Defining and Measuring the UK Digital Economy Phase 2

Final Report – May 2025

Department for
Science, Innovation
& Technology

cambridge
econometrics
clarity from complexity

Authorisation and version history

Version	Date	Authorised for release by	Description
1.2	07/05/2025	Adam Brown	Addresses final feedback
1.1	24/04/2025	Adam Brown	Addresses feedback to the first draft; addition of Executive Summary
1.0	31/03/2025	Adam Brown	First draft of final report

Contents page

Executive Summary

The Department for Science, Innovation & Technology (DSIT) is currently undertaking new research to define and measure the Digital Economy (DMDE), split into two phases of work. Phase 1, completed in May 2024, focused on developing a definition and application framework, informed by an in-depth review of relevant literature, existing frameworks, and extensive stakeholder engagement.

Introduction

The Phase 2 work, delivered collaboratively by Cambridge Econometrics, The Innovation and Research Caucus, The Data City, and DSIT, uses the framework and recommendations from the Phase 1 work to:

1) Identify and implement a new, overarching, Tier-based Standard Industrial Classification (SIC) definition of the Digital Economy, that allows comparability and consistency with national statistics and accompanying statistical products/datasets, and:

2) Integrate a web-scraping/keywords-based approach (using The Data City's Real-Time Industrial Classifications (RTIC) methodology) within this definition to identify subsectors/technologies (and associated groupings) within the Digital Economy not identifiable using SIC codes.

Using this new framework, the primary objective of the Phase 2 work has been to create an initial enterprise-level database and generate insights into the size and state of the UK Digital Economy.

Definitions and Methodology

Defining the Digital Economy

The Phase 1 framework defines the Digital Economy as "all economic activity derived from the production of Digital content, ICT goods, or ICT services." Within this framework, Digital economic activity is allocated into Tiers:

- **'Tier 1' of the Digital Economy** refers to economic activity derived from **enterprises that are primarily** involved in the **production of Digital content and/or ICT goods and/or ICT services**.

- **'Tier 2' of the Digital Economy** refers to economic activity derived from diversified **enterprises** which produce Digital content and/or ICT goods and/or ICT services as part of a broader **diversified product or service offer** and which have a **higher level of Digital intensity in their inputs than enterprises in the traditional economy.**

Digital economy SIC codes

To support the implementation of this framework, the Phase 2 consortium has worked collaboratively with DSIT to explore a new SIC-based definition of the Digital Economy informed by the Phase 1 Tier-based framework and recommendations, shown in the table below.

SIC Code(s)	SIC Description	Tier
18.20	Reproduction of recorded media	1
26.1 - 26.4	Manufacture of computer, electronic and communication equipment	1

58.14 & 58.2	Publishing of journals, periodicals and software	1
59	Motion picture, video and television programme production, sound recording and music publishing activities	1
60	Programming and broadcasting activities	1
61	Telecommunications	1
62	Computer programming, consultancy and related activities	1
63.11 - 63.91	Information service activities	1
95.11	Repair of computers and peripheral equipment	1
66	Activities auxiliary to financial services and insurance activities	2
69	Legal and accounting activities	2
92	Gambling and betting activities	2

Source: Department for Science, Innovation & Technology, Cambridge Econometrics, Innovation and Research Caucus, The Data City.

Using a clear and robust methodology underpinned by established, official data sources, our proposed approach classifies Digital enterprises according to their primary product. Tier 1 enterprises are identified as those within SIC codes that directly produce Digital content or ICT goods and services.

Tier 2 enterprises meanwhile are determined using input-output tables, which summarise the inputs to production for each industry. Enterprises in this category primarily produce non-ICT goods or services but have an input profile with a Digital intensity similar to Tier 1 enterprises.

Supporting this overarching SIC-based definition and ensuring the classification of key Digital industries and technologies, especially those not readily available using a standard SIC approach, is the integration of The Data City's proprietary RTIC framework to the database and economic estimates.

The taxonomy presented in   has been adopted for the Phase 2 work, which uses RTICs in conjunction with the SIC-based definition in Table 1 to provide a complete classification framework for the Digital Economy, while retaining the ability to 'deep-dive' into Digital technologies/subsectors of interest.

Table 2

Subsector taxonomy of the Digital Economy

	Enterprises with Tier 1 SICs	Enterprises with Tier 2 SICs	Enterprises without Tier 1 or 2 SICs
Enterprises with RTICs	**"Frontier dedicated"** These enterprises represent the frontier of Digital content generation and/or production of ICT goods/services. Compared to other Digital producers, they are more likely to be developing and/or leveraging emerging Digital technologies (such as AI, Cyber, and Quantum).	**"Agile diversified"** Includes those diversified enterprises at the Digital frontier, who are more likely to be developing and/or leveraging emerging Digital technologies as part of their diversified product or service.	Excluded, but reviewed for quality assurance/validation

Enterprises without RTICs	"Established dedicated" Captures all remaining enterprises involved in the dedicated production of Digital content and/or ICT goods/services, particularly those in established ICT industries that are typically less engaged with emerging Digital technologies.	"Traditional diversified" Represents all remaining diversified enterprises i.e. those that produce Digital content and/or ICT goods/services as part of a broader diversified product or service.	Excluded from analysis

Source: Department for Science, Innovation & Technology, Cambridge Econometrics, Innovation and Research Caucus, The Data City.

