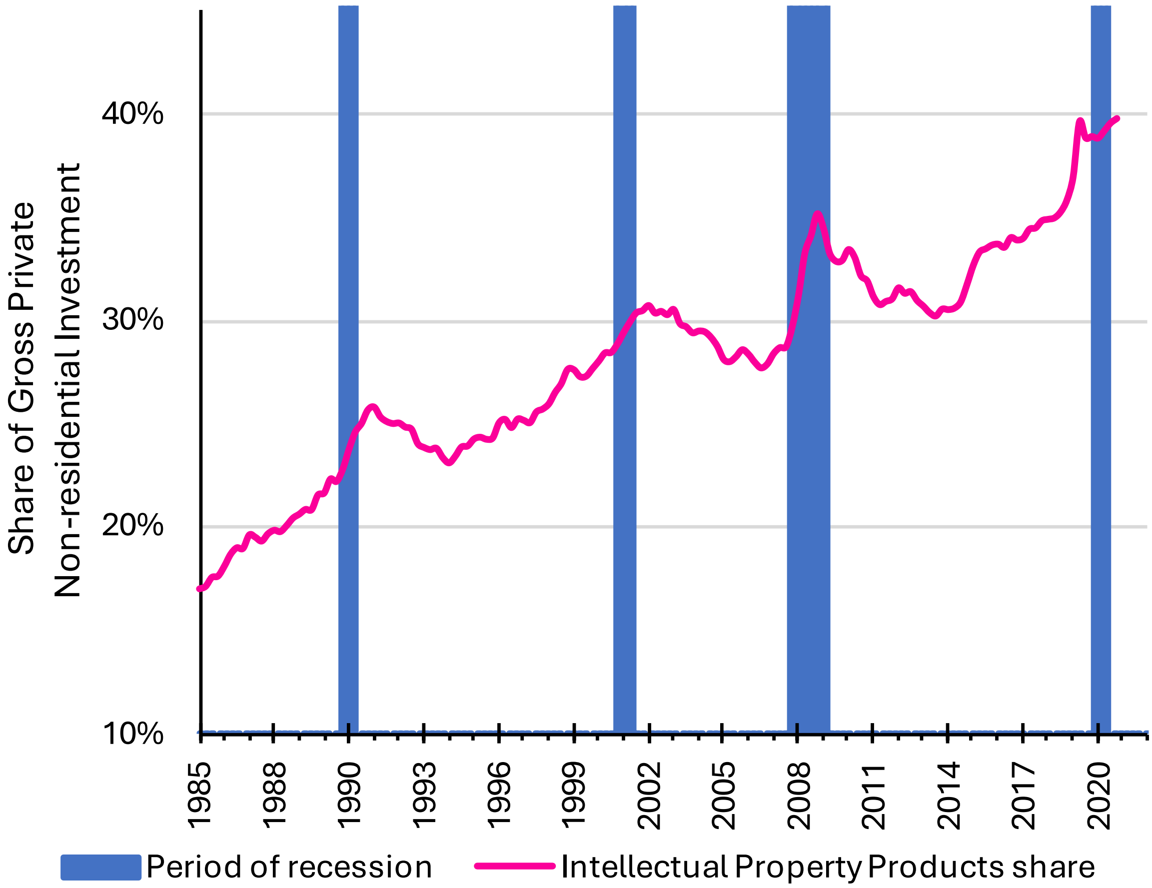
The data value chain framework, in which greater value added is created as raw data is processed and developed into insights and solutions, applies to data-driven development of engineering designs, customer platforms, and organizational practices, as well as to data-driven R&D processes. This suggests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

From this perspective, that is a knowledge-based or intangible capital perspective, the increased use of data assets derived from modern digital technologies is an “innovation in the method of innovation.” Modern data use fosters faster, more efficient experimentation and feedback in R&D processes, industrial production processes, marketing research, and business strategy and operating model development. This implies that, with increased use of data and application of digital technologies, the “productivity” of these activities improves, that is, that their resource cost per unit of final output falls.

The depiction of monetization as a capability in figure 2 refers to an organization’s capability to implement actions guided by data intelligence. The underlying idea is that the increased use of data in economies requires some adjustments in the use of existing primary factor inputs that is, labour and capital (tangible or intangible), eventually substituting or complementing them.

Though the primary focus of the analysis developed in this report is on how the increased use of data affects productivity growth, the rise of data capital as a strategic factor has the potential for altering cyclical patterns in macroeconomic data—patterns of investment and factor input demands, and perhaps the responsiveness of inflation to economic conditions in the short run. Though subjects for future research with more complete data, the partial incorporation of intangibles in quarterly GDP (see Figure 4 below) hints that there is indeed something different about the workings of the intangible macroeconomy.

Figure 4 - U.S. Intellectual Property Product Investment (1985Q1 - 2021Q4)



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

Figure 4 displays fluctuations in the intellectual property products (IPP) share of private non-residential investment using quarterly data from the U.S. national accounts, which suggests that these investments are the last category of capital spending cut during downturns.[17] Research on the formulation of investment demand argues that intangibles are less sensitive to changes in interest rates than tangibles due to their higher user cost and tendency to be less reliant on secured debt financing.[18] Businesses may view the acquisition of software (and other intangibles) as moves to increase efficiency that dampen the impact of workforce layoffs and cutbacks in customer demand, that is, that intangible capital (or some forms of it) may allow businesses to adjust production relatively rapidly to changes in economic conditions, with possible implications for inflation dynamics and monetary non neutrality.

---

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

[18] See, for example, Haskel and Westlake (2018, chapter 8), Crouzet and Eberly (2019), and Döttling and Ratnovski (2020) for further elaboration.

Furthermore, the most recent observations in figure 4 show that IPP investments remained relatively strong in the *recovery* from the economic downturn caused by the pandemic Corrado and others (2023).

The fact that intangible capital increasingly reflects knowledge built from the analysis of data may explain this persistence of relative strength. Half of the respondents in survey of companies administered by McKinsey & Company reported that the pandemic-induced economic downturn had *no effect* on their investments in AI, while 27 percent reported *increasing* them (AI Index Report 2021, page 103).

2.4 Data capital as Intangible capital

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

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

Table 2 - Intangible Investment: Major Categories and Asset Types

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

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

At first look 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 it may be seen in the itemized list in column 2 of table 2, only databases appear. This implies that national accounts' estimates of the value of investment in databases exclude the cost of acquiring or ingesting the data stores they contain; furthermore, as a matter of practice, outright purchases of data stores and databases are only included to the extent they are embedded or sold as software products.[19]

Consider now data intelligence, the most valuable, final stage of the data value chain as seen in figure 2. This is where the utility of the intangible capital framework becomes especially apparent. The knowledge created from data encompasses all modern, data-driven business, financial, marketing, engineering, and scientific intelligence. Intangible capital of investments in business operations, marketing, financial products, and engineering design (in addition to R&D and mineral exploration) is readily seen via lines 7, 8 and 9 of table 2.

An increase in the use of data capital in R&D activities (line 3), will cover novel forms of data-derived scientific intelligence (for example, the development of new AI techniques and certain bio-engineered substances or formulas). It will exclude, however, many uses of modern data-driven engineering design that yield improved industrial production systems, that are typically 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; they also are a component of line 9, which includes investments that design the re-engineering of in-house computer systems and computer network platforms to make use of cloud infrastructure services, data analytic services, and data.

The intangibles framework thus covers most, if not all, forms of data intelligence; virtually all assets in table 1 are potentially data driven. The perspective offered by the framework then informs the development of empirical estimates of data intelligence. Other approaches, including those that conceptualize data assets as a value chain, have missed certain 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.

Certain components of intangible capital are directly related to the fact that AI-based data systems involve the use of new software and cutting-edge computing systems hosted in the cloud. Byrne and Corrado (2017), show two series that arguably capture the data-driven demand for cloud services. The two series are business R&D in IT services and software development, and purchases of computer and network design consulting services – these are underlying components intangible investment categories listed on line 3 and 9 of Table 2: that is R&D and business process and organizational practices, respectively.

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

As it may be seen, these data and AI driven components of intangible investment have grown substantially, nearly tripling relative to private sector GDP over the period shown.[20]

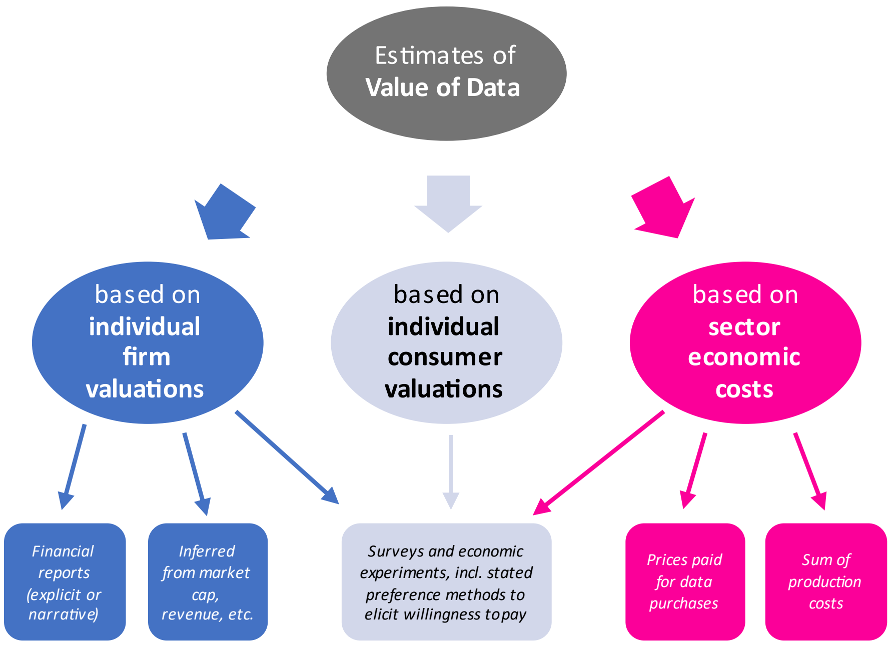
In summary, beyond the main message of this section that data capital is largely covered in intangible capital, key findings regarding the measurement of data capital are as follows:

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

3 Measuring Data: methodological overview

How much value do firms derive from data? And how is this related to consumers' valuations of their personal information or to costs sustained by firms to obtain the data that are used and/or created via the data stack?

Figure 5 - Approaches to the Valuation of Data



[20] This share relative to total GDP is 1.2 percent in 2018, which would *not* include public funding for AI research, suggesting that the true contribution of AI software research to total GDP is higher.

In addressing these questions, one encounters different perspectives and different measurement approaches to the valuation of data. The economics literature has taken three main directions to develop estimates of the value of data. As depicted in the middle panel of figure 6, these include approaches based on individual firm valuations, approaches based on consumers' valuations, and approaches based on sector (and/or industry) economic costs. The bottom panel of the figure indicates methods used under each approach. Surveys and economic experiments (the middle box in the bottom panel) are methods not unique to a given approach, as the figure indicates.

Approaches aimed at valuing consumers' personal information will not encompass the full data value chain of figure 2, as previously discussed. The review of methods is targeted at methods that can yield comprehensive coverage of data use in market activities in economies. Thus, first we discuss methods that have been used to estimate the value of data for individual businesses and/or based on individual business-level data. Then a discussion of stated preference methods applied to measure the value of data used in business is illustrated (Brynjolfsson, Collis, and Eggars, 2019). The sector-cost approach as deployed by national accountants is briefly summarized and it is then adopted to generate the estimates of the data capital for the UK economy as illustrated below.

3.1 Methods Based on Firm Valuations

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

3.1.1 Business reporting

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

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

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

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

3.1.2 Revenue-based approaches

Another interesting approach suggested by Nguyen and Paczos (2020) aims at capturing the value of data based on the revenue shares driven by data monetization across different types of firms (for example, manufacturers, utility providers, banks, or online platforms). Nguyen and Paczos (2020) adopt a stylized taxonomy of business models distinguishing two main categories: data-enhanced or data-enabled. This approach can be easily implemented even if it requires additional efforts from national statistical institutes to conduct ad-hoc economic surveys and coordinate internationally to guarantee comparable results across countries.

3.1.3 Demand-side approach

Coyle and Li (2021) develop a demand-side methodology for estimating the size of data markets using the recent finding that an online platform's entry can disrupt incumbent businesses' organizational capital by affecting its depreciation rate. They calculate the stocks of organizational capital based on before-entry and after-entry depreciation rates. This difference captures the loss caused by the failure of using data to cope with changes in competition due to the entry of an online platform. Thus, this methodology can be used to measure the potential size of the demand for data by incumbent businesses in the industry sectors disrupted by online platforms. In other words, they use the loss of value of incumbent firms' organizational capital to measure businesses' maximum willingness to pay for the access to data.

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

3.1.4 Market prices

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

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

It is also illustrative to examine financial indicators per record from companies that derive most (or all) of their income from advertising linked to personal data, for example, Facebook/Meta. Ahmad, Ribarsky, and Reinsdorf (2017) calculate a value equivalent to around 0.02 percent of global GDP for the user data collected by five major digital services (Facebook, Twitter, Instagram, LinkedIn, and Gmail) based on the number of active users and assumed prices of a user profile.

