UNIVERSITY OF CAMBRIDGE

Bennett School of Public Policy Cambridge

White paper

A shared agenda for the economics of data

Diane Coyle, Luca Gamberi, David Nguyen, Julia Schmidt

anmut

Suggested citation

Coyle, D., Gamberi, L., Nguyen, D., Schmidt, J. (2026). *A shared agenda for the economics of data*. Bennett School of Public Policy, Cambridge. DOI: https://doi.org/10.17863/CAM.136926

Contents

Preface 1

1. Why the economics of data matters now 2

2. Data, firms and productivity 4

3. Data access, public value and market structure 9

4. Data, AI and the changing production system 12

5. Measuring data in firms and the macroeconomy 15

6. An agenda on the economics of data 18

References 20

Preface

The purpose of this white paper is to contribute to the development of a shared agenda on the economics of data. Interest in data has increased in recent years, accelerated by the arrival of generative and now agentic AI. It is clear that data (in the form of digital records) in a fundamental sense fuels the creation and distribution of economic value in today's economy. Yet there remain significant gaps in understanding the economics of data, from the perspectives of businesses, public organisations, policymakers and official statisticians. This paper builds on a workshop held in Cambridge on 16 September 2026, with 40 senior participants from across multiple sectors and perspectives. The agenda that has emerged from the discussions covered the fundamental economics of data, business decision-making, market structure and public value, as well as measurement questions. This paper ends with a proposed agenda for research and policy, but this is a starting point. As it makes clear, many open questions remain.

We thank all the participants in the workshop. We are very grateful to Anmut for its support and organisation. We thank the Sloan Foundation for its support for the project on data policy at the Stanford Digital Economy Lab and Bennett School of Public Policy.

Diane Coyle, Bennett School of Public Policy, University of Cambridge

Luca Gamberi, Anmut

David Nguyen, Stanford Digital Economy Lab

Julia Schmidt, Université Paris Dauphine-PSL

1. Why the economics of data matters now

Advances in artificial intelligence (AI) are rapidly changing our economy. Data fuels this change: it provides the information used to train AI models, which are now shaping how firms and public services operate. Much of this data has accumulated within organisations over years and could previously be put to limited use because it was too costly to process. AI is now making that data easier to analyse and use in decision-making, while the data, in turn, expands what AI systems are able to produce.

The economic consequences of this rapid technological change extend well beyond AI adoption. Decisions about who produces and controls data, who can access it and on what terms shape the opportunities available to firms, public institutions and individuals. The challenge is to balance incentives for private investment in generating data with the wider societal and economic benefits that can arise from its wider use. Private data holders may have strong incentives to restrict access even where sharing would generate greater benefits across the economy. Public actors may, in turn, favour more open access, even where this risks strengthening already dominant players. Overcoming these challenges requires evidence-based policy frameworks that can strike this balance.

Designing appropriate data policy frameworks first requires a consistent measure of data investment. Unlike investment in physical assets, much investment in data is embedded in organisational activities that generate information as a by-product of production. Data is rarely identified separately in business accounts and often overlaps with investment in software, research and development, and other intangible assets. Even when the cost of producing data can be measured, this does not tell us what the data is worth. Its value depends on how it is used and the benefits that use generates. This makes it difficult to establish how much economies invest in data, what returns these investments generate, and how much data contributes to productivity and economic growth.

There is a unique window of opportunity for action now. In 2025, data was recognised as a produced asset in the System of National Accounts (SNA) [1]. Work is now underway to translate the new framework into economic statistics. In the UK, the Office for National Statistics is developing estimates of data investment, with initial experimental results expected in December 2026 and incorporation into the national accounts planned for 2030 [2]. Experimental estimates by the US Bureau of Economic Analysis suggest that capitalising internally produced data and databases would increase measured average annual GDP growth by approximately 0.07 percentage points over 2002–2024 [3]. Related growth-accounting estimates indicate that the contribution of IT-related capital to GDP growth would increase by about one-third [4].

Recognising data as an economic asset is an important step, but it is only a starting point. Understanding how data shapes productivity, competition and the opportunities created by AI requires a joint effort by researchers, policymakers and business leaders. Bringing together

these diverse perspectives and new evidence, this paper sets out a shared agenda for research and action, based on a common recognition that data is a foundational input to the digital economy. The aim is not simply to measure the value of data, but to realise more of it and to do so fairly.

Data as a distinctive asset

Data has several characteristics that distinguish it from conventional assets. These characteristics affect how data creates value, the incentives to invest in and share it, and the challenges involved in measuring its economic contribution.

Data and information. Data consists of recorded observations or representations, typically held in digital form. Its economic value arises through the information and insights they make possible, including better decisions, coordination and innovation. One important mechanism is the reduction of uncertainty, but the usefulness of data depends on what can be learned from it and how that knowledge is applied [5, 6].

Content, quality and context. The value of a dataset depends on its relevance to a particular use, its quality and timeliness, and whether it can be combined with other information. The same records may be valuable to one organisation but of limited use to another. Skills, software, computing resources and organisational capabilities are therefore important complements to data. The benefits of an activity combining these inputs cannot generally be attributed to data alone [7, 8].

Non-rivalry, access and externalities. Data is non-rival: its use by one organisation does not, in itself, prevent others from using the same records. Access can nevertheless be restricted through legal, contractual or technical arrangements. This creates a tension between private incentives to control data and the potential wider benefits from sharing and reuse. Data may also generate positive or negative externalities, including benefits from new applications and costs associated with privacy or security [8, 9].

Value and economic life. The costs of producing data need not correspond to the economic benefits it generates. Private value depends on the returns to a particular holder or user, while social value also includes effects on others. Data may have option value because new applications can emerge after collection. Some records become outdated quickly, while others remain useful for many years or gain new applications. Their economic life consequently depends on content, quality and use, complicating assumptions about depreciation [10, 11].

Working dimensions for describing data. No single classification captures all economically relevant differences. The following dimensions can be combined according to the question being examined.

Dimension	Distinction
Stock and flow	Existing data holdings versus records newly generated or acquired
Control and access	Public, private or shared control; exclusive, restricted or open access
Content and sensitivity	Personal or non-personal information; differences in subject, granularity and sensitivity
Processing and quality	Raw, cleaned, linked or analysed records; accuracy, timeliness and interoperability
Scope and use	General-purpose or specialised datasets; current and potential future applications

These distinctions provide a common vocabulary for examining data investment and firm behaviour, access and public value, AI, and economic measurement in the sections that follow.

2. Data, firms and productivity

Firms allocate resources and respond to market conditions using knowledge that shapes expectations about the future. The decisions that follow span a wide range of activities, from adjusting prices in response to customer demand to intervening on equipment at risk of failure. The predictions of future outcomes draw on accumulated experience, on the systematic analysis of digitalised observations (data), or a combination of the two. Managers who can act directly on the process then assess the implications of those predictions, weighing the expected benefits of adapting the process against the associated costs and risks, in light of their objectives [12]. This section sets out the conditions under which acquiring and analysing data is more advantageous for a firm than relying primarily on uncodified experience, and examines the evidence supporting that advantage.