The database developed as part of the Phase 2 work encompasses data collected from the following datasets, available as of January 2025:

Data sources and processing

- Inter-Departmental Business Register (IDBR) extract: 390,000 enterprises and accompanying local sites.

- Real-Time Industrial Classifications (RTIC) dataset from The Data City:

    o RTIC sectors identified as of January 2025.

    o Lightcast dataset on job postings: between January 2020 and January 2025.

    o Dealroom dataset on venture capital (VC) funding: between January 2010 and January 2025.

- UK Research and Innovation (UKRI): 101,000 projects successfully funded since January 2000.

- UK Data Service (UKDS) Quarterly Labour Force Survey (LFS) extract from January to December 2024.

- ONS Output per job estimates: by SIC section from 2023 (to estimate enterprise-level GVA)

These datasets have been linked together to arrive at a complete, up to date enterprise-level database and economic estimates for the UK Digital Economy.

To ensure these estimates meet the required statistical standards, extensive quality assurance and data validation processes have been embedded into the methodology and approach.

Key Findings

This has included prioritising official data sources; using established data processing and analysis methods; cross-validating estimates against related datasets/ publications, and finally; independent verification from an academic expert.

Summary economic estimates for the UK Digital Economy (2024)

The following estimates consider the economic contribution of the UK Digital Economy - using the new framework and definition adopted for the Phase 2 work - across key measures such as gross value added, productivity, employees (and their characteristics), and funding (both public and private).

Digital economy taxonomy	GVA (£m, current prices)	GVA per employee (£, current prices)	Employees	Enterprises
Frontier dedicated	60,518	89,007	679,927	25,563

Established dedicated	74,629	89,050	838,058	175,283
Total Tier 1	**135,147**	**89,031**	**1,517,985**	**200,846**
Agile diversified	40,938	111,997	365,528	5,591
Traditional diversified	110,240	105,082	1,049,080	105,359
Total Tier 2	**151,178**	**106,869**	**1,414,608**	**110,950**
Total Digital Economy	*286,325*	*97,636*	*2,932,593*	*311,796*
*DSIT Digital Sector Economic Estimates**	*153,483 (2023)*	*Not available*	*1,882,622 (2022)*	*Not available*

Source: Department for Science, Innovation & Technology, Cambridge Econometrics, Innovation and Research Caucus, The Data City. Note: * DSIT's current Digital Sector Economic Estimates are included here for reference. Due to the differences in underlying methods, definitions, data, and timeframes, these will be different to the economic estimates derived as part of the Phase 2 work and are only broadly comparable with Tier 1 estimates.

Gross value added (GVA) and productivity

- The UK Digital Economy is estimated to have total annual GVA of £286 billion (in current prices, as of 2024). This equates to approximately 13% of total annual UK national accounts GVA.

- Output (GVA) per employee in the UK Digital Economy is estimated at £97,600 as of 2024. This compares to a whole economy value of £62,300 (2023 estimate, the latest comparable value available).

- This means that the average Digital Economy employee is some two-thirds (57%) more productive than average UK employee.

Employees and employee characteristics

- Across the UK, it is estimated there are a total of 2.93 million employees working in the Digital Economy (as of 2024).

- Two-thirds (66%) of Digital Economy employees in the UK work in professional and associate professional occupations (as of 2024), compared to 39% in non-Digital Economy industries.

- Employees working in the UK Digital Economy are estimated to have a median hourly wage of £23.40 (in current prices, as of 2024) and median annual earnings of £39,500.

Business demography and funding

- This is 50% higher than the median hourly wage of £15.60 for employees working in non-Digital Economy industries, and 60% higher than the median annual earnings of £24,500.

- There are 311,800 enterprises currently active in the Digital Economy (as of 2024). In addition to this, there are 349,200 local sites associated with Digital Economy enterprises.

- A quarter (25%) of UK Digital Economy enterprises have an annual turnover below £100,000, which is significantly lower than the average of 34% for all UK industries during the same period. Meanwhile 10% generate turnovers over £900,000.

- Some 88% of UK Digital Economy enterprises employed fewer than 10 people in 2024, similar to the UK average for all industries of 89%, while 2% have more than 50 employees.

- Digital economy enterprises in the UK have received an estimated £72 million of UKRI funding per annum (2000-2025), and £4,065 million of VC funding per annum (2010-2025).

Introduction

The concept of the "Digital" sector encompasses a wide range of overlapping but conceptually distinct areas. Several definitions exist across government and industry, notably the 'Digital Sector Economic Estimates' published by the Department for Science, Innovation & Technology (DSIT).[1] As evaluating the impact of emerging industries and technologies across the economy becomes increasingly central to DSIT's policy agenda, a more comprehensive and granular approach to defining and measuring the Digital Economy (DMDE) is required.