3.2 Stated Preference Methods

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

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

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

3.3 Sector economic costs

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

If businesses purchased all or most data 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 (as well as other intangibles, including software and R&D) is not transacted on markets but produced in-house. Thus, national accounts compilers must come up with two components for intangible investments, own-account investment (when data are produced and used in-house) and purchased investment (when data are bought and sold in market transactions) to measure nominal investment flows in data assets. Consider now how each component might be estimated.

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

Imagine a business having a "software factory" or "R&D 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 (labour, capital, and intermediates). The key to accomplishing this task—called a "sum-of-costs" approach—is to identify the occupations of workers employed in the factory and to estimate their compensation. Based on knowledge of the compensation paid to these workers, the total payments made to all factors involved in the in-house production can be estimated. As a practical matter, the identified workers may not be involved in producing new assets their entire workday; for example, the conventional approach to measuring in-house software production in national accounts is to assume that software developers spend just 50 percent of their time working in their business's "software factory" to produce original code. In-house production of data assets can be estimated in a similar fashion.

The *System of National Accounts 2008 (2008 SNA*) explicitly recommends that national statistical offices use the sum-of-costs approach to estimate software and databases (unless produced for sale) and R&D (unless the market value of the R&D is observed directly) and the own-account component of any product for which it is not possible to find the price of a similar product. The INTAN-Invest and EUKLEMS & INTANProd databases use a sum-of-costs approach to estimate own account components of non-national accounts intangibles.

3.3.2 Purchased data assets

Purchased data should be valued at the transaction price. Although conceptually simple, measuring the purchased component of data investment is challenging because comprehensive data sources are very scant. 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. Ker and Mazzini (2020) considered business statistics sources and looked at the revenues generated by businesses that create explicit value from data (those collecting, compiling, and selling databases). But they found that focusing mainly on industry classifications is likely to generate an inexact identification of these activities. For example, Zillow sells its data on home real estate valuations, Nielsen sells its survey data, as do credit agencies such as Experian, but these businesses are in widely different industries. Also, monetizing databases is not necessarily the primary line of business for many businesses who charge for purchased databases or are in the business of producing data intelligence (for example, Gartner, McKinsey).

3.4 A sum-of-costs approach

A strict sum-of-costs approach does not quantify the difficult-to-measure purchased component of data investment or attempt to fold in information on trade flows, in effect counting all in-house production of data in the economy as investment. Despite this limitation, the approach produces conceptually comprehensive estimates that are a good proxy for total investment.

Statistics Canada (2019a, 2019b) set out a framework for measuring data using a strict sum-of-costs approach and prepared experimental estimates for Canada's total economy and major institutional sectors—non-financial corporations, financial corporations, non-profit institutions serving households, and governments. In the Statistics Canada application, occupational groups were selected from among those generally associated with converting observations 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 2, though Statistics Canada called the third category "data science" and viewed it as unmeasured R&D, for example, spending to develop new AI algorithms. Data and AI data tools are inextricably bound via feedback training data used to develop new tools, but the data stack or data value chain notion of how value is created from data does not end with the development of *new* algorithms. Novel science and engineering inventions are a rightful conceptual boundary for the funding (and measurement) of R&D in service of science policy goals. But this boundary is as limiting for thinking about data capital as it has been for intangible capital more broadly (see again figure 1 and table 2). Section 2's discussion of the data stack and its overlap with intangible investment suggests that value creation due to data intelligence occurs when existing AI tools or analytics are applied to existing (but perhaps newly digitized) data to obtain new 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 1.75 to 2.25 percent of the country's GDP in 2018. They further found that about 47 percent of the total was accounted for by non-financial corporations, 31 percent by financial corporations, 20 percent by governments, and 2 percent by non-profit institutions serving households. Finally, Statistics Canada produced estimates in volume terms (that is, adjusted to consider price changes) and data 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.

The choice of a lengthy service life of 25 years for data stores seems reasonable though it is an assumption requiring further study. When thinking about the valuation of data, Varian (2018) argued that data exhibit decreasing returns to scale, citing the example that an increase in the size of training data for AI algorithms yields diminishing returns in prediction accuracy. This is an important aspect of how the value of a given set of data depreciates with time, but like most intangibles, data exhibit economies of scope and the merging of two complementary datasets may produce more insights than possible from each alone. This suggests that the appropriate concept for data asset depreciation should be based on the observation that diminishing gains occur as new dimensions (or combinations) in use diminish.[21]

Goodridge at al. (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

---

[21] Li, Nirei, and Yamana (2019) explore this observation to estimate the influence of data assets on market valuations of digital platform companies. The fruits of combining data on the human genome (hardly new data) with "new" personal lifestyle data in applications for precision medicine solutions are another example of this observation.

fixed capital formation in software and databases by around 60 percent compared to own-account investment measured in EU official national accounts.

4 Implementing the cost-based approach to the UK data

In this section we illustrate the approach adopted to generate estimates of investment in data assets and corresponding capital stock for the UK both for the total economy and industries (SIC 2007 sections) over the years 2012-2020.[22] We adopted the cost-based approach because, due to a lack of data on business expenditure to purchase data assets, it is the only way to generate estimates of data investment that are at the same time comprehensive, comparable across different industries, and consistent with other national account variables.

As discussed in Section 3, investment should be measured as the sum of the purchased and the own-account component. EUKLEMS & INTANProd measures investment in non-national accounts intangible assets as the sum of the two components, but this approach is not feasible for data assets because, information on transactions in data stores, databases, and data intelligence, whether domestic or international trade, are not apparent in official statistics or any other exhaustive private source. In contrast, our measures of investment in data assets and software capture the value produced both in the total economy and at the industry level, regardless of whether the produced output is intended for its own final use or final sale.

Implicit in our measure is the assumption that output and investment coincide for data assets. Whether this assumption is realistic depends on how large the data assets market is compared to the in-house data production for internal use. Our view is that the most significant component of investment in data is through the production of data assets for own final use, and not through purchasing data asset in data markets. This is especially true for the total economy and major sectors (for example, when measuring data investment for the government sector and the business sector as a whole).

When measuring data investment for the total economy, the difference between investment and output only depends on the value of net exports of data assets, which is likely to be small. The difference at the level of major sectors also depends on how large data transactions between the government and the market sector are. Overall, they are likely to be small. On the one hand, it is very unusual that the government purchases data assets from the business sector. On the other hand, it is more common for the government to produce data assets, which are then made freely available to the private sector, either as open data or in other forms. However, according to the national accounts principles, when the government makes data assets freely available to businesses, remains the owner of the data and there are no investment transactions to be

[22] Our estimates are fully exhaustive and include business, public and third sector.

recorded.[23] Thus, it is correct to assume that any output of data assets produced by the government should be measured as government investment.

Our assumption that output and investment coincide is less robust when we want to measure data investment at the industry level. As a matter of fact, a business-to-business market for data exists. When there is an outright sale of data (that is, an explicit transaction that transfers the ownership of the data from one firm to another), the data producer is no longer the data owner, and our approach fails to assign investment to the proper industry (assuming that the data buyer and the data seller are two businesses classified in different industries).

However, the most common data transactions are data sharing and licenses to use data, not outright selling of data ownership. Our approach assigns data investment to the correct industry for most of this transaction. Data sharing is when a company decides to share its data as part of corporate partnerships. In this case, there is no transaction at all, and our approach correctly assigns investment (measured as the value of data produced) to the firm that has collected the data. Regarding licenses to use data, they should be recorded as an investment of the business that buys the license only if there is a multi-year contract.[24] Thus, our view is that overall data output is a good proxy for data investment, even at the industry level. On the other hand, collecting empirical evidence on how firms have access to data and in what form should be prioritized.

4.1 Measuring investment in data asset

The main assumption of the cost-based approach is that the value of an asset can be obtained as the sum of the costs sustained for producing it. The benchmark equation to be estimated is as follows:

$$Y_{bc}^{i} = COMP^{i} + IC^{i} + CK^{i} + T^{i} \tag{1}$$

where *i* = asset type, Y_{bc}^{i} is the value of the produced asset at basic prices, $COMP^i$ is the labour 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 (1) are not directly measurable, thus the sum of these unmeasurable components, set equal to α, is a factor that must be approximated. Hence equation (1) can be re-written as:

[23] The same principle is applied in national accounts with reference to the results of R&D activities conducted by the government sector and made freely available to the users.

[24] According to the 2008 SNA (paragraph 10.100), if the acquisition of a copy (of any intellectual property product) with a licence to use is purchased with regular payments over a multi-year contract and the licensee is judged to have acquired economic ownership of the copy, then it should be regarded as the acquisition of an asset (that is, as investment). However, if regular payments are made for a licence to use without a long-term contract, then the payments are treated as payments for a service (that is, as intermediate consumption and not an investment).

$$Y_{bc}^{i} = COMP^{i} + \alpha^{i} \quad (2)$$

where Y_{bc}^{i} can be measured directly by computing the compensation of employees ($COMP^i$) and finding a proxy for α. $COMP^i$ can be obtained as:

$$COMP^{i} = EMP_{tot}^{i} * W_{avg}^{i} * t^{i} \quad (3)$$

where EMP_{tot}^{i} is the total number of employees employed for producing the relevant asset, W_{avg}^{i} is the average remuneration (average wage) and t^i refers to the time spent on these activities (table B1, in appendix B, shows the time assumptions, t, underlying the calculations developed in this report). As different occupations spend different amount of working-time in data-producing activities, in practice equation 3 is applied to each occupational group separately and $COMP^i$ is obtained aggregating across the occupational groups.
Using equation (3) and substituting it in equation (2), where it is assumed that $\alpha = COMP^{i} * bp^{i} - COMP^{i}$, the value of the produced asset is determined as:

$$Y_{bc}^{i} = COMP^{i} * bp^{i} \quad (4)$$

where bp^i 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.

4.2 Implementing the procedure

The estimates of data assets illustrated in this report have been produced applying equation 4 across industries. 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 (t^i in equation 3 above); iii) data on the number of employees for the relevant occupations and their compensations (EMP^i and W^i in equation 3 above); 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 (bp^i in equation 4 above).

4.2.1 Selection of data-producing occupations

Table B1 in appendix B shows the list of occupations that are assumed to be engaged in producing data assets based on UK Standard Occupational Classification (SOC 2010) codes and of the time-use assumptions by asset. We have selected the relevant occupation starting from Corrado and others (2023), that identified the relevant occupations and the corresponding time-use assumptions at four-digit level of the US Standard Occupations Classification (SOC) and then these have been mapped to corresponding occupations in the UK SOC.

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, including only on data-science related occupation.

4.2.2 Data-producing occupations' wages

The main data source on wages of the selected -data-related occupation in the UK is the Annual Survey of Hours and Earnings (ASHE), which is the most comprehensive source of information on the structure and distribution of earnings in the UK. ASHE provides information about the levels, distribution and make-up of earnings and paid hours worked for employees in all industries and occupations.

From the ASHE survey we have calculated the value of wages cross-classified by industry (divisions from A to S of the Standard Industrial Classification, SIC 2007) and relevant occupation (79 occupations identified at the level of four-digit of the SOC 2010). The data cover the years from 2012 to 2020.[25]

The ASHE tabulations could not be immediately used to estimate data investment for two reasons. First, due to confidentiality issues, data for some relevant occupations at the industry level are not disclosed. Second, due to high estimates variability, there is large time-series variation in the results. Thus, we needed to set up an imputation procedure to fill blanked cells and a procedure to smooth the time-series dynamic. The imputation procedure was applied to estimate missing values of wages (by industry and four-digit occupation), while the smoothing procedure was applied to the shares of each four-digit occupation in total wages by industry.

The imputation approach was applied to estimate the level of wages and relied on the fact that data for each four-digit occupation for the total economy are always available and that data for each three-digit occupation at the industry level have very few missing cells. The first step was to calculate, for each four-digit occupation, the total value of wages to be imputed as the difference between total economy value and the sum of non-missing wages across industries. The basic idea of the imputation was to use as a preliminary estimate of wages of each missing four-digit occupation in each industry, the value of the corresponding three-digit occupation in the same industry. Then, the value of the imputed four-digit occupations were re-scaled to the total value of wages to be imputed.

One of the pillars of our methodology is to produce estimates of investment in data asset that are as consistent as possible with official national accounts. Therefore, we have adjusted the wages of

[25] Producing the tabulations requires ASHE microdata, which can be accessed only through the ONS Secure Research Service. Thus, the elaboration of ASHE data was a joint work of LUISS and NIESR. LUISS has identified the list of data-related occupations (at four and three digits of the UK SOC), and NIESR has produced tabulations of ASHE data on the number of employees and their compensation.

the data-producing occupations to make them consistent with compensation of employees (CoE) from national accounts.

The adjustment is as follows: first we have calculated the share of the wage of each relevant four-digit occupation in total industry's wages from ASHE; second, we considered 2019 shares as our benchmark, and we applied a winsorisation-like procedure to smooth yearly variations higher than 1.3 or lower than 0.8;[26] finally we have multiplied each (smoothed) share for the industry's CoE to get the CoE for each data-producing occupation.

