Muistio | Brief | 149

18.12.2024

Databarometer: An International Comparison of the Data Economy

Otto Kässi
ETLA Economic Research, Finland
otto.kassi@etla.fi

Suggested citation:
Kässi, Otto (18.12.2024). "Databarometer: An International Comparison of the Data Economy". ETLA Brief no. 149.
https://pub.etla.fi/ETLA-Muistio-Brief-149.pdf

Abstract

This Etla brief introduces a method for comparing the maturity of national data economies using a composite indicator. We assess the data economy from three perspectives: the prerequisites set by the public sector, the utilization of data, and the innovation impacts within the data economy. Combined, these three sub-indicators form the Databarometer. The chosen metrics draw from publicly available data sources and cover elements such as ICT expertise, the use of artificial intelligence technologies, and data driven start-ups. The various indicators are integrated into a single composite indicator using the Digibarometer calculation framework developed at ETLA. Both the framework and the underlying data are openly available and well-documented, ensuring that the calculations can be easily replicated.

Tiivistelmä

Databarometri: Datatalouden kansainvälinen vertailu

Esittelemme menetelmän, jolla verrataan maiden datatalouden kypsyyttä käyttämällä komposiitti-indikaattoria. Arvioimme datataloutta kolmesta eri näkökulmasta: datatalouden edellytyksiä, datan hyödyntämistä ja datatalouden innovaatiovaikutuksia. Yhdessä nämä kolme alaindikaattoria muodostavat Databarometrin. Mittarien valinta perustuu saatavilla oleviin julkisiin tietolähteisiin, ja ne kattavat mm. ICT-osaamisen, tekoälyteknologioiden käytön ja data-alan startupit. Eri indikaattorit yhdistetään yhteiseksi komposiitti-indikaattoriksi hyödyntämällä Etlassa kehitettyä Digibarometrin laskentakehikkoa. Laskentakehikko ja lähtöaineistot ovat avoimia ja siten dokumentoidut, että laskenta voidaan toistaa kevyesti.

Dr. Soc. Sci. **Otto Kässi** is a Researcher at ETLA Economic Research.

VTT **Otto Kässi** on Elinkeinoelämän tutkimuslaitoksen tutkija.

Acknowledgements: I am thankful to Heli Koski and the members of the steering group of the *Impacts of the Data Economy* VN TEAS project.

Kiitokset: Kiitän Heli Koskea ja *Datatalouden vaikuttavuus* VN TEAS -hankkeen ohjausryhmää kommenteista.

Key words: Data economy, Composite indicators, Digital innovation, Artificial intelligence, ICT capital

Avainsanat: Datatalous, Komposiitti-indikaattorit, Digitaaliset innovaatiot, Tekoäly, ICT-pääoma

JEL: O33, L86

1 Introduction

Current data sources for measuring the data economy do not allow for international comparisons of its productivity effects or value-added. As a result, we must turn to alternative methods to compare the maturity of data economies across countries.

We draw on the Digibarometer[1] methodology, developed at ETLA, to facilitate these comparisons. The Databarometer is a composite indicator that assesses a country's data economy preparedness from three perspectives: the preconditions established by the public sector, the extent of business adoption, and the innovation impacts observed. In this research note, we present the chosen indicators, the countries included, and the principles guiding their selection. We also explain how the various sub-indicators are combined into a single, comprehensive Databarometer index.

2 Databarometer calculation

2.1 Sub-indicators

The purpose of the Databarometer (Databarometri) is to capture the data economy from multiple perspectives. These perspectives—prerequisites, utilization, and impacts—reflect essential stages of the data value chain.

Where possible, the Databarometer uses publicly available data sources. Many such indicators are published by supranational organizations like the European Commission and the OECD. We prioritize indicators that are regularly updated, current, and internationally comparable. The indicators used in the Databarometer are presented in Table 1.

In selecting the indicators, we followed the approach used in ETLA's Digibarometer (Digibarometri) by mea-

Table 1 Databarometer: background variables

Variable	Dimension	Source	Year
Public-sector measures for leveraging machine learning	Prerequisites established by the public sector for the data economy	Tortoise.ai	2022
Openness of public data	Prerequisites established by the public sector for the data economy	European Commission (DESI)	2023
Share of ICT specialists in the workforce	Prerequisites established by the public sector for the data economy	Eurostat	2023
Use of AI technologies (share of companies)	Data utilization	Eurostat	2023
Use of data analytics (share of companies)	Data utilization	Eurostat	2022
Share of ICT-intensive jobs among all jobs	Data utilization	OECD	2023
IT investments as a share of gross capital formation	Data utilization	OECD	2022
Number of new data-sector startups	Innovation impacts of the data economy	Crunchbase	2022
Funding received by data startups	Innovation impacts of the data economy	Crunchbase	2022
Data and AI-related patents	Innovation impacts of the data economy	OECD	2021

suring the conditions fostered by the public sector, how data is utilized, and the measurable economic impacts of data. The chosen indicators aim to capture the state and development of these areas in a reliable and comparable manner. In addition, factors such as data availability, quality, and international comparability are key criteria for inclusion.

The selection of indicators involves several practical challenges, which we discuss in turn.