In response, DSIT commissioned new research on DMDE, structured into two phases. Phase 1, completed in May 2024, focused on developing a definition and application framework. This was informed by an in-depth review of relevant literature, existing frameworks, and extensive stakeholder engagement, culminating in the publication of a methodology and recommendations paper.[2]

Following these recommendations from Phase 1 and further analysis conducted by DSIT, the Phase 2 work – presented here – proposes a new Standard Industrial Classification (SIC)-based over-arching definition of the UK Digital Economy, allowing harmonisation with national statistics produced by the Office for National Statistics (ONS). Unlike previous definitions, these SICs have been organised into tiers of digital activity for greater clarity.

Although a SIC-based definition of the Digital Economy helps align with national statistics, SICs are updated infrequently and are not well-suited to capturing the economic impact of frontier and emerging industries. To address this, Phase 1 recommends a dynamic framework that integrates keyword-based classification (e.g., through enterprise-level data linking).

This has also been incorporated as part of the Phase 2 work and will enable the identification of rapidly evolving, niche, and novel Digital industries, ensuring DSIT's policy development remains adaptable and forward-looking.

Using this new framework, the primary objective of the Phase 2 work has been to create an initial enterprise-level database and generate insights into the size and state of the UK Digital Economy. The database and estimates are built upon the Phase 1 methodology and recommendations by integrating:

1) The implementation of an overarching, Tier-based SIC definition, that allows comparability and consistency with national statistics and accompanying statistical products/datasets, and:

2) The linking of a keywords-based approach (using The Data City's RTIC methodology) to identify subsectors/technologies (and associated groupings) within the Digital Economy not identifiable using SIC codes.

This report presents a technical summary of the methodology and approach used to produce estimates that

---

[1] See: Digital Sector Economic Estimates Series - GOV.UK. The DMDE Phase 1 report presents and appraises a wide range of related definitions.
[2] See: Defining and Measuring the UK Digital Economy

illustrate the size and state of the UK Digital Economy. It concludes with an initial review of emerging findings and analysis.

These estimates are based on the methodology and framework recommended in Phase 1, which has since been implemented and expanded by DSIT with the expert advice and direction of the Phase 2 consortium (comprising Cambridge Econometrics, The Innovation and Research Caucus, and The Data City). The Phase 1 report provides a detailed account of the process used to develop this framework, along with an assessment of its strengths and limitations.

The chapters in this report are structured as follows:

- **Methodology and Approach:** outlines the approach, data sources and definitions used to develop the enterprise-level database and estimates and associated quality assurance (QA) and validation undertaken.

- **Economic Estimates Summary:** presents summary estimates on the economic contribution of the UK Digital Economy, across key measures such as gross value added, productivity, employees (and their characteristics), and funding.

- **FAQs:** provides answers to some of the common technical queries related to the Phase 2 work.

- **Appendices:** includes supporting methodology detail and sources.

Abbreviations and Terminologies

Abbreviation	Definition
ABS	Annual Business Survey
BRES	Business Register and Employment Survey
CRN	Company Registration Number
DMDE	Defining and Measuring the Digital Economy
DSIT	Department for Science, Innovation & Technology
GVA	Gross Value Added
IDBR	Inter-Departmental Business Register
ITL	International Territorial Level
LFS	Labour Force Survey
ONS	Office for National Statistics
QA	Quality Assurance
RTIC	Real-Time Industry Classifications
SIC	Standard Industrial Classification*
SOC	Standard Occupational Classification*
TDC	The Data City

VC	Venture Capital

* for the purpose of the Phase 2 work, this refers specifically to Standard Industrial Classification 2007 / ISIC Rev. 4 and Standard Occupational Classification 2020.

Terminology	Definition
Company	A legal entity identified by its Company Registration Number (CRN).
Employees	A measure of full and part time employees (not workers).
Enterprise	A business entity recorded in the IDBR dataset, potentially linked to multiple companies.
GVA	Gross Value Added, a measure of economic output.
Head Site	The primary site of an enterprise where all turnover and employment are registered.
Local Site	Any site of an enterprise that is not the head site.
Productivity	Output (GVA) generated per employee.
Sites	Physical locations listed in the IDBR dataset as part of a single enterprise.

Methodology and Approach

This chapter describes the approach used to develop the enterprise-level database and estimates as part of the Phase 2 work. This includes a high-level summary of how the Digital Economy is defined, accompanying data sources, the data processing methods used, and data QA and validation undertaken. Supporting detail is provided in the Appendices of this report.

Reflecting the recommendations of the Phase 1 work, the definition of the Digital Economy applied here uses an overarching SIC definition, combined with a novel, keywords-based subsector taxonomy.

Defining the Digital Economy

Overarching SIC definition

The Phase 1 framework defines the Digital Economy as "all economic activity derived from the production of Digital content, ICT goods, or ICT services." Within this framework, Digital economic activity is allocated into Tiers:

- **'Tier 1' of the Digital Economy** refers to economic activity derived from **enterprises that are primarily** involved in the **production of Digital content and/or ICT goods and/or ICT services**.