4.2.3 Time-use factors

Determining what fraction of workers' time is spent creating data asset is difficult and the existing empirical evidence is very scant. For this reason, most of the existing literature relied on subjective assumptions. As stated in Statistics Canada (2019b), additional work is required in the future to collect data about both the specific occupational groups that engage in data production activities and the shares of their labour inputs associated with the activity.

The Survey conducted by Ipsos (2023) about data usage within businesses ran as part of this project is a first step to collect data on the time that selected occupations spend in data asset-producing activities on average.[27] To the best of our knowledge, this was the first-time information on time spent producing data asset is collected through a specific survey.

On the other hand, the survey did not help to select the relevant occupational groups, because it collected information for 13 broadly defined occupations (approximately to the level of two-digit SOC). Thus, to match the survey's results with the four-digit relevant occupation it has been necessary to make further elaborations: for each four-digit occupation we used the time-percentages of the corresponding broad occupational group from the survey. The selected relevant occupations also include some occupations for which there is no matching with the broad occupational group of the survey. In this case, we have used the same time-use factor used in Corrado and others (2023).

4.2.4 Blow-up factors

The bp^i factor in equation (4) is an essential element for generating a measure of output consistent with national accounts. In this report we have used the blow-up factors estimated by Corrado and others (2023). The bp^i for each asset is 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

---

[26] Winsorisation is when values above or below a specific value are set to that value, restricting the range of a distribution of results. More precisely we edited growth rates of wage shares as follows. Let WSl,j,t be the share for workers type l in industry j at time t and $varWSl,j,t = WSl,j,t / WSl,j,t\text{-}1$. If $varWSl,j,t > 1.3$ then $adj\_varWSl,j,t = varWSl,j,t*0.2 + 1.3*0.8$. If $varWSl,j,t < 0.8$ then $adj\_varWSl,j,t = varWSl,j,t*0.2 + 0.8*0.8$.

[27] Ipsos run the survey, while developing the questionnaire was a joint effort by Ipsos, LUISS and NIESR.

in our framework. This adjustment is relevant, especially for those industries producing a sizable share of intangibles whose production structure is assumed to be 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^i of the relevant industries averaged over 1997-2020 are then applied to each data asset, they are set equal to 1.7 for data intelligence and 1.8 for the other assets.

4.3 Capital stocks 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.

In particular, we use the so-called geometric model, which defines the real stock of data asset *j* in industry *i* at the end of year *t* ($Kq_{i,tj}$) as:

$$Kq_{i,tj} = Kq_{i,(t-1)j} * (1 - \delta_i) + \mathrm{Iq}_{\mathrm{i,tj}} \qquad (5)$$

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 and others, 2023).

The depreciation rates are defined as follows: 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).

4.4 Main results

We produced three alternative set of estimates of investment in data assets for the UK based on the following time-use factors: 1) A baseline version that uses time-use factors as in Corrado and others (2023) (labelled "baseline"); 2) a version based on the Survey's time factor which, consistently with version 1, uses the time-factors estimated for the total economy across all industries ("survey_TOT"); 3) a version that uses industry-specific time factors from the Survey ("survey_IND"). In option 3) we have used the Survey's total economy time factors for the industries not covered by the survey (for example, public administration).

We believe that the best approach to setting the time-use percentages should be based on information collected directly from businesses and consider cross-industry heterogeneity in the importance of data assets. Thus, overall, our preferred results are those based on the industry-specific time factors derived from the Survey.

At the same time, it should be noted that the Survey's results should be considered very preliminary. It was the first time a time-use survey was conducted on the time spent producing data assets. The analysis of the answers showed that respondents sometimes misunderstood the question regarding the time percentages and declared a total percentage of time spent on producing data assets and software higher than 100 percent. In addition, for each occupational group, the declared percentages of time spent producing the three data asset types are very similar (table B1, appendix B). Again, this might signal that respondents did not fully understand the differences across the data assets and that the time percentages should only refer to the time spent in activities related to producing data assets (and not using existing data or producing any type of data). Finally, sample sizes for some roles are very small, with wide confidence intervals.

Therefore, it is crucial that future waves of the survey will collect information on time factors, refining the sample and the questionnaire based on the results of this first wave. In addition, future waves of the Survey might ask respondents to list the occupations more involved in producing data assets (instead of only providing time percentages for a pre-selected list of occupations). In this way, survey results could also inform the selection of the relevant occupations.

Based on the industry-specific time-use percentages derived from the Survey, in 2020 (the last year available in our estimates), the UK invested 222,932 million pounds sterling in data assets, compared with 213,977 million in 2019. The market economy accounted for 190,516 million (85 percent of total investment), and the non-market economy accounted for 32,416 million (15 percent) (**Error! Reference source not found.**). The shares of the market and non-market economy were very stable in the whole period from 2012 to 2020.

Table 3 - Investment in Data Assets in the UK, Millions of Pounds Sterling, 2012-2020

	2012	2013	2014	2015	2016	2017	2018	2019	2020
Market Economy	129,158	136,196	145,258	148,222	153,480	163,952	172,077	181,760	190,516
Non-market Economy	23,714	23,952	25,002	25,185	26,793	27,300	29,675	32,217	32,416
Total Economy	152,872	160,148	170,260	173,407	180,273	191,252	201,752	213,977	222,932

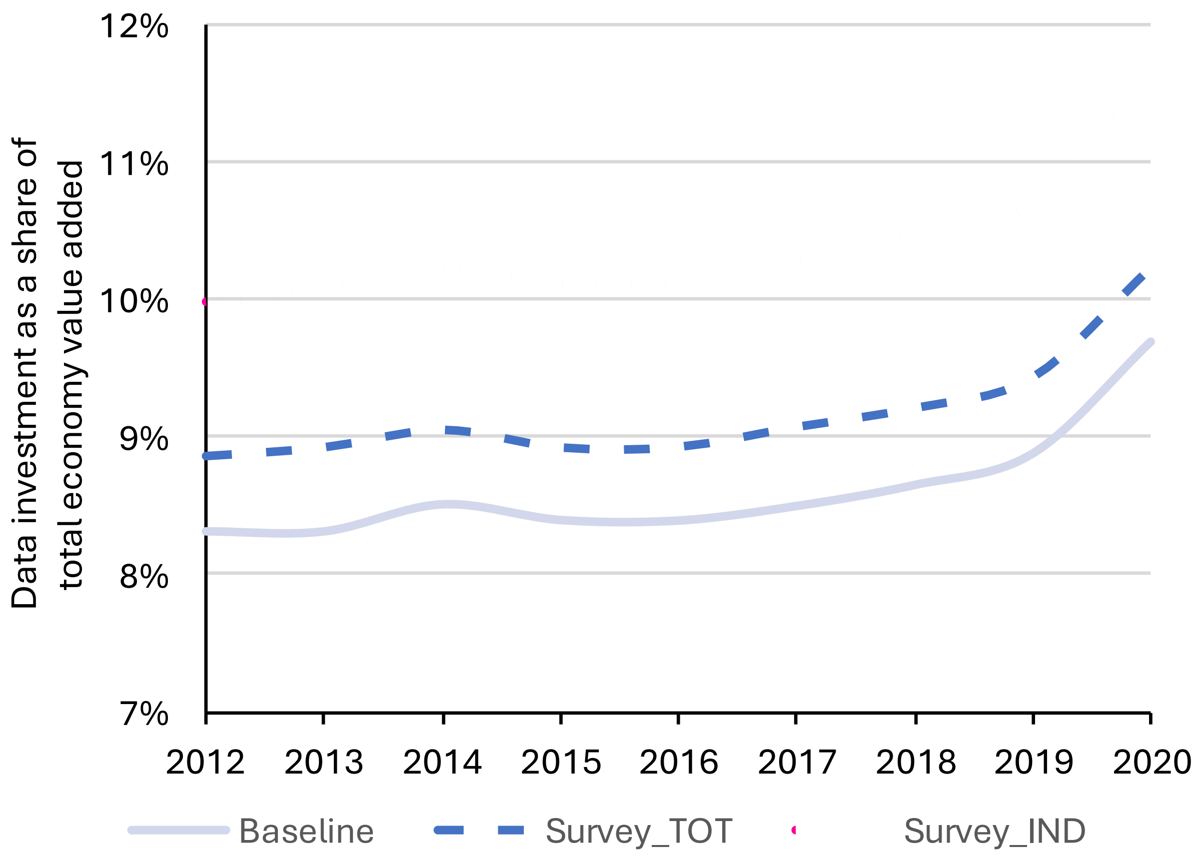
Source: Authors' elaborations on the UK Annual Survey of Hours and Employment and EUKLEMS & INTANProd.
Note: Market economy includes all industries except 2007 SIC divisions L, O, P, Q, and T. Non-market economy includes 2007 SIC divisions L, O, P, Q, and T.

Figure 6 shows that data investment as a share of total economy value added remained stable at about 10.1 percent from 2012 until 2016 and then showed a steadily growing trend afterward, with a value of 10.7 percent in 2019. The significant increase in 2020 (11.7 percent of total value added) reflects the combined effect of a further increase in data investment and the decrease in total value added during the first year of the COVID-19 pandemic.

Using time factors that are not industry-specific does not significantly affect the dynamic of data investment over time but reduces the level (Figure 6). Measures of investment in data assets based on total economy time factors estimated from the survey are about 12 percent lower than those based on industry-specific information. Baseline measures are about 18 percent lower than the industry specific estimates.

However, a relevant finding of the project is that the two estimates based on time factors that are not industry-specific are consistent in terms of dynamics and levels (with only a six percent difference). Although the two estimates use the same list of occupational groups, they are derived using partially different data sources (the ASHE survey for "Industry_TOT" and the EU Structure of Earning Survey for "Baseline") and time factors based on entirely independent sources (the survey for "Industry_TOT" and based on subjective assumptions in "Baseline"). Getting consistent results from two independent estimates reassures on the accuracy of the results.

Figure 6 - Investment in data asset in the UK, investment share in value added, 2012-2020



Source: Authors' elaborations on the UK Annual Survey of Hours and Employment and EUKLEMS & INTANProd.

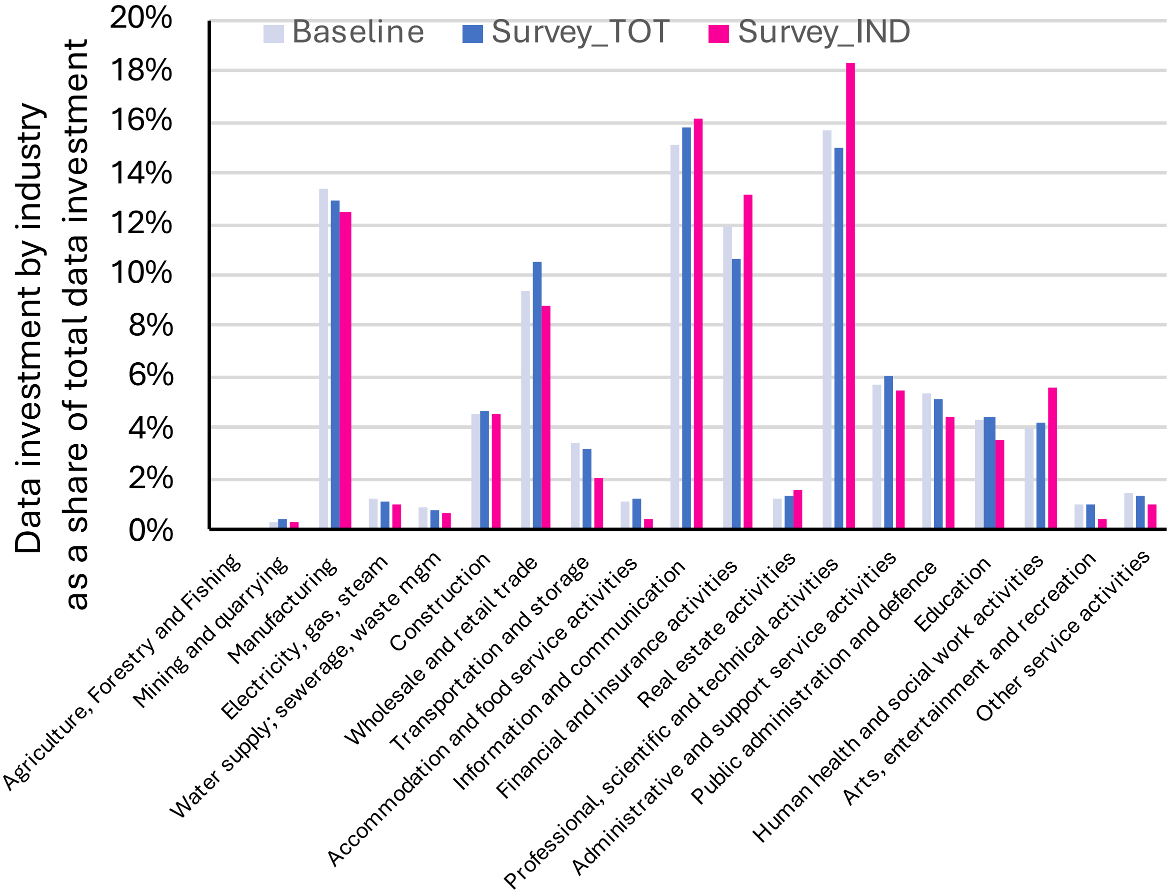
4.4.1 *What are the industries that contribute more to UK investment in data assets?*

Figure 7 shows that five industries are the main drivers of total investment in data assets. In 2019, manufacturing (SIC section C), trade (G), communications (J), financial services (K), and professional business services (M) accounted for 69% of total investment.

