# The rising value of data:

## How data is reshaping the Dutch economy

# Table of contents

# Main findings

## The various growing roles of Data in the economy

**Data as a signal of value in motion and an opportunity for reinvention**
A defining trend in modern economies is the rapidly growing volume and value of data. Once merely a record of information, data has become the foundation upon which modern economies and societies are built. It now underpins almost every economic activity, from storing information and driving data-informed decision-making to training AI models. As data grows richer and more abundant, and our capacity to extract insights from it deepens, it increasingly signals where economic value is being created and where it is shifting.

Tracking where investments in various data products, such as data (collection), databases (organisation and storage) and especially data science (generation of insights), take place, reveals which occupations and industries are positioning themselves for future value creation.

**The economic value of data is difficult to measure, yet demonstrably large and growing**
Measuring the economic value of data is inherently challenging, as data is typically embedded in processes, systems, and services rather than traded as a standalone product. Although this value cannot be directly observed, it can be approximated by estimating the inputs involved, such as labour and other costs related to data collection, organisation and storage, as well as the application of data science to generate insights.

With this in mind, we used large language models to analyse the task profiles of 923 occupations, identifying which tasks relate to the production of data, databases and data science products, and what proportion of working time each task represents. This occupation-level analysis forms the foundation for estimating the value of data across occupations, industries and the Dutch economy as a whole.

> “Once merely a record of information, data has become the foundation upon which modern economies and societies are built.”

## Results

**The value of data products for the Dutch economy reached 9.6% of Dutch GDP in 2024**

We estimate that in 2024 total investment in data products reached €108 billion – equivalent to 9.6% of Dutch GDP – up from 9.0% in 2013. Data science accounted for 37.0% of this amount in 2024, up from 28.9% in 2013, while the share of data fell from 58.3% to 51.0% over the same period. This shift reflects a broader trend: although data collection still represents the largest share of total investment, data science has grown rapidly, expanding at a minimum of 5.5% annually since 2017 and steadily closing the gap.

Notably, because investments in data products accumulate over time, their stock of value grows substantially year on year. We calculate that investments made between 2013 and 2024, accounting for depreciation, have accumulated to a total stock value of €466bn for the Dutch economy.

**Healthcare & social work and specialised business services led in total data product investment, while financial services & real estate, and information & communication were the most data-intensive in 2024**
Looking across industries, healthcare & social work and specialised business services were the largest investors in data products in 2024, with €19.5 billion and €14.7 billion invested, respectively.

Since industries vary considerably in total employment, adjusting for that, financial services & real estate and information & communication emerged as the most data-intensive industries, each investing more than €20,500 per full-time equivalent employee in 2024.

Over the 2013–2024 period, service sector industries recorded the strongest growth in data product investment, in both absolute and relative terms – reflecting where the largest shifts in data value are occurring.

**Higher educated ICT specialists and higher educated business & administration specialists are the occupational segments that generate most data product value**
We find large differences in data intensity across occupations. While more than half of occupations spend at least 10% of their time on data product-related tasks, only a tenth spend more than 30%.

Higher educated ICT and business & administration specialists were the most data-intensive occupational segments, each generating more than €30,000 in data product value per job in 2024.

## Boosting future value of data for the Dutch economy

While data already plays a large role in the Dutch economy, we highlight three areas where future value could be unlocked: tackling labour shortages, increasing data intensity to boost economic growth, and using AI as an enabler to harness greater value from data.

**Addressing labour shortages in the most data-intensive occupations could unlock greater investment in data products and generate more value for the Dutch economy**
Labour shortages represent a key bottleneck in capturing greater value from data products. The occupations that generate the most data product value, such as highly educated ICT, business & administration, and healthcare specialists, are also those facing the tightest labour markets. Addressing these shortages could therefore unlock substantially greater value for the Dutch economy. Beyond resolving labour shortages, two additional levers could unlock further value: shifting employment from occupations with looser to tighter labour market conditions and upskilling workers into more data product-intensive roles.

**Increasing data product intensity to boost economic growth**
We find that a positive and statistically significant relationship: a 1% increase in total data product investment per worker is associated with 0.13 percentage points higher economic growth in the following year, suggesting that industries investing more intensely in data products tend to experience faster economic growth.

**AI as an enabler to harness more value from data products**
AI is transforming how occupations perform and expand their tasks, altering the value derived from data products. Depending on each occupation's data product intensity and task set, AI accelerates existing core tasks and expands task frontiers. The most data product-intensive occupations, such as ICT professionals, are disproportionately represented among AI users, reflecting their greater capacity to absorb and deploy these technologies. We find that a 1% increase in AI usage is associated with a 0.38% increase in data product intensity per worker. While current AI use is concentrated in data product-intensive roles, its transformative potential is likely to extend progressively across the occupational spectrum.

## Policy recommendations

**A better understanding of the role data products play in the economy is needed**
Data should be treated as core economic infrastructure, requiring more precise measurement of its value and contribution, particularly in national statistics.

**Work on aligning incentives for widespread access to data and data products**
First, public institutions should ensure fair and widespread access to the foundational infrastructure of the intelligence economy, including electricity, data centres, computing power, secure internet connectivity, and the software and technologies required to participate fully in an increasingly data-driven economy, particularly for firms that cannot provide these capabilities themselves.

Second, while public sector data already enables significant economic value, it does not always translate into improved productivity and service quality, particularly at the executive organisation level. By improving data integration, interoperability, uptake of data analytics, and data-driven decision-making, the public sector can substantially raise productivity, service quality, and policy effectiveness.

Third, public institutions should encourage broader sharing of data and analytical capabilities in a way that rewards first movers and preserves competitive incentives while delivering broader economic benefits.

**As much of the value of data products is enabled by people, it is necessary to address labour and skills shortages by lowering barriers to entry and upskilling the existing workforce**
Policymakers, working together with companies, should tackle structural bottlenecks in labour supply and skills availability to unlock the full potential of data products. Such measures could include lifelong learning programmes and active labour market policies aimed at upskilling the existing workforce and lowering barriers to entry for new participants in key data-driven roles.

## Recommendations for organisations

**Organisations should identify and strengthen the weakest links in their data value chain, as these ultimately determine the ceiling on productivity and innovation gains**
Too few firms, particularly small and medium-sized enterprises (SMEs), make use of data products, despite their proven potential to improve productivity, enhance innovation, and enable new business models. This gap is especially pronounced in the Netherlands, where this gap between large and small firms is the biggest across the OECD. Companies need to enable data capabilities across the entire value chain to fully unlock the benefits in terms of efficiency and innovation.

**To unlock the full value of data products, organisations must first get their 'data house' in order**
Data is the starting point. To fully capture the value of data products, companies need the full stack of supporting capabilities in place: cybersecurity, cloud computing, data storage, and data analysis skills. This also includes labour and skills, for which organisations should build internal capability through upskilling, role redesign, and active employer initiatives to boost digitalisation and AI uptake within occupations, rather than relying solely on external hiring.

Taken together, these interdependencies suggest that much like weak links in a supply chain, gaps in an organisation's data house create bottlenecks that slow overall AI and digital adoption and limit efficiency gains.

"To fully capture the value of data products, companies need the full stack of supporting capabilities in place: cybersecurity, cloud computing, data storage, and data analysis skills."

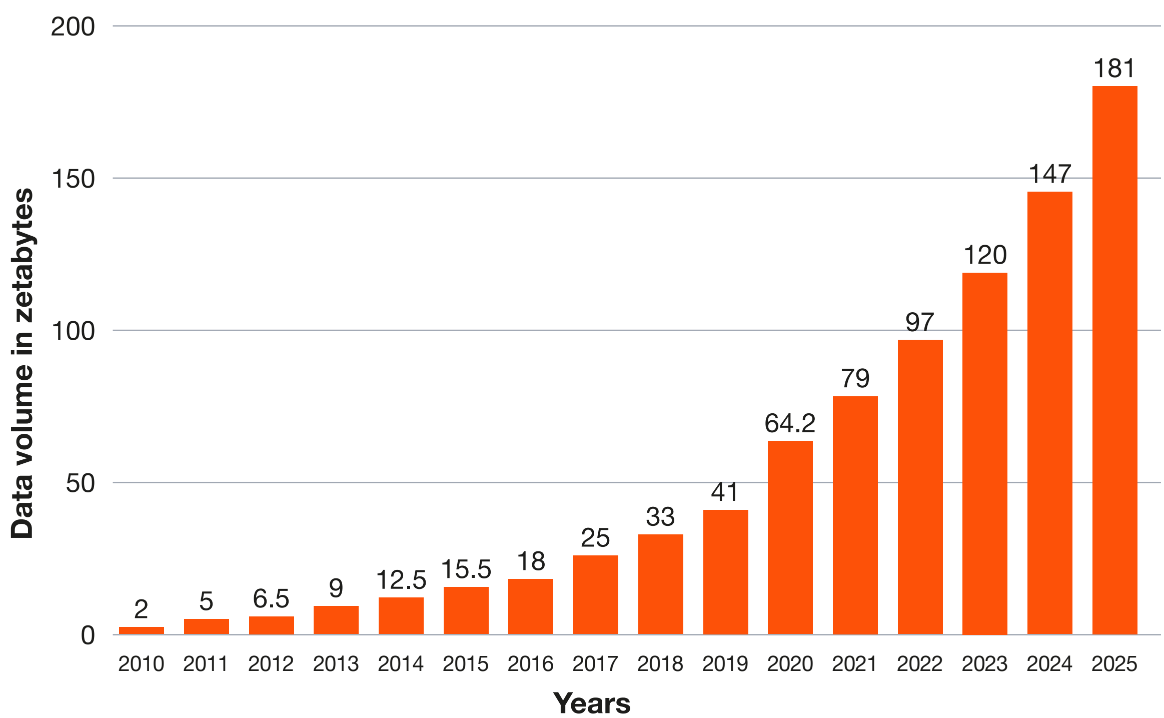
# Introduction: the various growing roles of data in the economy

## Data as a signal of value in motion and an opportunity for reinvention

The way we live and work is changing rapidly. Emerging technologies, shifting climate patterns, and evolving geopolitical forces are reshaping customer demand, unlocking new markets, and enabling innovative business models. Together, these forces are putting significant value in motion, attracting new competitors and collaborators, blurring industry boundaries, and reorganising economic activity around fundamental human needs.

A defining trend in modern economies is the rapidly growing volume and value of data. The scale of this growth is staggering, as global data volumes are estimated to have reached 181 zettabytes by the end of 2025,[1] with total data volume doubling every four years.[2]

**Figure 1: Volume of data produced and consumed worldwide (2010-2025)**




Source: Romano: Synthetic geospatial data and fake geography: A case study on the implications of AI-derived data in a data-intensive society.

This growth is driven by proliferation of data across virtually every domain: IoT devices and smart technologies; real-time data processing; digitalisation and cloud storage; AI and user-generated content; social media; enterprise and transactional data; e-commerce; scientific research; streaming; and public statistics.[3] Before, data was just information. Now it is the start of everything.

The modern data economy was born from innovations in computing, storage, and data science, which made it possible to store vast datasets cheaply and extract meaningful insights at scale. As machine learning and AI capabilities advanced, data's value surged, giving rise to the now-familiar notion that 'data is the new oil.'[4] As data becomes richer and more abundant, and our capacity to derive insights from it grows, data increasingly signals where economic value is shifting. Where high-value companies once built their competitive advantage on physical assets – factories, offices, and equipment – data has emerged as a defining asset and factor of production in the modern economy, enabling companies to reinvent their business models and unlock new capabilities and value. A key indicator of this shift is that the world's most valuable companies now derive much of their worth from data rather than physical assets, making data and its application a leading indicator of broader economic activity.

For established companies in particular, reinvention is the defining organisational response to these shifts. Research consistently shows that data-intensive companies outperform their less data-intensive counterparts across a range of key dimensions. By redesigning their business models around data and digital capabilities, companies position themselves where future value will be created rather than where it has historically resided. Those that invest in data capabilities gain predictive insight, enhance productivity, and build the agility needed to compete in rapidly changing markets.[5] By combining macro-level evidence on value flows with micro-level business model renewal, reinvention translates value in motion into concrete, measurable outcomes – driving growth, strengthening resilience, and securing long-term competitiveness.

We calculate the value of data in the Dutch economy. Before setting out the empirical approach, we examine the various roles that make data a distinctive economic factor and why placing a value on it presents particular challenges.

> "Before, data was just information. Now it is the start of everything."

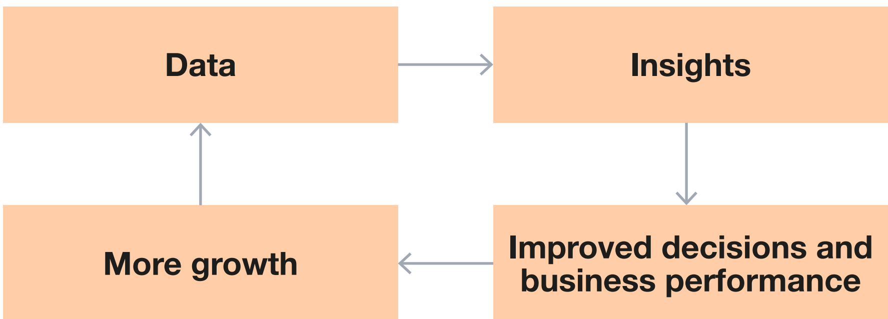
## Data as a special economic factor

**Data as an input into predictions to reduce uncertainty, make better decisions, and improve processes**

At its core, data is digitised information. Claude Shannon, the pioneer of information theory, mathematically defined information as the reduction in uncertainty. In other words, information is the signal that improves understanding and enables better decision-making. The fact that uncertainty is costly is precisely why organisations collect data – digitally stored information – and place such value on it.[6]

Combined with other economic inputs – labour, capital (including computer equipment and software), and complementary resources – data can be organised in databases and transformed into actionable insights through data science. Such insights make the future more predictable, leading to better decision-making and improved business performance across every dimension – from finding customers and suppliers to managing inventory and refining products. Companies that consistently make better decisions gain a measurable competitive advantage, raising revenue, cutting costs, and generating more data in the process. This is the feedback loop of data-intensive production:[7] a self-reinforcing cycle that puts value into motion (Figure 2).

**Figure 2: The feedback loop of data-intensive production: how companies capture value from data**




**Data as a by-product of economic activity**

A further distinctive feature of data is that vast amounts arise naturally as a by-product of economic activity and everyday life. Every transaction, commute, website visit, or app interaction leaves a digital trace – and unlike physical assets such as machinery, buildings, or natural resources, data does not decay or deplete. While the utility of older data may diminish over time, data as a whole accumulates rather than erodes. Multiplied across billions of people and trillions of daily interactions, the resulting volume is staggeringly large. These digital footprints, once collected and stored, serve as a primary source of business intelligence.[8]

**Data as a non-exhaustible good**
Data fundamentally differs from physical goods because of its non-rival nature: one party can use the same data at the same time as another party. Alice can sell data to Bob and keep a copy of the same data to herself. In this respect, data more closely resembles ideas or knowledge – resources that can be used by many people concurrently without being exhausted – than it does human capital, which is acquired through experience and learning and remains tied to the individual.[9]

The following section addresses the conceptual challenges inherent in valuing data before setting out the approach that we use to quantify the value of data for the Dutch economy.

## Challenges in valuing data

**Most data is not traded, revealing its value only implicitly**
The primary challenge in directly capturing the value of data is that it functions as an input in production – much like labour and capital – whose value is embedded in the prices of the goods and services it helps produce. As a result, most data is not directly traded and therefore has no observable market price.

There are, however, several indicators that point to the value of data. First, practically every modern company and institution uses data in some capacity – whether as a strategic asset or an operational resource – suggesting that organisations across the board find data valuable in pursuing their goals. For example, American companies spent over $19bn in 2018 acquiring and analysing consumer data.[10] In addition, according to Tucker and Neuman (2020), collecting and selling customer data generates as much as $200bn in value.[11] Yet these visible estimates and investments represent only the tip of the iceberg – the true total value of data runs far deeper.

A second signal comes from the substantial market valuations of data-intensive companies – reflecting widespread recognition that the value of data is both enormous and growing. In many cases, leading technology companies provide goods and services free of charge, asking only for users' data and platform engagement in return. These exchanges are real and economically significant, yet largely invisible in the official statistics upon which we rely to understand our economy.[12] Two studies underscore the scale of this hidden value. One values ten free digital goods at $2.5tn in consumer welfare across 13 countries – equivalent to 5.5% of US GDP, yet entirely excluded from official measures.[13] While another estimates that free internet goods generate $38bn annually in US consumer surplus, equivalent to 0.3% of GDP.[14] Together, these findings suggest that conventional economic statistics are materially understating the true economic value of data.

### The value of data depends on the user, use case, context and complementary technologies

Beyond its limited tradability, the value of data is largely dependent on who uses it, in what context, and for what purpose. Four economic concepts illustrate this: increasing returns to scale, diminishing marginal returns, market context, and complementary technologies.

#### Increasing returns to scale

First, data exhibits increasing returns to scale – as data volumes grow, their analytical and predictive value tends to rise disproportionately relative to other inputs. Greater data availability enables companies to operate more efficiently, offer more specialised products and services, and lower average costs. Companies that merge or combine datasets can amplify this effect further.[15] As a result, even two companies operating in identical markets may derive fundamentally different value from the same dataset, with the better-resourced firm able to extract significantly more insights and, in turn, gain a lasting competitive advantage.

#### Diminishing marginal returns

Like other innovations, data benefits from returns to scale because information is expensive to collect but cheap to replicate – once gathered, it can be copied at near-zero marginal cost. Sharing data, however, introduces important strategic considerations. When competing companies access the same dataset, the informational advantage it confers is diluted, reducing its value to each. Data can derive considerable value from exclusivity and scarcity; once that exclusivity is lost, so too is a significant portion of its competitive worth.