The prerequisites of the data economy, the indicators focus on public-sector measures that support the use of data and artificial intelligence in various forms. However, many initiatives that strengthen the conditions for data-driven business are not unambiguously positive for citizens. A concrete example is the European data protection regulation (GDPR), which has improved the privacy of European citizens but, according to some studies, may have hindered the growth of European data-driven business (Blind et al. 2022). Similarly, in many authoritarian countries, the prerequisites for the data economy—viewed in terms of the availability and exploitation of personal data—are excellent, but from the citizens' perspective, such conditions are hardly desirable.

A key challenge related to the dimension of data utilization is that it is difficult to separate the data economy from the broader software business, as they complement each other. It is hard to find examples of data-economy innovations that are not built on software activities. Consequently, companies that create business from data typically invest heavily in software development and employ numerous software developers. For this reason, the national accounting framework underlying GDP calculation records both software and data investments under the same subcategory, "Computer software and databases." Thus, in traditional national economic accounting, it is not possible to distinguish the value of software investments from that of the datasets held by firms.

Furthermore, both software and data are widely considered general-purpose technologies[2] that can be utilized across nearly all sectors. Limiting the analysis to IT and data-sector firms, patents, or financing alone may therefore provide an incomplete picture of the entire data economy. In this research note, we aim to measure the data economy directly, but where no direct data-economy indicators are available, we rely on general IT-industry measures.

The challenges for innovation indicators mirror those for utilization. Differentiating the data economy from general software business is not straightforward. For patents and startups, the distinction is inevitably somewhat arbitrary. Nonetheless, the selected indicators are designed to measure the data economy from multiple angles. By averaging a large number of indicators, we gain a better understanding of the state of the data economy across countries than would be possible using individual measures alone.

Considering the challenges described above, the Databarometer should be viewed more as an indicative, ordinal measure than as a final truth about the data economies of the countries under comparison.

2.2 Selection of comparison countries

The countries included in the comparison were chosen based on three criteria. First, we selected small, high-income countries (e.g., the Netherlands and Austria) and Finland's neighbors (Sweden, Estonia). In addition, the comparison includes the three largest EU economies (Italy, France, Germany) as well as established large industrial nations (Japan, the United Kingdom, the United States, and China).

The above-mentioned selection of indicators heavily influences the choice of comparison countries. While we strive to include as many countries as possible, not all indicators are available for major economies—especially the United States and China. Due to missing data points, countries such as India, South Korea, and those in South America have been excluded from this analysis.

2.3 Aggregation of Indicators

The aggregation of various measures into dimensions and, ultimately, into a composite indicator follows the same approach used in ETLA's Digibarometer. We employ z-scores to standardize the variables. A z-score is calculated by subtracting the mean of all countries' values from a given country's value and then dividing by the standard deviation. This standardization sets each variable's mean to 0 and its standard deviation to 1.

To prevent extreme values from distorting the results, we apply winsorization. In this procedure, outliers that exceed +/- 2 standard deviations are set to -2 or +2. After this adjustment, the z-scores are recalculated based on the revised values.

Dimension scores are formed by summing the standardized variables and then scaling these results to a range of 1–100. A country that performs worst in all variables receives a score of 1, while a country that performs best in all variables receives a score of 100.

The index value is calculated as follows:

$$Index\ value\ =\ 99\ x\ \frac{(Sum\ of\ variables\ per\ country - Sum\ of\ minimum\ values)}{(Sum\ of\ maximum\ values - Sum\ of\ minimum\ values)} + 1.$$

If a country excels in all dimensions, the numerator and denominator are identical, yielding an index value of 100. Conversely, if a country is the poorest performer across all dimensions, the index value is 1.

3 Databarometer results

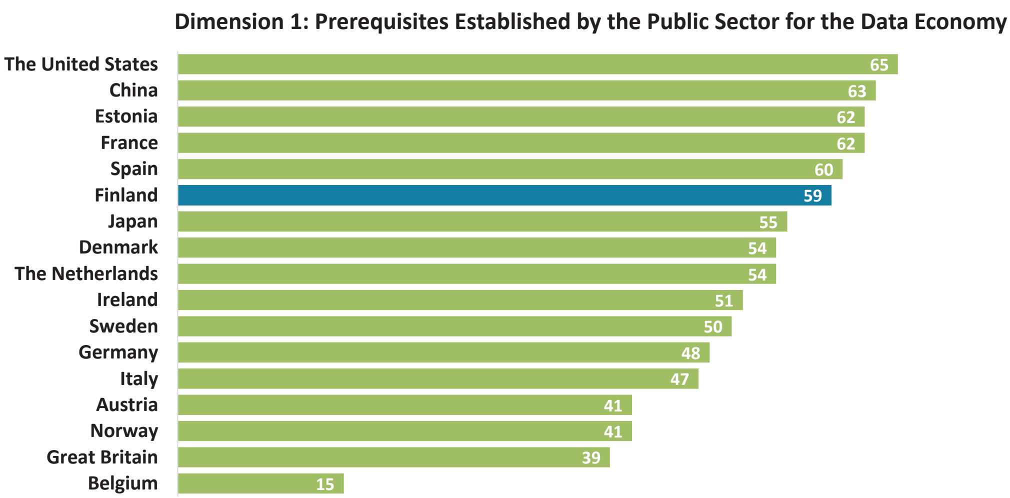
Figure 1 presents each dimension of the Databarometer separately. The United States leads in the prerequisites of the data economy, reflecting the high share of skilled professionals in its workforce and public initiatives through which the U.S. has strengthened its position in recent years. In terms of prerequisites, Finland scores well by European standards, nearly at the top in Europe.