Other relevant contributors to total data investment are construction (F), administrative and support services (N), public administration (O), education (P), and health services (Q), which account for between four and six percent of total data investment each.

Using the total economy time factors confirms the prominent role of these five industries (65 percent of total investment), with only a minor reduction of the contribution from financial services (K) and professional business services (M).

Figure 7 - Data investment by industry in the UK, industry share in total data investment, 2019



Source: Authors' elaborations on the UK Annual Survey of Hours and Employment and EUKLEMS & INTANProd.

4.4.2 *In What Data Asset Types Does the UK Invest More?*

Figure 8 shows that when using industry specific time factors, data stores, databases, and data intelligence account for about the same share of total data investment in 2019.

Using total economy time factors from the survey increases the share of data stores from 34 to 38 percent, which is consistent with what we get using the baseline time percentages.

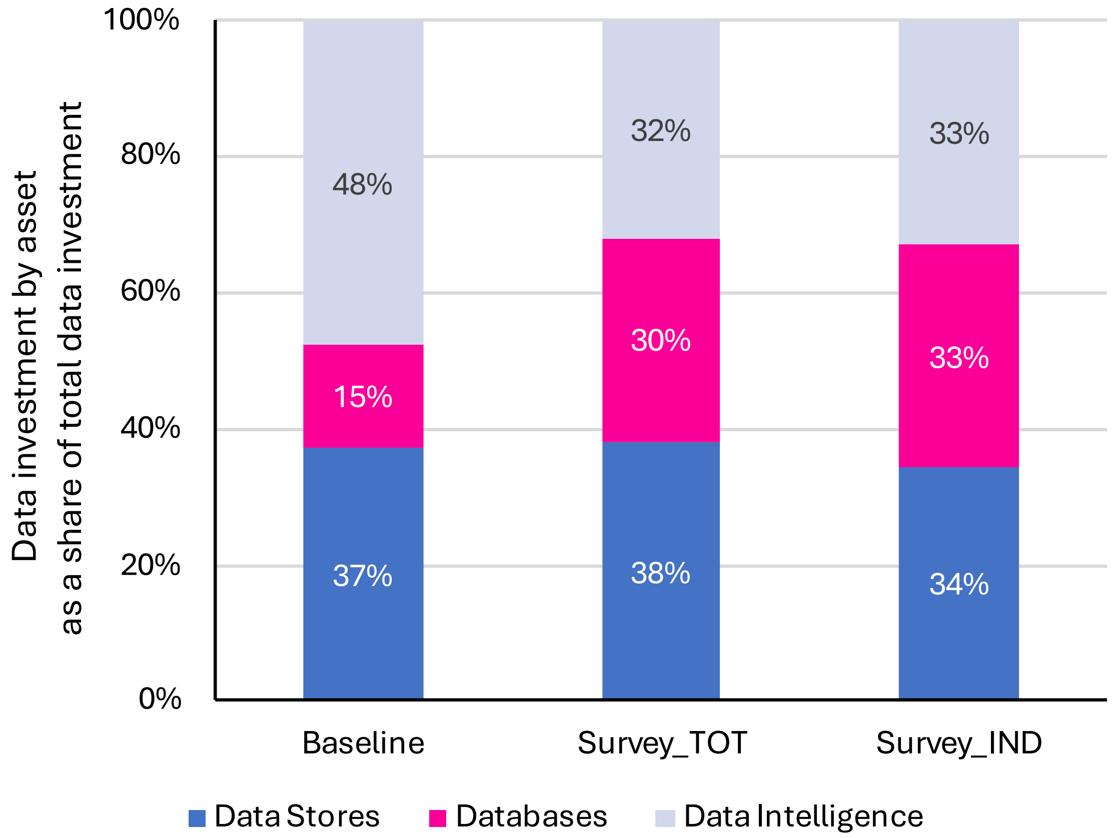
The major difference between the baseline results and the two versions derived from the survey-based time factors is the relative importance of databases and data intelligence. In the baseline results, data intelligence accounted for 48 percent of data investment, while databases only 15 percent.

The small share of databases in the baseline estimates is consistent with how this asset is defined in national accounts. The *2008 SNA* recommends measuring the value of databases produced for own final use by the sum-of-costs approach only including the cost of preparing data in the appropriate format but not the cost of acquiring or producing the data (which, in our framework, is captured by data stores). In addition, preparing a database is an entirely different activity than using the database itself to generate data intelligence. This implies that the total value of databases should be relatively lower than that of data stores (as the major cost is likely related to obtaining the data and not preparing them in the appropriate format to be analysed and data intelligence).

The survey questionnaire defined databases consistently with the *2008 SNA*.[28] Still, the relatively large investment value in databases derived from the survey-based time factors might signal that respondents might not have fully understood the questionnaire, especially regarding the differences in the three data assets. On the other hand, it might also be that the distinction between the creation of raw data and databases is easier to define in theory than practice and that empirical measurement should focus on raw data and databases as only one asset.

[28] Questionnaire description and survey results are illustrated in Ipsos (2023).

Figure 8 - Data investment by asset in the UK, asset share in total data investment, 2019



Source: Authors' elaborations on the UK Annual Survey of Hours and Employment and EUKLEMS & INTANProd.

5 Theoretical framework: Intangibles, GDP, and Innovation

The aggregate effects of the rise of data capital are analysed using the upstream/downstream two-sector model summarized in Corrado and others (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 big data 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 big data/AI versus the "appropriability effect" that constrains Total Factor Productivity (TFP) growth—operate in a two-sector model with data/intangible capital.

5.1 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, for example, 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.

Appendix A 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.

To the extent that 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 appendix A), 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 (for example, Awano and others, 2010).

5.2 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:

$$P^Q Q = P^Y Y + P^N N = P^C C + P^I I + P^N N \quad (6)$$

$$= P^L L + P^K K + P^R R \quad (7)$$

where $P^Q Q$ is the value of GDP; $P^Y Y$ is the downstream output reflecting the production of (tangible) investment and consumer goods; $P^N N$ is investment in data assets; $P^C C$ is consumption; $P^I I$ is investment; $P^L L$ is the investment in labour; $P^K K$ is tangible capital investment; $P^R R$ is the value of intangible capital, defined as its replacement cost. As seen in (6) to the right, investment in final demand is expanded to include data value creation and thus GDP is larger. Factor income, the second line (7), 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:

$$dq = \sigma_Q^X dx + \sigma_Q^R dr + da \quad (8)$$

where σ_Q^X is the combined factor income share for conventional inputs relating to labour, L, and capital, K, in total production and σ_Q^R is the factor income share attributed to owners of data/intangible capital.[29] 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 (including data capital) $\sigma_Q^R dr$, 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**.

5.3 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, the taxi company Lyft was able to duplicate and compete against Uber's

[29] The decomposition is obtained via the usual log differentiation of (7) 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.

innovative, data-enabled ride-sharing business model.[30] 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 (for instance, Amazon's very efficient delivery system, Google's targeted advertising systems), 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, that is, unless 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.[31]

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

$$da = s_Q^Y da^Y + s_Q^N da^N \qquad \textit{(9)}$$

That is, 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 (9). Estimated production shares in for a sample of EU economies in Corrado and others (2023) 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.

5.4 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 (9), 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

---

[30] Lyft is taxi company (https://www.lyft.com/) adopting similar business model as Uber (https://www.uber.com/it/en/).

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

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 decline in relative prices reflects 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.[32]

The main idea is that changes in prices of intangible assets may partly capture the relative productivity of data capital (Corrado and others, 2023).

5.5 Relative efficiency of data capital: Some evidence

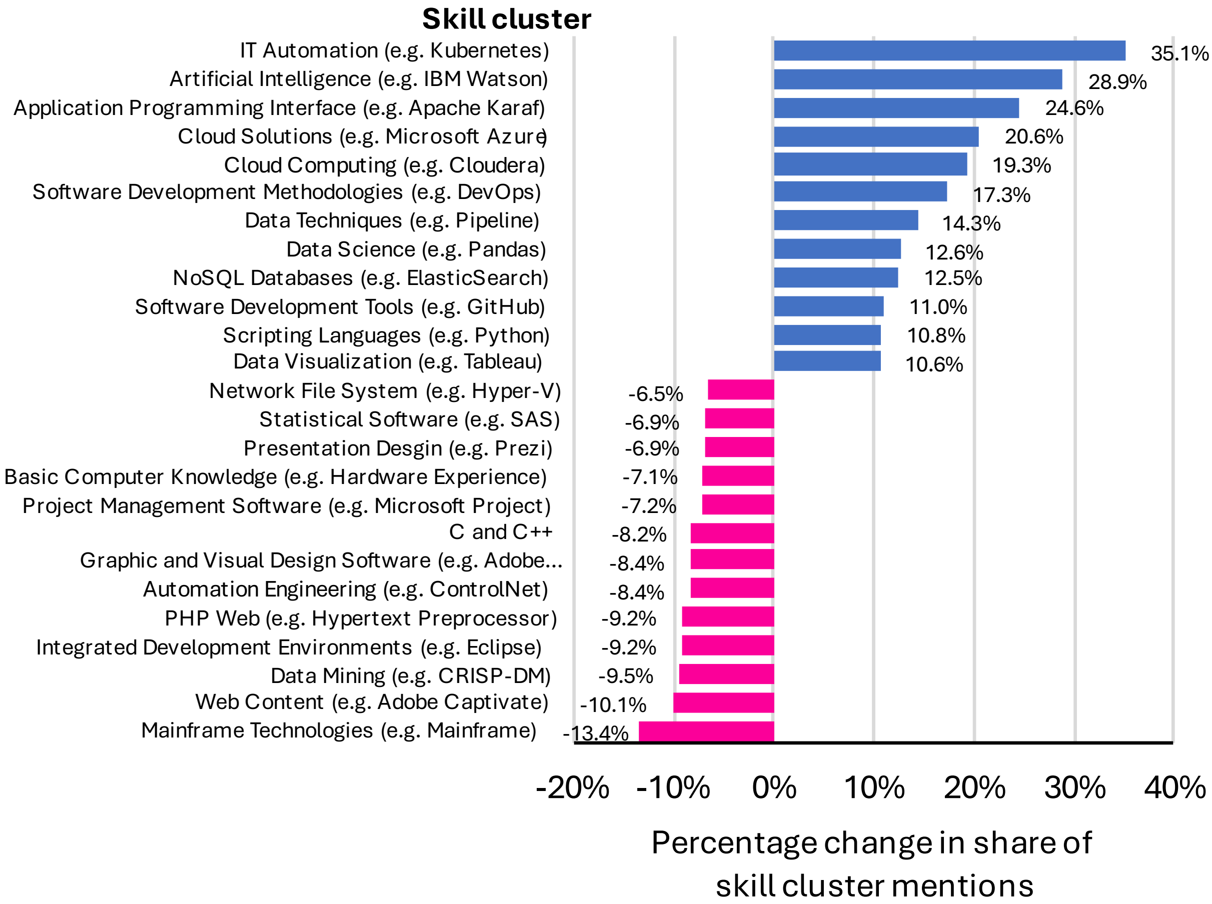
Evidence that the impact of data technologies on intangible capital asset price change might be rather powerful include: (a) strong relative demand for skills used in the production of data capital, (b) stunning growth in the availability of open-source software based on data technologies, and (c) direct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

5.5.1 AI/Cloud systems skill demand

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 9, 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 suggests that workforce composition changes associated with increased data use may have significantly offset wage pressures on asset prices for data capital.

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

Figure 9 - Emerging skill clusters including Artificial Intelligence and Cloud Solutions relative to other tech occupations



Source: The Conference Board (2021) based on The Conference Board®-Burning Glass® Help Wanted OnLine® (HWOL) data series. Notes: Percentage 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

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

5.5.2 *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 and others (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.[33] Their estimates of person-months of effort in value creation for the nine European countries for which data capital production estimates were reported in Corrado and others (2023) 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, for example, the software that runs industrial Internet-of-Things (IoT) installations and advanced robots.

5.5.3 Cloud and other efficiencies

Many studies document improvements in the efficiency of modern cloud systems to ingest, store, process and analyse large quantities of data (for example, 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.