Second, data is subject to diminishing marginal returns. Early data can substantially improve predictions and processes – even feedback from a handful of customers, for instance, can yield meaningful improvements to services. However, each additional unit of similar or repetitive data yields progressively less insight; the thousandth similar data point adds far less value than the first. This has an important competitive implication: smaller companies with limited datasets often gain proportionally more from adding data than larger companies already operating at scale.[16] Equally, a company that accumulates large volumes of data without the analytical capacity to organise and interpret it will capture only a fraction of its potential value.

#### Market context

Third, the value of data is highly context-dependent. Detailed weather data for the Netherlands may be largely irrelevant to farmers in South America yet highly valuable to bar and restaurant owners in the Netherlands planning to open their terraces for the summer season – the same data, differently situated, carries entirely different worth.

**Complementary technologies**
Fourth, advances in complementary technologies can unlock value in data that was previously unusable or overlooked. AI is the most prominent example: vast quantities of video footage and publicly available text have found new value as primary training material for large-scale AI models.

Taken together, these factors illustrate why establishing the objective value of data is inherently difficult – it depends critically on the user, the use case, market context, and complementary technologies. Importantly, the fact that data's value is not intrinsic but determined by external actors and factors also means that data can be exploited for harmful purposes, such as manipulation, coercion, breaches of privacy, or warfare. Data on its own, therefore, has limited value; its true value arises only when it is organised into databases, analysed through data science, and deployed responsibly and ethically. This is the foundation upon which our approach to quantifying the value of data in the Dutch economy is built.

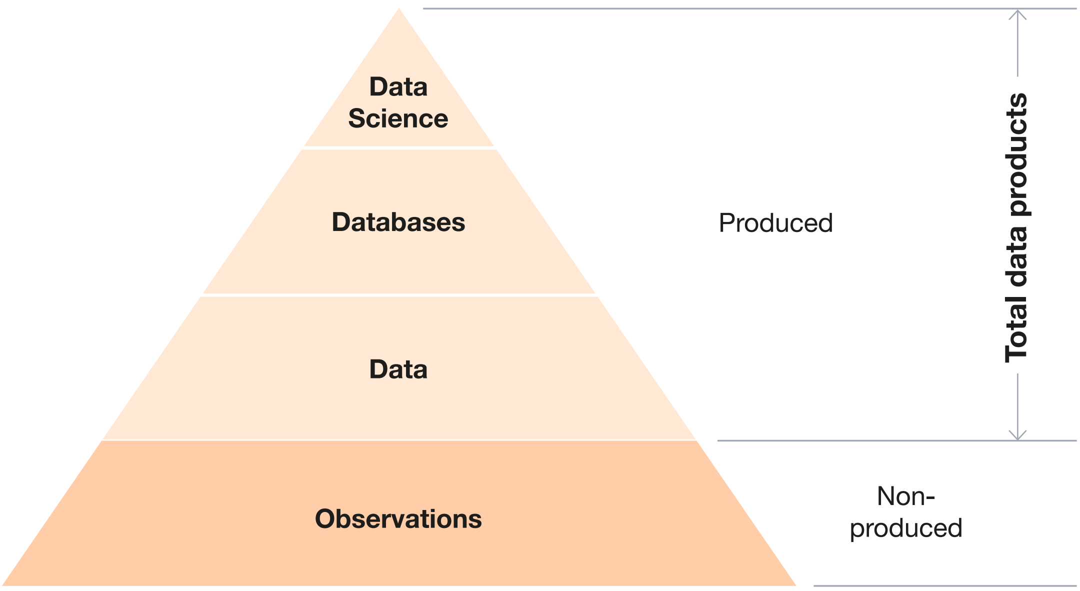
# Methodology: empirical approach to capture the value of data for the Dutch economy

## Information value chain and cycle

As established earlier, data is, at its core, digitally stored information. This definition does not imply that everything digitised is necessarily data. A song converted into digital format – or indeed recorded in digital format from the outset – remains a song; it is not redefined as data merely because a digital representation exists. The process of digitising that song, however, may well involve the use of data, databases, and data science.[17]

To illustrate how we attempt to quantify the value of total data products, it helps to think of the information value chain (Figure 3).[18]

**Figure 3: The information value chain**




Source: Statistics Canada.

“As established earlier, data is, at its core, digitally stored information. This definition does not imply that everything digitised is necessarily data.”

The information value chain comprises four distinct and separable stages: **observations**, **data**, **databases**, and **data science**.

**Observations: raw output**

At the base of the chain are observations – the raw outputs of human activity, physical phenomena, and the environment. Continuous, fleeting, and often intangible, they range from temperature readings to behavioural patterns such as commuting habits and daily routines. Crucially, observations do not require human perception to exist; the world generates them constantly, whether anyone is watching or not. Most go unrecorded, yet collectively they represent the full breadth of activity occurring in the world at any given moment.[19]

**Data: capturing and recording information**

Observations worth preserving are recorded and collected – their conversion into digital form constitutes the second layer of the value chain: data. This process varies considerably in cost and complexity. Where ledgers and handwritten records once served this purpose, keyboards, sensors, and electronic storage devices now do so with far greater efficiency and at far greater scale. Yet even in an increasingly automated world, data collection continues to require meaningful human input.[20]

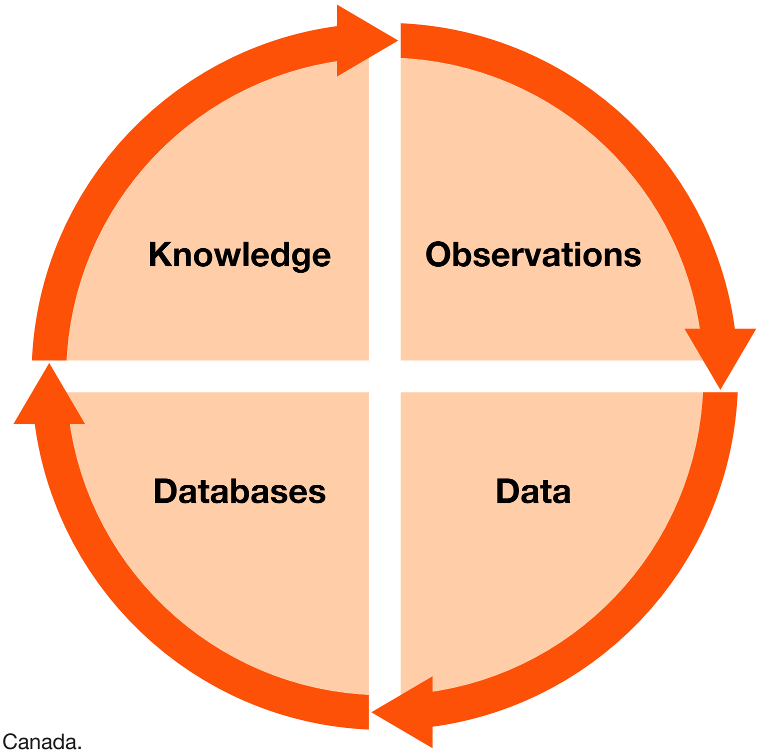
**Databases: structuring and organising data**

Data does not arise spontaneously – it requires deliberate action. Someone must identify what is worth recording and establish the systems needed to capture and store it. Once collected, raw data can yield insights in its unstructured form through modern analytical techniques, but organising and structuring it creates additional value. High-quality, domain-specific, and well-labelled data serves as a foundation for the modern intelligence economy.[21] This is where databases come in: organised repositories designed to store data in a way that makes it readily retrievable and analysable. The boundary between raw data and structured databases, however, is not always clearly defined in practice.[22]

**Data science: generating insights from data**
The fourth tier of the information value chain is the extraction of insights and knowledge from data. While each individual data point carries some inherent information, the true power lies in analysing large volumes collectively – patterns, relationships, and insights that are invisible at the individual level emerge only when data is examined as a whole. It is this process of transforming raw, organised data into actionable knowledge that this study defines as data science. Realising this value requires a combination of analytical techniques, specialised skills and expertise from people, technology, software, and equipment.

**Data feedback loop: the information value chain becomes an information cycle**
These steps describe a single iteration of the information value chain. In practice, the process is cyclical: observations become data, data is stored in databases, insights are extracted through data science, and those insights feed back into economic activity that generates new observations – some of which become data once more, setting the cycle in motion again. This cycle does not operate in isolation; it continuously interacts with countless others like it, collectively driving the broader data economy (Figure 4).[23]

**Figure 4: Information cycle**




Source: Statistics Canada.

## Calculating the total value of data products for the Dutch economy

Once the value of data, databases, and data science is aggregated, this study refers to the combined total as the ‘total value of data products’. To estimate this, a fundamental question must first be answered: what economic role do data and other data products play? Is data a capital asset – a store of value deployed repeatedly in the production of goods and services? An intermediate input, fully consumed in a single production cycle? Or a final consumption good, used directly by households, governments, and non-profit organisations?

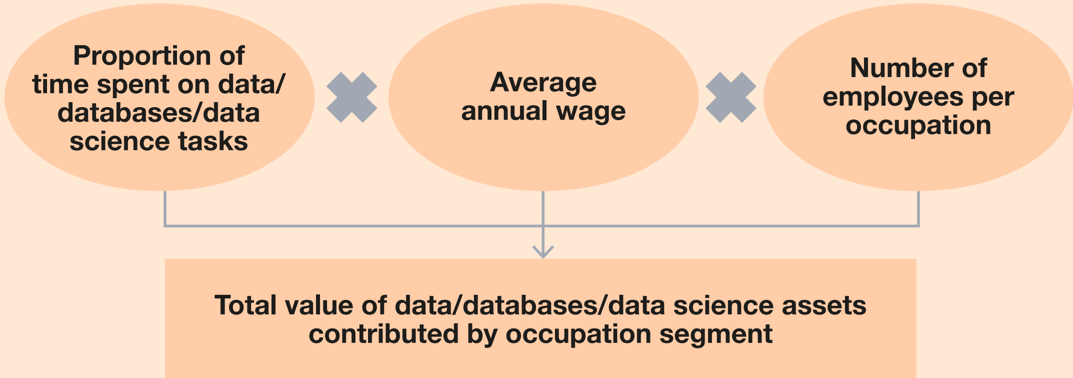
While data can plausibly function as either an intermediate or final consumption good, these roles are likely modest and interlinked to its primary and most economically significant function: that of a capital asset.[24] According to previous research by Statistics Canada and Statistics Netherlands, this seems to be the most intuitive approach, motivated by the sum-of-costs methodology. This methodology determines economic value by totalling all expenditures – such as labour, materials, and equipment – required to produce or acquire an asset.

The reasoning is straightforward: zettabytes of data already exist, and this volume is growing continuously. If such quantities of data are being produced, there must be producers employing labour and capital to capture it, organise it into databases, and extract insights from it through data science. Even where the resulting assets are never sold on the market, they nonetheless generate real economic value for their producers.

This study applies the same logic, quantifying the value of each category of data product based on the wages earned by employees involved in the production of such assets. Following the methodology of Statistics Canada and Statistics Netherlands, this estimate also incorporates an additional capital expenditure of 60% of labour costs and a 3% capital surcharge on top of the non-labour costs.[25, 26]

Figure 5 shows conceptually that there are three channels through which the total value of data, databases, and data science assets can be generated – the proportion of time spent on tasks related to data, databases, and data science, multiplied by the average wage, and the number of employees per occupation.

**Figure 5: Conceptual methodology for calculating the total value of data products**

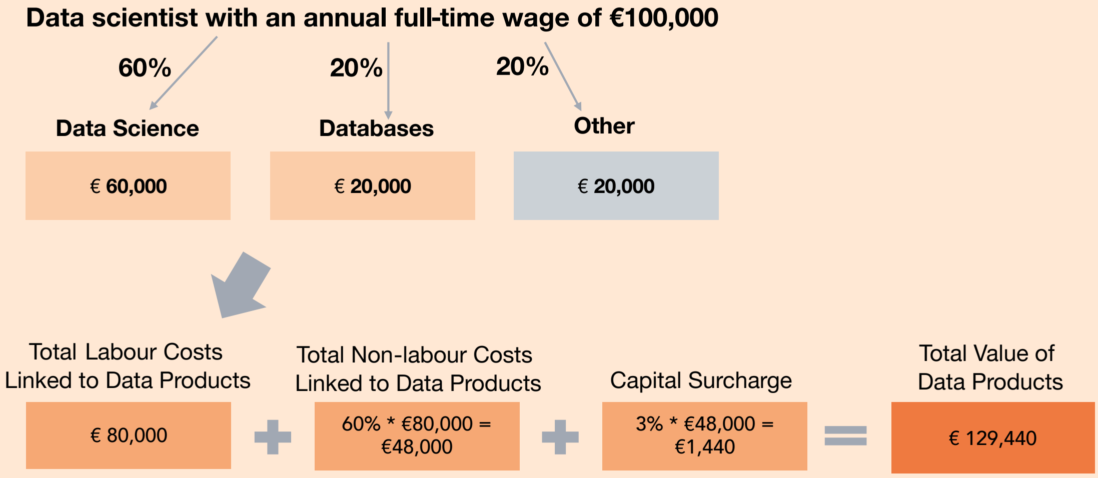



We start our analysis by looking at the labour market at a granular level, covering 923 occupations and using large language models to assess whether their associated tasks relate to the production of data, databases, and data science products and what proportion of time is devoted to each. Occupation-level analysis is the foundation of our analysis that further shows results across occupational segments, industries, and the whole Dutch economy. To balance analytical granularity with presentability, these 923 occupations have been aggregated into 41 occupational segments (or 112 occupation groups) used to classify the Dutch labour market.

To illustrate this, for example, the occupational segment 'higher-educated ICT specialists' encompasses a wide range of occupations, including software and application developers – such as systems analysts, software developers, web and multimedia developers, and application programmers – and database and network specialists, such as data warehousing specialists, computer network support specialists, network and computer systems administrators, and digital forensics analysts. Even within a single occupational segment such as this, there is considerable variation in the types of roles and the extent to which they involve working with data, databases, and data science products, as well as the wages received. Further variation can arise within the same role depending on the company and industry in question. The results presented in this study therefore represent averages across all roles within each occupational segment.

Next, Figure 6 illustrates example of calculations for one occupation – data scientist.

**Figure 6: Illustrative example of the annual data product value generated by a data scientist**




Consider a company that employs a data scientist on an annual salary of €100,000. If that person spends 60% of their time on tasks classifiable as data science activities and 20% on tasks related to databases, this implies the production of data science and database products worth €60,000 and €20,000, respectively. These estimates are based on wages paid, proportional to the time spent on each activity, and together constitute the total labour cost attributed to the production of each type of data product.

To this labour cost, this study adds non-labour expenditures – software, hardware, and related capital costs – estimated at 60% of total labour costs (€80,000), yielding an additional €48,000. In addition, a 3% capital surcharge, estimated from the non-labour costs, is included. Hence, in total, this figure produces a final estimated asset value of €129,440.

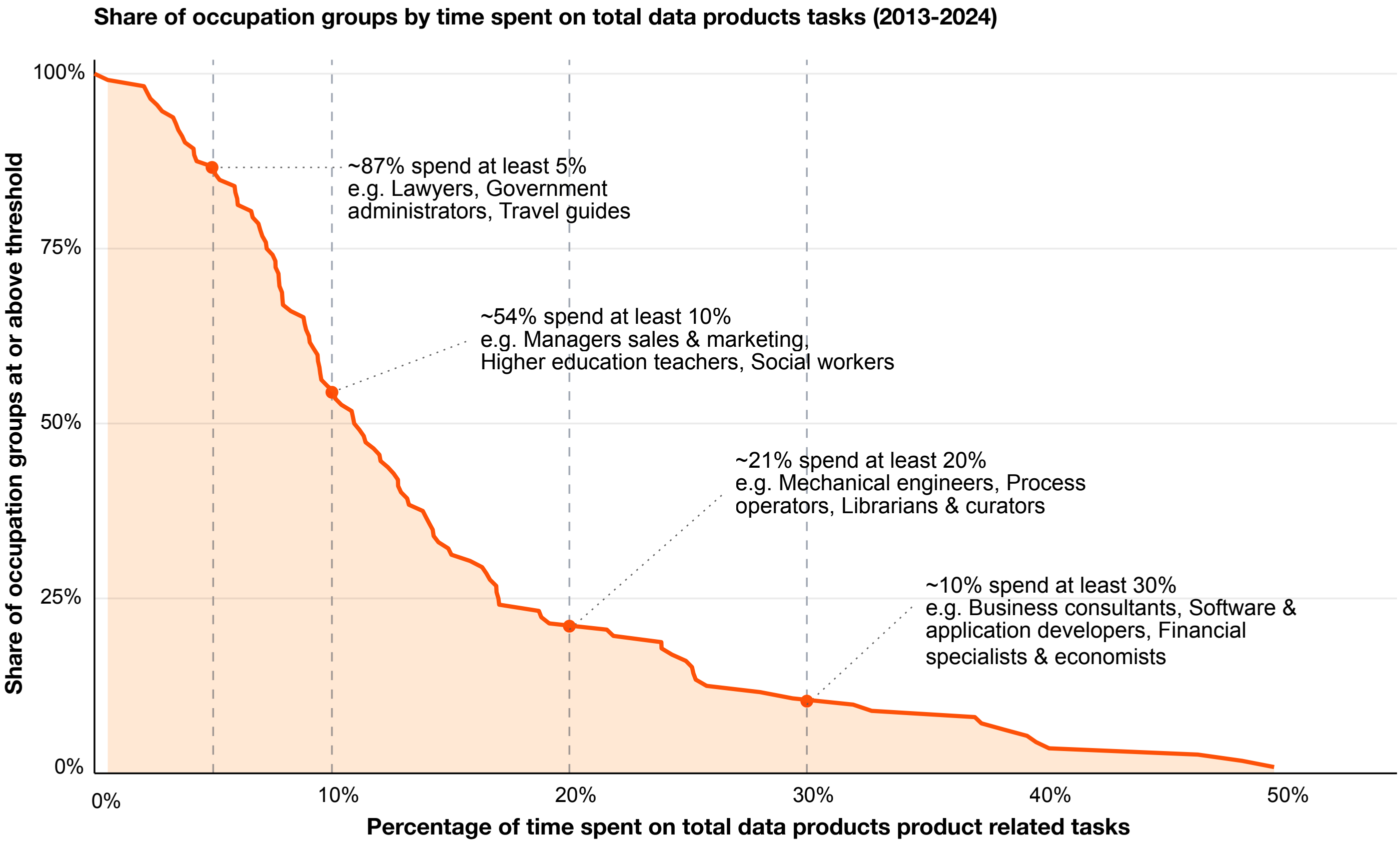
This study applies the same approach across all occupational segments in the Dutch labour market over the 2013-2024 period, calculating the total value of data, database, and data science products generated by each occupation segment. Drawing on industry-level employment data, it also captures the value of these data products across industries and for the Dutch economy as a whole. The full methodology is set out in the Appendix on page 46.