Figure 2 illustrates data utilization. On these measures, European economies fare relatively well; Finland ranks third among the comparison countries. Generally, countries that excel in data utilization are small, high-skill economies.

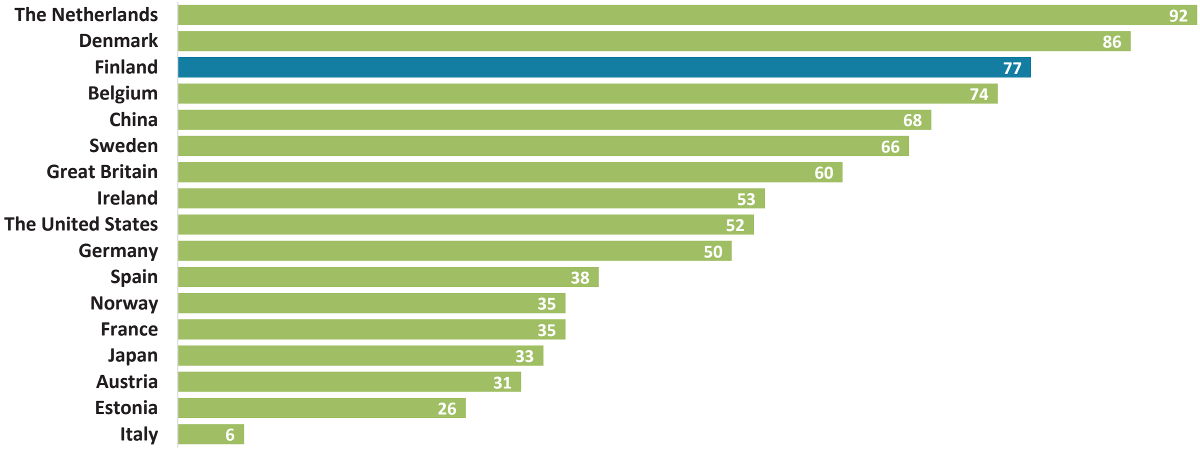
When it comes to the innovation impacts of the data economy, the United States stands in a league of its own. This is largely because the lion's share of data-sector startups and their funding originate in the U.S. As for patents, China and the United States are roughly on par.

Figure 2 shows the aggregated results. The United States leads overall, followed by China and the Netherlands. Finland also scores well in the overall index, maintain-

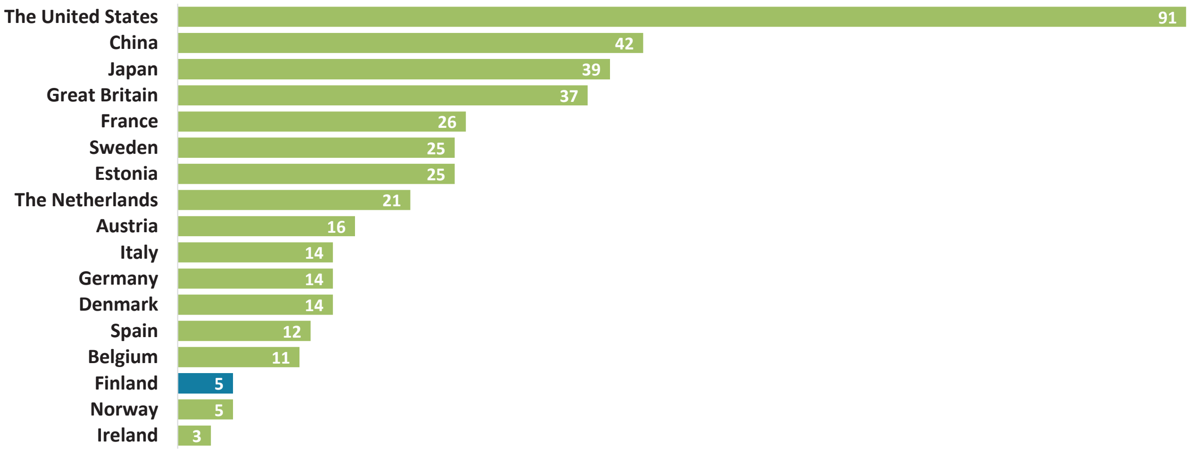
Figure 1 Databarometer country comparison by dimension



Dimension 2: Utilization of data

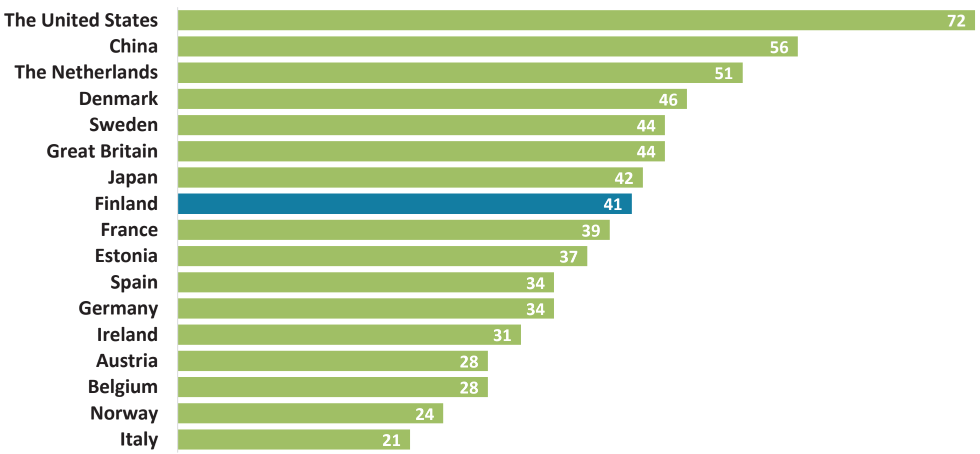


Dimension 3: Innovation Impacts of the Data Economy



Source: Authors' calculations.