- **'Tier 2' of the Digital Economy** refers to economic activity derived from diversified **enterprises** which produce Digital content and/or ICT goods and/or ICT services as part of a broader **diversified product or service offer** and which have a **higher level of Digital intensity in their inputs than enterprises in the traditional economy.**

To support the implementation of this framework, the Phase 2 consortium has worked collaboratively with DSIT to explore a new SIC-based definition of the Digital Economy informed by the Phase 1 Tier-based framework and recommendations.

Using a clear and robust methodology underpinned by established, official data sources, our proposed approach classifies Digital enterprises according to their primary product. Tier 1 enterprises are identified as those within SIC codes that our analysis has found to directly produce Digital content or ICT goods and services.

Table 1

Tier 2 enterprises meanwhile are determined using input-output tables, which summarise the inputs to production for each industry. Enterprises in this category primarily produce non-ICT goods or services but have an input profile with a Digital intensity similar to Tier 1 enterprises.

This classification forms the basis of the Tier-based SIC definition of the Digital Economy, presented in below, which has been applied in Phase 2 to develop the enterprise-level database and estimates. Additional detail on the theoretical basis, supporting data and calculations, and QA undertaken for the SIC-based definition can be found in the Appendices.

As highlighted in the Phase 1 report, this definition excludes a potential third Tier: Digital adopters – enterprises that rely on Digital inputs but do not produce Digital content, ICT goods, or ICT services. Since these enterprises fall outside the scope of the Phase 1 defined Digital Economy, they have also been excluded from the Phase 2 work.

SIC Code(s)	SIC Description	Tier
18.20	Reproduction of recorded media	1
26.1 - 26.4	Manufacture of computer, electronic and communication equipment	1
58.14 & 58.2	Publishing of journals, periodicals and software	1
59	Motion picture, video and television programme production, sound recording and music publishing activities	1
60	Programming and broadcasting activities	1
61	Telecommunications	1
62	Computer programming, consultancy and related activities	1
63.11 - 63.91	Information service activities	1
95.11	Repair of computers and peripheral equipment	1
66	Activities auxiliary to financial services and insurance activities	2
69	Legal and accounting activities	2
92	Gambling and betting activities	2

Source: Department for Science, Innovation & Technology, Cambridge Econometrics, Innovation and Research Caucus, The Data City. See Appendices for additional detail on the theoretical basis, supporting data and calculations, and QA undertaken.

Subsector taxonomy

As outlined in the Phase 1 report, a "key strength of the revised approach is that it allows for granular measurement of key industries and technologies in a flexible manner, using a classification that will be updated periodically as the nature of the Digital Economy evolves".

The classification of key Digital industries and technologies, especially those not readily available using a standard SIC approach, has been addressed by the Phase 2 consortium by integrating The Data City's proprietary RTIC framework to the database and economic estimates.

What are Real Time Industrial Classifications (RTICs)?

The Data City's industrial classifications (Real Time Industrial Classifications or RTICs) are generated with a proprietary methodology that groups companies considering the way they describe themselves on their website. The Data City has built a unique platform (The Data Explorer) that combines company financials from Companies House, CreditSafe, Dealroom, InnovateUK and 360Giving with up to 75 pages of website text per company for more than 1.6 million companies.

RTICs provide a means of capturing 'true' company activity, especially for frontier sectors and technologies.

For additional information on RTICs and their underlying method, sources, and QA, as well as a full list of currently defined RTICs, see the Appendices.

The use of an RTIC framework has been prioritised over other keywords-based approaches for a number of reasons, including:

- its immediate availability and existing quality assurance and validation;

- widespread use and understanding in DSIT and other government departments (especially in areas relating to Digital subsectors and technologies), and;

- ability to quickly adapt and respond to future emerging digital subsector and technologies.

The Data City currently have more than 60 RTICs defined, quality assured and validated (many done so in conjunction with central government departments such as DSIT). This includes RTICs representing Digital industries and technologies of significant policy interest to DSIT, including Artificial Intelligence (AI), Cyber, Data and Software (a full list of RTICs used in the Phase 2 work is provided in Table A.5 in the Appendices).

RTICs are by design focused on frontier companies in emerging industries and technologies. Such companies are traditionally overlooked by other classifications such as SICs yet are of significant economic and policy interest. This means companies with an RTIC, by their nature, represent only a small number of total companies, even within the broader Digital Economy.

Though these RTICs can be reported individually (to better understand individual subsectors/technologies), they will not provide a complete profile of every company's activity in the broader Digital Economy. This follows from companies not necessarily belonging to an RTIC. The taxonomy presented in Table 2 has therefore been adopted for the Phase 2 work, which uses RTICs in conjunction with the SIC-based definition to provide a complete classification framework for the Digital Economy, while retaining the ability to 'deep-dive' into Digital frontier subsectors/technologies of interest.