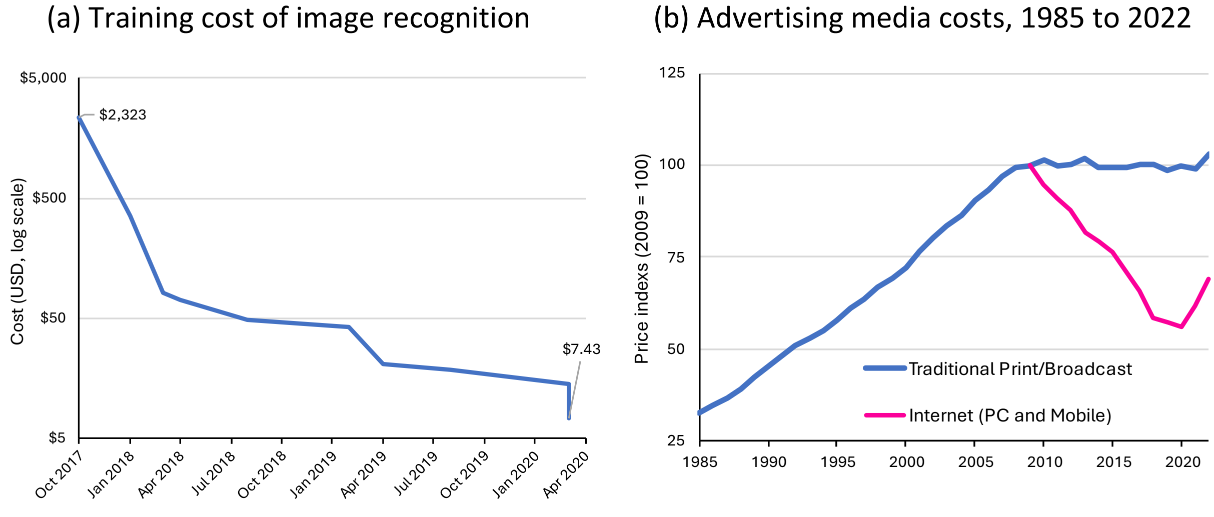
Indicators of these cost efficiencies are shown in figure 10. According to tests shown in the AI Index Report (Zhang and others, 2021, page 49), the costs of training a contemporary image recognition system was "a few dollars in 2020, down by around 150 times from costs in 2017" (figure 10, left panel). 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 10, 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).[34]

---

[33] 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.

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

Figure 10 - Data-driven Cost Efficiencies



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

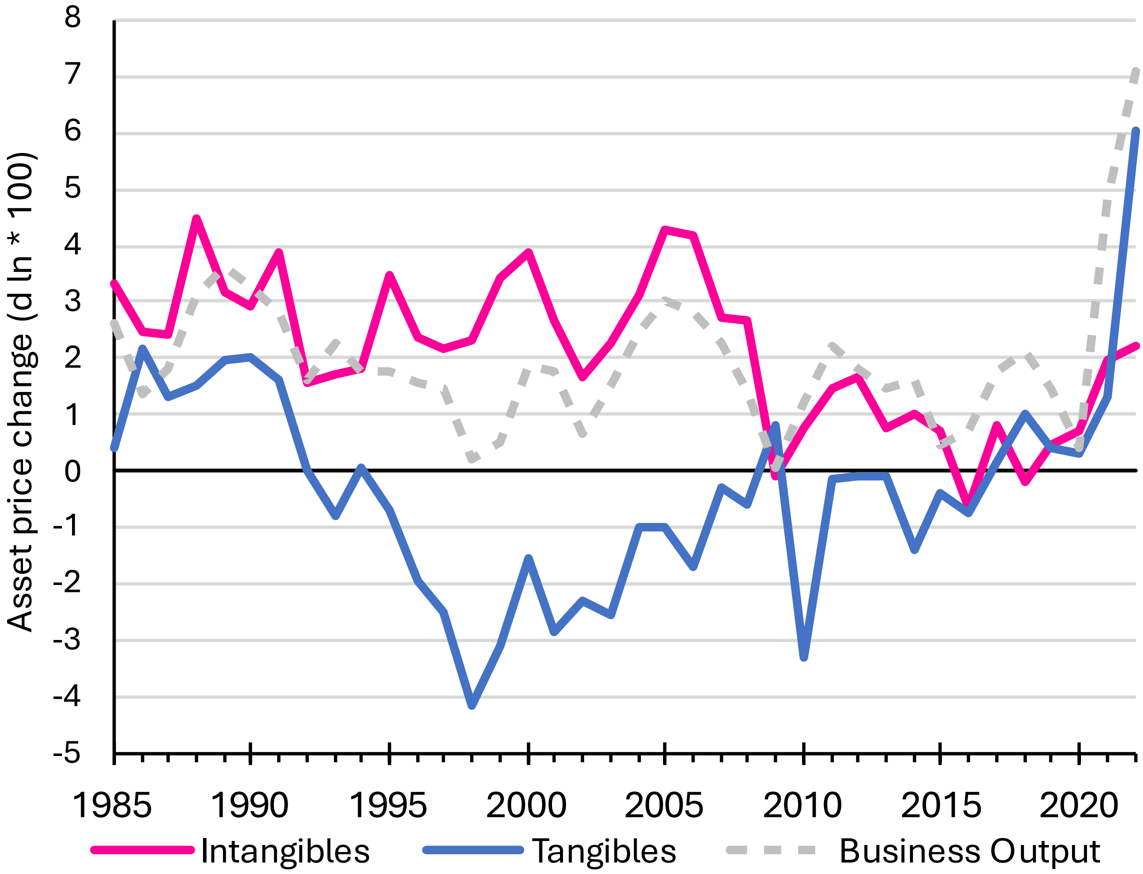
Hard-to-measure services price research typically does not address intangible asset-producing activities—R&D labs, marketing teams, engineering design projects—nor do assessments of productivity mismeasurement view these activities as hotbeds of rapid quality change missed by price collectors. That said, 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.

5.5.4 *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 10 (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 11 and table 4.

Figure 11 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 4, lines 5 and 6).

Figure 11 - U.S. Intangible and Tangible Asset Price Change, 1985 to 2022



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

Table 4 - U.S. Intangible and Tangible Asset Price Change, selected periods

Asset group	1995 to 2009 (1)	2009 to 2019 (2)	2019 to 2022 (3)
1. Intangibles	2.7	0.7	1.6
2. Tangibles	-1.7	-0.4	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	-0.8	1.0
6. Organization process investment	3.0	-0.7	0.8
Relative price change (asset price/business output price):			
7. Intangibles, total	1.2	-0.8	-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

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). Note: Private non-residential assets. Natural log changes, annualized.

From 2009 to 2019, the relative price of total intangible assets fell 0.8 percent per year in the United States (table 4, line 7) and the relative price of the data-intensive components, investments in branding, marketing and organizational process change (lines 9 and 10), fell 2.25 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 (for example, see the 2023 AI Index Report, Maslej and others, 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, that is, data capital, of about 2 to 4 percent per year.

5.6 Relative asset prices and "potential" growth in labour 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 (that is, 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 5 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, that is, 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 labour productivity by making assumptions about data capital's relative productivity and income share, assuming labour'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 (Appendix A, equations A1-1 and A1-2) exhibit constant returns and differ only by their "A" terms and that there is faster TFP growth in the data capital-producing sector. 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 labour productivity in this solution is the sum of a "use" or "investment" effect plus a "production" effect that may be expressed as follows:

$$\underbrace{\bar{\sigma}^R/_{\bar{\sigma}^L}\ (productivity\ advantage)}_{investment\ or\ use\ effect} + \underbrace{\bar{\omega}^D(productivity\ advantage)}_{production\ effect} \qquad (10)$$

The “overbar” notation in (10) denotes steady-state solution values. Thus, $\bar{\sigma}^R$ and $\bar{\sigma}^L$ represent steady state income shares of data capital and labour, 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, that is, $da^N > da^Y$, 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,

$$productivity\ advantage = -(d\ln p^N - d\ln p^Y) \qquad (11)$$

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

Table 5 presents alternative scenarios for potential labour productivity growth using equations (10) and (11). 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 above.

The capital input shares of data capital are assumed to range from 5 to 10 percent, that is, a bit above the approximate band about the estimates for the penetration of data capital in intangible capital as discussed above. Production shares are assumed to be the capital income share plus or minus 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, see Corrado and others, 2023).

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 2. They are set at 2 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, for example, 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 labour's share of total income $\bar{\sigma}^L$ equals 0.7.

Table 5 - Productivity Scenarios: Contribution of data capital to potential labour productivity growth (percentage points)

	Productivity advantage (Relative asset price growth differential)	
Income and production share	Narrow edge 2 percentage points	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 next 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 labour productivity growth range by more than a factor of 5—from 0.23 percentage points per year to 1.26 percentage points per year. The range highlights the synergies among data capital efficiency and an economy's capability for digital transformation of its production processes.

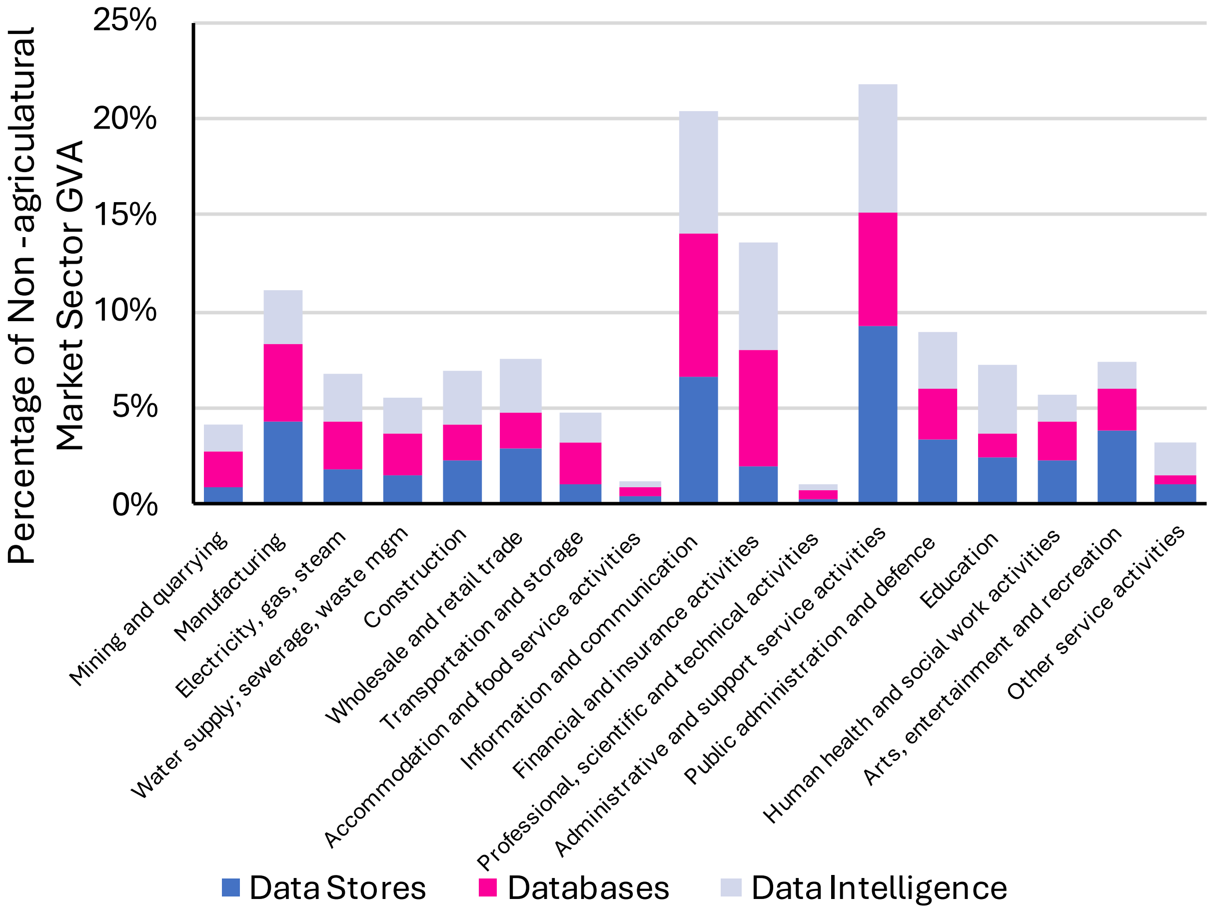
Having established that data capital has considerable potential for impacting labour productivity growth, let us now empirically address how data capital affects measured total factor productivity da.

6 Empirical analysis

In this section we illustrate the estimates of data assets obtained implementing the approach described in section 4 to UK data and then the main findings from the econometric analysis of the impact of data capital on productivity growth. The econometric model aims at estimating the productivity impact of data capital assuming that it can be treated as a capital input in a production function as tangible and intangible capital.

Figures 12 and 13 show the estimates of data asset in the UK economy providing evidence on the relative size and growth of market sector data asset production. Data asset production is shown according to the three segments in the data value chain revealing that the total data value chain averages 7.6 percent relative to adjusted gross value added (GVA) in the UK industries over the sample years.[35] Professional, scientific, and technical activities and Information and communication are the most data intensive sectors in the UK ranging between 22 to 20 percent of GVA while Agriculture, Accommodation and food service activities and Real estate are the least (1.0 – 1.2 percent relative to GVA).