Figure 2 Databarometer country comparison composite index



Source: Authors' calculations.

ing a solid European level. From a Finnish perspective, however, it is concerning that Finland ranks particularly low in the innovation impacts dimension. This dimension is crucial for both national productivity and citizens' well-being.

4 Conclusions

There is no unambiguous measure for assessing the maturity of the data economy across countries. Any choice of indicators is subject to interpretation. In this research note, we construct a composite indicator composed of several sub-indicators that measure the data economy's maturity from different angles. By using multiple sub-indicators, we aim to achieve a more comprehensive and balanced picture of the data economy's state, since no single measure can capture all the essential factors.

Summarizing the country comparisons, European countries—including Finland—perform reasonably well in terms of the prerequisites for the data economy and data utilization. When looking at innovation-oriented indicators, however, the United States stands apart, especially regarding the number of data-sector startups and the funding they attract. This suggests that beyond technology and infrastructure, regulatory and funding conditions are also critical to success in the data economy.

It is also worth noting that there is no single "correct" way to aggregate different dimensions. For example, Finland outperforms Sweden in terms of the public-sector conditions enabling the data economy and in companies' data use. Yet from the perspective of citizens' well-being, innovation, and productivity, innovation impacts may be more important than prerequisites and utilization. Moreover, a single snapshot does not allow us to draw far-reaching conclusions about the extent to which prerequisites and utilization support innovation. For this reason, comparing the sub-indices is generally more informative than focusing solely on the overall index.

Endnotes

1 ETLA has published the Digibarometer ("Digibarometri"), a comparison of countries' digital capabilities, for a total of ten years. More information is available at http://www.etla.fi/digibarometri (accessed 30 September 2024).

2 General-purpose technologies (GPTs) are innovations that have broad application potential and a significant impact across multiple industries and on overall economic growth (Bresnahan & Trajtenberg, 1995). Examples include electricity and microprocessors.

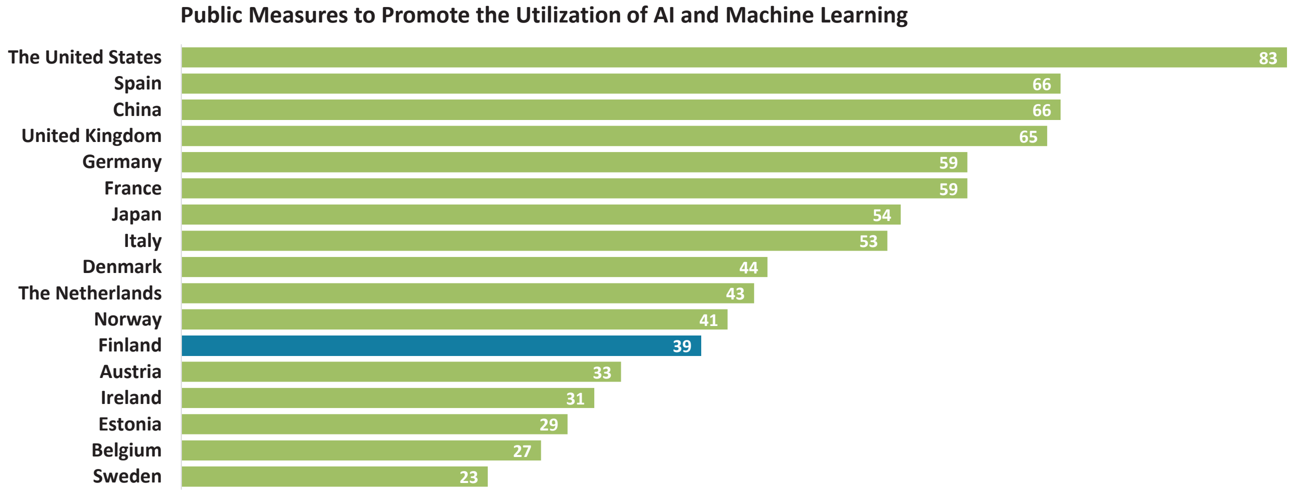
References

Blind, K., Niebel, C., & Rammer, C. (2022). The impact of the EU General Data Protection Regulation on innovation in firms. *ZEW-Centre for European Economic Research Discussion Paper*, (22-047).

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? *Journal of econometrics, 65*(1), 83–108.