Experience helps decision makers recognise situations they have encountered and draw parallels to interpret the broader context. Its predictive capacity depends on the extent to which past experience provides useful information about current conditions, and on the quality of feedback from previous decisions. Learning these relationships can be harder in complex systems, where outcomes depend on multiple factors, especially when interacting non-linearly. These features can make relevant patterns difficult to identify through experience alone without a systematic analysis [13]. Statistical analysis of operational records, transactions and

other observations can help estimate how outcomes and their probability of occurring vary across several factors [6, 12], giving management a more specific basis for deciding where and how to intervene.

2.1 Extracting value from data assets

Consider a firm that inspects equipment on a fixed rotation. Condition records may reveal which units are deteriorating, while production records show where failure would cause the largest losses. Combining these with environmental observations can identify patterns in failure and the conditions associated with them. The firm can then prioritise inspections and maintenance where the expected reduction in failures is greatest relative to the cost of intervention. The maintenance schedule is worth adapting in response to the acquired information if the gains over existing practice also cover the costs of processing the information and the adjustment cost of implementing the new schedule.

Organisational and process complexity, together with outcome uncertainty, provide a basis for guiding these investment decisions. Complexity arises from the number and diversity of activities and their interdependence [14]. Uncertainty concerns current and future conditions, the relationships determining outcomes and the range of outcomes that may follow. Data is particularly useful when existing routines provide insufficient information and better evidence reveals differences that change an allocation or intervention.

Data can help target inspections and maintenance where failures would be most costly [15], or allocate inventory where demand is expected to exceed available stock. Manufacturing evidence connects records of defects and downtime with routines for identifying causes and taking corrective action. For example, a randomised management intervention in Indian textile plants raised productivity by 17 per cent in its first year through improvements in quality and efficiency [16]. Retail experiments similarly show that using transaction records to identify suspected shelf shortages helped retailers target audits, improving stock availability and increasing sales; programme running costs were a small fraction of benefits after a transitional period [17]. A controlled inventory-allocation pilot combining demand forecasts with stock constraints estimated sales gains of 3–4 per cent [18].

Customer targeting and monitoring depend on recognising differences in preferences and responses to product offerings that aggregate measures of behaviour may conceal. A randomised dashboard rollout increased online sellers' revenues by 3.6 per cent on average, partly through more active performance monitoring [19]. Withholding personal data from a recommendation algorithm reduced engagement and transactions [20]. For retention spending, analysis of two randomised campaigns found that targeting predicted responsiveness to an offer could improve retention relative to targeting customers most likely to leave [21].

Firms can also generate information to learn how their choices affect outcomes, using observations from one decision to improve the next and so on [22]. In an agricultural field experiment, experienced producers changed practices after receiving summaries showing how a

previously overlooked input characteristic affected output [23]. Another example of the use of data in a learning feedback loop comes from dynamic pricing. Where demand depends on customer behaviour and competitors across thousands of products, random price variation can help estimate those responses. A pricing method using such estimates increased revenue by over 15 per cent and profit by over 10 per cent in a controlled field experiment [24].

Evidence of this kind in general identifies three settings in which an advantage is plausible: allocating scarce resources across differing conditions, tailoring commercial decisions to customer and market differences, and learning the effects of choices before repeating or expanding them. These studies provide evidence that better information can improve specific decisions. However, they generally evaluate changes in information, technology or management practices together, rather than identifying the economic return to data assets separately. To date, experiments provide limited direct tests of data-driven against experience-based predictions, or of how gains vary with complexity and uncertainty, although collectively they provide indicative evidence of this pattern. The findings therefore help identify decisions worth investigating, but whether investment pays depends on both the improvement achieved and the full cost of changing those decisions.

2.2 Firm productivity

Productivity measures how much output firms produce relative to the inputs used. Attributing firm-level productivity gains to data is difficult because revenue-based estimates also reflect selling prices [25], while data assets themselves rarely have observable prices and skills and other intangible inputs may be incompletely or inconsistently measured [26]. Section 5 discusses these measurement issues in more detail.

Firm-level studies nevertheless associate greater use of data-driven decision making with higher productivity [27, 28]. In US manufacturing, plants adopting these practices subsequently recorded around 3 per cent higher value added than non-adopters, controlling for production inputs [29]. Evidence on predictive analytics similarly indicates 1–3 per cent higher plant-level productivity, measured using revenues and controlling for production inputs [30]. Improvements to a narrow process may have a modest effect on overall productivity, although firms that saw gains in the trial extended the practices to their other plants [16].

The gains also depend on firms' capacity to use the information. The manufacturing evidence identifies complementarities with existing IT and workforce skills [30], while UK evidence on firms active online associates the use of online data with larger productivity gains where employees have greater decision-making autonomy [28]. A study of business analytics across 814 firms finds productivity improvements concentrated in IT-intensive or highly competitive industries, with no statistically detectable improvement elsewhere [31].

There is broader evidence also pointing to uneven productivity gains from improved digital and data capabilities. A European industry-level study shows digital adoption is associated with larger productivity gains among already more productive firms and smaller gains where skills

are scarce [32]. Digital assets are also disproportionately concentrated among a small group of "superstar" firms [33]. The falling cost of IT and intangible investments that raise fixed costs while lowering marginal costs can favour expansion by leading firms and contribute to greater concentration and market power [34, 35]. These findings concern digital capabilities more broadly and do not isolate the separate contribution of data assets.

Assessing data's contribution therefore requires matching the outcomes improved with the resources used to achieve them, including staff time, software, computing services and the services provided by accumulated data assets.

2.3 Data as investment

Creating a data asset requires collecting, cleaning, processing and documenting information so that it can support future uses. Experimental estimates suggest that firms devote substantial resources to these data-related activities, amounting to 5–6.5 per cent of market-sector gross value added in the European economies studied over 2010–2018, although differences in definitions and coverage make international comparisons difficult [26]. Section 5.2 expands on the boundaries of data assets and what is recognised as investment in data, given the implications for measurement.

For firms, this means tracing the full data value chain. The investment case must cover not only the data itself, but also all complementary assets needed to use it, the gaps in working practices that must be addressed to support decision-making and any resulting recurring expenditure.

Shared datasets and interoperable systems can allow firms to reuse information across departments and enable new applications. Each intended use still requires work to select the relevant data, validate predictions and incorporate them into local decisions. Where this depends on detailed knowledge of operating conditions, implementation may require close involvement from the teams responsible for those decisions. Shared systems can support this local implementation by making information available across operating units. Evidence from publicly traded firms associates analytics capabilities with greater productivity benefits where innovation is organised through decentralised groups [36].

Training employees and redesigning processes are adjustment costs that belong in the investment assessment [37]. Qualitative UK evidence documents difficulties with legacy systems and limited staff training resources [38]. The investment case should therefore relate expenditure on shared datasets and systems to the processes they are expected to improve, including the costs of implementation and continued use.