Table 2: Subsector taxonomy for the Digital Economy

	Enterprises with Tier 1 SICs	Enterprises with Tier 2 SICs	Enterprises without Tier 1 or 2 SICs
Enterprises with RTICs	"Frontier dedicated"	"Agile diversified"	Excluded, but reviewed for QA/validation
Enterprises without RTICs	"Established dedicated"	"Traditional diversified"	Excluded from analysis

Source: Department for Science, Innovation & Technology, Cambridge Econometrics, Innovation and Research Caucus, The Data City.

This taxonomy, which has been informed by the framework and terminology identified in the Phase 1 report, categorises any enterprise with a Digital Economy SIC code according to the following:

- **Frontier dedicated:** these enterprises represent the frontier of Digital content generation and/or production of ICT goods/services. Compared to other Digital producers, these enterprises are more likely to be developing and/or leveraging emerging Digital technologies (such as AI, Cyber, and

Quantum). *Defined as any enterprise associated with <u>both</u> a Tier 1 SIC and any RTIC.*

- **Established dedicated:** captures all remaining enterprises involved in the dedicated production of Digital content and/or ICT goods/services, particularly those in established ICT industries that are typically less engaged with emerging Digital technologies. *Defined as any enterprise associated with a Tier 1 SIC but currently without an RTIC.*

- **Agile diversified:** includes those diversified enterprises at the Digital frontier, who are more likely to be developing and/or leveraging emerging Digital technologies as part of their diversified product or service. *Defined as any enterprise associated with <u>both</u> a Tier 2 SIC and any RTIC.*

- **Traditional diversified:** represents all remaining diversified enterprises i.e. those that produce Digital content and/or ICT goods/services as part of a broader diversified product or service. *Defined as any enterprise associated with a Tier 2 SIC but currently without an RTIC.*

Any enterprise that does not meet the above categorisations are currently excluded from the database and economic estimates (in the Phase 1 framework, such enterprises are regarded as being within the 'traditional economy').

The taxonomy is mutually exclusive, and collectively exhaustive; that is, an enterprise can only appear in one categorisation at any given time, so enterprises **estimates are not double counted within the taxonomy**. The methodology and sources for how an enterprise is assigned an SIC code and RTIC is provided later in this chapter.

Defining the Digital Economy moving forward

Given the ever evolving and expanding nature of the Digital Economy, it is expected that both the SIC definition and subsector/technology taxonomy will be reviewed on a regular basis. In particular, this will include revisiting the SIC code list (especially those identified as 'edge cases': see the Appendices for more information) and reviewing those codes that may require removal/inclusion as the Digital Economy evolves, as well as defining or adapting RTICs to better capture new subsectors/technologies of interest. Both the SIC framework and subsector taxonomy have been designed in a way to account for this longer-term adaptability.

Data sources

The database developed as part of the Phase 2 work encompasses data collected from the following datasets, available as of January 2025:

- Inter-Departmental Business Register (IDBR) extract: 390,000 enterprises and accompanying local sites.

- Real-Time Industrial Classifications (RTIC) dataset from The Data City:
  - RTIC sector identified as of January 2025.
  - Lightcast dataset on job postings: between January 2020 and January 2025.
  - Dealroom dataset on venture capital (VC) funding: between January 2010 and January 2025.

- UK Research and Innovation (UKRI): 101,000 projects successfully funded since January 2000.

- UK Data Service (UKDS) Quarterly Labour Force Survey (LFS) extract from January to December 2024.

- ONS Output per job estimates: by SIC section from 2023 (to estimate enterprise-level GVA)

The Inter-Departmental Business Register (IDBR) serves as the foundation of the analyses, offering detailed enterprise-level data, including company names, primary and secondary Standard Industrial Classification (SIC) codes, employment figures, turnover, and postcodes. It also provides site-level data, including SIC codes, postcodes, and employment figures. For details on how the IDBR dataset is processed to ensure accurate identification of local sites and correct allocation of turnover, please refer to the Appendices.

The Real-Time Industry Classification (RTIC) system, developed by The Data City, identifies companies within emerging sectors based on information from their websites. RTICs therefore provide valuable insights into companies that may not be accurately reflected by traditional classifications such as SIC sectors. The database also provides estimates for key business metrics, such as job postings and median salary based on Lightcast data, and Venture Capital funding based on Dealroom data. These estimates are available at the SIC code level and enhances the summary of the Digital Economy database. Estimates of turnover, value added and employment in the database also serves as a tool for QA and cross-validation.

UK Research and Innovation (UKRI) successful project funding applications identifies public funding received by companies. If funding allocation among collaborators is unavailable, we distribute the total funding equally among them. **Quarterly Labour Force Survey (LFS) data in 2023 and 2024 identifies median annual and hourly earnings** of full-time and part-time employees in the Digital Economy across occupations and Tiers.

Linking Data

To build the Digital Economy core database, we supplement the IDBR with additional company-level datasets, such as RTICs from Data City and UKRI funding applications. **We link the various datasets by matching companies using their Company Registration Numbers (CRNs).** The IDBR dataset includes a mapping of enterprises to CRNs, but approximately 17% of enterprises lack CRNs. We identify additional CRNs by using fuzzy company name and postcode matching based on the Company House dataset[3], requested through its API. This improved the matching by 4 percentage points, meaning about 13% of enterprises lack CRNs. For more information, please refer to the Appendices.