Figure 12 - Data value chain in the UK industries (average GVA shares 2010-2020)



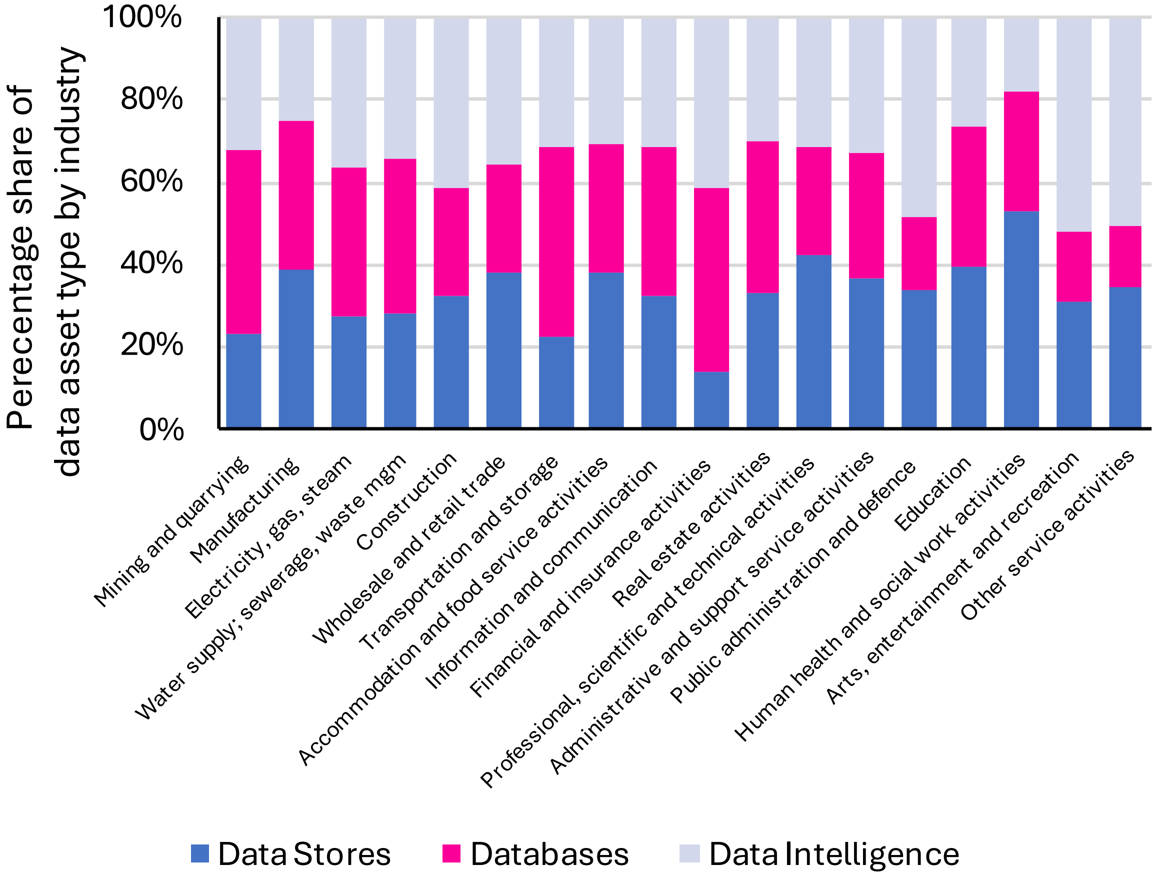
The largest component among data assets is data intelligence (Figure 13) accounting for 35 percent of data asset followed by data stores (33 percent) and databases (32 percent). Data intelligence accounts for around 40 percent of total investment in data assets in Financial and insurance activities, public administration, Arts, entertainment and recreation and Other services. Data storage and databases varies across sectors following a similar pattern besides in Human health and social work activities for which data store accounts for 50 percent of total data asset.

Table 6 shows the shares of data capital investment (column 1) compared with intangible investment (column 2) in some industry groups. The most data intensive industry groups consist

[35] This refers to Gross Value Added adjusted to take into account the capitalization of intangible assets (Corrado and others, 2009).

of professional, scientific, and technical activities; information and communication services; and finance and insurance activities sectors (lines 1-3).

Figure 13 - Components of data assets in the UK industries (average shares 2010-2020)



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, 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. Comparable evidence is illustrated in Corrado and others (2023) for a sample of European Economies over the same years.

Table 6 - Sectoral distribution of investment in data asset and total intangibles, percentages of sector gross value added (average 2010-2019)

Industries	**Data asset (1)**	**Total Intangibles (2)**
Professional, scientific and technical activities	21.9	20.3
Information and communication	20.4	26.0
Financial and insurance activities	13.6	21.6
Manufacturing	11.1	17.9
Administrative and support service activities	9.0	13.0

Wholesale and retail trade	7.5	11.0
Human health and social work activities	7.4	6.0
Public administration and defence; compulsory social security	7.2	14.1
Construction	7.0	10.6
Electricity, gas, steam	6.8	10.7
Other service activities	6.3	12.5
Education	5.7	15.5
Water supply; sewerage, waste management	5.6	7.1
Transportation and storage	4.7	8.9
Mining and quarrying	4.1	9.4
Arts, entertainment and recreation	3.3	13.9
Accommodation and food service activities	1.2	6.7
Total non-agriculture market sectors	**8.4**	**13.2**

6.1 Evaluating the contribution of data assets to UK productivity growth

Our discussion so far 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 assumption has been tested in Corrado and others (2023) who found evidence of substantial overlapping between intangibles and data assets in a sample of nine European economies. To check if this evidence is supported by the UK data, this section provides an econometric analysis of the contribution of data asset and intangibles to labour productivity growth and their possible synergies/overlapping looking at the UK over the years 2010-2020.

The idea of a possible overlap between data asset and total intangibles is supported by simple correlation tests on UK figures indicating that they are positively linked (correlation coefficient is 0.9). Further, when looking specifically at the correlation with data-intensive intangibles it turns out that this is rather strong (correlation coefficient is 0.8).

To further explore their possible synergies and overlapping, first we check the linkages between data asset and productivity growth and then we test a regression model to further explore this relationship within a simple production function framework with controls for country, industry, and time fixed effects.

Figure 14 confirms a positive correlation between capital data assets and industry productivity growth suggesting further exploration is warranted.

Figure 14 - Data asset and industry productivity growth in the UK (2010-2019)

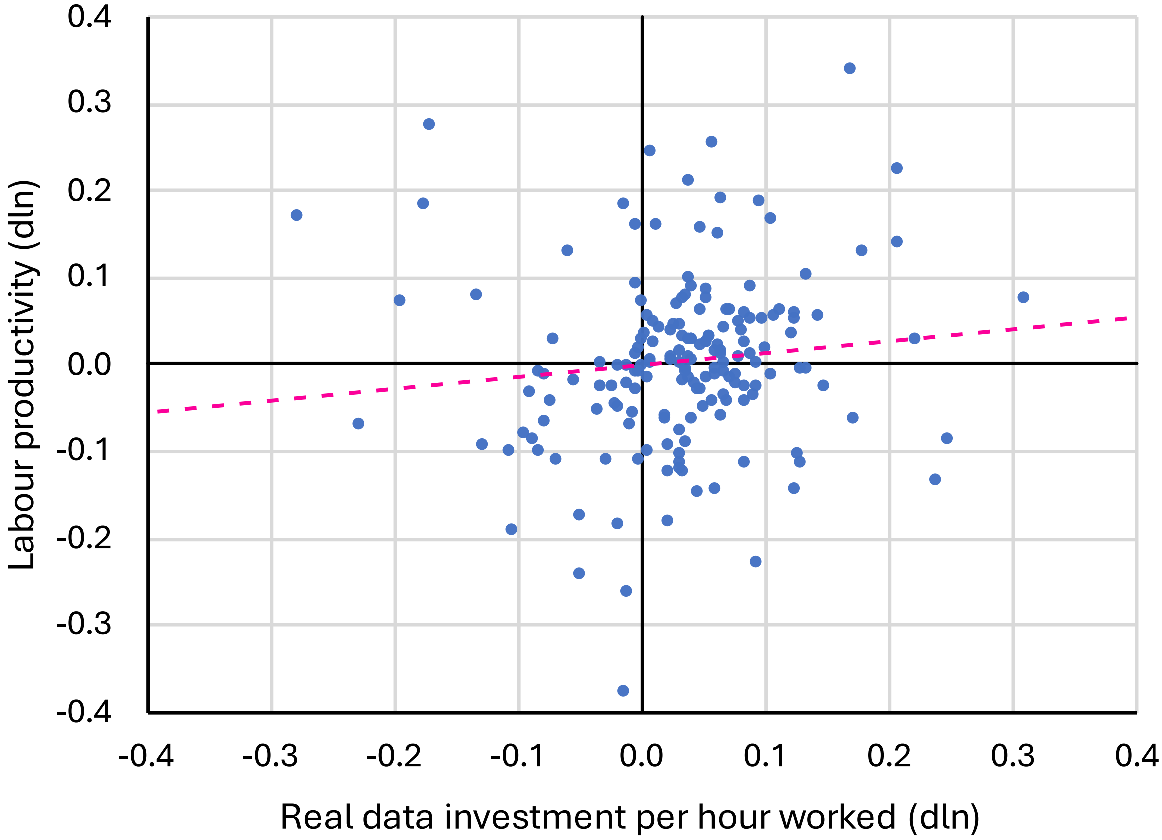


Table 7 reports the results of the econometric analysis. The sample refers to UK industry data from EUKLEMS & INTANProd over the years 1998-2019. Column 1 is our benchmark specification showing that intangible and tangible capital are statistically significantly associated with labour productivity growth. Adding data capital in column 2 reduces the coefficient of intangibles by more than 50 percent and renders both intangible and 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 5 percent significance statistical level, implying that both capitals contribute to explaining labour productivity growth. We than examine the overlap between data and intangible assets in terms of two different components of intangibles distinguishing between data-intensive intangibles and other intangibles, with the latter further split into 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 include data assets in our model, data intensive intangibles lose their statistical significance. This evidence suggests that the data-intensive group of intangibles and data capital capture similar factors affecting productivity growth. We then test the impact of data assets with lags to control for possible endogeneity biases affecting data capital inputs in Column 5. Estimation results support this idea suggesting adopting an Instrumental Variable approach (IV) to better control for endogeneity. Finally, column 6 looks at the individual categories of data capital, revealing that the high statistical significance of the relationship between data capital and labour productivity growth is mainly driven by data intelligence.

Table 7 - Regression model estimates of the contribution of the growth in data and intangible capital deepening to labour productivity growth

	(1)	(2)	(3)	(4)	(5)	(6)
					Data capital (t-1)	
Intangible capital	.266***	.011				
	(.080)	(.096)				
Tangible capital	.325***	-.092	.304***	-.0375	.381***	-.117
	(.085)	(.125)	(.110)	(.121)	(.119)	(.121)
Data capital		.748***		1.34***	.271***	
		(.168)		(.246)	(.098)	
Data-intensive intangibles			.599***	-.248	.461**	-.049
			(.164)	(.220)	(.179)	(.222)
National accounts intangibles			.118**	.095*	.248***	.087*
			(.054)	(.0522)	(.0907)	(.051)
Training			-.049	-.134	-.0878	-.065
			(.107)	(.102)	(.106)	(.103)
Databases						-.617
						(.448)
Data stores						.330
						(.444)
Data intelligence						1.44***
						(.220)
Observations	274	274	274	274	274	274

*Notes: Standard errors are in parentheses; significance levels indicated by *** p<0.01, ** p<0.05, * p<0.1; Dependent variable is labour 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, that is, per hour, delta natural log terms. Estimates are generalized least squares. All columns include time, industry, and country fixed effects.*

Summing up, econometric results suggest that the intangible investments and the “data stack”-inspired estimates of data investment strongly overlap especially 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.

According to the model developed in Corrado and others (2023) and consistently with their findings, the UK estimates support the theoretical assumption that there are likely two possible effects from the use of data capital on productivity: an efficiency and an appropriability effect. The estimated model does not allow us to measure directly these effects, but the empirical findings can be interpreted following the conceptual analysis in section 5. Therefore, the estimated coefficients in table 7 suggest there are first-order impacts on productivity via the use of data capital that boosts labour productivity growth **(the efficiency effect)** but, at the same time, the increased data intensity of intangibles weakens commercial knowledge diffusion and diminishes TFP growth **(the appropriability effect).**

7 Conclusions and policy implications

In this report, we have illustrated how an intangible assets approach can be helpful for addressing the question of how the increased use of data in economies affects productivity Corrado and others (2023). We have explored the role of data capital in the UK economy and their contribution to productivity growth. Our results show that the main driver of investment in data asset is data intelligence accounting for 35 percent of data capital in the UK and that it is mostly concentrated in the services sector. The analysis for the UK corroborates the evidence, shown by Corrado and others (2023) for a sample of European economies, of a strong and positive correlation between data assets and intangibles especially 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 data-intensive intangibles. Similarly, the UK estimates confirm that the first-order impacts of data capital on productivity are that the use of data capital boosts labour productivity growth **(the efficiency effect)**, but that the increased data intensity of intangibles weakens commercial knowledge diffusion and diminishes TFP growth **(the appropriability effect)**. The interaction of these effects might be further investigated to improve our understanding of the drivers of the UK productivity slowdown.