2.4 Building the investment case for data initiatives

For firms, the main difficulty in assessing data investment is that the uncertainty and complexity creating the opportunity for better-informed decisions also make the returns harder

to estimate. Realising those returns depends on changes to the decisions and working practices the investment is intended to support, often across several parts of the organisation. An investment case should specify which decisions are expected to improve, what information is missing and how better-informed action would affect costs, revenues or risk. It should also set out the adoption plans and changes in operational practices required across the business units expected to benefit, especially for shared capabilities. The overall expected benefits can then be compared with the full costs of producing and using the information.

For centralised data investments, the assessment should be developed jointly with the affected business areas and should identify opportunities to reuse data and capabilities across different uses. Trials focused on specific processes can test whether better information improves decisions before the firm commits to wider implementation. Where feasible, these should compare the proposed decision process with existing practice, measuring both operational outcomes and the resources required to achieve them. Separating the contribution of data from that of complementary assets remains difficult: when data, software and working practices change together, the comparison estimates the return to the combined intervention. Estimating data's separate contribution requires evidence that varies the information available while holding other capabilities sufficiently constant. Trials can establish whether the contribution of data is significant under the conditions tested, but wider adoption can blur the boundaries between investment classes and requires further assessment of implementation costs and operating performance across units.

Our review finds limited comparable firm-level evidence linking the costs of data investment to realised benefits. Voluntary disclosure by firms could help document the costs of producing and using information and clarify which expenditure constitutes investment in a data asset. Reporting operating outcomes as processes are adapted, and following adoption across operating units, could help assess whether improvements extend beyond the initial setting. Common measures of change relative to existing practice, such as output per hour or downtime, would allow comparisons across similar processes and, with information on operating volumes and costs, estimates of financial benefits. A separate question is whether moving from local capabilities to centrally managed, shared provision reduces the total cost of achieving these improvements. The evidence reviewed leaves this question open: savings from reuse need to be assessed against transition and coordination costs, and any change in local effectiveness.

3. Data access, public value and market structure

3.1 Private versus social value

The use of data can generate externalities: benefits or costs affecting people other than those who control the data. One important negative externality arises when sharing personal data reveals information not only about the individual concerned but also about others connected to them [39].

However, there can also be large positive externalities, often arising from the scope to link data. In fact, the same data can support uses across several organisations and help fill gaps in the information available to each, although private holders often restrict access to their data to protect their advantage, even where doing so leads to inefficiencies across the economy [9]. At the same time, restrictions on access may reflect legitimate concerns about privacy, confidentiality or the need to recover the costs of producing and maintaining data. The policy challenge is therefore to balance incentives for data investment with the wider benefits of sharing and reuse.

Public bodies can help address these inefficiencies by funding data provision and enabling wider access which is the economic basis for the extensive provision of public data, such as official statistics. Access to and sharing of public sector data is estimated to generate benefits worth 0.1 to 1.5 per cent of GDP [40]. For the UK alone, a recent estimate put the potential value of public sector data in the range £15-200 billion [41], and the welfare gains from switching from pricing data at cost to an open charging policy between £1.6 billion and £2 billion a year [42]. So the benefits of data use can accrue downstream and diffuse outside the organisation that funds the data; private value can diverge sharply from social value [43].

Evidence from specific datasets shows how charging and access conditions affect downstream use. Lower prices and relaxed sharing restrictions for Landsat imagery from 1995 increased its use in public research, particularly among researchers at lower-ranked institutions and in developing countries [44]. A study of architectural and engineering firms across 15 countries associated free or marginal-cost access to public geographic data with annual real sales growth about 15 percentage points higher than under cost-recovery pricing [45]. The Met Office, the UK's national weather service, produces and shares weather and climate information, and its services are estimated to generate £18.80 in benefits to the UK for every £1 of public funding over 2024–2033 [46]. Denmark's free address-data agreement was estimated to generate direct financial benefits of about €62 million over 2005–2009, against approximately €2 million in costs of implementing and operating the agreement [47].

3.2 Public data and the infrastructure question

Some data has public-good characteristics, while widely used datasets and the systems that provide access to them can also serve an infrastructure-like role. The analogy is useful, but it is

important to distinguish the data itself from the infrastructure needed to collect, maintain and share it.

The digital stack contains different infrastructure layers, some conventional such as data centres and communications networks. Others are less so, including access infrastructure and foundational or reference datasets such as land registry, identity data and broader geospatial or economic data [48]. Their infrastructure-like role arises not simply from their availability, but from the reliance of multiple users and services on their quality, continuity, interoperability and accessibility.

One question that emerges is when data or data flows become economically essential.

3.3 Access, charging and benefit-sharing

The public good character of some data and the existence of externalities raise questions about:

(a) the form and amount of financing for investment and maintenance of data (including cybersecurity); and
(b) the terms of access and use.

These are areas where more clarity about the most appropriate models is needed. Public investment needs methods of appraisal for the downstream benefits of data access and use. These could draw on approaches used to assess support for research and innovation, where the benefits of shared knowledge extend beyond the organisation funding its production. Appraisal should consider whether wider access creates public benefits or reduces risks to essential services, and whether investment can reduce dependence on dominant data holders alongside competition policy. These questions also arise in AI, where access to data can support innovation while reinforcing the position of firms best placed to exploit it. Access and public governance arrangements will influence how widely these benefits are shared [49, 50], an issue considered further in Section 4. Yet despite the evident importance of data for the AI economy, public funding has often been limited by general fiscal constraints. This points to a corresponding need for appropriate business models for charging for some public data: which datasets are reference data that should be freely available open access as of right, where might value-added data products be charged for to recover costs (for example, API fees versus free CSV downloads). Charging arrangements also involve a trade-off: they may help finance the continued provision and maintenance of data, but can discourage reuse and reduce the wider benefits generated by access.

Private investment will occur where businesses see commercial potential. Here the questions are the extent to which some open data sharing or access on fair, reasonable and non-discriminatory (FRAND) terms might be mandated. One example is England's Bus Open Data Service, which requires operators of local bus services outside London to publish timetable, fares and vehicle-location data in standardised formats [51]. Safety-relevant data

from self-driving vehicles is another area where road safety considerations would lead to enforced data sharing between operators and regulators [52].

3.4 Critical datasets

These considerations point to the desirability of identifying datasets whose disruption or restricted access may have especially large effects. These include data fundamental to the role of the state, such as identity and cadastral records [48], as well as data supporting essential services related to health and energy systems. Credit reporting data is economically critical to lending decisions [53].

Criticality is therefore not an inherent property of a dataset: it depends on the functions it supports, the availability of substitutes and the potential for disruption to spread across interconnected services.

Established approaches to assessing critical infrastructure and its interdependencies provide a useful starting point [54]. Requirements for resilience, cybersecurity, data quality and access should reflect the consequences of disruption, while balancing the risks of wider sharing against those of restricted access.