Appendix: Databarometer underlying variables

Prerequisites established by the public sector for the data economy

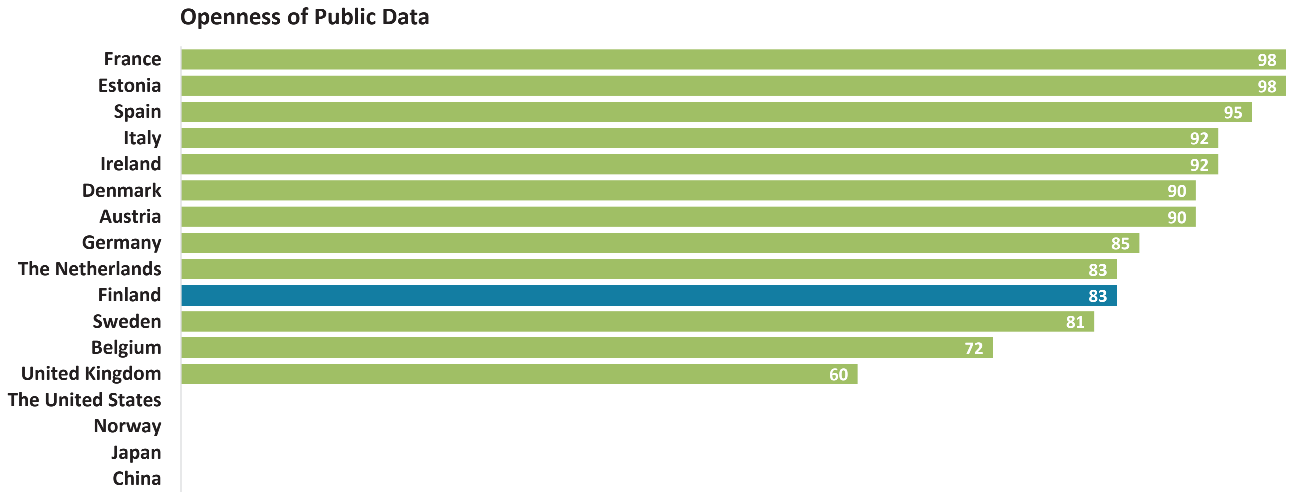


Index value from zero to one hundred, where a higher score indicates that public policies and initiatives consistently support the integration of AI and machine learning into the economy.

The source is the Global AI Index and its Government Strategy sub-index, compiled by Tortoise Media. Year: 2022.

Source: Tortoise Media (https://www.tortoisemedia.com/intelligence/global-ai/)

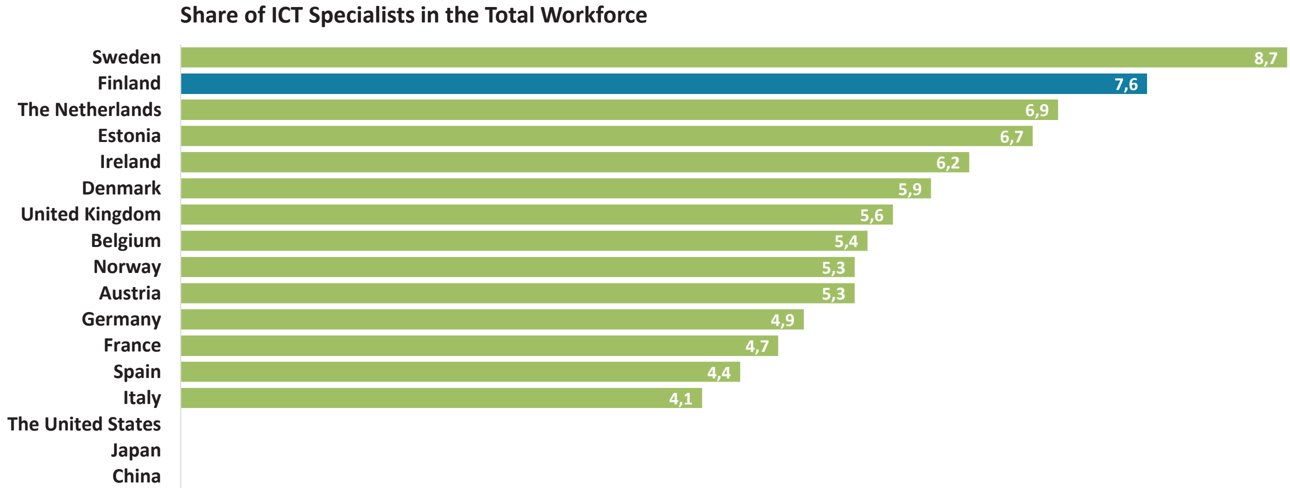
Prerequisites established by the public sector for the data economy



Index value from zero to ten, where a higher value indicates a stronger public-sector commitment to open data.