Many enterprises in the IDBR dataset have multiple CRNs which do not correspond in any way to individual sites of that enterprise. When linking datasets based on CRNs, **we adopt an approach that prioritises completeness, ensuring that the activities of all companies (CRNs) are associated with the correct enterprise.** For example, if an enterprise has two CRNs, any funding received by both CRNs will be credited to the enterprise as a whole. We then distribute the enterprise activities based on employment shares across its sites or distribute activities equally across enterprise sites if employment shares are unavailable.

We append the identified RTIC sectors to the IDBR dataset based on the CRNs of the enterprise. For enterprises with multiple CRNs, and consequently multiple assigned RTIC sectors, **we ensure that the enterprise's activities are included only once in each associated RTIC sector.** Because an enterprise can be associated with multiple RTICs, any estimates at the RTIC level should not be aggregated together, to prevent double counting of enterprise. When aggregating estimates at the Digital Economy-level, we ensure each enterprise's activities are only included once.

A match for approximately 15% of enterprises is found in The Data City's RTIC database. Not every company

---

[3] See: Companies House API Overview

in The Data City's database has an RTIC. RTICs by nature are focussed on the frontier economy, as they have been created to provide insights where conventional SICs cannot. The matched enterprises are then classified as part of frontier sectors, in either frontier dedicated, or agile diversified. As shown in the following chapter, companies with RTICs contribute a significant part to the Digital Economy, given their count.

The total UKRI funding for each company is aggregated based on their CRNs. **For an enterprise with multiple CRNs, any funding received by those companies/CRNs will be credited to the enterprise as a whole.** Since there is no way of identifying the amount of funding received by each local site, we distribute the enterprise's total funding equally across its local sites. For instance, if an enterprise with five sites receives £10,000 in funding, each site is assumed to receive £2,000. This prevents double counting when aggregating across sectors and regions.

We map site postcodes into UK regions using data from National Statistics Postcode Lookup to analyse regional distribution of business activities. [4] The UK is divided into 12 regions: the 9 regions of England (North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East of England, London, South East, and South West), along with Scotland, Wales, and Northern Ireland. Note that the sum of enterprise counts across all regions is higher than the total number of enterprises because an enterprise's local sites can appear in different regions.

Table 3 Dataset linking and data loss mitigation

Dataset linking steps	Number of enterprises	Data loss mitigation
Extract of IDBR dataset	390,285 in IDBR extract*	
Filter for enterprise in Digital Economy	311,676 in Digital Economy	Checked if either primary or secondary SIC codes are in the Digital Economy (as defined in Table 1).
Append with GVA per job by SIC sector	311,676 with estimated GVA	Ensured every Digital Economy SIC code has an accompanying GVA per job estimate.
Identify ITL regions from postcodes	311,676 with regions	For postcodes without a region, identified region using first 3-digits of postcode.
Identify CRNs of enterprises	269,974 with CRNs	Fuzzy match by enterprise/company names. See Appendices for additional detail on this process.
Identify RTICs based on CRNs	43,526 with RTICs	Ensured over 90% of CRNs of Digital Economy enterprises are identified in The Data City database.
Merge with UKRI funding based on CRNs	7,673 with funding information	See above for funding allocation to collaborators and across enterprise's sites.

Source: Cambridge Econometrics. * the overall IDBR extract provided by the DSIT sponsor for the Phase 2 work contained records for enterprises in the databases with Tier 1 and Tier 2 SIC codes, as well as those in 'edge case' SIC codes. See the Appendices for more information on 'edge case' SICs.

In addition to company-level data, **we incorporate additional business metrics for Digital Economy companies from The Data City Data Explorer platform.** These metrics include VC funding, job posting

---

4 See: National Statistics Postcode Lookup UK - data.gov.uk

counts, and median salaries, aggregated at both the SIC sector and taxonomy levels. We identify Digital Economy companies using their CRNs, extract sector- and taxonomy-level statistics, and integrate the results into our database summaries.

Similarly, **we enrich our database summaries with labour force statistics from the LFS dataset.** These include average earnings and hours worked for full-time permanent employees at the SIC sector level. The resulting tables provide a more comprehensive view of the Digital Economy labour market and business environment. However, we acknowledge recent concerns over the representativeness of the LFS dataset.[5]

To control for this during the Phase 2 work, when using the LFS dataset, only aggregate estimates by Tier and 2-digit SOC are provided. In future, it is recommended that the DMDE work considers other approaches to understanding employee characteristics in the Digital sector, such as through the use of the Annual Survey of Hours and Earnings (ASHE).[6]

Experimental estimates of gross value added (GVA) have also been produced, using an identical methodology to DSIT's GVA estimates for the UK Data Driven Market.[7] This approach applies official estimates (2023 values, the latest currently available) of employment-to-GVA ratio at the SIC section level to convert site level employment into GVA (see Table A.6 in the Appendices for more information, including consideration of the limitations of this approach, and recommendations for improvement).