Access and governance conditions should vary across these datasets and consider both the risks of wider sharing and the consequences of restricting access.

3.5 Market power, governance and trust

Beyond helping firms improve decisions and create economic value, data can also be a source of market power. Access to information accumulated through customer interactions may give established firms an advantage over competitors that have not yet developed comparable data resources [55]. Data can form a barrier to entry to new competitors which have not yet built up customer data insights, and creates a self-reinforcing loop whereby more data means more revenue, more investment and more customers, which in turn creates more data. This loop can benefit customers but harms competition and innovation over time, especially where rivals cannot obtain comparable information [55].

Competition authorities and regulators therefore need to be able to assess the proportionality of imposing data sharing mandates (such as the UK's Open Banking regime), as there is a trade-off between private commercial value and the public value of broader access. Depending on the type of data involved, any mandates need to construct a secure access and control regime.

Any intermediaries involved need to establish trust both with the public and with businesses which may be called on to share data they have produced and maintained. Public trust in the use of data whether by public bodies or private businesses is essential and potentially fragile.

4. Data, AI and the changing production system

The value of a data asset depends on what can be learned from it, and at what cost. AI is changing both. It has reduced the cost of extracting information from records that were previously too unstructured or not linked. In other words, data that organisations already hold can become more valuable without the records themselves changing. It has also made new data cheaper to produce, as collection, cleaning and analysis become easier to automate. A third effect concerns the distribution of returns. Where models, computing capacity or specialised skills are scarce, the organisation holding the data may not be the one capturing most of the economic benefits.

These effects may work in opposing ways, and the implications for investment are not obvious. It is possible that the private incentive to invest in new data may weaken even as the social return to doing so rises. This section takes each mechanism in turn and asks what each implies for access and competition. The measurement questions are left to Section 5.

4.1 AI as an amplifier of data value

Most organisations hold considerably more recorded material than they use. With the right systems in place, structured data could be queried, while unstructured text, spreadsheets, images, audio, and video largely could not until recently, at least not at a scale that justified the effort. For many organisations, a major obstacle has been the cost of turning these digital records into usable information, alongside difficulties of access and integration.

Extracting anything typically meant assembling labelled examples and building a structured pipeline for a particular task, which came with a fixed cost only the most valuable data could carry. AI models pretrained on broad corpora have largely removed this constraint [56]. Clinical notes, maintenance logs, call transcripts, inspection reports, and correspondence, for example, are now feasible inputs to analysis, yet the organisations holding them did not have to do anything for that to happen.

There are direct and somewhat unexpected implications for data valuation. First, a data asset can become more valuable without any change in the underlying records, in the holder's capabilities, or in the uses for which it was originally collected. What has changed is the cost of extraction. Second, the assumption that an intangible asset loses usefulness over a fixed service life does not hold. In fact, the asset could appreciate, sometimes overnight. And thirdly, cost-based estimates of data value, which approximate what the data cost to produce rather than what can now be learned from it, are limited as the cost could be zero in many cases. The real options literature anticipates some of the problems by discussing how a dataset may have value due to uses that are not yet feasible, and a fall in extraction costs is what converts that option into a realised gain [57].

A second channel runs through data linkage. Combining datasets can be limited by regulations or preferences, but a key factor has often been cost. Some of the barriers have been

non-matching identifiers, units, or definitions, storage formats and coverage changing over time. AI-based entity resolution and matching reduce that cost substantially [56]. Previously the benefits of linking datasets have been known in principle but were hard to estimate. As linkage becomes routine and automatic, the gains from combining sources become an empirical question instead.

How far any of the issues discussed here raises the value of data in general depends on the degree to which AI rewards scale or specificity. Scale lowers marginal cost, and in many cases, patterns cannot be detected in small samples. This favours holders of large datasets and large single data markets. At the same time the public stock of digital human-generated text could be fully used for model training between 2026 and 2032 [58], and verified domain-specific inputs appear increasingly scarce. This has direct implications for policy. If scale dominates, there is a case for broad access to large holdings, while if specificity dominates, the case is for sectoral institutions and verified public data, more demanding to build.

4.2 How AI changes data production

The second change concerns the cost of producing data rather than the cost of using it. Agentic systems can now carry out much of the work of collecting, documenting and maintaining records with limited supervision. If processing costs fall towards zero, the constraint moves upstream to the observation of real-world phenomena. What becomes scarce is not the capacity to work with data records but the fact of having made the record at all. The focus of data investment shifts from processing to creating the observation in the first place, which includes instrumenting workflows and processes so they record what passes through them, running surveys and experiments to generate observations that would not otherwise occur, and negotiating access to material held by others.

Synthetic data complicates this for two reasons. First, generated data can potentially substitute for observed data and thereby reduce the premium for holding the latter. Second, there can be a degradation of data quality as generated data is drawn from a narrower distribution than it was trained on. This is because it under-represents rare events and since data generation is tuned towards likely output it is pushed away from the tails [59].

Moreover, since much public data used in model training was an unpriced by-product of everyday activity, there was never a clear investment decision in the first place. It is not clear who will make that decision as AI agents increasingly substitute for the activity that helped train it. Meanwhile, previously openly available material is being closed off, with restrictions on automated collection (robots.txt), particularly among the highest-quality sources [60]. In a recent study, del Rio-Chanona et al. [61] estimated a 25 per cent decline in activity on Stack Overflow within six months of ChatGPT's release.

4.3 Complementarity, scarcity and the distribution of returns

Data is not productive on its own. It requires models, computing capacity, software, and people who can use it. Section 2 discussed what this means for a firm building an investment case, but there is a broader and less obvious implication: where inputs are complements, returns tend to accrue to whichever is scarcest and hardest to substitute, rather than being shared in proportion to contribution. There are two competing views about whether this is the data.

One view is that it is not obvious that data is the binding constraint. Access to the most capable models, advanced chips, electricity, and specialised staff is expensive, while the cost of storing and processing records continues to fall. If the scarce inputs are the complements rather than the records themselves, the organisation holding a dataset is not necessarily the party capturing most of the return from it.

The opposite case is that model capabilities are converging, open weights are widely available and only a few months behind the leading models, and the cost of deploying AI models is falling. If models become something close to a commodity, what distinguishes one application from another moves back towards proprietary data and the context in which it is used.

Where returns accrue is also a commercial outcome, and one that determines whether a potential price can be observed which matters for measurement as discussed in Section 5. A data holder can sell raw records, a curated dataset, or a service built on top of them. Selling raw records creates competitors, while selling a service captures more of the margin and makes the value of the underlying records invisible from outside. For most of the period in which data has been discussed as an asset, very little was sold in the first form, which is why market prices were so rarely available.

The market in training data has begun to change this. Deals concluded since 2020 show several distinct pricing mechanisms and, notably, that most exclude the original creators of the material from any share of the proceeds [62]. Large enterprise deals mostly take the form of recurring, time-limited access, which makes the arrangement closer to a subscription than a sale. Reported values should be treated with care, since most terms are not disclosed.