Source: European Commission Open Data Maturity Report, data from 2023, except for United Kingdom (2020). (https://data.europa.eu/en/publications/open-data-maturity/)

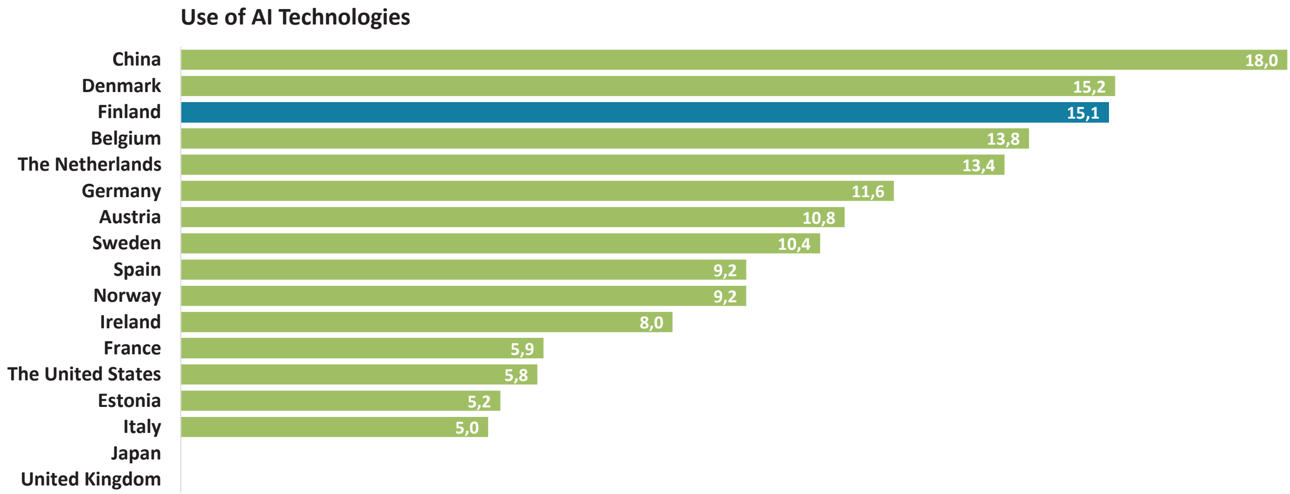
Prerequisites established by the public sector for the data economy



Proportion of ICT specialists in the total workforce, as reported by labor force surveys.

Source: Eurostat, data from 2023, except for United Kingdom (2020).

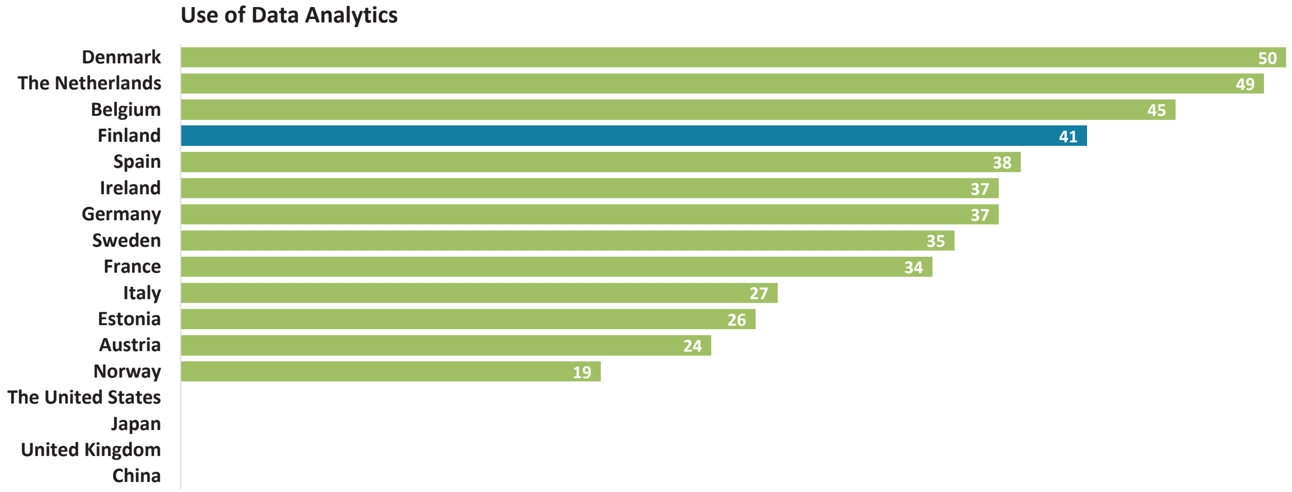
Data utilization



Share of companies that use AI technologies.

Source: For EU countries Eurostat, for the United States, McElheran et al. (2023) (https://www.nber.org/papers/w31788); and for China, SAS Institute (https://www.sas.com/pt_pt/news/press-releases/2024/july/genai-research-study-global.html). Data is from 2023 for the EU and the United States, and from 2024 for China.