# Results

## Total value contribution to data products across occupations

**Occupations that spend most of their time on producing data products**
This section begins by examining the proportion of working time that different occupations devote to tasks related to data, databases, and data science products. Figure 7 illustrates the distribution of occupations at the most granular level available for the Netherlands, showing the proportion of working time devoted to data production-related tasks across the economy.

**Figure 7: Ten percent of occupations spend at least 30% of their time on such data product tasks**

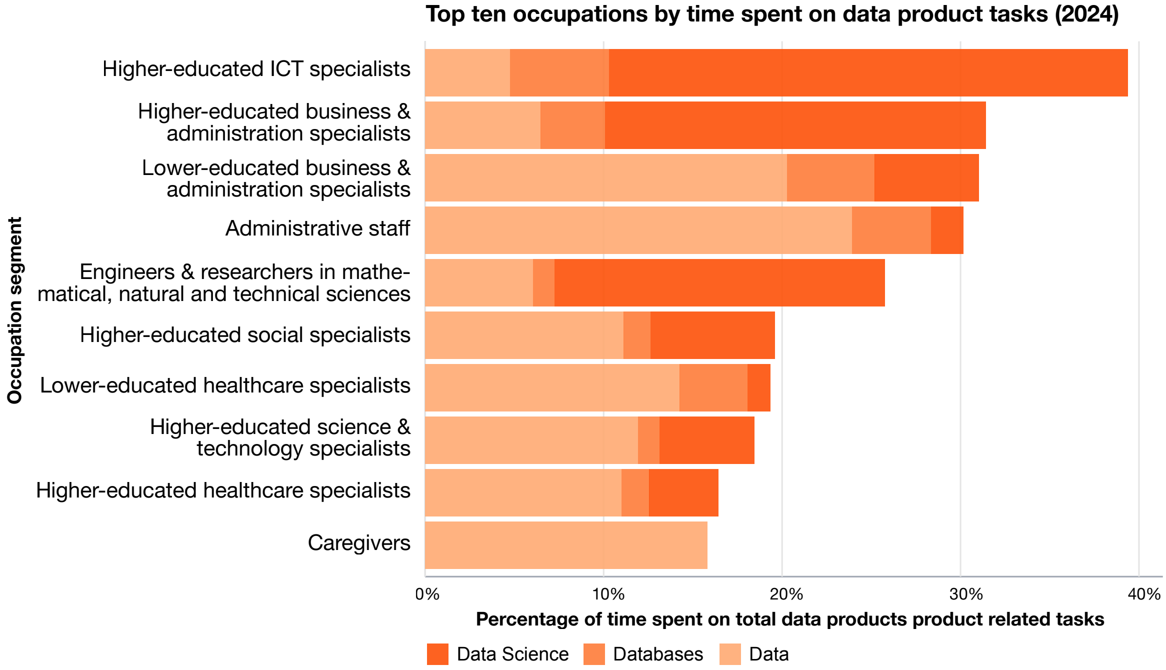



Sources: O*NET, CBS, ROA, PwC analysis.

There is considerable variability across occupations. For instance, 54% of occupations – including sales and marketing managers, higher education teachers, and social workers – devote at least 10% of their working time to data product-related tasks. At the other end of the spectrum, only 10% of occupations – among them financial specialists & economists, software & application developers, and business consultants – devote at least 30% of their working time to such tasks.

Figure 8 illustrates the ten occupational segments that devote the greatest proportion of their working time to data product-related tasks in 2024, presented at a more aggregated level than the occupation groups shown in Figure 7.

**Figure 8: Higher educated ICT specialists spent relatively the largest share of their time in 2024 on data, database, and data science tasks**




Sources: O*NET, CBS, ROA, PwC analysis.

Higher-educated ICT specialists and both higher- and lower-educated business & administration specialists are the most data-intensive occupational segments, each devoting more than 30% of their working time to tasks related to data products.

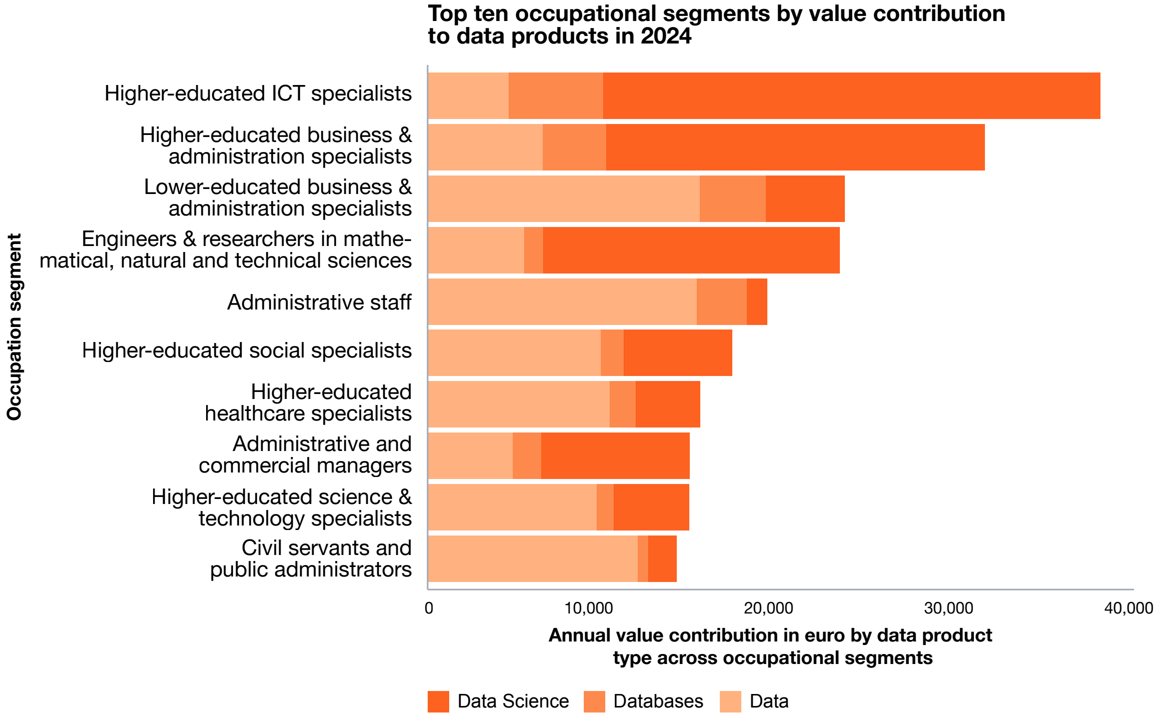
Notably, there are pronounced differences in which stages of the data production chain different occupations tend to focus on. Occupations such as administrative staff, lower-educated business & administration specialists, and caregivers devote a comparatively larger share of their time to data-related tasks, whereas higher-educated ICT and business & administration specialists – along with engineers & researchers in the mathematical, natural, and technical sciences – are far more heavily involved in data science work.

Hence, those occupations that devote the greatest share of their time to data product-related tasks contribute disproportionately to the overall creation of data products. These findings are consistent with a comparable analysis of data-intensive occupations in the United Kingdom and Canada, which found that roles that fit within the 'higher-educated ICT specialists' segment exhibit the highest data intensity.[27]

**Total value of data products contributed by occupation segment**

Having established the proportion of time that different occupational segments devote to data, database, and data science tasks, we apply employment and wage data at the occupational segment level to calculate the total value contribution of each segment to data products in 2024. These findings are presented in Figure 9.

### Figure 9: Higher-educated ICT specialists generated the highest value in terms of data products in 2024 out of all occupational segments




Sources: O*NET, ROA, CBS, Statistics Canada, PwC analysis.

Higher-educated ICT specialists and higher-educated business & administration specialists both generated more than €30,000 in data products in 2024. Furthermore, there are differences in which types of data products the different occupational segments focus on. For example, higher-educated ICT and business & administration specialists, and engineers & researchers in mathematical, natural, and technical sciences generate the bulk of their value contribution through data science products, while lower-educated business & administration specialists and administrative staff contribute predominantly through data products. Databases account for a relatively small share of total value production across all occupational segments.

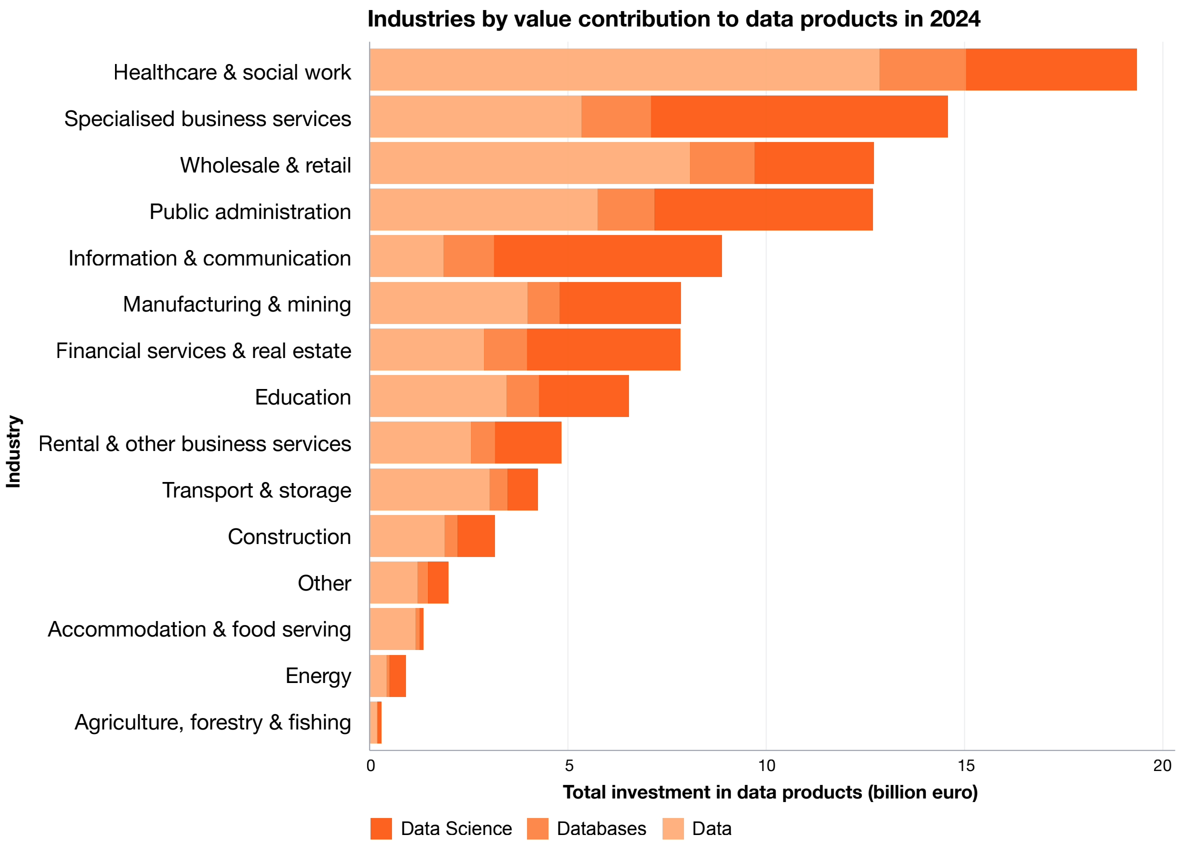
# Total value contribution to data products across industries

**Total investments in 2024 across industries**

Having examined the total value of data product production from the perspective of occupational segments, now we look at the value contribution to data products across industries. The analysis follows the Statistics Netherlands industry classification.[28]

Figure 10 shows the average annual investment in the production of total data products by industry.

**Figure 10: Healthcare, specialised business services, and wholesale & retail, lead in total data product value generation**




Sources: O*NET, ROA, CBS, Statistics Canada, PwC analysis.

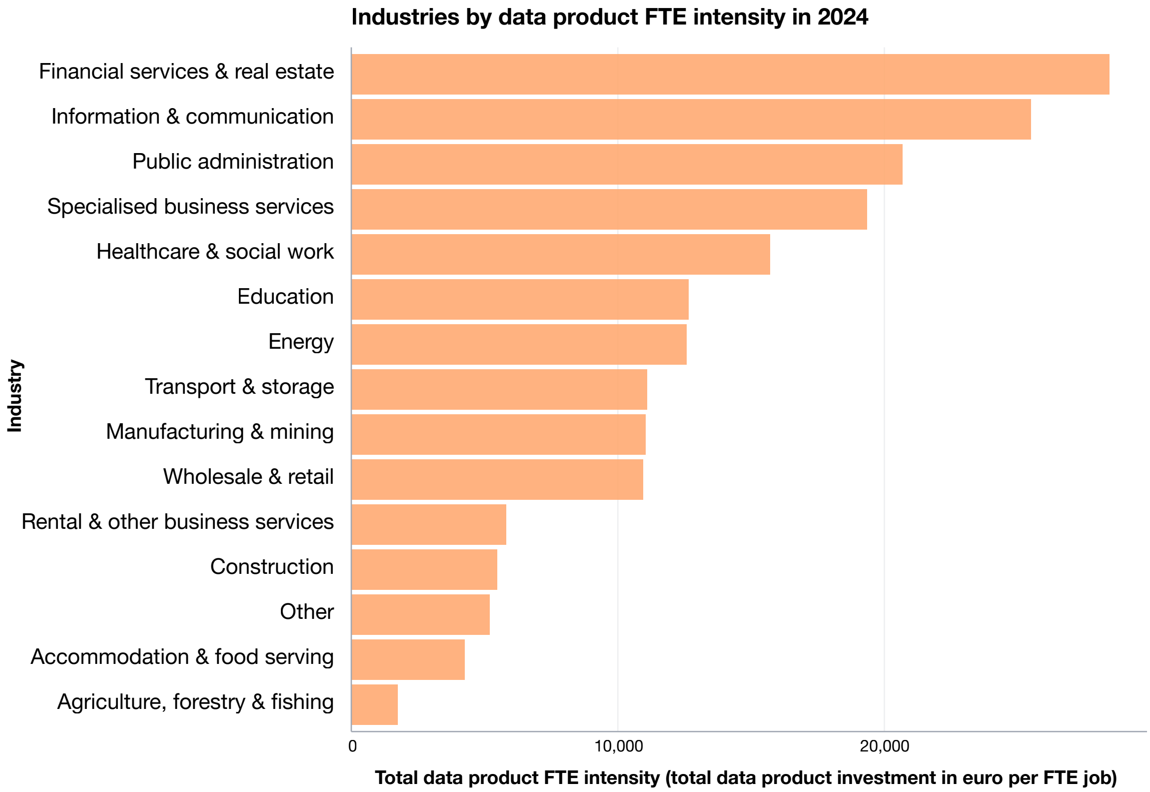
In 2024, the healthcare and social work industry generated approximately €19.5bn in total data product value, while it was approximately €14.7bn for specialised business services, and €12.8bn for wholesale and retail.

Industries such as specialised business services, information & communication, and financial services & real estate invest relatively more heavily in data science products, while healthcare & social work and wholesale & retail invest comparatively more in data. Across all industries, investment in databases accounts for the smallest share of total data product value. As with occupational segments, there are substantial differences across industries in the value of total data products generated.

### Total data product intensity in 2024

While absolute spending figures are informative, they obscure differences in the relative size of industries. Therefore, we adjust for industry size by examining total data investment per full-time equivalent (FTE) employee to capture data product intensity across industries in Figure 11 (in the Appendix on page 51 we also include a measure as a share of gross value added (GVA)).