4.4 Access, competition and distribution

The right to access data for AI training is not settled within existing law in most jurisdictions. The market has formed ahead of the law. At the same time, Longpre et al. [60] document a sharp rise from 2023 in restrictions on automated collection, concentrated among the highest-quality sources, while licensing agreements have emerged in parallel to supply what can no longer simply be taken for free.

The distributional consequences can be large and follow mainly from timing. First, developers who collected material while the web was open hold corpora that cannot now be assembled as easily or at all. This constitutes an advantage being there early rather than by anything they

built. Second, licensing favours those able to negotiate, which means large publishers on one side and well-capitalised labs on the other, while the individuals who produced the material are mostly excluded from the proceeds. Early-mover data advantages can raise barriers to entry for later developers [63].

5. Measuring data in firms and the macroeconomy

Firms use data to improve products, make decisions, organise production and develop new services. Data is also increasingly an essential input into artificial intelligence. Yet much of the investment in producing data assets remains difficult to identify separately in economic statistics.

Recognising data as a produced asset in the 2025 System of National Accounts (SNA) is an important step towards understanding its contribution to economic growth [64]. What is counted as investment affects estimates of capital accumulation and how we interpret the contributions of capital and productivity to economic growth [65]. Yet much data is produced within organisations, often alongside other activities, and rarely has a separately observable market price. The challenge is to develop consistent estimates of data investment and the resulting capital stock, while understanding the economic benefits these assets generate.

5.1 What should be measured?

Different questions require different measures. Investment captures the resources devoted to producing data assets; the capital stock reflects the accumulated investment that remains useful; capital services capture the productive contribution of these assets over time; and measures of economic returns seek to establish what firms and society gain from using data.

Consider a retailer collecting information about customer purchases. Its production costs may be relatively straightforward to estimate, but its value depends on whether the retailer can use it to improve stock management, forecast demand or develop new services. Another business with similar information may derive quite different benefits, depending on its skills, technologies and organisational capabilities.

National accountants need measures of investment and asset stocks, businesses are interested in commercial returns, and policymakers must also consider benefits beyond the organisation holding the data [10]. Data can be reused for different purposes, and new applications may emerge long after its collection, complicating assumptions about depreciation and the value of future uses [57].

5.2 Incorporating data into national accounts

The 2025 SNA recognises data as a new category of produced assets, following years of experimental work by national statistical offices and international organisations [5, 66]. The Eurostat–IMF handbook provides practical guidance for implementing the new framework [67].

Recognising data as an asset requires defining what qualifies as data production and distinguishing it from other forms of investment. The boundary with software, databases and research and development is particularly important, as some of the expenditure associated with creating data assets may already be recorded under these categories. Without clear boundaries, estimates risk double-counting the same investment.

An important criterion is that the data is expected to provide economic benefits for more than one year, distinguishing longer-lived assets from information consumed in current operations [68].

For data produced within organisations, the absence of separately observable market prices makes a cost-based approach the practical starting point. Statistical offices estimate the labour and other resources devoted to collecting, organising and maintaining information, often by identifying data-related occupations and the proportion of working time devoted to producing data assets.

Experimental estimates of data investment have been developed across several countries, including Canada [69] and the United States. Santiago Calderón and Rassier [70], for example, use machine learning and online job advertisements to estimate data-related working time, finding that annual investment in internally produced data assets in the US business sector rose from $84 billion in 2002 to $186 billion in 2021. Extending this approach internationally, Schmidt et al. [71] use natural language processing of online job advertisements to identify data-intensive occupations in the United Kingdom, Canada and the United States, highlighting differences in data-related work across countries. Turning to Europe, Goodridge et al. [72] estimate investment in software and data assets using harmonised labour force data, showing how extending the asset boundary substantially increases measured own-account investment.

Recent BEA research extends the US estimates of data and database investment, depreciation and capital stocks to 1997–2024 [3]. Samuels et al. [4] find that capitalising internally produced data raises the contribution of IT-related capital to US GDP growth by approximately one-third over 2002–2024, with substantial differences across industries. The treatment of data thus matters for understanding the sources of measured economic growth, not only the level of investment.

Implementation is also advancing in the UK. The Office for National Statistics is developing methods to estimate data production and the stock of existing data assets, with a proposed timetable of 2029–2030 for incorporating these estimates into its core national accounts and fiscal statistics [68].

Important measurement questions nevertheless remain, even in this initial implementation of the new standard. Occupational classifications do not consistently identify data-related activities, and working-time assumptions are difficult to validate. Estimating depreciation presents a further challenge. Some data becomes obsolete quickly, while other datasets remain useful for many years or acquire new applications when combined with additional information. Assumptions about useful lives and depreciation therefore affect estimates of data capital stocks and the capital services they provide, with consequences for measured productivity growth.

Differences in occupational coverage, asset boundaries and depreciation assumptions can lead to divergent estimates of investment and capital stocks, limiting international comparability [5, 8]. Capitalising data also changes the allocation of measured growth between capital deepening and total factor productivity, without necessarily reflecting a change in underlying economic performance. Consistent implementation will therefore require better empirical evidence, transparent methodologies and closer coordination across statistical offices.

5.3 What can firm-level evidence tell us?

National accounts provide aggregate estimates of data investment, but improving these estimates requires better information about how firms produce, maintain and use data assets. Firm surveys and business records could identify the time employees devote to data production, expenditure on purchased data and related services, and the types of datasets maintained. Evidence on how datasets are updated, shared across business units and eventually retired could also inform assumptions about useful lives and depreciation.

Beyond production costs, firm-level evidence can help establish how data assets are used and what economic outcomes are associated with them. This requires connecting investment and data-related activities to subsequent changes in productivity, costs or revenues, while recognising the role of software, skills and organisational capabilities. Existing approaches to data valuation provide a basis for exploring these links, including the value of possible future uses [10, 57].

Targeted business surveys and reporting pilots, developed with statistical offices and firms, could test how reliably such information can be collected and compared, while limiting reporting burdens and protecting commercially sensitive information.

5.4 How does agentic AI change the measurement of data?

AI is changing both the production and use of data. Measurement must distinguish these effects: lower expenditure on data production may accompany larger quantities of data, improvements in quality or greater economic value from its use.

Existing estimates often identify workers involved in collecting, processing and maintaining information, then assign a share of their working time to data production. Yet AI can

increasingly perform these tasks, while more autonomous agents may combine them into longer workflows. Shao et al. [73] examine the potential for AI agents to automate or augment activities across occupations, illustrating how the division of work between employees and AI systems may change.

As data-related tasks shift from employees to AI systems, statistical estimates will need to capture changes in both labour costs and expenditure on software, computing services and other inputs. Lower production costs may reflect genuine efficiency gains, but estimates relying primarily on occupations and working time may fail to capture changes in how data is produced.