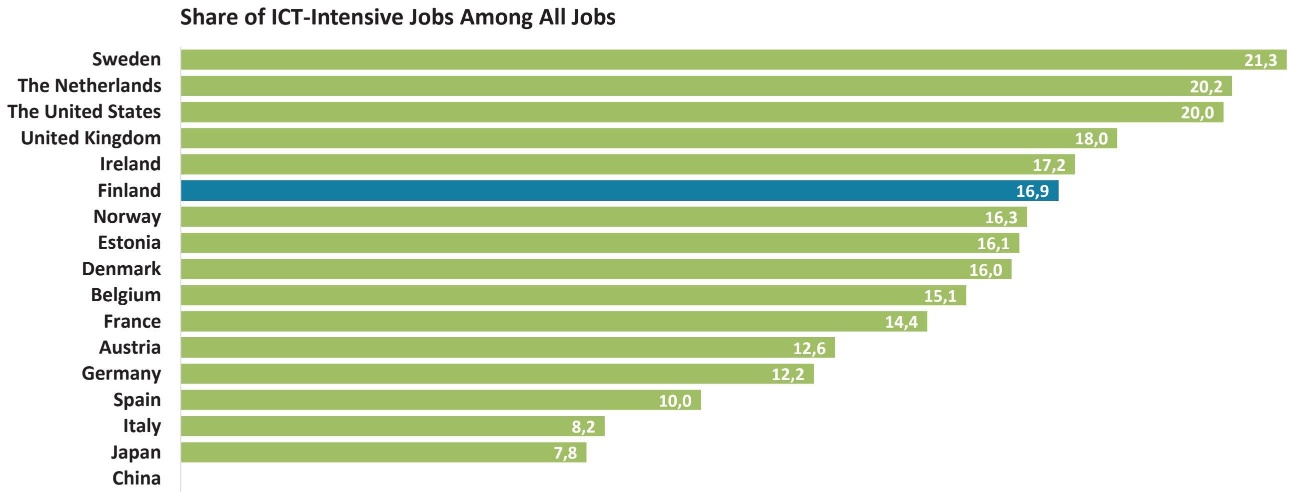
Data utilization



Percentage of companies that utilize big data analytics.

Source: Eurostat, data from 2023.

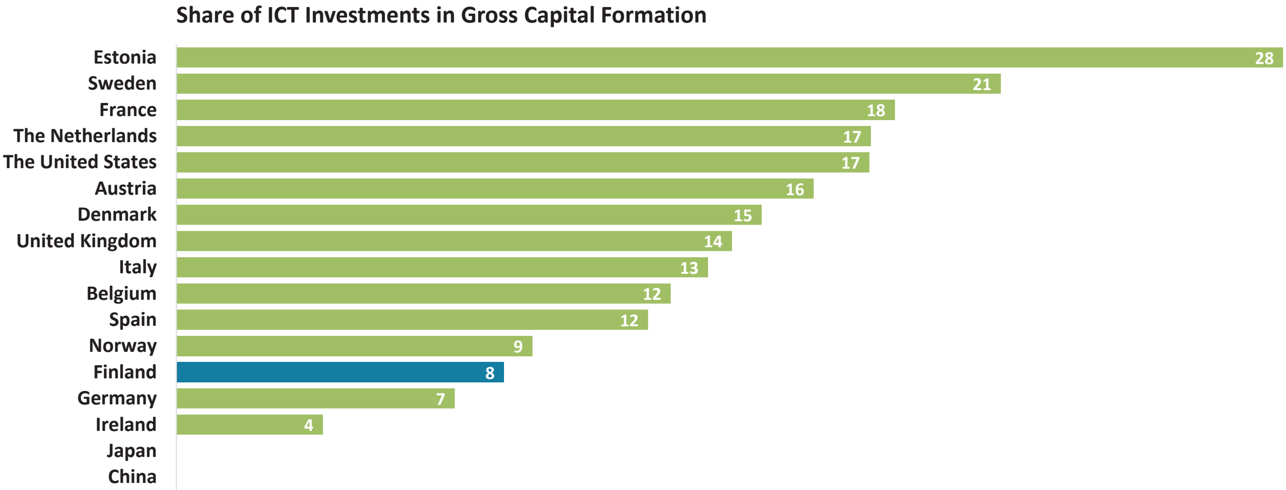
Data utilization



Proportion of ICT-intensive occupations among all employees.

Source: OECD, data from 2023.

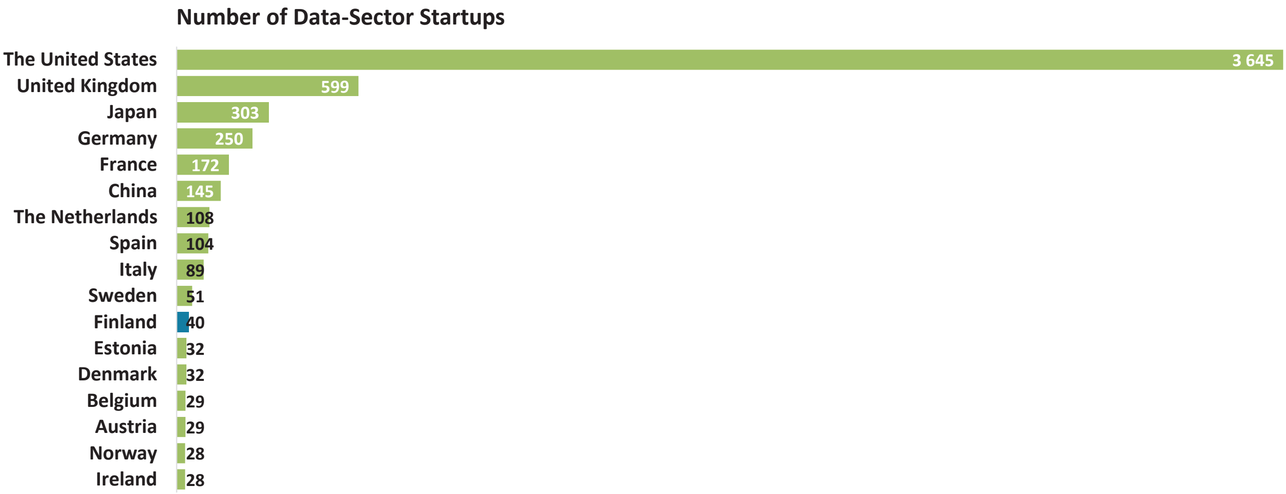
Innovation impacts of the data economy



Proportion of ICT investments in total investments, net of depreciation.