### Figure 11: Financial services & real estate, information & communication, and public administration led in per FTE data product intensity in 2024



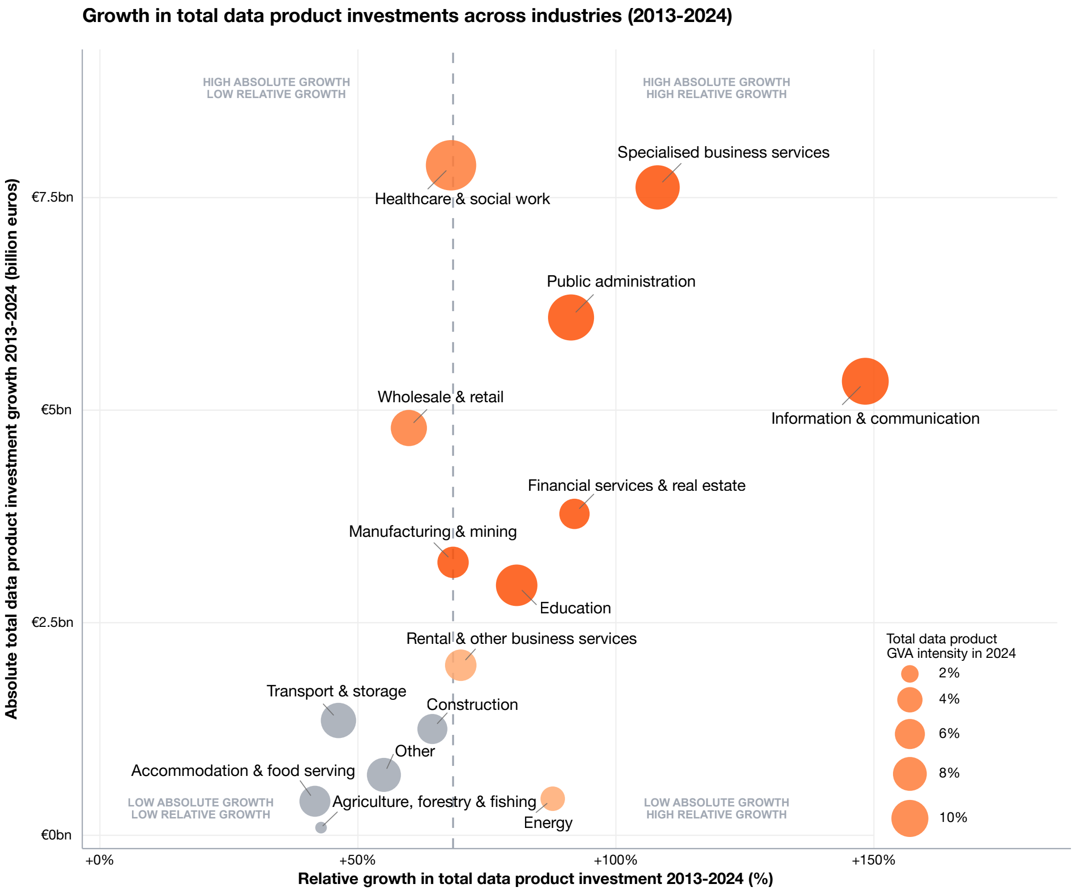

Sources: O*NET, ROA, CBS, Statistics Canada, PwC analysis.

In this case, financial services & real estate, information & communication, and public administration led in data product FTE intensity in 2024. For financial services & real estate, the total data product investment per FTE was €28.5k; for information & communication, it was €25.5k; and for public administration, it was €20.7k.

### Growth in total data product investments across industries 2013-2024

Next, we assess how total data product investments have grown over the 2013–2024 period, both in absolute terms and in relative terms indexed to each industry's investment in 2013. The available data does not permit a direct comparison of how the time devoted to data, database, and data science tasks has changed over this period, as only a single measure can be estimated for the entire period. It is possible, however, to track changes in the number of jobs and wage levels across occupational segments and industries, which together capture the value in motion, driven by the total data product value shifts across the economy over time. Hence, Figure 12 shows the growth in total data product investments from 2013 to 2024.

**Figure 12: Services sector industries have grown their total data product investments the most both in absolute and relative terms**




Sources: O*NET, ROA, CBS, Statistics Canada, PwC analysis.
The dashed line is the median absolute or relative growth out of all industries.

It makes clear that the highest-growing industries in terms of total data product growth in both absolute and relative terms are from the services sector, such as specialised business services, information & communication, and public administration. These are also among the industries that had the highest total data product intensity in 2024.
In the next section, we aggregate the outcomes one step further, looking at the total data product value for the Dutch economy over time.

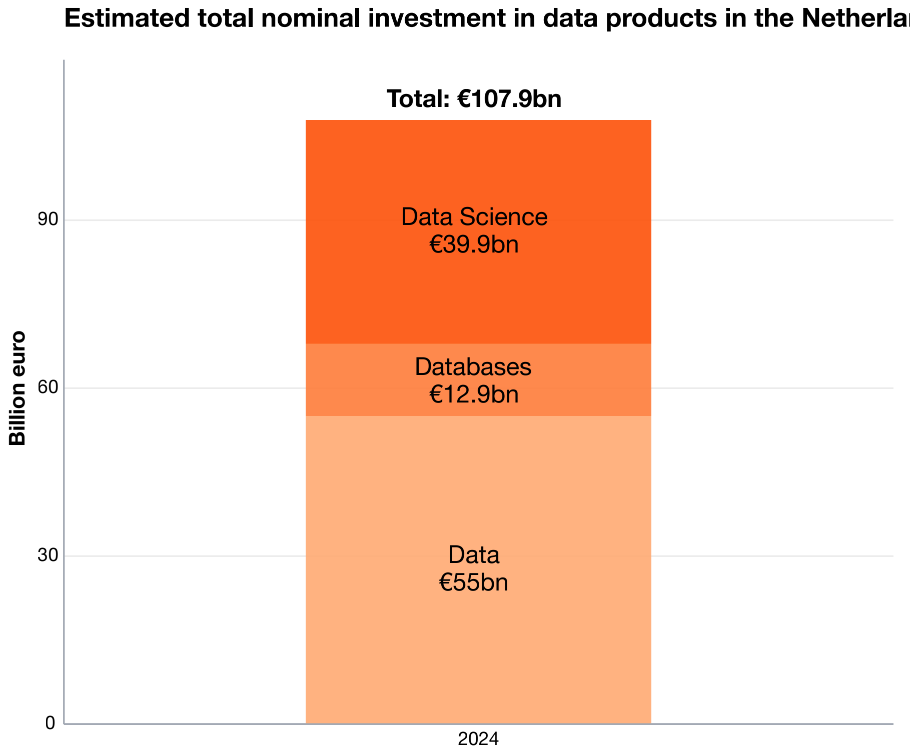
“The highest growing industries in terms of total data product growth in both absolute and relative terms are from the services sector.”

## Total data product value for the Dutch economy

**Annual investments**

We start this section by showing the annual total data product investment value for the whole Dutch economy in 2024 in Figure 13.

**Figure 13: Total data product investment amounted close to €108bn in 2024**

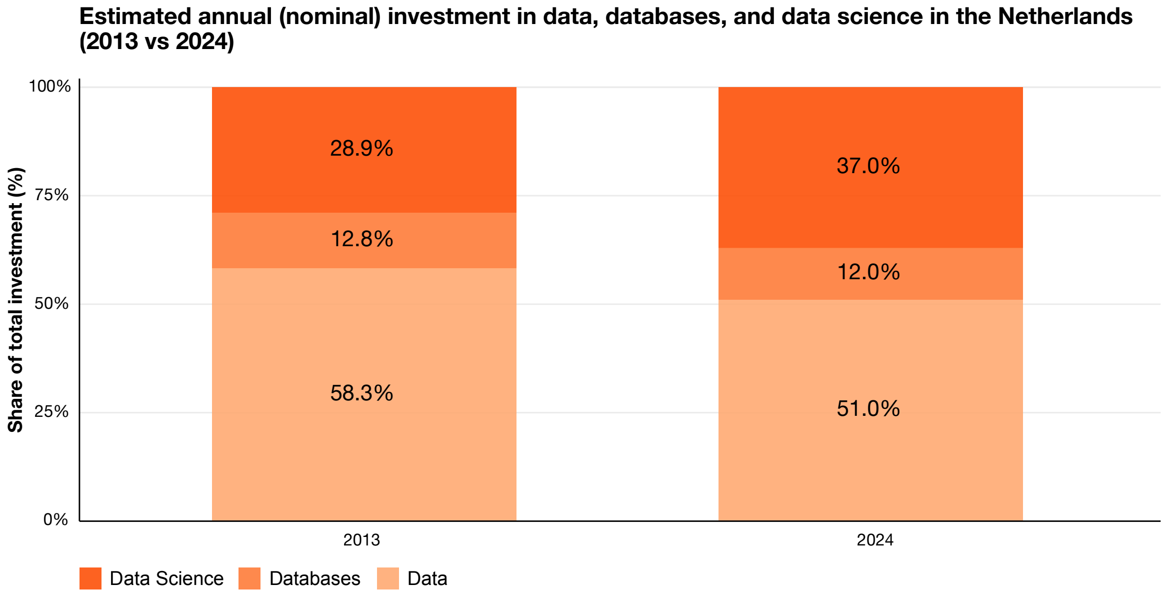



Sources: O*NET, CBS, Statistics Canada, PwC analysis.

In total, €107.9bn was invested in total data products, with €55bn in data, €12.9bn in databases, and €39.9bn in data science.

Figure 14 contrasts the relative investment shares by different data products over time.

**Figure 14: Data science investment share out of all data products has risen since 2013**

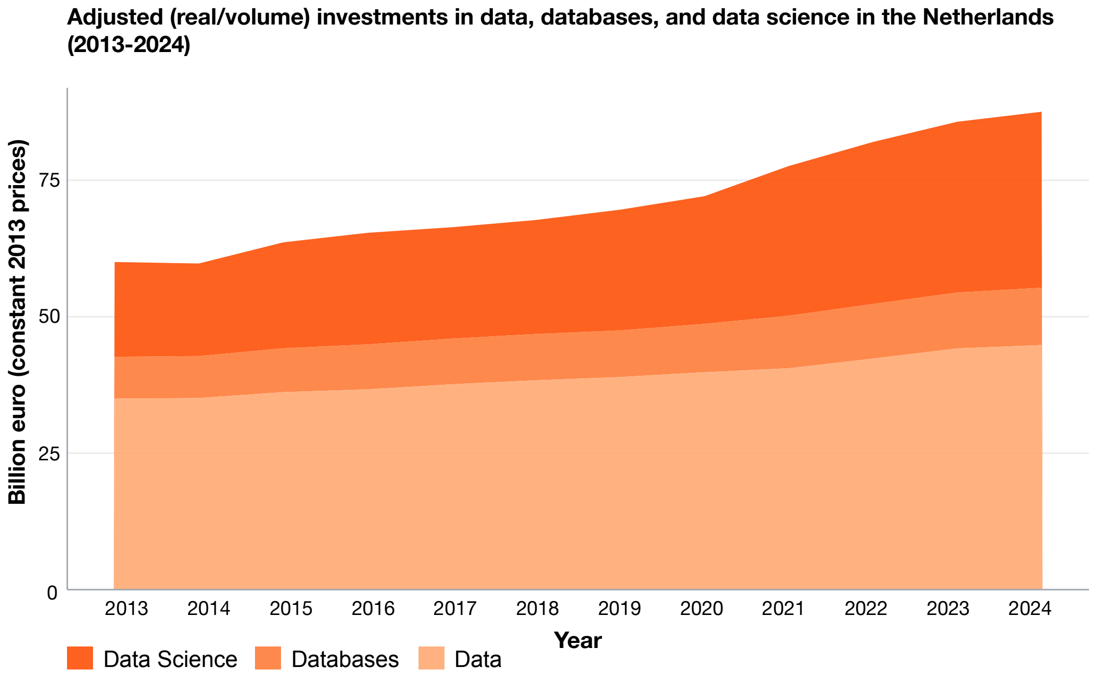



Sources: O*NET, CBS, Statistics Canada, PwC analysis.

Based on the composition of total data products in 2013, the bulk of the investment was in data (58.3%), followed by data science (28.9%) and databases (12.8%). In 2024, data took up the largest investment share, with its relative share (51.0%) falling, while for databases it decreased slightly (12.0%), and for data science, it increased substantially (37.0%). Hence, data science had the biggest improvement in investment shares.

In Figure 15, we compare the investment growth in different types of data products over time, adjusting for price effects to capture real or volume growth in investments using constant 2013 prices.

**Figure 15: Even adjusting for price effects, data science investments have grown more than investments in the other two data product types over the 2013-2024 period**

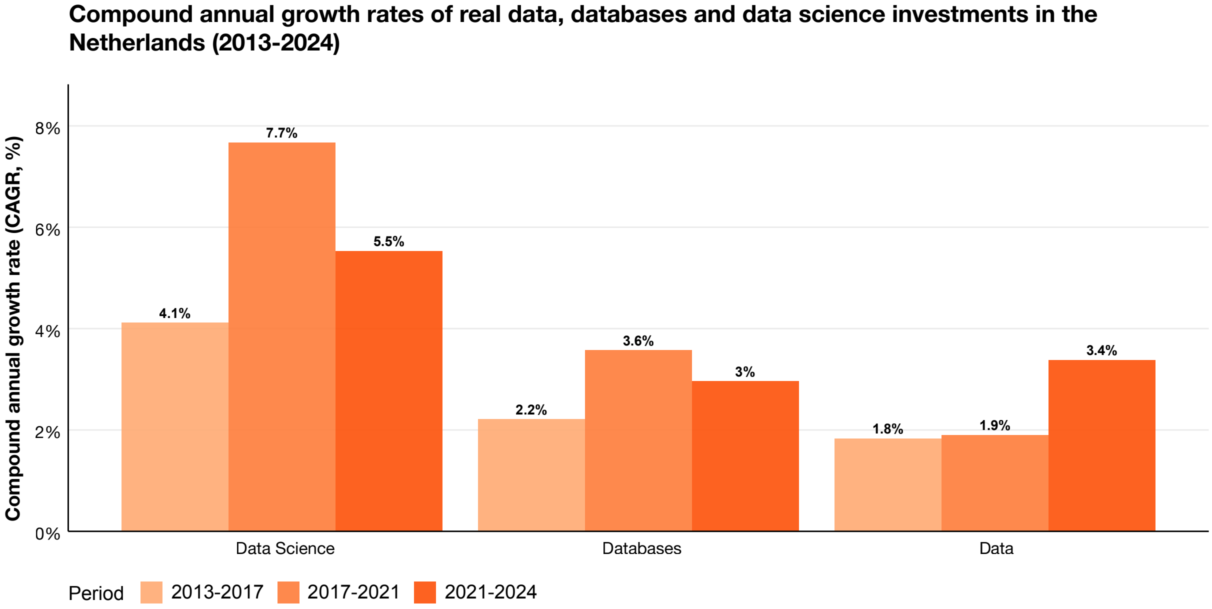



Sources: O*NET, CBS, Statistics Canada, PwC analysis.

It shows that the total data product investment amount grew from €60.0bn in 2013 to €87.6bn in 2024 – an increase of 45.9%. Comparing the growth rates of each data product type, annual investment volume for data increased from €35bn in 2013 to €44.8bn in 2024 (28% increase), for databases from €7.7bn in 2013 to €10.5bn in 2024 (37% increase), and for data science from €17.4bn in 2013 to €32.2bn in 2024 (86% increase). This indicates that data science products had the largest cumulative growth over the 2013-2024 period.

Finally, we compare the average growth rates of each asset class over three time periods: 2013-2017, 2017-2021, and 2021-2024. Figure 16 shows the compound annual growth rates (CAGR) over the three continuous time periods.

**Figure 16: Data science investments have achieved compound annual growth rates of at least 5.5% since 2017**



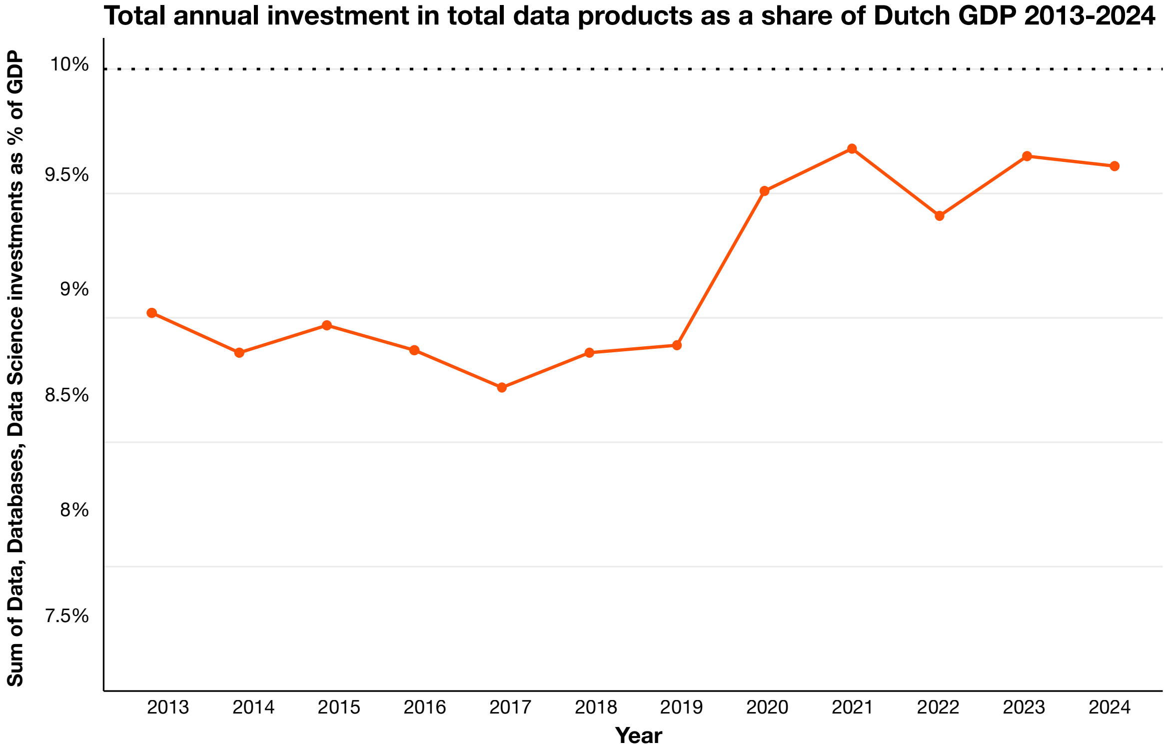

Sources: O*NET, CBS, Statistics Canada, PwC analysis.

Already in the 2013-2017 period, data science recorded the highest CAGR among the three data product types (4.1%), ahead of databases (2.2%) and data (1.8%). Growth accelerated in the 2017-2021 period across all asset classes except data (1.9%), with data science growing at 7.7%, and databases at 3.6%. In the most recent 2021-2024 period, growth rates remained elevated but moderated slightly, with data science having a CAGR of 5.5%, followed by databases (3%) and data (3.4%). Hence, while investments in the production of all three asset classes have continued to grow, data science has been the highest-growing data product type, sustaining a CAGR of at least 5.5% since 2017.