Policy implications

The analysis developed in this report shows that treating data as an asset has consequences on both economic theory and policy as already stressed by Corrado and others (2023). On the policy side, this requires a re-thinking of many different but interrelated areas ranging from competition policy, trade and international coordination policies, as well as monetary policy and pricing rules. Renewed policy schemes should be designed building on increased and fresh cooperation between national agencies such as competition authorities, consumer protection and privacy agencies, statistical institutes, and financial regulators.

Businesses, regulators and policymakers are all affected by the challenges posed by the data transformation. The economic characteristics of data and their abundance in the economy associated with insufficient sharing creates pervasive information asymmetries and inefficiencies. When data is considered an independent asset there are consequences for economic policies as their goals and effectiveness might be affected. Data use might generate large economic benefits to the society but there are relevant policy trade-offs in both public and private sector. This ranges from extensive data availability and use on the one hand and the identification of proper incentives for investments in its creation and provision on the other, or between long term economic and social gain from open data access and short-term economic benefits from private sector exclusive data access.

Thus, in the data economy, the main goal for policymakers becomes finding the balance between potential economic benefits of wide data use and sharing and the concerns arising for data holders in terms of possible loss of commercial advantages and privacy. This entails revising current legislation on intellectual property rights for data, establishing a trustworthy regulatory framework where also the privacy concerns can be overcome and a re-thinking of competition policy rules.

Further, the potential for data capital to contribute to productivity growth is substantial and highly dependent on factors influenced by policy settings, for example, policies that build digital skills and capabilities, as well as the factors affecting the inherent capacity for data to be shared.

Investigating the role of data capital as a source of productivity growth and more specifically of the productivity slowdown in the advanced economies including the UK, suggests alternative scenarios that policymakers need to consider:

- It is possible that productivity growth slowed because the spillover effects of data-intensive intangibles are weaker due to higher replicability costs (that is, off-patent blueprints are free, but data-enabled business models are costly to replicate). It is also possible that the impacts of a more data-intensive economy are yet to come, though when they do, economies should expect a wide range of outcomes depending on the synergies between data capital efficiency and an economy's breadth of use and capability for digital transformation.
- It is also possible that the impacts of a more data-intensive economy are obscured due to mismeasurement. Some important salutary impacts of data capital use and digitization (for example, shift to online sales) are in fact missing in our price and productivity statistics, due in part to shortcomings of the frameworks used to develop them. Quantifying these and other impacts from the increase in the use of data, for example, on aggregate productivity via spillovers and on within-industry performance dispersion, is a priority for future work.

As discussed above, the possibility for data to generate positive productivity spillovers is limited by the tendency to exclusive data access. In this respect, policymakers need to identify areas where defining clear rules for larger data access might favour innovation, competition, and productivity growth. In this respect, competition policy can be used to distribute the value of data-driven markets to a larger number of providers.

At the very least policy makers need timely assessments of data capital penetration, data capital sharing, and data capital contributions to industry output growth to monitor the balances that policies must strike. The costs of data/intangible capital also are an ongoing concern in the analysis of data-intensive industries for competition policies. The framework set out in this report aims at facilitating these assessments, though for them to be timelier and more effective requires more timely measurements.

Finally, a prompt policy response to the new data driven challenges should be focused on the definition of clear rules favouring data sharing to foster competition and innovation considering the complementarities between data and other intangible assets. The pandemic has demonstrated that an organized and controlled use of data can generate high quality outcomes and strongly increase efficiency of processes producing social benefits. As an example, sharing data among hospitals and health research institutions across the globe has accelerated the discovery of a vaccine for COVID-19 producing a big value for the society.

8 References

Abdirahman, M and others (2020). 'A Comparison of Deflators for Telecommunications Services Output', *Economics and Statistics*, volume 517-518-519, pages 103–122.

Association of Chartered Certified Accountants (2019). *Business Reporting of Intangibles: Realistic proposals – Comments from ACCA*. Accessed on 15 September 2021.

Zhang, D and others (2021). *The AI Index 2021 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford (CA).

Akcigit, U and Ates, ST (2021) 'Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory' *American Economic Journal: Macroeconomics*, volume 12, issue 1, pages 257-98.

Nadim, A and others (2017). *Can potential mismeasurement of the digital economy explain the post-crisis slowdown in GDP and productivity growth?* , OECD Statistics Working Papers, number 2017/09. OECD Publishing, Paris.

Andrews, D and others (2016). *The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy*, OECD Productivity Working Papers, number 5. OECD Publishing, Paris.

Bontadini, F and others (2023). *EUKLEMS & INTANProd: Industry Productivity Accounts with Intangibles*. LUISS Lab of European Economics, Rome.

Brynjolfsson, E, & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies,* first edition. W. W. Norton & Company, New York (NY).

Brynjolfsson, E and others (2019). 'Using Massive Online Choice Experiments to Measure Changes in Well-being', *Proceedings of the National Academy of Sciences*, volume 116, issue 15, pages 7250-7255.

Byrne, D and Corrado, C (2017). *ICT Prices and ICT Services: What do they tell us about Productivity and Technology?*, Finance and Economics Discussion Series (FEDS), number 2017-15. Federal Reserve Board, Washington.

Byrne, D and Corrado, C (2020). 'The Increasing Deflationary Influence of Consumer Digital Access Services', *Economic Letters*, volume 196, November 2020, article 109447.

Byrne, D and Corrado, C (2021). 'Accounting for Innovations in Consumer Digital Services: IT Still Matters', In Corrado, C and others (editors), *Measuring and Accounting for Innovation in the Twenty-First Century*, pages 471- 517.

Carson, R and others (2001). 'Contingent Valuation: Controversies and Evidence', *Environmental & Resource Economics*, volume 19, issue 2, pages 173-210.

Chiarello, F and others (2021). 'Data Science for Engineering Design: State of the Art and Future Directions' *Computers in Industry*, volume 129, article 103447.

Clarkson, KW (1996). 'The Effects of Research and Promotion on Rates of Return.' In Helms, B (editor) *Competitive Strategies in the Pharmaceutical Industry*, pages 238-268. American Enterprise Institute for Public Policy Research, Washington (DC).

Corrado, C (2022). *Intangible Capital: What it is and why it matters.* [Unpublished chapter in forthcoming Brookings-Hutchins Centre volume, available upon request from the author].

Corrado, C and others (2005). 'Measuring Capital and Technology: An Expanded Framework,' in Corrado, C and others (eds.), *Measuring Capital in the New Economy*, chapter 1, pages 11–46. University of Chicago Press, Chicago.

Corrado, C and others (2009) 'Intangible Capital and U.S. Economic Growth', *Review of Income and Wealth*, volume 55, issue 3, pages 661–685.

Corrado, C and others (2011) 'Constructing a Price Deflator for R&D: Calculating the Price of Knowledge Investments as a Residual' (August 30, 2011).

Corrado, C and others (2017) 'Knowledge Spillovers, ICT, and Productivity Growth', *Oxford Bulletin of Economics and Statistics*, volume 79, issue 4, pages 592-618.

Corrado, C and others (2021). *New Evidence on Intangibles, Diffusion and Productivity*, OECD STI Working Paper 2021/10. OECD Publishing, Paris.

Corrado, C and others (2022). 'Intangible Capital and Modern Economies', *Journal of Economic Perspectives,* volume 36, number 3, pages 3-28.

Corrado, C and others (2023). 'Data, Intangible Capital, and Productivity', in Basu, S and others (editors) *Technology, Productivity, and Economic Growth*. University of Chicago Press, Chicago (IL).

Coyle, D and Li, WCY (2021). *The Data Economy: Market Size and Global Trade*, Economic Statistics Centre of Excellence (ESCoE) Discussion Paper number 2021-09. Economic Statistics Centre of Excellence (ESCoE), London.

Coyle, D and Nguyen, D (2018). *Cloud Computing and National Accounting*, Economic Statistics Centre of Excellence (ESCoE) Discussion Paper number 2018-19. Economic Statistics Centre of Excellence (ESCoE), London.

Crouzet, N and Eberly, J (2019). *Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles*. NBER working paper number 25869. National Bureau of Economic Research, Cambridge (MA).

Deloitte (2020). *Data Valuation: Understanding the Value of your Data Assets*. Deloitte, London.

De Ridder, M (2024). 'Market Power and Innovation in the Intangible Economy', *American Economic Review,* volume 114, issue 1, pages 199-251.

Döttling, R and Ratnovski, L (2020). *Monetary Policy and Intangible Investment*, IMF Working Papers, number 2020/160. International Monetary Fund, Washington (DC).

The Economist (2017). *The world's most valuable resource is no longer oil, but data*, *The Economist Magazine* (6 May).

Foster, L and others (2001). 'Aggregate Productivity Growth: Lessons from Microeconomic Evidence', in Hulten, C and others (editors), *New Developments in Productivity,* pages 303–63. University of Chicago Press, Chicago.

Goodridge, P and others (2021), 'We See Data Everywhere Except in the Productivity Statistics', *Review of Income and Wealth*, volume 68, issue 4, pages 862-894.

Granstrand, O and Holgersson, M (2020). 'Innovation ecosystems: A conceptual review and a new definition' *Technovation*, volumes 90–91, article 102098.

Griliches, Z (1992). 'The Search for R&D Spillovers' *Scandinavian Journal of Economics*¸ volume 94 (supplement), pages S29-47.

Haskel, J and Westlake, S (2018). *Capitalism without Capital: The Rise of the Intangible Economy*. Princeton University Press, Princeton (NJ).

Hopenhayn, H and Rogerson, R (1993). 'Job Turnover and Policy Evaluation: A General Equilibrium Analysis', *Journal of Political Economy,* volume 101, issue 5, pages 915–38.

Inklaar, R & Timmer, MP (2008). *GGDC Productivity Level Database: International Comparisons of Output, Inputs and Productivity at the Industry Level*, GGDC Working Papers, number GD-104, Groningen Growth and Development Centre. University of Groningen, Groningen.

Ipsos (2023). 'DSIT Data Use and Productivity Longitudinal Survey – Technical Report'.

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

Jones, CI (2005). 'Growth and Ideas', In Aghion, P & Durlauf, S (editors), *Handbook of Economic Growth*, edition 1, volume 1B, chapter 16, pages 1063-1111.

Jones, CI. and Tonetti, C (2019). *Nonrivalry and the Economics of Data*, NBER working paper number 26260. National Bureau of Economic Research, Cambridge (MA).

Jorgensen, D (1963). *Capital Theory and Investment Behaviour*, *American Economic Review*, volume 53, issue 2, pages 247-259.

Ker, D & Mazzini, E (2020). *Perspectives on the value of data and data flows*, OECD Digital Economy Papers, number 299. OECD Publishing, Paris.

Kornfeld, R (2019). *Measuring Data in the National Accounts*, Presentation to Bureau of Economic Analysis Advisory Committee.

Lev, B. (2019), 'Ending the Accounting-for-Intangibles Status Quo', European Accounting Review, volume 28, issue 4, pages 713-736.

Li, WCY. and Chi, PJ (2021), *Online Platforms' Creative 'Disruption' in Organizational Capital – The Accumulated Information of the Firm*, Moon Economics Institute Discussion Paper Series, Number 1 (2021). Moon Economics Institute, Irvine (CA).

Mandel, M (2019). 'Building the New Manufacturing Stack', Forbes Magazine. Forbes, Jersey City (NJ).

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

McKinsey Global Institute (2016). *The Age of Analytics: Competing in a Data-Driven World*, MGI Research Report. McKinsey, New York (NY).

McFadden, D and Train, K (2017). *Contingent Valuation of Environmental Goods. A Comprehensive Critique*. Elgar Online, Cheltenham.

Miller, HH and others (2013). *Users, Uses, and Value of Landsat Satellite Imagery: Results from the 2012 Survey of Users*, Open-File Report, number 2013-1269. US Geological Survey, Reston (VA).

Murciano-Goroff, R and others (2021). *Hidden Software and Veiled Value Creation: Illustrations From Server Software Usage*, NBER working paper number 28738. National Bureau of Economic Research, Cambridge (MA).

Nelson, RR and Rosenberg, N (1993). 'Technical Innovation and National Systems', in Nelson, RR (editor), *National Innovation Systems: A Comparative Analysis*. Oxford University Press, Oxford.

Nguyen, D and Paczos, M (2020). *Measuring the Economic Value of Data and Cross-border Data flows: A business Perspective*, OECD Digital Economy Papers, number 297. OECD Publishing, Paris.

Open Data Institute (2016), *Research: The economic value of open versus paid data* (20 April).

Oulton, N (2012). 'How To Measure Living Standards and Productivity', *Review of Income and Wealth*, volume 58, issue 3, pages 424-456.

Prince, J and Wallsten, S (2020). 'How Much is Privacy Worth Around the World and Across Platforms?', TPRC48: The 48th Research Conference on Communication, Information and Internet Policy.