Source: OECD.

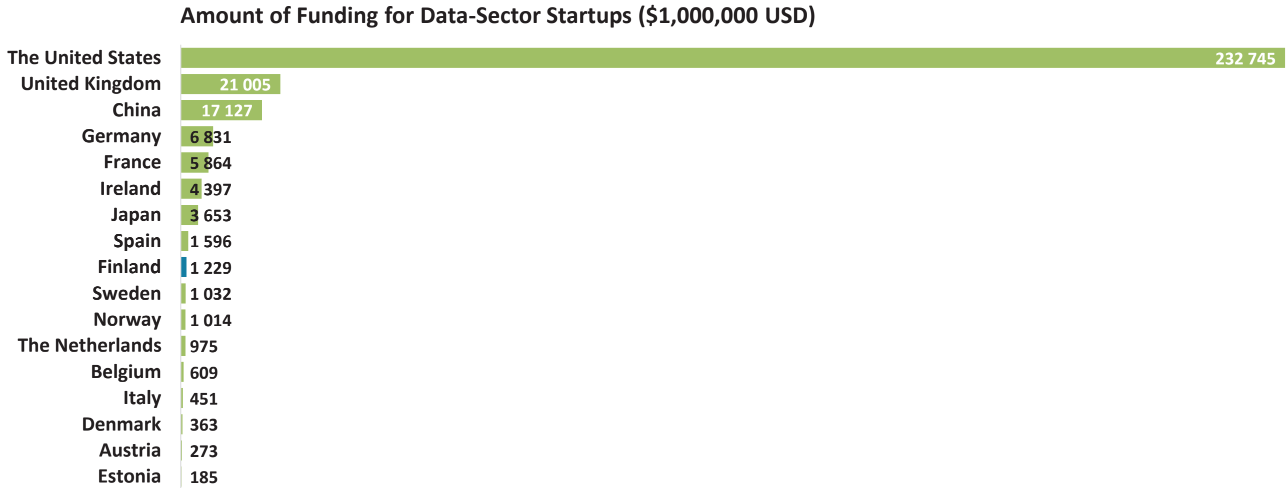
Innovation impacts of the data economy



Number of startups founded in 2023 in categories related to artificial intelligence, analytics, big data, data mining, machine learning, and natural language processing.

Source: Crunchbase.

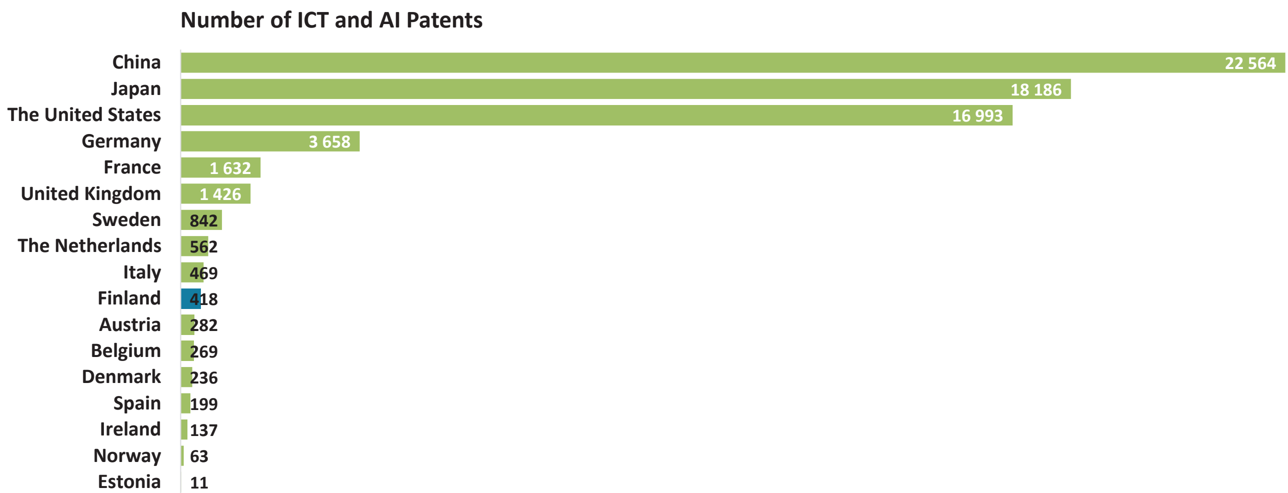
Innovation impacts of the data economy



Amount of funding received by startups in 2023 in categories related to artificial intelligence, analytics, big data, data mining, machine learning, and natural language processing, in U.S. dollars.

Source: Crunchbase.

Innovation impacts of the data economy



Number of international patents in the IP5 patent family in 2023. Includes patents in the fields of ICT and artificial intelligence.

Source: OECD.

Elinkeinoelämän tutkimuslaitos

ETLA Economic Research

ISSN-L 2323-2463
ISSN 2323-2463

Publisher: Taloustieto Oy

Tel. +358-9-609 900
www.etla.fi
firstname.lastname@etla.fi

Arkadiankatu 23 B
FIN-00100 Helsinki