### Total data product value as a share of GDP

To put values in context, we compare the estimated amounts as a share of GDP in Figure 17. The total value of data products amounted to 9.6% of Dutch GDP in 2024, up from 9.0% in 2013, representing a meaningful increase in the contribution of total data products to the Dutch economy over this period.

**Figure 17: Total data product value GDP share was to 9.6% in 2024**




Sources: O*NET, CBS, Statistics Canada, PwC analysis.

Furthermore, we compare our estimates against findings from other studies. Statistics Canada's approach, which forms the basis of the methodology adopted here, included only 12 occupations in its calculations for 2018 – excluding highly relevant occupations such as software developers and data scientists – and found that total data product investment as a share of GDP amounted to 1.8%.[29]

Statistics Netherlands similarly adopted an approach based on the Statistics Canada methodology, including ten relevant occupational groups but excluding occupations such as software developers and data scientists. It found that in 2017 the total value of data products as a share of Dutch GDP was between 2.2% and 2.7%. Furthermore, Statistics Netherlands excluded all non-commercial sector occupations, which account for a substantial share of investments in our analysis.[30]

Additionally, another study estimated for six major European countries (France, Germany, Italy, Spain, and the United Kingdom) that data and data intelligence-related investments consisted of 5% to 6.5% of the market sector GVA in 2010-2018.[31]

The present study's estimate of 8.7% of GDP in 2017 is therefore broadly consistent with these findings, given that it incorporates a more comprehensive range of industries and occupations alongside a different methodology – specifically, analysis based on O*NET data using large language models rather than surveys – in estimating the proportion of working time devoted to the production of data, database, and data science products.

**Accumulated stock of total data products**

In the previous section we focused on annual investments. However, as with other types of assets, annual investments accumulate. The stock represents the accumulated value of past investments that are still economically relevant today.

To move from flows to stocks, we apply the perpetual inventory method (PIM).[32] This method adds new investments each year to the existing stock, while accounting for economic depreciation and obsolescence of older data products. The result is an estimate of the stock of total data products, capturing both newly created and previously accumulated total data products. This shows that data behaves like capital: its economic value builds up over time and forms an increasingly important productive base of the economy.

The PIM method calculates the total gross stock value of annual real investments. Then, because some data products are consumed in the production of new data products and they lose value over time, the total gross stock value is adjusted,[33] based on each data product type's useful life estimates: 25 years for data, five years for databases, and six years for data science products.[34] Finally, the net capital stock of each data product type is calculated by, in each year, adding new investments on top of the surviving total data product stock from previous years.

Table 1 shows an illustrative example of how the calculations would look for the years 2013 and 2014, with an assumed 2013 investment amount of €10bn in data, €3bn in databases, and €7bn in data science, following the assumed useful life estimates.

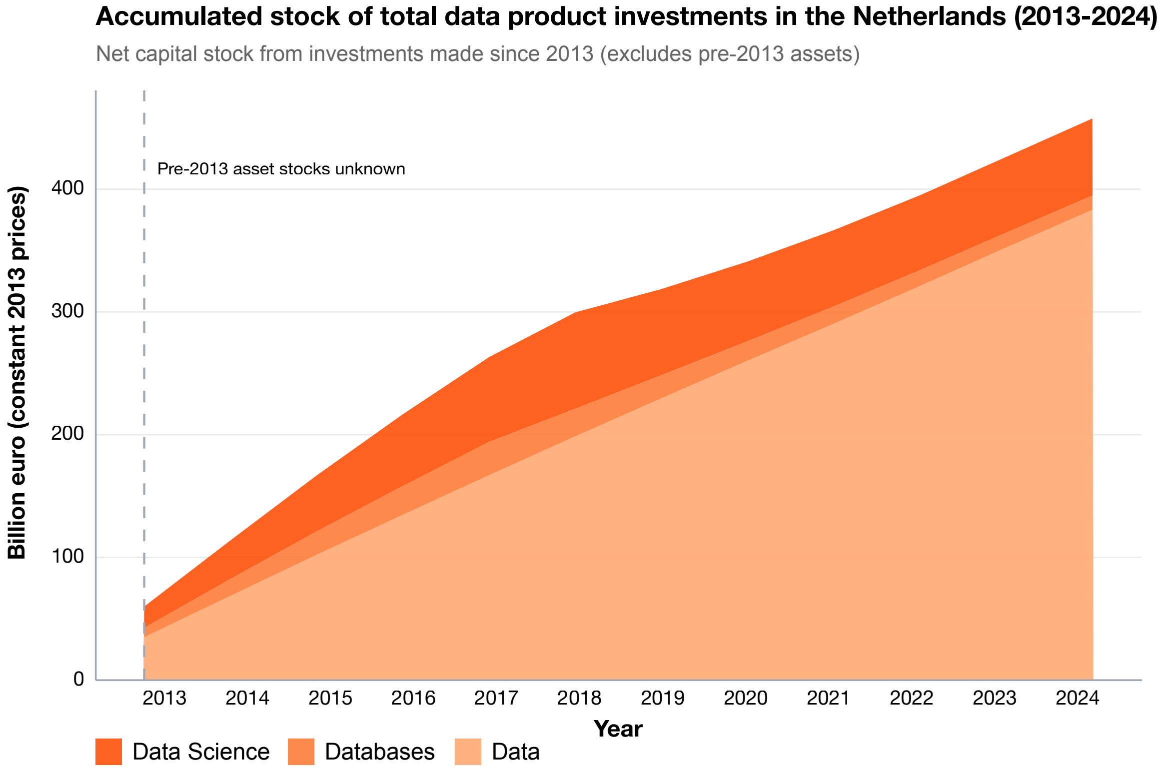
**Table 1: Example of total data product stock accumulation 2013–2014**

| | Data | Databases | Data science | Total |
|---|---|---|---|---|
| 2013 investment | €10bn | €3bn | €7bn | €20bn |
| 2013 net stock | €10bn | €3bn | €7bn | €20bn |
| Depreciation rates based on useful life estimates | 4% | 20% | 16.67% | |
| 2014 value of 2013 products, due to depreciation | €9.6bn | €2.4bn | €5.8bn | €17.8bn |
| 2014 investment | €10bn | €3bn | €7bn | €20bn |
| 2014 net stock | €19.6bn | €5.4bn | €12.8bn | €37.8bn |

Investments are made in 2013. Then, at the end of this year, some part of the newly created investment stock has depreciated based on each type of data product's useful life estimates (25 years for data or 4% per year, five years for databases or 20% per year, and six years for data science or 16.67% per year). Investments are made again in 2014, and the newly created investments are added to the depreciation-adjusted 2013 stock. In this way, investments accumulate over time for the 2013-2024 period.

Figure 18 shows the newly accumulated stock of total data products starting from 2013. Note that as data does not allow us to calculate the accumulated stock of investments before 2013, we only include the newly accumulated stock of assets from 2013, which is likely an underestimation, as in reality there would be stocks of data products that would have been 'carried forward' from before 2013.

**Figure 18: Only considering the newly created investments in data, databases, and data science from 2013 to 2024, they have accumulated to more than €457bn by 2024**




Note: Stock values represent cumulative investments since 2013, net of depreciation (PIM with geometric rates: data δ=0.04, databases δ=0.20, data science δ=0.167). Pre-2013 assets are not included due to data limitations. Databases (5yr) and data science (6yr) assets from before 2013 would have fully depreciated by 2018/2019.Data assets (25yr lifespan) from before 2013 may still hold economic value but are not captured here.Sources: O*NET, CBS, Statistics Canada, PwC analysis.

This makes clear that investments add up substantially over time. With roughly €76.8bn being invested in total data products each year, even accounting for price effects and depreciation, the newly created total data product stock in the 2013-2024 period accumulated to over €457bn by 2024.

# Boosting future value of data for the Dutch economy

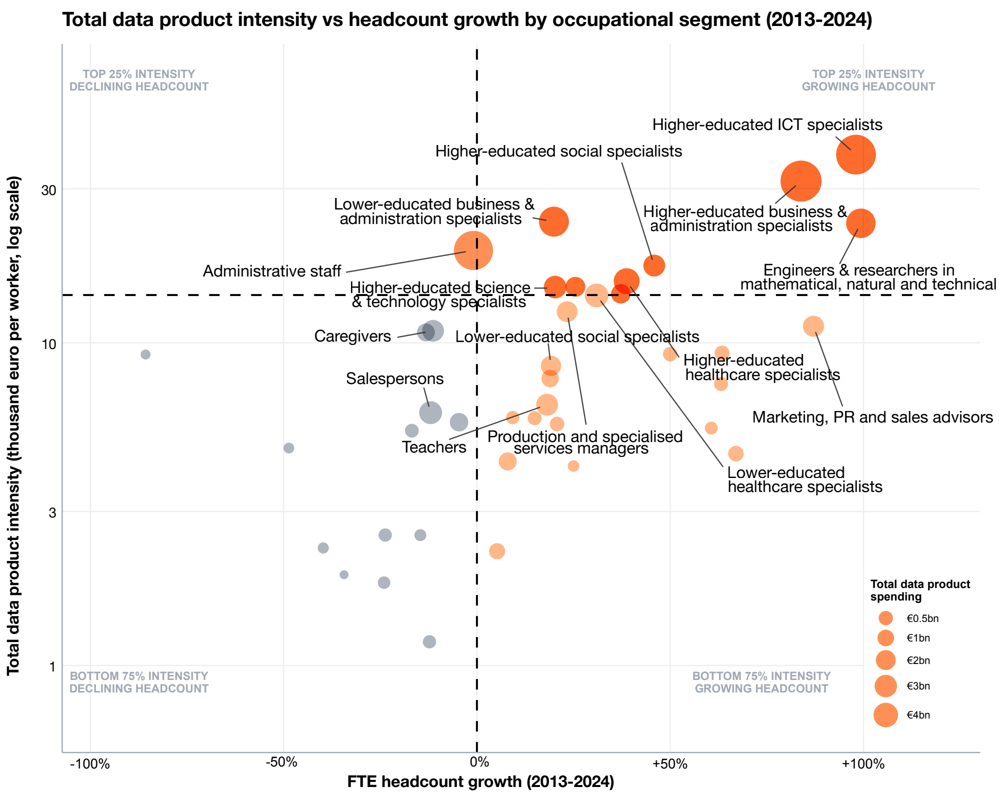
In the previous sections, we demonstrated that many occupations and industries already contribute greatly to the total data product value of the Dutch economy. In the following sections, we examine three ways in which this value could be increased in the future: tackling labour shortages, increasing data intensity to boost economic growth, and leveraging AI to unlock greater value from data products.

## Labour shortages as a bottleneck to capture more value of data products

The Dutch economy has been subject to labour shortages for some time.[35] Here we assess the potential additional value that could be generated in key data-product occupational segments if these shortages were resolved.

Figure 19 visualises all occupational segments by headcount growth over the 2013-2024 period and total data-product intensity.[36]

**Figure 19: The highest-growing occupational segments are also the most intensive in terms of total data product value contribution**

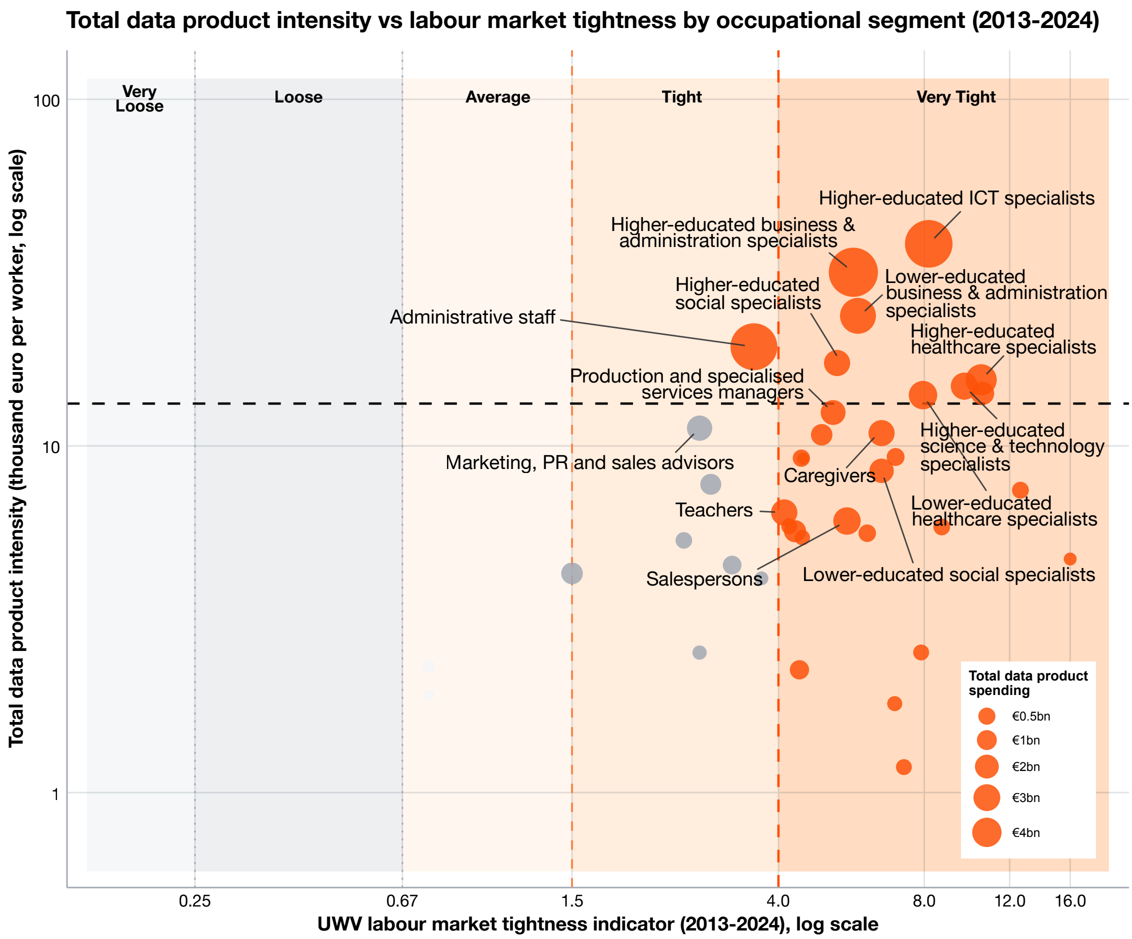



Sources: O*NET, ROA, CBS, Statistics Canada, PwC analysis.
Y-axis log scale. X threshold: 0% growth. Y threshold: 75th percentile intensity = €14.1k per worker (top 25%).
Only occupational segments exceeding €2bn in total data product spending are labelled.

Notably, many of the occupations that have grown the most over the 2013–2024 period, such as higher-educated ICT and business & administration specialists and engineers & researchers in the mathematical, natural, and technical sciences, are also those with the highest data-product intensity. At the same time, administrative staff and lower-educated business & administration specialists, while also making relatively high contributions to data products, have grown considerably less. These trends suggest that, at least up to 2024, occupations most engaged in the production of data products have been among the fastest-growing segments of the labour market, likely without reaching their full growth potential due to labour shortages.

Indeed, Figure 20 confirms this relationship between labour market tightness, measured by the UWV labour market tightness indicator,[37] and total data product intensity.[38]

**Figure 20: The occupations contributing the most to data product value are also those most constrained by labour shortages**




Sources: O*NET, ROA, CBS, UWV (2024 Q4), Statistics Canada, PwC analysis.
Both axes log scale. UWV zones: Very Loose (<0.25), Loose (0.25–0.67), Average (0.67–1.5), Tight (1.5–4.0), Very Tight (≥4.0).
Intensity threshold: 75th percentile = €13.3k per worker (top 25%).
Only occupational segments exceeding €2bn in total data product spending are labelled.

The UWV labour market tightness indicator measures the ratio of job vacancies to unemployed workers within each occupation, where values above four indicate a ‘very tight’ labour market for this occupation.

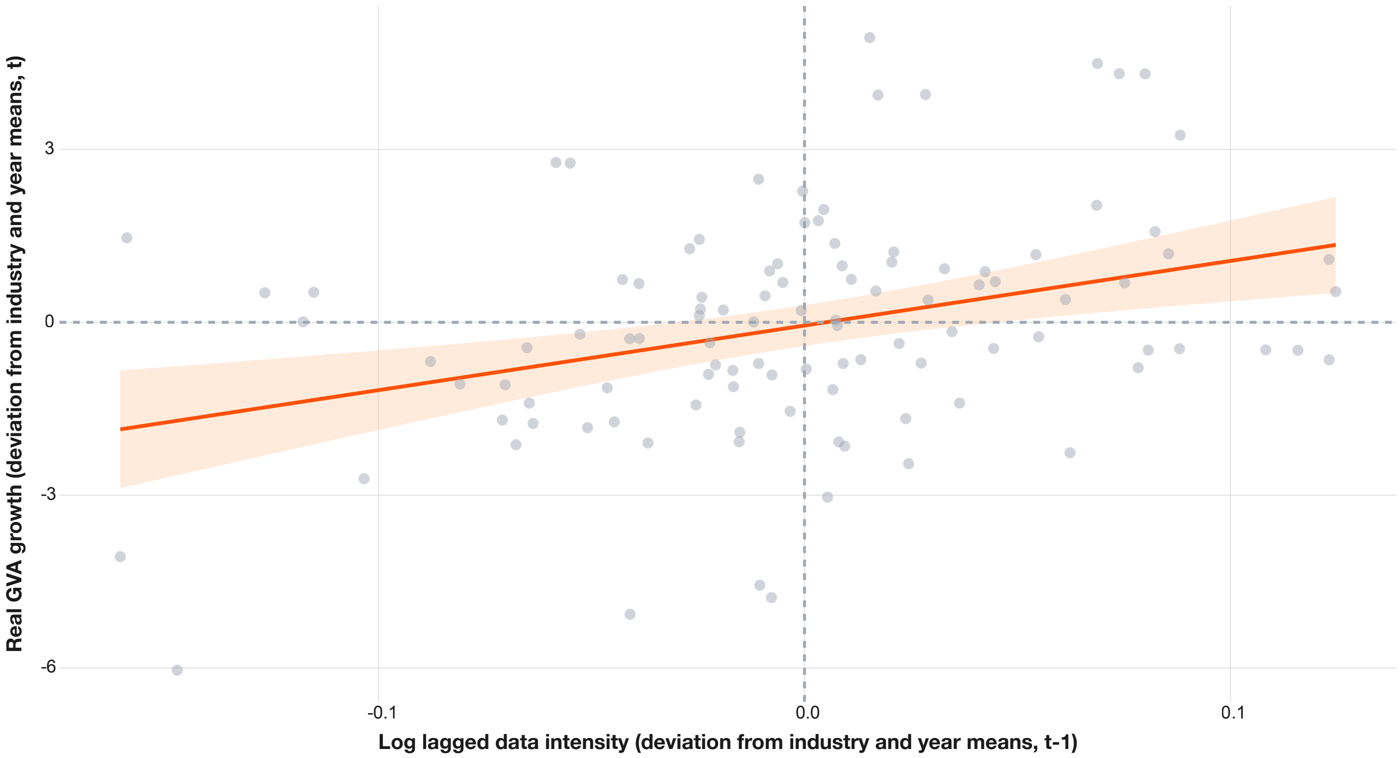
Figure 20 shows that all occupations that contribute a large part of total data products, such as higher-educated ICT and business & administration specialists, and engineers & researchers, are in the ‘very tight’ part of the labour market tightness indicator, indicating that there indeed are potential labour supply bottlenecks. If labour supply constraints are limiting the expansion of these occupations engaged in the production of data products, addressing these bottlenecks could unlock substantially greater value for the Dutch economy. At the same time, shifting employment from occupations with relatively loose to tight labour market conditions, as well as upskilling from low- to high-data-product intensity, are additional means to get more value.