QA and data validation

To ensure the economic estimates of the Digital Economy meet the required statistical standards, extensive QA and data validation processes have been embedded into the methodology and approach. This includes:

- **Prioritising official data sources:** the primary data source, the IDBR database, is an official and comprehensive source of company activity, maintained by the ONS. It is used extensively across government for statistical work and research. Meanwhile, many of the non-IDBR data sources used in the Phase 2 work are based on official datasets (e.g. the LFS), and all have a strong track record use for related research and analysis. Additionally, all data sources have undergone extensive QA and validation separate to that undertaken here.

- **Using established data processing and analysis methods:** this includes employing advanced matching techniques to retain data integrity; preserving data granularity during aggregation; and conducting rigorous testing throughout the merging process. Such methods are widely used in the production of related data and analysis and help to ensure accurate integration of datasets.

- **Cross-validation against related datasets/publications:** throughout the development of the database, estimates have been regularly cross-validated against those from related datasets. This includes official economic estimates of Digital activity from DSITs Digital Sector and Data Driven Markets research, ONS BRES, ABS and Business Count datasets, and The Data City's company-level database.[8] Though it should be emphasised our estimates are unlikely to precisely match such

---

[5] See Statistics from the Labour Force Survey – Office for Statistics Regulation for more information.

[6] DSITs Digital Sector Economic Estimates currently uses ASHE to derive earnings estimates. The access time for the ASHE dataset, which unlike the LFS is not open access, made it unviable to use for this iteration of DMDE work.

7 See 5.3 Gross Value Added in: The UK Data Driven Market - GOV.UK

[8] The Data City database includes estimates for company count, GVA, turnover, and employment at both the SIC sector and taxonomy levels, derived from company registers. However, not all companies report employees every year, and turnover figures may include global revenues for UK-registered multinational organisations. In addition, these aggregate estimates should be interpreted with caution, as they do not account for company

estimates, due to the difference in underlying methods, data, timeframes etc.

- **Independent verification from an academic expert:** Professor Jen Nelles, from the Innovation and Research Caucus, has provided independent, expert QA and validation during the implementation of the Phase 1 framework, and the development and interpretation of the database and estimates.

The QA log for the development of the Phase 2 database and accompanying estimates is provided in the Appendices. The Appendices also details the QA checks and data validation undertaken to define the Tier 1 and 2 SICs, as well those included as part of the RTIC methodology.

group structures or the allocation of turnover by site location. Given these limitations, The Data City database serves as a secondary data source for validating our IDBR database.

Economic Estimates Summary

The following chapter presents estimates on the economic contribution of the UK Digital Economy, across key measures such as gross value added, productivity, employees (and their characteristics), and funding (both public and private). These estimates have been developed using the methodology and sources outlined in the previous chapter, and have undergone extensive QA and validation.

Due to the differences in underlying methods, definitions, data, and timeframes (which are detailed in the previous chapter), the below estimates will be different to those presented in related publications on the Digital Economy (such as DSIT's Digital Sector Economic Estimates). Unless stated otherwise, the below estimates are for 2024, using IDBR enterprise records up to Quarter 4 (Q4) of that year.

£286bn

13%

£97,600

digital economy GVA (2024)

of national accounts GVA (2024)

output (GVA) per job (2024)

By Tier and Taxonomy

Table 4: GVA by Digital Economy taxonomy (2024)

Digital economy taxonomy	GVA (£m, current prices)	Share of UK Digital Economy
Frontier dedicated	60,518	21%
Established dedicated	74,629	26%
Total Tier 1	**135,147**	**47%**
Agile diversified	40,938	14%
Traditional diversified	110,240	39%
Total Tier 2	**151,178**	**53%**
Total Digital Economy	*286,325*	*-*

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics.

- The UK Digital Economy is estimated to have total annual GVA of £286,325 million (in current prices, as of 2024).

- This equates to approximately 13% of total UK national accounts GVA (in current prices) for the same

period (2024).[9]

- As   shows, this is relatively evenly distributed across the Digital Economy with £135,147 million (47%) of GVA in Tier 1, and £151,178 million (53%) in Tier 2.

- In terms of the Digital Economy taxonomy subsectors, traditional diversified enterprises generate 39% of GVA, more than any other category. Frontier dedicated enterprises account for 21% of GVA, despite representing only 8% of enterprises.

Table 4

Digital economy taxonomy	GVA per Employee (£, current prices)	Relative to All Industries*
Frontier dedicated	89,007	143%
Established dedicated	89,050	143%
Total Tier 1	**89,031**	**143%**
Agile diversified	111,997	180%
Traditional diversified	105,082	169%
Total Tier 2	**106,869**	**171%**
Total Digital Economy	*97,636*	*157%*

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics. * this column shows taxonomy productivity relative to UK average for all industries (excluding imputed rental, 2023 values).

- Output (GVA) per employee in the UK Digital Economy is estimated at £97,600 (in current prices) as of 2024. This compares to the UK average of £62,300 for all industries in 2023 (the latest comparable value available). [10]

- Table 5 This means that the average Digital Economy employee is 57% more productive than the average UK employee.