AI may also increase the usefulness of existing data by reducing the costs of extracting information, combining sources and carrying out analysis. This creates measurement challenges beyond data production itself. Coyle and Poquiz [74] argue that understanding AI's economic effects will require better evidence on tasks, business processes and outcomes. Makridis and Brynjolfsson [75] similarly propose integrating information on AI investment and actual use into national statistics.

Cost-based measurement remains necessary for national accounts, but statistical methods will need to adapt to changes in data production. In particular, they must distinguish the creation of new data assets from the processing and reuse of existing information.

5.5 Beyond national accounts

Even comprehensive estimates of business investment will not capture all the economic benefits of data. Data generated through household and non-market activities may create value without corresponding transactions, while sharing and reuse can generate benefits beyond the organisation that originally produced the data [76].

Cross-border activities introduce further difficulties. Data may be collected in one country, processed in another and used to generate revenues elsewhere, complicating the measurement of production, ownership and international flows [8, 77].

The inclusion of data as an asset in the 2025 SNA provides an important foundation for economic measurement. Yet understanding its wider contribution will require complementary evidence on data use, economic returns and social benefits. The research and statistical priorities arising from these gaps are taken up in the shared agenda in Section 6.

6. An agenda on the economics of data

This overview of the economics of data presents a set of connected questions that no single discipline or institution can resolve. The preceding sections have identified important recent advances in understanding how data creates value, how firms invest in it, and how access, governance and AI affect their use. Yet the evidence remains fragmented, and many of the choices facing businesses and policymakers lack a sufficiently robust economic basis.

A shared agenda should focus on questions where further evidence can improve decisions about investment, access, governance and measurement. Five priorities emerge.

6.1 Proposed priorities

1. Firms' investment and productivity. Firms could contribute stronger evidence on the realised returns to investments in producing and acquiring data. Particular gaps concern differences across firms and sectors, the role of complementary capabilities, and the relationship between data investment, innovation and productivity. Policymakers could support this work through targeted surveys and common standards for voluntary reporting. Both could also cover AI-supported data production to track changes in fixed production costs.

2. Access, sharing and public value. Researchers and policymakers could focus on identifying the conditions under which wider data access generates net social benefits, taking account of investment incentives, privacy, competition and the costs of maintaining data. This includes comparing the economic consequences of different access, pricing, licensing and benefit-sharing arrangements. Public bodies could also test methods for appraising investment in data within frameworks such as the UK's Green Book, through pilots focused on data provision and use. These pilots could trace how reductions in uncertainty change decisions and generate benefits beyond the funding body by estimating their value added to the economy.

3. Critical data and infrastructure. Policymakers could develop criteria for identifying data resources whose restricted availability or disruption could have substantial economic consequences, and assess the condition of those resources. This work would need to consider dependencies, substitutability, resilience and the respective responsibilities of public and private actors.

4. Data in the AI economy. Researchers could examine how AI, including agentic systems, changes the value of existing data and the costs and incentives for producing new observations. They could also assess how changes in the relative scarcity of data and complementary inputs affect training-data markets, competition and the distribution of economic returns.

5. Measurement systems fit for the data economy. Statistical agencies could prioritise the development of internationally comparable methods for implementing the 2025 SNA treatment of data assets, covering production boundaries, investment, depreciation and capital services.

They could complement cost-based measures with evidence on economic returns and non-market benefits, while improving the measurement of cross-border data flows.

6.2 From shared questions to a common evidence base

These priorities call for different types of evidence. Some require theoretical and empirical research to identify the mechanisms through which data creates value. Others require access to firm-level information, evaluation of existing data-sharing arrangements, or statistical experimentation. The relevant evidence is often dispersed across researchers, businesses, public institutions and statistical agencies.

Greater comparability would make emerging evidence more useful. Shared definitions and research questions, transparent methods, and comparable case studies would help distinguish findings that generalise across settings from those that depend on particular technologies, institutions or sectors. Private-sector experience is essential, both for understanding how data is used in practice and for evaluating the costs and benefits of alternative investment and access arrangements.

International cooperation is particularly valuable where data crosses borders, markets are interconnected, or statistical methods must remain comparable. But coordination need not imply a uniform approach: the appropriate response may be further research, improved business practice, statistical guidance or a reconsideration of existing governance arrangements.

The purpose of this agenda is to build agreement about developing an economic evidence base that supports better decisions about data: what to invest in, what to share, how to govern access and how to assess and track the benefits that follow.

References

[1] European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, & World Bank. (2025). *System of National Accounts 2025*. United Nations. https://unstats.un.org/unsd/econstats/national-accounts/sna2025

[2] UK Statistics Authority. (2026). *Advisory Panel on National Accounts minutes: 17 September 2026*. https://uksa.statisticsauthority.gov.uk/publications/advisory-panel-on-national-accounts-minutes-17-september-2026/

[3] Santiago Calderón, J. B. (2026). *Toward economics statistics on: Own-account data and databases assets for 1997–2024* (BEA Working Paper No. 2026-4). U.S. Bureau of Economic Analysis.

https://www.bea.gov/research/papers/2026/toward-economics-statistics-own-account-data-and-databases-assets-1997-2024

[4] Samuels, J. D., Santiago Calderón, J. B., & Garner, C. (2026). *The impact of capitalizing data on productivity growth in the U.S.* (BEA Working Paper No. 2026-7). U.S. Bureau of Economic Analysis. https://www.bea.gov/research/papers/2026/impact-capitalizing-data-productivity-growth-us

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

[6] Veldkamp, L. (2023). Valuing data as an asset. *Review of Finance, 27*(5), 1545–1562. https://doi.org/10.1093/rof/rfac073

[7] Coyle, D., Diepeveen, S., Wdowin, J., Tennison, J., & Kay, L. (2020). *The value of data: Summary report 2020.* Bennett Institute for Public Policy and Open Data Institute. https://doi.org/10.17863/CAM.127050

[8] Organisation for Economic Co-operation and Development. (2022). *Measuring the value of data and data flows* (OECD Digital Economy Papers No. 345). OECD Publishing. https://doi.org/10.1787/923230a6-en

[9] Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review, 110*(9), 2819–2858. https://doi.org/10.1257/aer.20191330

[10] Coyle, D., & Manley, A. (2024). What is the value of data? A review of empirical methods. *Journal of Economic Surveys, 38*(4), 1317–1337. https://doi.org/10.1111/joes.12585

[11] Lange, S., & Spiezia, V. (2026). The value of data in the economy. In S. Zambon & L. Girella (Eds.), *Handbook on intangibles: Accounting, reporting and valuation* (pp. 641–668). Edward Elgar Publishing. https://doi.org/10.4337/9781035306374.00045

[12] Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy, 47*, 1–6. https://doi.org/10.1016/j.infoecopol.2019.05.001

[13] Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526. https://doi.org/10.1037/a0016755

[14] Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

[15] Nguyen, D., & Paczos, M. (2020). *Measuring the economic value of data and cross-border data flows: A business perspective* (OECD Digital Economy Papers No. 297). OECD Publishing. https://doi.org/10.1787/6345995e-en