# Increasing data product intensity to boost economic growth

Next, we assess the relationship between data intensity and GVA growth in Figure 21.

**Figure 21: Industries with higher data product intensity demonstrate significant positive correlation with economic growth rates the following year**

**The relationship between previous year's total data product intensity and GVA growth**




The industries that are excluded are transport, accommodation, other services, and the years are excluding 2020–2021.
Data intensity is measured as the log of real data spending per worker, expressed in 2013 prices.
Sources: CBS, ROA, PwC analysis. 107 industry-year observations (1 outliers excluded, >2.5×IQR).
Double-demeaned (industry + year). Full-sample r = 0.38 (p = 0.000). Panel FE □ = 13.06, clustered p = 0.022. Outliers removed: ICT (J) 2022 (Y).

We find that a positive and statistically significant relationship: a 1% increase in total data product investment per worker is associated with 0.13 percentage points higher GVA growth in the following year, suggesting that industries investing more intensely in data products tend to experience faster economic growth (for more details, see the Appendix on page 48). Hence, industries that are able to increase their data product intensity can achieve positive GVA growth. In that sense, data product investments can act as a leading indicator of future economic growth.

## AI as an enabler to harness more value from data products

AI is transforming how different types of occupations capture value from data products. In a stylised example, we distinguish between two types of occupations to demonstrate how AI can have varying effects depending on the tasks and data product intensity (Figure 22).

**Figure 22: Expansion and enhancement through AI: differential impact across occupations in terms of their data product intensity**

| | Low data product intensity | | High data product intensity | |
|---|---|---|---|---|
| **Expansion** | **AI-expanded occupations**<br>Roles where AI expands into adjacent technical or existing analytical tasks | | **AI-frontier occupations**<br>Roles where AI both accelerates existing data-heavy work and creates new work categories | |
| | **Examples:**<br>- Marketing managers,<br>- Product managers,<br>- HR professionals,<br>- Designers,<br>- Consultants,<br>- Small business operators. | **AI enables:**<br>- Lightweight coding,<br>- Dashboarding,<br>- Prototype building,<br>- Automation,<br>- Deeper research. | **Examples:**<br>- Data engineers in orchestrating AI agents and systems,<br>- Software engineers building AI agents,<br>- Quantitative researchers and analysts using AI in existing tasks and unlocking new capabilities. | **AI enables:**<br>- Prompt workflow design,<br>- Agentic systems,<br>- Multimodal prototyping,<br>- Faster iteration cycles,<br>- Model oversight. |
| **Acceleration** | **AI-supported occupations**<br>Roles where AI mainly supports existing tasks | | **AI-accelerated occupations**<br>Roles where AI mainly accelerates already data-centric workflows | |
| | **Examples:**<br>- Customer service representatives,<br>- Sales staff,<br>- Recruiters,<br>- Teachers,<br>- Administrative assistants. | **AI improves:**<br>- Writing,<br>- Summarising,<br>- Searching,<br>- Comunication,<br>- Scheduling. | **Examples:**<br>- Software developers,<br>- Data scientists,<br>- BI analysts,<br>- Financial analysts,<br>- Actuaries. | **AI improves:**<br>- Coding,<br>- Debugging,<br>- Data collection,<br>- Modelling,<br>- Analysis. |

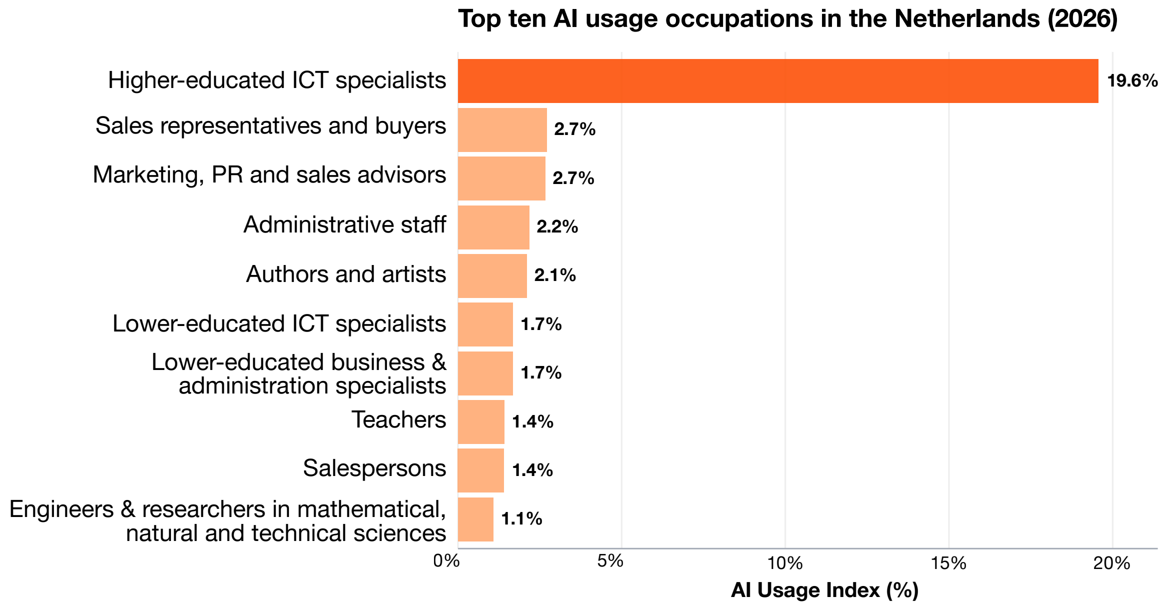
To understand how AI enters the picture when looking at data product intensity, we can think of two channels through which AI impacts different occupation groups: **expansion** and **acceleration**. Starting from the top left quadrant and moving clockwise in Figure 22:

- **AI-expanded** occupations (low data product intensity, high expansion): roles such as marketing and product managers, where AI not only improves existing task performance but also enables movement into adjacent technical and analytical functions, including deeper research, basic coding, dashboarding, and building digital prototypes.
- **AI-frontier** occupations (high data product intensity, high expansion): highly data product-intensive roles at the forefront of AI adoption, such as data engineers orchestrating AI agent systems and software engineers building AI agents, where AI both accelerates existing tasks and unlocks entirely new task capabilities.

- **AI-accelerated** occupations (high data product intensity, moderate expansion): roles such as software developers and data scientists, focused on established technical functions, including data analysis and software development. AI delivers significant productivity and quality gains, enabling faster and higher-quality output, but has limited impact on expanding the scope of job function.
- **AI-supported** occupations (low data product intensity, moderate expansion): roles such as customer service representatives and teachers, where AI enhances performance in core existing tasks, including writing, communication, and scheduling, without substantially expanding task boundaries.

Current evidence suggests that so far, at least in professional settings, the most transformative impact and intense AI usage has taken place within the higher-educated ICT specialists occupational segment (Figure 23).

**Figure 23: Higher-educated ICT specialists have been the main users of AI in the Netherlands so far**



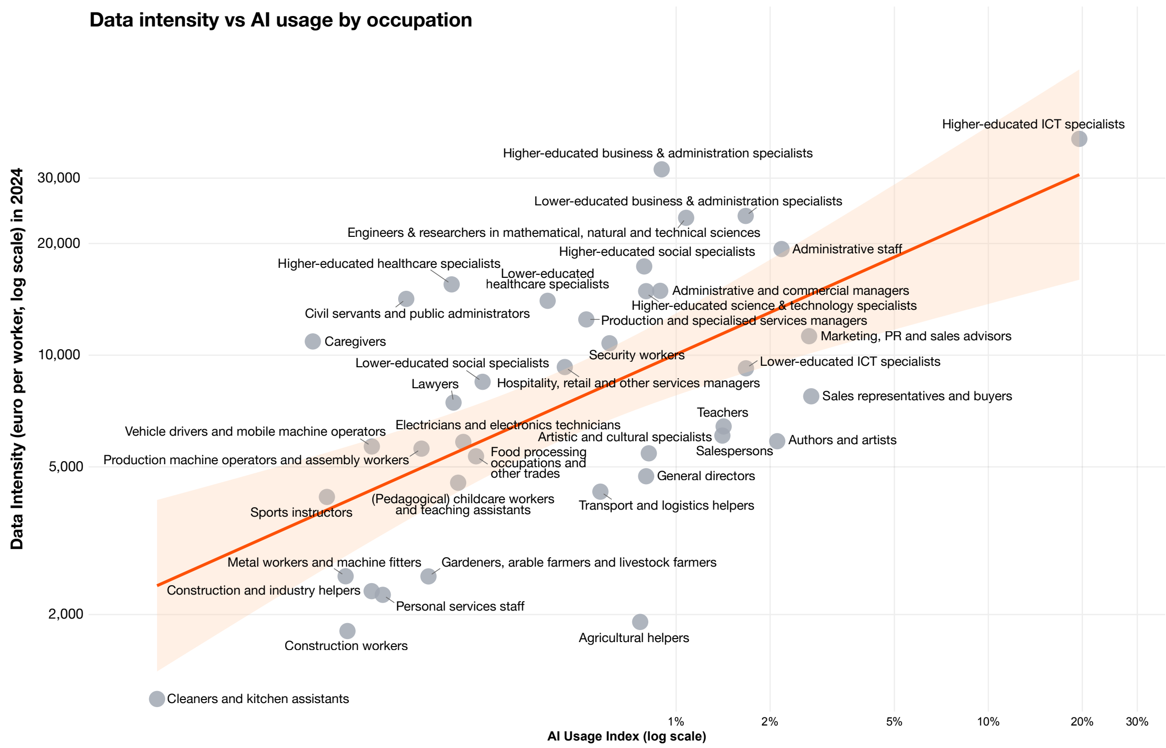

Sources: O*NET, Anthropic Economic Index (Global, Feb 2026), ROA, CBS, PwC analysis.
AI usage = share of tasks with LLM capability × average capability level.

This figure is based on the Anthropic Economic Index, which collects prompts from direct API users,[39] and maps them to occupational tasks and job categories (for more information see the Appendix on page 49). Applying this framework to the Netherlands, the data show that ICT specialists with higher education account for the largest share of AI usage, or 19.6% of all recorded uses, substantially ahead of other occupation groups.

This pattern is somewhat less pronounced among non-API users of the consumer-facing interface (Claude.ai), where usage is more evenly distributed across occupation groups; however, computer and mathematical occupations still account for more than three times the usage of the second-largest group. Consistent findings from other leading AI models, such as OpenAI's ChatGPT[40], point to the same trend of ICT professionals being the primary users.

Indeed, combining Anthropic's Economic Index data with our analysis of data product intensity across Dutch occupational segments, we also find that the most data product-intensive occupations are the leading AI users (Figure 24).[41]

### Figure 24: Data product intensity and AI usage exhibit a strong positive correlation




Log scale | r = 0.61, p < 0.001 | 39 occupations with AI usage > 0 | Global AEI data
Sources: O*NET, Anthropic Economic Index (Global, Feb 2026), ROA, CBS, PwC analysis.
AI usage = share of tasks with LLM capability × average capability level.
Data intensity measured using 2024 data; AI usage from AEI extracted for 2026.

We find that a 1% increase in the AI usage index is associated with a 0.38% increase in total data product intensity per worker. These findings highlight two aspects. First, while AI is arguably the fastest-spreading general-purpose technology, we are still in a relatively early phase of AI adoption. It is, therefore, logical that the most prominent initial impact can be found at the AI-frontier occupations, such as ICT professionals who have high data-product intensity and high potential for AI to accelerate and expand tasks. Second, for other occupations, who are further away from frontier developments and lower on the data product-intensity scale, AI is enabling a wider group of occupations to unlock new data-driven capabilities that previously were not possible. Hence, AI acts as both an accelerator for existing core tasks and an enabler of new task frontiers, and the impact of both of those channels is highly dependent on the nature of each occupation.

These findings highlight two key patterns. First, while AI is spreading rapidly, faster than many previous general-purpose technologies,[42] adoption remains at a relatively early stage at the overall labour market level, except for highly data product-intensive and AI-frontier adjacent occupational segments. It is therefore unsurprising that the most prominent initial impact is concentrated among AI-frontier occupations, such as ICT professionals, who combine high data product intensity with high potential for AI to both accelerate and expand their tasks.[43]

Second, for occupations that are less data product-intensive and further from the AI frontier, AI is enabling a broader group of workers to unlock new data-driven capabilities previously out of reach and transforming jobs through changes in tasks.[44] In combination these patterns suggest that AI acts as both an accelerator of existing core tasks and an enabler of new task frontiers – with the balance between these two roles shaped by each occupation's data-product intensity, proximity to the AI frontier, and the task composition.

Alongside the potential benefits of widespread AI adoption, including productivity gains and access to capabilities previously out of reach, significant challenges and uncertainties remain.

First, as with data product intensity, AI-readiness varies considerably across industries. There seems to be a symbiotic relationship between AI, computing power, and data access: without sufficient AI capability to process it, more raw data can lead to information overload and diminishing returns, while data and computing power act as preconditions for AI to function effectively.[45] These capabilities are highly skewed, with some companies and industries at a much higher readiness level in terms of data, skills, and AI capabilities.

Second, while some roles stand to gain new task frontiers through AI, others face a longer-term risk of significant task automation – potentially displacing a substantial share of existing job functions. Third is the risk to talent development pipelines. Many of the tasks that AI might be capable of automating could precisely be those performed by junior employees as part of their professional development. If these entry-level tasks are displaced, it remains unclear how future workers will acquire the foundational skills and experience needed to progress into more senior roles.

# Recommendations for policy-makers and organisations

Driven by the growing use of data and rapid advances in AI, the economy is transitioning from a knowledge economy to an intelligence economy.[46] Tracking where investments in data, databases and data science concentrate therefore reveals which sectors and occupations are positioning themselves for future value creation. This matters because it allows policymakers, investors and companies to identify growth engines early, often years before these shifts materialise in GDP, productivity or employment figures, making data investment a leading indicator of structural economic change. We now turn to additional recommendations for policymakers and companies.

## Policy recommendations

**A better understanding of the role data products play in the economy is needed**
Data should be treated as core economic infrastructure, and more precise measurement of its economic value and contribution is necessary. Our estimates suggest that annual data product investment reached €108 billion in 2024, equivalent to 9.6% of GDP, which is a substantial figure that underscores the macroeconomic significance of data in the Dutch economy. While national statistics frameworks sometimes capture data and database investments as intangible capital, the role and scale of investment in data intelligence tools remain largely uncaptured.[47]

**Work on aligning incentives for widespread access to data and data products**
First, access to data products should be treated like access to electricity: as a basic service, not a privilege. Public institutions should ensure fair and widespread access to the foundational infrastructure of the intelligence economy for firms that cannot provide these capabilities themselves, including electricity, data centres, computing power, reliable and secure internet connectivity, and the software and technologies required to participate fully in an increasingly intelligence-driven economy.

Second, while public sector data already enables a lot of economic value, it does not always translate into improved productivity and service quality, particularly at the executive organisation level.[48] By improving data integration, interoperability, uptake of data analytics, and data-driven decision-making, the public sector can substantially raise productivity, service quality, and policy effectiveness.[49]

Third, public institutions should encourage broader sharing of data and analytical capabilities in a way that rewards first movers and preserves competitive incentives, while preventing valuable data products from remaining siloed within private organisations to the detriment of the wider economy. Combining proprietary data and tools across organisational boundaries could unlock additional value by generating insights and capabilities that no single actor could achieve alone. A good example is the Dutch Consumer Price Index, for which supermarket chains provide data to Statistics Netherlands, enabling more accurate and timely measurement of inflation, benefiting not only the companies themselves but also the broader economy.[50] Hence, policymakers should identify similar cases where data held by companies can be used for wider social benefit, while rewarding the companies for their intellectual property.