PriceWaterhouseCoopers LLP (2019). *Putting a value on data*. PriceWaterhouseCoopers LLP, London.

Reinsdorf, M and Ribarsky, J (2019). *Measuring the Digital Economy in Macroeconomic Statistics: The Role of Data*, paper presented at the ASSA 2020 Annual Meeting (January, San Diego Marriott Marquis & Marina, San Diego).

Reinsdorf, M and Schreyer, P (2019), *Measuring consumer inflation in a digital economy*, OECD Statistics Working Papers, number 2019/01. OECD Publishing, Paris.

Robbins, C and others (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, C (2021a). *How to build a modern data stock from the ground up* (8 July).

Roca, C (2021b). *What is the modern data stack?* (5 October).

Romer, PM (1990). 'Endogenous Technological Change', *Journal of Political Economy*, volume 95, issue 5, pages S71– S102.

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

Statistics Canada (2019b). *The Value of Data in Canada: Experimental Estimates*. Statistics Canada, Ottawa.

Timmer, M and others (2011). 'Productivity and Economic Growth in Europe: A Comparative Industry Perspective', *International Productivity Monitor,* volume 21, pages 3-23. Centre for the Study of Living Standards, Ottawa.

UNECE (2018), *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics*. United Nations Economic Commission for Europe, Geneva.

UK FRC (2019). 'Discussion Paper - Business Reporting of Intangibles: Realistic proposals' (February). United Kingdom Financial Reporting Council, London.

Varian, H (2019). 'Artificial Intelligence, Economics, and Industrial Organization', in Agarwal, A and others (editors), *The Economics of Artificial Intelligence: An Agenda.* Chicago University Press, Chicago (IL).

Zéghal, D and Maaloul, A (2011). 'The accounting treatment of intangibles – A critical review of the literature', *Accounting Forum*, volume 35, issue 4, pages 262– 274.

Appendix A – A Model of an Economy with Intangibles

In a simplified model of the economy with intangibles, production can be divided 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.

Sectoral activity is described and denoted as follows:

- Upstream output reflects the production of new commercial knowledge. This is also intangible investment, which is in volume terms N and in nominal terms is $P^N N$, where P^N is a price index for intangible assets. (We assume no trade in intangibles, which can however be easily relaxed with no major change in model implications.)
- Downstream output reflects the production of (tangible) investment and consumer goods, $P^Y Y$, or $P^I I + P^C C$, in nominal terms. (Note that for simplicity, intermediates are ignored, an assumption that also is easily relaxed.)
- The outstanding stock of commercially valuable knowledge, which reflects the accumulation of the upstream sector output after adjusting for losses due to economic depreciation (that is, ageing), is the stock of intangible capital, R.
- Freely available basic knowledge, scientific or otherwise, is represented by R^{Basic}. It is an input to upstream production, for example, open-source software, which we assume is produced outside the model. (Another assumption that is easily 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 $P^K K$. Labour inputs and their price are L and P^L, respectively.

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, that is, 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 labour and tangible capital.

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

(A1-1) $$N = A^N F^N(L^N, K^N, R^{Basic}); \quad P^N N = P^L L^N + P^K K^N + \pi^N$$

(A1-2) $$Y = A^Y F^Y(L^Y, K^Y, R); \quad P^Y Y = P^L L^Y + P^K K^Y + P^R R$$

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 $P^R = (r + \delta - \dot{p})P^N$. The user cost of tangible capital is similarly linked to it asset price. The model is closed via arbitrage of returns

(r) across sectors, that is, 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 (that is, π^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.

Appendix B – Occupations producing data assets and time-use factor

Table B1 - Relevant occupations producing data assets and time-use factor

		Survey results				Baseline assumptions			
SOC (4 digit)	Occupation title	Data stores	Data intelligence	Data bases	Software	Data stores	Data intelligence	Data bases	Software
1132	Marketing and sales directors	0.10	0.09	0.08	0.04	0.00	0.10	0.00	0.00
1134	Advertising and public relations directors	0.10	0.09	0.08	0.04	0.00	0.10	0.00	0.00
1136	Information technology and telecommunications directors	0.10	0.09	0.08	0.04	0.10	0.10	0.10	0.00
2111	Chemical scientists	0.12	0.08	0.17	0.00	0.10	0.40	0.00	0.00
2112	Biological scientists and biochemists	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2113	Physical scientists	0.12	0.08	0.17	0.00	0.10	0.40	0.00	0.00
2114	Social and humanities scientists	0.12	0.08	0.17	0.00	0.00	0.10	0.00	0.00
2119	Natural and social science professionals n.e.c.	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2121	Civil engineers	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2122	Mechanical engineers	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2123	Electrical engineers	0.12	0.08	0.17	0.00	0.10	0.40	0.00	0.00
2124	Electronics engineers	0.12	0.08	0.17	0.00	0.10	0.45	0.00	0.00
2126	Design and development engineers	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2127	Production and process engineers	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2129	Engineering professionals n.e.c.	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00

SOC (4 digit)	Occupation title	Survey results: Data stores	Survey results: Data intelligence	Survey results: Data bases	Survey results: Software	Baseline assumptions: Data stores	Baseline assumptions: Data intelligence	Baseline assumptions: Data bases	Baseline assumptions: Software
2133	IT specialist managers	0.15	0.11	0.13	0.11	0.20	0.20	0.20	0.20
2134	IT project and programme managers	0.15	0.11	0.13	0.11	0.20	0.20	0.20	0.20
2135	IT business analysts, architects and systems designers	0.15	0.11	0.13	0.11	0.10	0.00	0.10	0.50
2136	Programmers and software development professionals	0.15	0.11	0.13	0.11	0.00	0.00	0.10	0.50
2137	Web design and development professionals	0.15	0.11	0.13	0.11	0.00	0.00	0.10	0.50
2139	Information technology and telecommunications professionals n.e.c.	0.15	0.11	0.13	0.11	0.20	0.20	0.20	0.20
2141	Conservation professionals	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2142	Environment professionals	0.12	0.08	0.17	0.00	0.10	0.25	0.00	0.00
2150	Research and development managers	0.12	0.08	0.17	0.00	0.00	0.10	0.00	0.00
2213	Pharmacists	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2216	Veterinarians	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2217	Medical radiographers	0.03	0.16	0.22	0.01	0.20	0.00	0.00	0.00
2421	Chartered and certified accountants	0.14	0.13	0.12	0.05	0.10	0.25	0.00	0.00
2423	Management consultants and business analysts	0.14	0.13	0.12	0.05	0.20	0.20	0.20	0.00
2424	Business and financial project management professionals	0.14	0.13	0.12	0.05	0.20	0.20	0.20	0.00
2425	Actuaries, economists and statisticians	0.14	0.13	0.12	0.05	0.10	0.25	0.00	0.00

SOC (4 digit)	Occupation title	Survey results				Baseline assumptions			
		Data stores	Data intelligence	Data bases	Software	Data stores	Data intelligence	Data bases	Software
2426	Business and related research professionals	0.14	0.13	0.12	0.05	0.10	0.10	0.10	0.00
2429	Business, research and administrative professionals n.e.c.	0.14	0.13	0.12	0.05	0.10	0.10	0.10	0.00
2431	Architects	0.40	0.10	0.00	0.00	0.40	0.10	0.00	0.00
2433	Quantity surveyors	0.10	0.25	0.00	0.00	0.10	0.25	0.00	0.00
2435	Chartered architectural technologists	0.40	0.10	0.00	0.00	0.40	0.10	0.00	0.00
2452	Archivists and curators	0.08	0.02	0.01	0.01	0.20	0.00	0.00	0.00
2461	Quality control and planning engineers	0.08	0.02	0.01	0.01	0.10	0.25	0.00	0.00
2462	Quality assurance and regulatory professionals	0.08	0.02	0.01	0.01	0.20	0.20	0.20	0.00
2463	Environmental health professionals	0.08	0.02	0.01	0.01	0.20	0.20	0.00	0.00
2473	Advertising accounts managers and creative directors	0.08	0.02	0.01	0.01	0.20	0.30	0.00	0.00
3111	Laboratory technicians	0.11	0.08	0.10	0.09	0.20	0.10	0.00	0.00
3112	Electrical and electronics technicians	0.11	0.08	0.10	0.09	0.10	0.25	0.00	0.00
3113	Engineering technicians	0.11	0.08	0.10	0.09	0.20	0.10	0.00	0.00
3114	Building and civil engineering technicians	0.11	0.08	0.10	0.09	0.20	0.10	0.00	0.00
3115	Quality assurance technicians	0.11	0.08	0.10	0.09	0.20	0.00	0.00	0.00
3116	Planning, process and production technicians	0.11	0.08	0.10	0.09	0.20	0.10	0.00	0.00
3119	Science, engineering and production technicians n.e.c.	0.11	0.08	0.10	0.09	0.20	0.10	0.00	0.00
3121	Architectural and town planning technicians	0.11	0.08	0.10	0.09	0.10	0.00	0.00	0.00

SOC (4 digit)	Occupation title	Survey results: Data stores	Survey results: Data intelligence	Survey results: Data bases	Survey results: Software	Baseline assumptions: Data stores	Baseline assumptions: Data intelligence	Baseline assumptions: Data bases	Baseline assumptions: Software
3122	Draughtspersons	0.11	0.08	0.10	0.09	0.10	0.00	0.00	0.00
3422	Product, clothing and related designers	0.10	0.10	0.00	0.00	0.10	0.10	0.00	0.00
3531	Estimators, valuers and assessors	0.12	0.11	0.07	0.03	0.10	0.25	0.00	0.00
3532	Brokers	0.12	0.11	0.07	0.03	0.10	0.10	0.00	0.00
3534	Finance and investment analysts and advisers	0.12	0.11	0.07	0.03	0.10	0.25	0.00	0.00
3535	Taxation experts	0.12	0.11	0.07	0.03	0.10	0.25	0.00	0.00
3536	Importers and exporters	0.12	0.11	0.07	0.03	0.10	0.10	0.00	0.00
3537	Financial and accounting technicians	0.12	0.11	0.07	0.03	0.10	0.25	0.00	0.00
3538	Financial accounts managers	0.12	0.11	0.07	0.03	0.10	0.25	0.00	0.00
3539	Business and related associate professionals n.e.c.	0.12	0.11	0.07	0.03	0.10	0.40	0.00	0.00
3542	Business sales executives	0.12	0.11	0.07	0.03	0.10	0.10	0.00	0.00
3543	Marketing associate professionals	0.12	0.11	0.07	0.03	0.20	0.30	0.00	0.00
3545	Sales accounts and business development managers	0.12	0.11	0.07	0.03	0.10	0.10	0.00	0.00
3550	Conservation and environmental associate professionals	0.12	0.11	0.07	0.03	0.10	0.00	0.00	0.00
3562	Human resources and industrial relations officers	0.12	0.11	0.07	0.03	0.00	0.10	0.00	0.00
3563	Vocational and industrial trainers and instructors	0.12	0.11	0.07	0.03	0.00	0.10	0.00	0.00
3564	Careers advisers and vocational guidance specialists	0.12	0.11	0.07	0.03	0.00	0.10	0.00	0.00

SOC (4 digit)	Occupation title	Survey results				Baseline assumptions			
		Data stores	Data intelligence	Data bases	Software	Data stores	Data intelligence	Data bases	Software
3565	Inspectors of standards and regulations	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3567	Health and safety officers	0.12	0.11	0.07	0.03	0.10	0.00	0.00	0.00
4113	Local government administrative occupations	0.10	0.11	0.07	0.04	0.05	0.00	0.00	0.00
4114	Officers of non-governmental organisations	0.10	0.11	0.07	0.04	0.05	0.00	0.00	0.00
4121	Credit controllers	0.10	0.11	0.07	0.04	0.10	0.40	0.00	0.00
4122	Book-keepers, payroll managers and wages clerks	0.10	0.11	0.07	0.04	0.20	0.00	0.00	0.00
4123	Bank and post office clerks	0.10	0.11	0.07	0.04	0.10	0.00	0.00	0.00
4124	Finance officers	0.10	0.11	0.07	0.04	0.20	0.00	0.00	0.00
4131	Records clerks and assistants	0.10	0.11	0.07	0.04	0.05	0.00	0.00	0.00
4132	Pensions and insurance clerks and assistants	0.10	0.11	0.07	0.04	0.10	0.00	0.00	0.00
4134	Transport and distribution clerks and assistants	0.10	0.11	0.07	0.04	0.05	0.00	0.00	0.00
4135	Library clerks and assistants	0.10	0.11	0.07	0.04	0.10	0.00	0.00	0.00
4138	Human resources administrative occupations	0.10	0.11	0.07	0.04	0.10	0.00	0.00	0.00
4217	Typists and related keyboard occupations	0.10	0.11	0.07	0.04	0.80	0.00	0.00	0.00
7211	Call and contact centre occupations	0.10	0.11	0.07	0.04	0.05	0.00	0.00	0.00
7215	Market research interviewers	0.10	0.11	0.07	0.04	0.05	0.00	0.00	0.00