[16] Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., & Roberts, J. (2013). Does management matter? Evidence from India. *Quarterly Journal of Economics, 128*(1), 1–51. https://doi.org/10.1093/qje/qjs044

[17] Chuang, H. H.-C., Oliva, R., & Liu, S. (2016). On-shelf availability, retail performance, and external audits: A field experiment. *Production and Operations Management, 25*(5), 935–951. https://doi.org/10.1111/poms.12519

[18] Caro, F., & Gallien, J. (2010). Inventory management of a fast-fashion retail network. *Operations Research, 58*(2), 257–273. https://doi.org/10.1287/opre.1090.0698

[19] Bar-Gill, S., Brynjolfsson, E., & Hak, N. (2024). Helping small businesses become more data-driven: A field experiment on eBay. *Management Science, 70*(11), 7345–7372. https://doi.org/10.1287/mnsc.2021.02026

[20] Sun, T., Yuan, Z., Li, C., Zhang, K., & Xu, J. (2024). The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy. *Management Science, 70*(4), 2645–2660. https://doi.org/10.1287/mnsc.2023.4828

[21] Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research, 55*(1), 80–98. https://doi.org/10.1509/jmr.16.0163

[22] Baley, I., & Veldkamp, L. L. (2025). *The Data Economy: Tools and Applications*. Princeton University Press.

[23] Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *Quarterly Journal of Economics, 129*(3), 1311–1353. https://doi.org/10.1093/qje/qju015

[24] Jiang, Z., & Li, J. (2025). Instrumenting while experimenting: An empirical method for competitive pricing at scale. *Operations Research, 73*(5), 2477–2495. https://doi.org/10.1287/opre.2022.0157

[25] Foster, L., Haltiwanger, J., & Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review, 98*(1), 394–425. https://doi.org/10.1257/aer.98.1.394

[26] Corrado, C., Haskel, J., Iommi, M., & Jona-Lasinio, C. (2022). *The value of data in digital-based business models: Measurement and economic policy implications* (OECD Economics Department Working Papers No. 1723). OECD Publishing. https://doi.org/10.1787/d960a10c-en

[27] Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). *Strength in numbers: How does data-driven decisionmaking affect firm performance?* [Working paper]. https://doi.org/10.2139/ssrn.1819486

[28] Bakhshi, H., Bravo-Biosca, A., & Mateos-Garcia, J. (2014). *The analytical firm: Estimating the effect of data and online analytics on firm performance* (Nesta Working Paper No. 14/05). Nesta.

[29] Brynjolfsson, E., & McElheran, K. (2016). *Data in action: Data-driven decision making in U.S. manufacturing* (CES Working Paper No. 16-06). U.S. Census Bureau, Center for Economic Studies.

[30] Brynjolfsson, E., Jin, W., & McElheran, K. (2021). The power of prediction: Predictive analytics, workplace complements, and business performance. *Business Economics, 56*(4), 217–239. https://doi.org/10.1057/s11369-021-00224-5

[31] Müller, O., Fay, M., & vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems, 35*(2), 488–509. https://doi.org/10.1080/07421222.2018.1451955

[32] Gal, P., Nicoletti, G., Renault, T., Sorbe, S., & Timiliotis, C. (2019). *Digitalisation and productivity: In search of the holy grail. Firm-level empirical evidence from EU countries* (OECD Economics Department Working Papers No. 1533). OECD Publishing. https://doi.org/10.1787/5080f4b6-en

[33] Tambe, P., Hitt, L., Rock, D., & Brynjolfsson, E. (2020). *Digital capital and superstar firms* (NBER Working Paper No. 28285). National Bureau of Economic Research. https://doi.org/10.3386/w28285

[34] Lashkari, D., Bauer, A., & Boussard, J. (2024). Information technology and returns to scale. *American Economic Review, 114*(6), 1769–1815. https://doi.org/10.1257/aer.20220522

[35] De Ridder, M. (2024). Market power and innovation in the intangible economy. *American Economic Review, 114*(1), 199–251. https://doi.org/10.1257/aer.20201079

[36] Wu, L., Lou, B., & Hitt, L. M. (2019). Data analytics supports decentralized innovation. *Management Science, 65*(10), 4863–4877. https://doi.org/10.1287/mnsc.2019.3344

[37] Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives, 14*(4), 23–48. https://doi.org/10.1257/jep.14.4.23

[38] Department for Science, Innovation and Technology. (2026). *Business data use and productivity study (wave 2): Qualitative research report*. UK Government.

[39] Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics, 14*(4), 218–256. https://doi.org/10.1257/mic.20200200

[40] Organisation for Economic Co-operation and Development. (2019). *Enhancing access to and sharing of data: Reconciling risks and benefits for data re-use across societies*. OECD Publishing. https://doi.org/10.1787/276aaca8-en

[41] Department for Science, Innovation and Technology. (2025, October 24). *Value of public sector data estimate*. UK Government.

https://www.gov.uk/government/publications/value-of-public-sector-data-estimate/value-of-public-sector-data-estimate

[42] Pollock, R. (2011). *Welfare gains from opening up public sector information in the UK* [Working paper]. University of Cambridge. https://web.archive.org/web/20160322084116id_/http://rufuspollock.org/papers/psi_openness_gains.pdf

[43] Coyle, D., Diepeveen, S., Wdowin, J., Kay, L., & Tennison, J. (2020). *The value of data: Policy implications*. Bennett Institute for Public Policy, University of Cambridge, and Open Data Institute. https://doi.org/10.17863/CAM.127048

[44] Nagaraj, A., Shears, E., & de Vaan, M. (2020). Improving data access democratizes and diversifies science. *Proceedings of the National Academy of Sciences, 117*(38), 23490–23498. https://doi.org/10.1073/pnas.2001682117

[45] Koski, H. (2011). *Does marginal cost pricing of public sector information spur firm growth?* (ETLA Discussion Paper No. 1260). The Research Institute of the Finnish Economy. https://www.etla.fi/wp-content/uploads/2012/09/dp1260.pdf

[46] London Economics. (2024). *The economic value of the Met Office: Updated evaluation study of the economic value of the Met Office's activities to the UK over the next ten years* (Final report). Met Office. https://www.metoffice.gov.uk/about-us/governance/economics

[47] Danish Enterprise and Construction Authority. (2010). *The value of Danish address data: Social benefits from the 2002 agreement on procuring address data etc. free of charge*. https://opendatatoolkit.worldbank.org/content/dam/sites/data/odk/documents/value-assessment-danish-address-data-uk-2010-07-07b.pdf

[48] Organisation for Economic Co-operation and Development. (2024). *Digital public infrastructure for digital governments* (OECD Public Governance Policy Papers No. 68). OECD Publishing. https://doi.org/10.1787/ff525dc8-en

[49] Goos, M., & Savona, M. (2024). The governance of artificial intelligence: Harnessing opportunities and mitigating challenges. *Research Policy, 53*(3), Article 104928. https://doi.org/10.1016/j.respol.2023.104928

[50] Savona, M. (2026). *Data value and data governance* (LEAP Working Paper No. 3/2026). Luiss Institute for European Analysis and Policy.