However, there is no one-size-fits-all model for data sharing. Effective implementation requires sector-specific analysis to determine which combination of open standards, public interfaces, data pools, and trusted intermediaries could best deliver economy-wide returns.[51] In addition, regulatory strategies must also account for geopolitical risks, as openly shared data, while valuable domestically, could inadvertently empower foreign competitors. Balancing these objectives demands policies that encourage open innovation while safeguarding intellectual property.[52]

The alternative danger is concentration. If data remains siloed within organisations and access to AI tools remains reserved for those who can afford and utilise them, a new dividing line emerges based on who has the data, technology, and talent.[53]

**As much of the value of data products is enabled by people, it is necessary to address labour and skills shortages by lowering barriers to entry and upskilling the existing workforce**

Despite widespread concern about AI-driven job losses, there is little evidence that this is a pressing issue at present.[54] More immediately, we have identified a set of occupations that already serve, and will increasingly become the primary channels through which companies unlock the value of data products. Yet precisely these occupations remain scarce in the current labour market and represent one of the main concerns for Dutch companies. Unless addressed, these labour shortages risk leaving substantial economic potential untapped. Policymakers, working together with companies, should therefore tackle these structural bottlenecks through targeted measures such as lifelong learning programmes and active labour market policies aimed at upskilling the existing workforce and lowering barriers to entry for new participants in key data-driven roles.

## Recommendations for organisations

**Organisations should identify and strengthen the weakest links in their data value chain, as these ultimately determine the ceiling on productivity and innovation gains**

While public institutions are mandated to collect, work with, and disseminate data, many private organisations either fail to recognise the strategic value of data products and AI or lack the capacity to fully exploit what these tools offer. Too few firms, particularly small and medium-sized enterprises (SMEs), make use of data products, despite their proven potential to improve productivity, enhance innovation, and enable new business models.[55] This gap is especially pronounced in the Netherlands: large firms are three times more likely to perform big data analysis than their smaller counterparts, and the Netherlands has the largest such gap among all OECD countries.[56] Moreover, uneven adoption of data products across supply chain partners can dampen macroeconomic productivity growth and undermine innovation. Practices such as ASML's open innovation ecosystem illustrate the importance of enabling entire value chains to develop the same data-driven capabilities.[57]

**To unlock the full value of data products, organisations must first get their 'data house' in order**

Data is the starting point. To fully capture the value of data products, companies need the full stack of supporting capabilities in place: cybersecurity, cloud computing, data storage, and data analysis skills. PwC analysis across nine pillars of AI fitness found that the most AI-ready companies achieve a 7.2 times greater AI-driven performance boost, combining revenue growth and cost reductions, compared to their peers. These companies are also more than twice as likely to have eliminated outdated and costly IT systems and infrastructure.[58]

The challenge for organisations lies not only in keeping pace with rapid technological change but also in the fact that realising data- and AI-driven capabilities requires all pillars to be in place simultaneously. Data science cannot function without the underlying data, databases, and technology; nor can digital tools and data be fully utilised if privacy and cybersecurity vulnerabilities remain unaddressed. Labour and skills shortages add a further layer of complexity: even in the age of AI, these shortages remain acute and are concentrated precisely in data-intensive roles. Organisations should therefore build internal capability through upskilling and role redesign, rather than relying solely on external hiring. Evidence supports this approach: active employer initiatives, such as upskilling programmes and encouraging employees to adopt data and AI tools, are associated with greater tool usage and higher reported benefits, including time savings, improved quality, enhanced creativity, and greater job satisfaction.[59]

Employees, too, stand to gain by moving up the data value stack: becoming more proficient with data and AI tools can improve personal productivity, open new task frontiers, and strengthen employability. This requires not only foundational digital literacy but also a commitment to continuous lifelong learning to become increasingly digitally native.

Taken together, these interdependencies suggest that, much like weak links in a supply chain, gaps in an organisation's data house create bottlenecks that slow AI and digital adoption and limit efficiency gains.[60]

# Appendix

## Methodology notes

**Measuring data product intensity at the occupation level**

To estimate the share of time each occupation spends on data, database, and data science tasks, we drew on the O*NET 30.1 database, which contains 18,796 tasks across 923 occupations. We classified each task using three large language models (GPT-4o mini, GPT-4.1 mini, and Gemini 2.5 Flash), each prompted to assign one of four labels – data, database, data science, or none – along with a confidence score between 0 and 1. We adopted the following definitions, based on Statistics Canada's approach:

- **Data:** The activity of converting observations into digital format (digitisation) and producing structured information from raw or unstructured sources. This is about capturing and recording information.
- **Databases:** The activity of structuring, organising, storing, and managing data within database systems or information repositories for practical retrieval and use. This is about organising and managing information systems.
- **Data Science:** Research, analysis, computational development, and engineering activities that use data and code to build systems, test hypotheses, generate insights, make predictions, automate processes, evaluate performance, or yield valuable new capabilities. This category encompasses the full spectrum of technical work that creates value from data or through software, including data analysis, statistical modelling, computer programming, software engineering, software development, data engineering, machine learning, and algorithmic design. This is about analysing data, generating knowledge, and building computational tools.

The prompts defined the core activities associated with each data product category, provided examples, flagged common classification mistakes, and established rules to help the models distinguish between task types. These rules included taking occupational context into account, assigning lower confidence scores to short or vague task descriptions, and downweighing tasks where data-related work is not the primary objective.

We then assessed agreement across models as follows. Where all three models assigned the same label and each reported a confidence score above 0.9, we accepted that classification directly. This applied to 10,037 tasks, representing 53.4% of the dataset. The remaining 8,759 tasks, where models either disagreed or agreed with insufficient confidence, were flagged for review by a more powerful model (GPT-4.1). This model evaluated each response in anonymised form, applying a conservative classification rule that gives greater weight to cases where at least two of the three models agreed and where associated confidence scores were higher. In total, we identified 1,891 tasks classified as data (851), databases (211), or data science (829).

We also incorporated task frequency ratings from O*NET, which capture how much time workers spend on each task. From this, we constructed a frequency-weighted measure estimating both the share of tasks and the share of time falling into each of four categories: data, databases, data science, and none. This served as our core measure of data product intensity at the occupational level. Since ONET updates task ratings periodically rather than annually, we applied a single set of occupational estimates across the full 2013–2024 period.

Of the 923 occupations in the O*NET task ratings data, 83 had incomplete information – either missing tasks or missing frequency ratings for some or all tasks. To address these gaps, we identified the three most similar occupations for each affected occupation using ONET's occupational similarity scores and imputed the missing values as a weighted average of those three occupations' ratings.

Several limitations of the O*NET task list are worth acknowledging. While it represents the most comprehensive available measure of occupational task content, it does not capture all tasks in the economy nor other important dimensions of work, such as tacit knowledge, interpersonal relationships, and judgement under uncertainty, or the interdependencies between tasks.[61] Additionally, O*NET does not distinguish between task profiles across different firms or industries within the same occupation, and each version is static, meaning task estimates are not updated simultaneously across all occupations.[62] Finally, we assume that the O*NET task list, which is based on the United States' labour market, serves as a reasonable proxy for the task profiles of Dutch workers in equivalent occupations.

**Adjusting occupation names**

Occupation segment names follow the BRC 2014 classification, with several labels adapted from the original ROA nomenclature for clarity. In the ROA classification, the suffix "(specialised)" (Dutch: vakspecialisten) denotes occupational segments at a lower educational level – typically MBO or associate degree level – within the same broad domain, while the base label without this suffix denotes the higher-educated segment – typically HBO or WO level. To make this distinction more intuitive for an international audience, we relabelled these pairs as "lower-educated" and "higher-educated":

- "Business management and administration specialists" and its "(specialised)" counterpart were relabelled as "higher-educated business & administration specialists" and "lower-educated business & administration specialists."
- "ICT specialists" and "ICT specialists (specialised)" were relabelled as "higher-educated ICT specialists" and "lower-educated ICT specialists."
- "Specialists in science and technology" was relabelled as "higher-educated science & technology specialists."
- "Physicians, therapists and specialised nurses" and "healthcare specialists" were relabelled as "higher-educated healthcare specialists" and "lower-educated healthcare specialists."

- “Social specialists” and “social workers, group and residential counsellors” were relabelled as “higher-educated social specialists” and “lower-educated social specialists.”
- All remaining segment names are unchanged from the original ROA/CBS classification.

**Estimating data product value for the Netherlands**

To apply our estimated data product intensity to the Dutch labour market, we mapped O*NET-SOC occupation codes to ISCO-08 codes using a SOC-ISCO crosswalk then converted these to Dutch BRC 2014 occupation codes at the 3-digit level (beroepssegment) using an ISCO-BRC crosswalk. Data intensity estimates were aggregated accordingly at each step.

We then merged the data product intensity estimates with Statistics Netherlands data on median hourly gross wages and average hours worked per occupation, accounting for different part-time work rates. Multiplying these by the number of working weeks yielded an estimate of annual labour costs at the occupational level. We subsequently multiplied this by the occupation-level data product intensity to isolate the share of annual labour costs attributable to producing data, database, or data science products. Finally, using CBS and ROA data on the number of FTE jobs per occupation and their distribution across industries, we scaled these occupational estimates up to obtain economy-wide results.

**Regression estimates of data intensity and economic growth**

To examine this, we estimate the relationship between lagged data intensity – defined as total data product investment per FTE worker – and real GVA growth (at 2013 prices). As a robustness check, we also test an alternative measure defined as total data product investment as a share of GVA; the results align closely. However, for three industries (accommodation and food, other services, and transportation and logistics) the two measures diverge (Figure 29), and we therefore exclude these industries from the regression model. We additionally exclude 2020 and 2021, as these years were dominated by Covid-19-related volatility (Figure 30), and tested the exclusion of 2022 as a partial recovery year; the results do not change substantially.

We lag data intensity by one year for two reasons: first, investment takes time to materialise in output growth; and second, lagging addresses simultaneity arising from GVA appearing in the denominator of the data intensity measure.

We apply a two-way fixed effects approach, controlling for time-invariant industry characteristics and economy-wide annual shocks. The regression therefore identifies within-industry variation – specifically, whether industries tend to grow faster in years when their data intensity exceeds their own historical average. We test the robustness of our results using both industry-clustered and heteroskedasticity-robust standard errors.

Even with a fixed effects approach, we cannot fully establish causality, as reverse causality remains possible: faster-growing industries may increase data product investment rather than investment driving growth. We test for this by estimating the relationship between lagged GVA growth and data intensity, finding no statistically significant results and providing some reassurance against reverse causality.

**Table 2: GVA growth and lagged data intensity regression results**

| Regression results | Model 1: GVA growth ~ Log lagged data intensity (total data product spending per worker) | Model 2: GVA growth ~ Lagged data intensity (total data product spending as a share of GVA) |
|---|---|---|
| (Intercept) | -91.58 (35.03)* | -0.71 (0.91) |
| Lagged data intensity | 13.06 (4.88)* | 1.58 (0.38)** |
| Fixed effect cross-sectional (industry) | Yes | Yes |
| Fixed effect time | Yes | Yes |
| R squared | 0.60 | 0.66 |
| Adjusted R-squared | 0.50 | 0.59 |
| Time periods | 9 | 9 |
| Clusters (industries) | 12 | 12 |
| Total observations | 108 | 108 |

*indicates significance at $p < 0.05$ level; ** at $p < 0.01$ level, *** at $p < 0.001$ level. Clustered standard errors in brackets.

**Calculating AI usage for occupations in the Netherlands**

We used raw data from the Anthropic Economic Index (AEI) 2026, which tracks conversations users have with Claude via direct API, classified by the occupational tasks being performed. For example, if a user asked Claude to ‘develop a model for software testing’, Anthropic classified this as a specific O*NET occupational task. The data provide the percentage of Claude conversations involving occupational tasks globally. As Dutch-specific data are unavailable, global task shares were used as a proxy for Dutch task shares.

O*NET task statement data provided a comprehensive list of tasks for each occupation, enabling us to map AI usage to specific occupational tasks using the AEI data. For each occupation, we calculated the share of tasks that were AI-exposed – the extensive margin. For example, if a software developer has 10 tasks in total and 8 were matched to AEI data indicating Claude usage, the extensive margin is 0.8, meaning 80% of that occupation’s task portfolio is AI-exposed.

We also calculated the average AI intensity across those AI-exposed tasks (the intensive margin), representing the mean AEI usage share among AI-exposed tasks. Continuing the example, we computed the average AI usage score across the eight exposed tasks, reflecting the average level of Claude traffic those tasks generated.

$$\textbf{alpha = extensive margin x intensive margin}$$

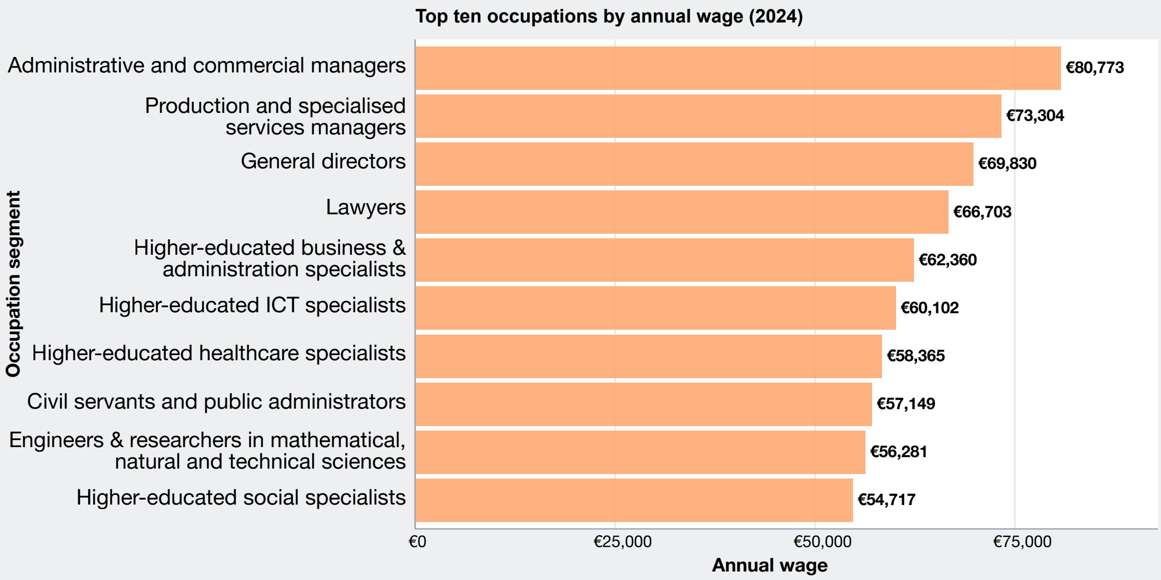
Next, we mapped O*NET-SOC occupation codes to ISCO-08 codes using a SOC-ISCO crosswalk and then converted these to Dutch BRC 2014 occupation codes at the 3-digit level (beroepssegment) using an ISCO-BRC crosswalk. Data intensity estimates were aggregated accordingly at each step.

We restricted our sample to occupations with an extensive margin above zero, retaining only those with at least one AI-exposed task. We additionally extracted Statistics Netherlands data on headcount and median hourly wages by BRC occupation segment, alongside occupation-level data intensity estimates, to analyse the relationship between AI exposure and data intensity for 2024.

Two additional limitations are worth acknowledging. First, our AI exposure estimates draw on data from a single provider (Anthropic), covering API and business use only. While we believe this dataset is comprehensive and overlaps reasonably well with broader AI model usage, actual AI exposure may be understated if occupations make use of multiple AI tools or Claude's consumer chat interface. Second, this analysis assumes that global AEI task percentages serve as a reasonable proxy for the Netherlands, meaning that Dutch-specific patterns of AI adoption may not be fully reflected in our estimates.

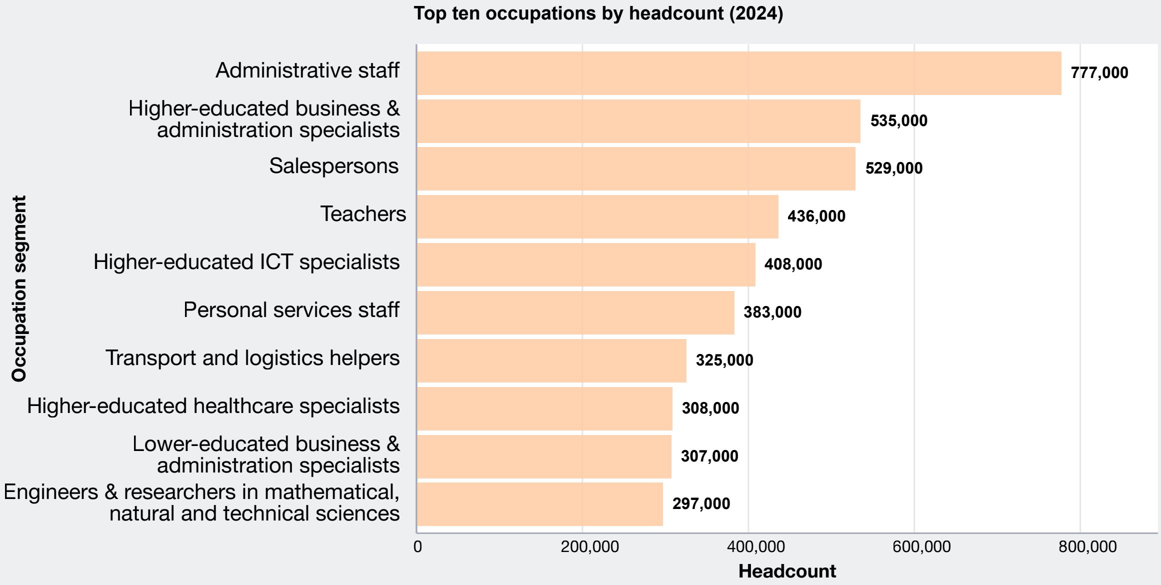
## Extended results

### Figure 25: Highest paid occupational segments



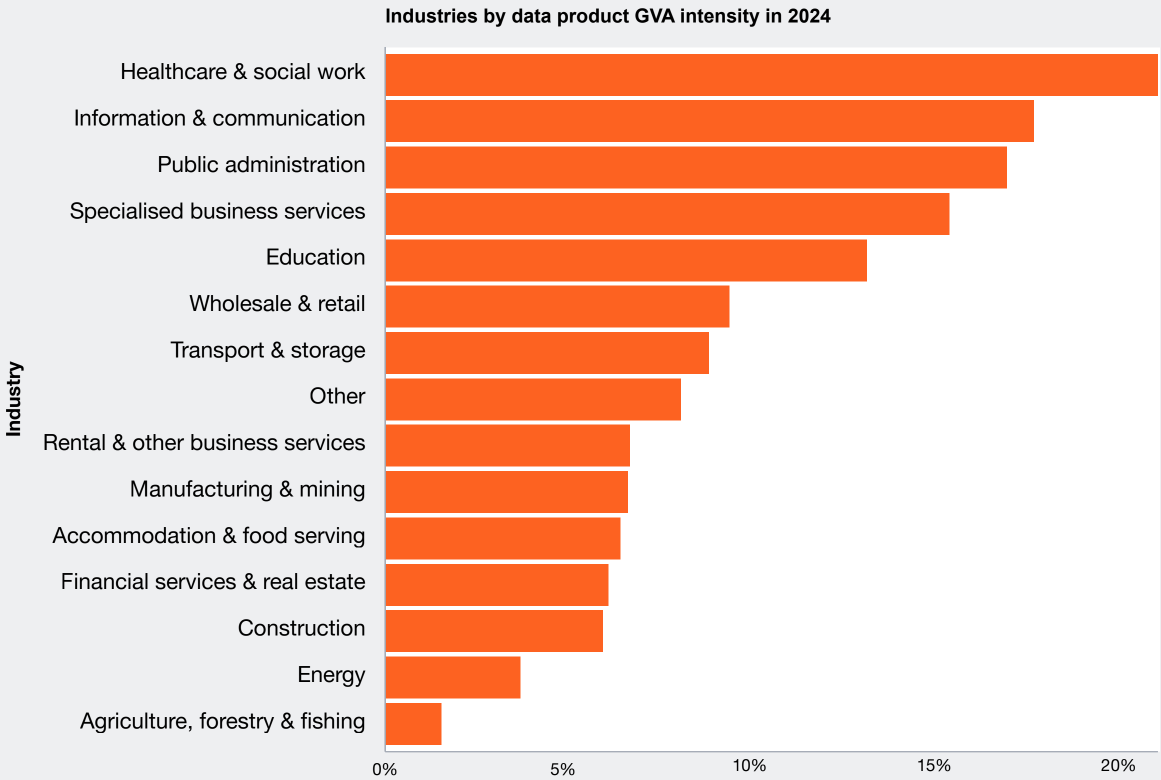

Sources: O*NET, ROA, CBS, PwC analysis.

## Figure 26: Largest number of jobs per occupation segment




Sources: O*NET, ROA, CBS, PwC analysis.

## Figure 27: Healthcare & social work, information & communication and public administration had the highest total data product GVA intensity in 2024

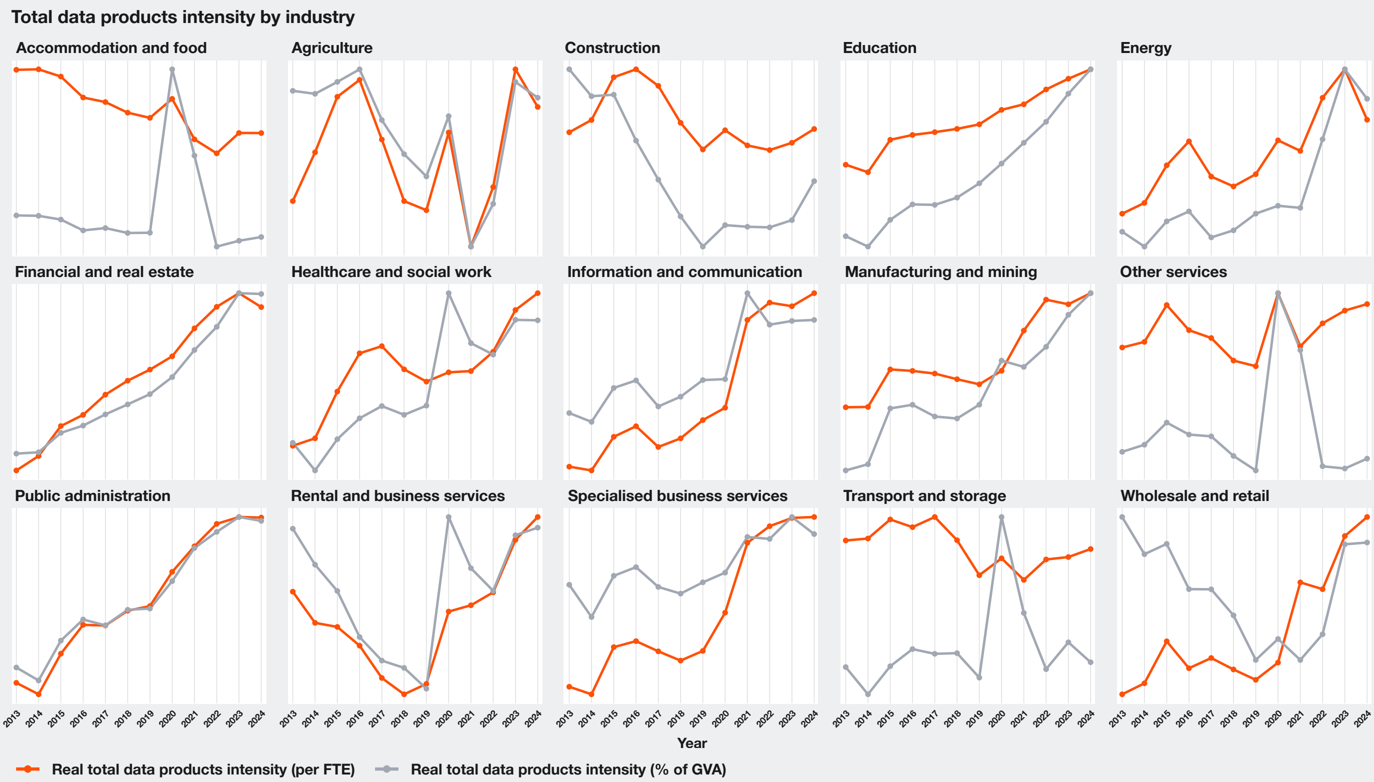



Sources: O*NET, ROA, CBS, PwC analysis.

## Figure 28: Alternative data intensity measure: industries with higher data product intensity demonstrate significant positive correlation with economic growth rates the following year

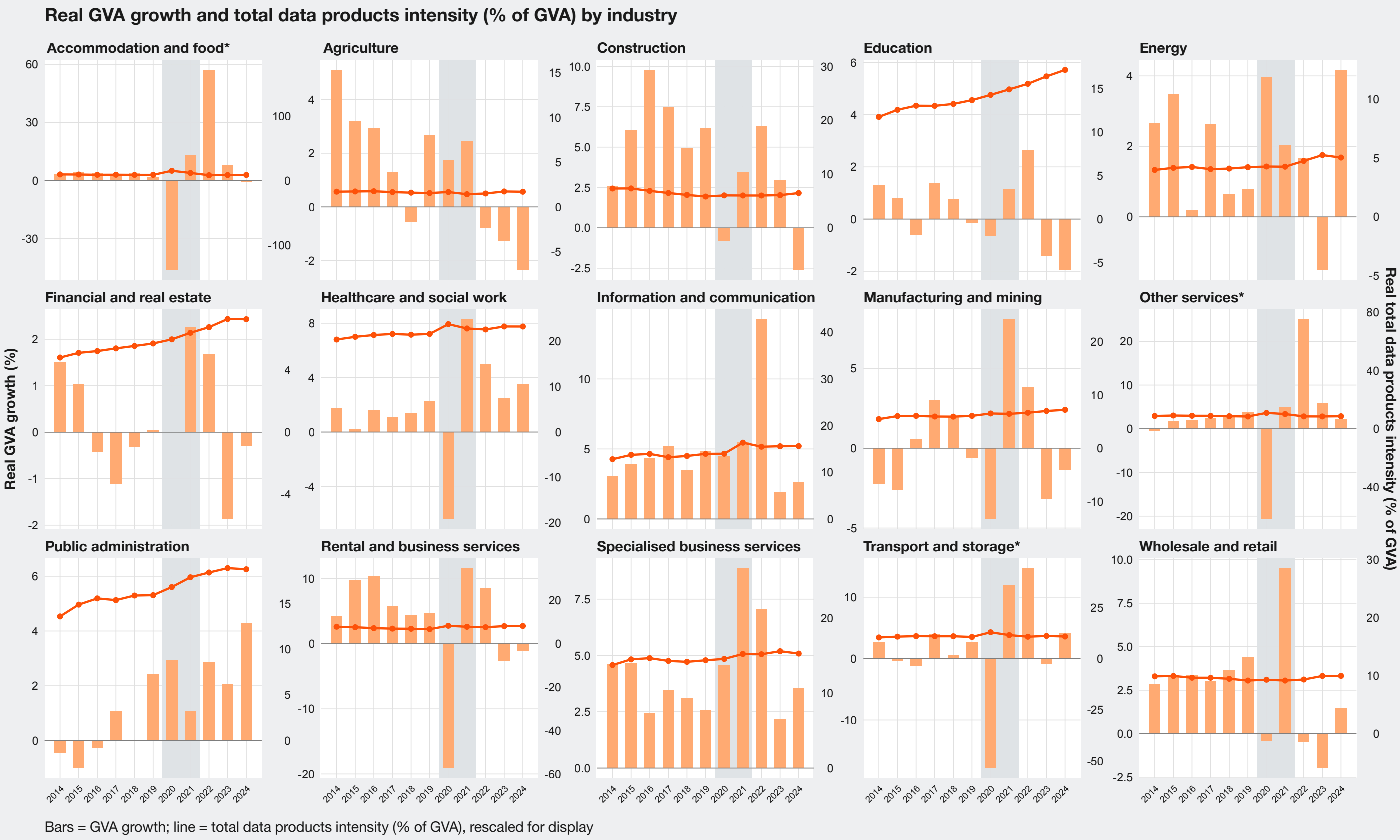

The industries that are excluded are transport, accommodation, other services, and the years are excluding 2020–2021
Data intensity is measured as real data spending divided by real GVA, with both expressed in 2013 prices.
Sources: CBS, ROA, PwC analysis. 107 industry-year observations (1 outliers excluded, >2.5×IQR). Double-demeaned (industry + year). Full-sample r = 0.54 (p = 0.000). Panel FE β = 1.58, clustered p = 0.002.
Outliers removed: ICT (J) 2022 (Y).

## Figure 29: Two different measures for industry level total data product intensity growth




The % of GVA series is rescaled within each industry for visual comparison only.
No common y-axis scale is shown because the two measures are in different units.

## Figure 30: Industry-level relationship between real GVA growth and lagged data product intensity




Shaded area indicates 2020–2021, excluded from the main regressions. * denotes industries excluded from the restricted-sample robustness regressions. The line is scaled by 1/3 for visual comparability; the figure is descriptive and not intended for direct level comparison between the two series.

# Endnotes

**1** To put it in perspective, a zettabyte equals 1 sextillion bytes (1,000,000,000,000,000,000,000 bytes), or the equivalent of storing 250 billion DVDs.
**2** Romano (December 2024): Synthetic geospatial data and fake geography: A case study on the implications of AI-derived data in a data-intensive society.
**3** Rivery (May 2025): Big data statistics: How much data is there in the world?
**4** Baley & Veldkamp (November 2025): The Data Economy: Tools and Applications.
**5** OECD (June 2024): Towards a better understanding of data-intensive firms in the United Kingdom.
**6** Baley & Veldkamp (November 2025): The Data Economy: Tools and Applications.
**7** Ibid.
**8** Ibid.
**9** Ibid.
**10** Brookings (2024): The partial Data barter trades of the digital economy.
**11** Ibid.
**12** Statistics Canada (June 2019): Measuring investment in Data, Databases and Data science: Conceptual framework.
**13** Brookings (2024): The partial Data barter trades of the digital economy.
**14** Ibid.
**15** Baley & Veldkamp (November 2025): The Data Economy: Tools and Applications.
**16** Ibid.
**17** Statistics Canada (June 2019): Measuring investment in data, databases and data science: Conceptual framework.
**18** Ibid.
**19** Ibid.
**20** Ibid.
**21** Agenda Nieuw Amsterdam (March 2026).
**22** Statistics Canada (June 2019): Measuring investment in data, databases and data science: Conceptual framework.
**23** Ibid.
**24** Ibid.
**25** Adding a capital surcharge stem from the methodology of Statistics Canada and Statistics Netherlands. It is customary to attribute these profits to market participants, as they expect an excess return on their investment. In practice, this does not necessarily occur in all cases, but it is assumed that the average investment yields a return.
**26** Statistics Canada (June 2019): Measuring investment in data, databases and data science: Conceptual framework.
**27** Schmidt et al. (October 2023): The role of Data skills in the modern labour market.
**28** CBS: Standard Industrial Classifications. In addition, we group together financial services (K) and real estate (L), and manufacturing (C) and mining (B) industries. The group 'Other' includes R-U industries.
**29** Statistics Canada (March 2025): Data, Intangible Capital and Economic Growth in Canada.
**30** Statistics Netherlands (January 2021): De waarde van Data 2001-2017.
**31** Corrado et al. (2022): The value of data in digital-based business models: Measurement and economic policy implications.
**32** Dey-Chowdhury (September 2008): Methods explained: Perpetual Inventory Method (PIM).
**33** Following other methodologies, we use a geometric depreciation rate, when depreciation every year happens at the same rate instead of the same amount, which would be arithmetic depreciation.
**34** Statistics Canada (July 2019): The value of Data in Canada: Experimental estimates.
**35** De Nederlandsche Bank: Labour market.
**36** Note that the y-axis uses a logarithmic scale, where each gridline represents a tenfold increase in absolute growth. A log scale means that each step on the axis represents a multiplication rather than an addition: instead of going up by equal amounts (100, 200, 300...), it goes up by equal multiples (10, 100, 1,000, 10,000...). This allows industries with both small and large absolute growth to be compared in the same chart. However, that means that equal visual distances do not represent equal absolute differences.

**37** UWV Dashboard Spanningsindicator.
**38** Note that both the y and x axis use a logarithmic scale, where each gridline represents a tenfold increase in absolute growth. A log scale means that each step on the axis represents a multiplication rather than an addition: instead of going up by equal amounts (100, 200, 300...), it goes up by equal multiples (10, 100, 1,000, 10,000...). This allows industries with both small and large absolute growth to be compared in the same chart. However, that means that equal visual distances do not represent equal absolute differences.
**39** First-party API or 1P API refers to developer traffic routed directly through Anthropic's own programming interface, which is distinct from both Anthropic's consumer-facing Claude.ai application and third-party platforms such as Amazon Bedrock or Google Cloud Vertex.
**40** OpenAI (September 2025): ChatGPT usage and adoption patterns at work.
**41** Note that both the y and x axis use a logarithmic scale, where each gridline represents a tenfold increase in absolute growth. A log scale means that each step on the axis represents a multiplication rather than an addition: instead of going up by equal amounts (100, 200, 300...), it goes up by equal multiples (10, 100, 1,000, 10,000...). This allows industries with both small and large absolute growth to be compared in the same chart. However, that means that equal visual distances do not represent equal absolute differences.
**42** Microsoft (November 2025): AI diffusion report: Mapping global AI adoption and innovation.
**43** Anthropic (January 2026): The Anthropic Economic Index report: Economic primitives.
**44** Humlum & Vestergaard (March 2026): Still waters, rapid currents: Early labour market transformation under generative AI
**45** Mihet et al. (January 2026): Is it AI or data that drives firm market power?
**46** Agenda Nieuw Amsterdam.
**47** Corrado et al. (June 2022): Intangible capital and modern economies; Corrado et al. (November 2022): Measuring data as an asset: Framework, methods and preliminary estimates.
**48** PwC: A closer look on the productivity of executive organisations.
**49** World Bank Group (2022): Interoperability: Towards a data-driven public sector.
**50** CPB (May 2021): Policy options for the data economy - a literature review.
**51** Ibid.
**52** Mihet et al. (January 2026): Is it AI or data that drives firm market power?
**53** Ibid.
**54** Brookings (October 2025): New data show no AI jobs apocalypse–for now.
**55** OECD (December 2022): Data shaping firms and markets.
**56** Ibid.
**57** ASML: Innovation ecosystem.
**58** PwC (April 2026): Decoding ROI from AI.
**59** Humlum & Vestergaard (March 2026): Still waters, rapid currents: Early labour market transformation under generative AI.
**60** Jones (January 2026): A.I. and our economic future.
**61** Anthropic (November 2025): Estimating AI productivity gains from Claude conversations.
**62** Ibid.

# Contacts


**Veronique Roos-Emonds**
Reinvention Leader,
PwC Netherlands
T: +31 (0)6 23 78 78 25
E: veronique.roos@pwc.com

**Barbara Baarsma**
Chief Economist,
PwC Netherlands
T: +31 (0)6 24 20 47 07
E: barbara.baarsma@pwc.com


## Acknowledgements


**Authors** (alphabetically by surname order):

| | | |
|---|---|---|
| Barbara Baarsma | Chief Economist | PwC Netherlands |
| Guntars Upis | Chief Economist Office | PwC Netherlands |
| Clemens van der Veen | Chief Economist Office | PwC Netherlands |
| Eu Zhang | Chief Economist Office | PwC Netherlands |

The authors would like to thank Femke Alsemgeest, Robin van den Akker, Remko Blom, Saar van Dijk, Brian Jakobs, Corrina Klefken-Schoordijk, Yentl Knossenburg, Ricardo Ribas Santolim, Richard Volbeda, and Astrid van der Werf for their valuable feedback and contributions in the making and dissemination of this report.