[51] Department for Transport. (2020). *Bus open data implementation guide*. UK Government. https://www.gov.uk/government/publications/bus-open-data-implementation-guide/bus-open-data-implementation-guide

[52] Centre for Data Ethics and Innovation. (2022). *Responsible innovation in self-driving vehicles*. UK Government.

https://www.gov.uk/government/publications/responsible-innovation-in-self-driving-vehicles

[53] World Bank. (2011). *General principles for credit reporting*. World Bank. https://documents1.worldbank.org/curated/en/662161468147557554/pdf/70193-2014-CR-General-Principles-Web-Ready.pdf

[54] Organisation for Economic Co-operation and Development. (2019). *Good governance for critical infrastructure resilience* (OECD Reviews of Risk Management Policies). OECD Publishing. https://doi.org/10.1787/02f0e5a0-en

[55] Digital Competition Expert Panel. (2019). *Unlocking digital competition: Report of the Digital Competition Expert Panel*. HM Treasury. https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel

[56] Wang, T., Chen, X., Lin, H., Chen, X., Han, X., Sun, L., Wang, H., & Zeng, Z. (2025). Match, compare, or select? An investigation of large language models for entity matching. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 96–109). Association for Computational Linguistics. https://aclanthology.org/2025.coling-main.8/

[57] Coyle, D., & Gamberi, L. (2024). A real options approach to data valuation. *Business Economics, 59*(4), 227–234. https://doi.org/10.1057/s11369-024-00374-2

[58] Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). Position: Will we run out of data? Limits of LLM scaling based on human-generated data. In *Proceedings of the 41st International Conference on Machine Learning* (Proceedings of Machine Learning Research, Vol. 235, pp. 49523–49544). PMLR. https://proceedings.mlr.press/v235/villalobos24a.html

[59] Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature, 631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

[60] Longpre, S., Mahari, R., Lee, A., Lund, C., Oderinwale, H., Brannon, W., Saxena, N., Obeng-Marnu, N., South, T., Hunter, C., Klyman, K., Klamm, C., Schoelkopf, H., Singh, N., Cherep, M., Anis, A., Dinh, A., Chitongo, C., Yin, D., … Pentland, S. (2024). Consent in crisis: The rapid decline of the AI data commons. *Advances in Neural Information Processing Systems, 37*, 108042–108087. https://doi.org/10.52202/079017-3431

[61] del Rio-Chanona, R. M., Laurentsyeva, N., & Wachs, J. (2024). Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS Nexus, 3*(9), pgae400. https://doi.org/10.1093/pnasnexus/pgae400

[62] Oderinwale, H., & Kazlauskas, A. (2025). *The economics of AI training data: A research agenda* [Preprint]. arXiv. https://arxiv.org/abs/2510.24990

[63] Vipra, J., & Korinek, A. (2023). *Market concentration implications of foundation models: The invisible hand of ChatGPT* (Brookings Center on Regulation and Markets Working Paper No. 9). Brookings Institution. https://doi.org/10.48550/arXiv.2311.01550

[64] Corrado, C., Haskel, J., Iommi, M., & Jona-Lasinio, C. (2022). *Measuring data as an asset: Framework, methods and preliminary estimates* (OECD Economics Department Working Papers No. 1731). OECD Publishing. https://doi.org/10.1787/b840fb01-en

[65] Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and U.S. economic growth. *Review of Income and Wealth, 55*(3), 661–685. https://doi.org/10.1111/j.1475-4991.2009.00343.x

[66] Rassier, D. G., Kornfeld, R. J., & Strassner, E. H. (2019). *Treatment of data in national accounts* [Paper prepared for the BEA Advisory Committee]. U.S. Bureau of Economic Analysis. https://www.bea.gov/sites/default/files/2019-05/Paper-on-Treatment-of-Data-BEA-ACM.pdf

[67] Eurostat, & International Monetary Fund. (2026). *Handbook on measuring data in the System of National Accounts* (2025 ed., Manuals and Guidelines). Publications Office of the European Union. https://doi.org/10.2785/2811750

[68] Office for National Statistics. (2025, June 27). *Looking ahead: Developments in public sector finance statistics, 2025*. https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/publicsectorfinance/articles/lookingaheaddevelopmentsinpublicsectorfinancestatistics/2025

[69] Statistics Canada. (2019). *The value of data in Canada: Experimental estimates*. Latest Developments in the Canadian Economic Accounts. https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00009-eng.htm

[70] Santiago Calderón, J. B., & Rassier, D. G. (2022). *Valuing the U.S. data economy using machine learning and online job postings* (BEA Working Paper No. 2022-13). U.S. Bureau of Economic Analysis. https://www.bea.gov/research/papers/2022/valuing-us-data-economy-using-machine-learning-and-online-job-postings

[71] Schmidt, J., Pilgrim, G., & Mourougane, A. (2026). Mapping labor demand for data-intensive work in the United Kingdom, Canada, and the United States. *Review of Income and Wealth, 72*(1), Article e70047. https://doi.org/10.1111/roiw.70047

[72] Goodridge, P., Haskel, J., & Edquist, H. (2022). We see data everywhere except in the productivity statistics. *Review of Income and Wealth, 68*(4), 862–894. https://doi.org/10.1111/roiw.12542

[73] Shao, Y., Zope, H., Jiang, Y., Pei, J., Nguyen, D., Brynjolfsson, E., & Yang, D. (2025). *Future of work with AI agents: Auditing automation and augmentation potential across the U.S. workforce* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.06576

[74] Coyle, D., & Poquiz, J. L. S. (2025). *Making AI count: The next measurement frontier* (NBER Working Paper No. 34330). National Bureau of Economic Research. https://doi.org/10.3386/w34330

[75] Makridis, C. A., & Brynjolfsson, E. (2026, January 15). *Counting AI: A blueprint to integrate AI investment and use data into US national statistics*. Brookings Institution. https://www.brookings.edu/articles/counting-ai-a-blueprint-to-integrate-ai-investment-and-use-data-into-us-national-statistics/

[76] Coyle, D., & Manley, A. (2021). *Potential social value from data: An application of discrete choice analysis* (ESCoE Discussion Paper No. 2021-17). Economic Statistics Centre of Excellence. https://www.escoe.ac.uk/publications/potential-social-value-from-data-an-application-of-discrete-choice-analysis/

[77] Coyle, D., & Nguyen, D. (2019). Cloud computing, cross-border data flows and new challenges for measurement in economics. *National Institute Economic Review, 249*(1), R30–R38. https://doi.org/10.1177/002795011924900112

UNIVERSITY OF CAMBRIDGE | Bennett School of Public Policy

Bennett School for Public Policy
University of Cambridge
The Mond Building
New Museums Site
Free School Lane
Cambridge CB3 9DT

bennettschool.cam.ac.uk
enquiries@bennettschool.cam.ac.uk