- As   shows, output per employee is higher in Tier 2 of the Digital Economy (£106,900), compared to Tier 1 (£89,000), largely reflecting significantly higher output per employee in finance-related industries present in Tier 2.

By Region
- Given the methodology for calculating GVA, there is less dispersion in output per employee between Table 6 taxonomy subsectors (see the Appendices for more information on why subsectors will show less productivity dispersion under our methodology).

- shows the distribution of Digital Economy GVA by ITL Region. London generates more Digital Economy GVA (£105,375 million) than any other ITL region (37% of UK total) as of 2024.

---

[9] National accounts GVA from Table 1 in: GDP output approach – low-level aggregates - Office for National Statistics. Note this value excludes imputed rental.
[10] UK all industry productivity estimates from Table_12 in: Output per job, UK - Office for National Statistics. Note this value excludes imputed rental.

- This is followed by the South East (14% of total) and the North West (10%). The South East and North East are the only ITL regions to have more of their Digital Economy GVA within Tier 1 than Tier 2.

- London's share of UK Digital Economy GVA (37%) is notably higher than its share of UK total GVA (for all industries, excluding imputed rental), which stood at 24% in 2023 (the latest comparable value available).[11]

ITL Region	GVA (£m, current prices)				
	Tier 1 GVA	Tier 2 GVA	Tier 1 & 2 GVA	Share of UK Digital Economy	Share of UK Total*
East Midlands	5,120	5,306	10,426	4%	6%
East of England	9,192	10,576	19,768	7%	8%
London	49,503	55,872	105,375	37%	24%
North East	3,111	2,933	6,045	2%	3%
North West	10,642	17,376	28,017	10%	10%
Northern Ireland	2,733	2,589	5,323	2%	2%
Scotland	7,562	8,632	16,194	6%	8%
South East	22,278	17,214	39,492	14%	14%
South West	7,818	10,088	17,906	6%	7%
Wales	2,991	4,201	7,191	3%	3%
West Midlands	7,548	7,958	15,506	5%	7%
Yorkshire and The Humber	6,648	8,435	15,083	5%	7%

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics. * this column shows the regions share of UK total GVA for all industries (excluding imputed rental), for 2023.

[11] Region total GVA shares from Table 1c in: Regional gross value added (balanced) by industry: all ITL regions - Office for National Statistics. Note these values exclude imputed rental.

2.93m   66%   £39,500

Employees and Employee Chains

employees (2024)   **work in professional and associate professional occupations** (2024)   **median annual earnings** (2024)

By

Table

Taxonomy	Employees	Share of UK Digital Economy
Frontier dedicated	679,927	23%
Established dedicated	838,058	29%
Total Tier 1	**1,517,985**	**52%**
Agile diversified	365,528	12%
Traditional diversified	1,049,080	36%
Total Tier 2	**1,414,608**	**48%**
Total Digital Economy	*2,932,593*	*-*

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics.

- Table 7 Across the UK, it is estimated there are a total of 2.93 million employees working in the Digital Economy (as of 2024).

- As   shows, this is relatively evenly distributed across the Tiers of the Digital Economy, with slightly more - 1.52 million (52% of total) - employees engaged in Tier 1 than Tier 2 (1.41 million, 48% of total).

Table 8. Employee occupations in the Digital Economy (2024)

- In terms of the Digital Economy taxonomy subsectors, traditional diversified enterprises represent the greatest share (36%) of employees. Frontier dedicated enterprises account for 23% of employees, despite representing only 8% of enterprises.

SOC occupation group	Share of employees (%)			
	Tier 1	Tier 2	Total Digital Economy	Non-Digital Economy
Managers, Directors And Senior Officials	15%	10%	**13%**	*11%*
Professional Occupations	51%	44%	**48%**	*25%*
Associate Professional Occupations	19%	17%	**18%**	*14%*

Administrative And Secretarial Occupations	6%	25%	**14%**	*10%*
Skilled Trades Occupations	4%	0%	**2%**	*10%*
Caring, Leisure And Other Service Occupations	0%	1%	**0%**	*9%*
Sales And Customer Service Occupations	3%	2%	**2%**	*6%*
Process, Plant And Machine Operatives	1%	0%	**1%**	*6%*
Elementary Occupations	1%	1%	**1%**	*9%*

Source: LFS, calculated by Cambridge Econometrics.

- The Labour Force Survey (see   above) shows thattwo-thirds (66%) of Digital Economy employees in the UK work in professional and associate professional occupations (as of 2024), compared to 39% in non-Digital Economy industries.

- Within Tier 1, this share increases to 70%, while Tier 2 has a notable overrepresentation of employees working in administrative and secretarial occupations, at 25%, relative to 10% in non-Digital Economy industries.

Table 9. Median annual and hourly earnings by Digital Economy Tier (2024)

Digital economy Tier	Median annual earnings (£, current prices)	Median hourly wage (£, current prices)
Tier 1	£39,500	£25.05
Tier 2	£39,250	£21.29
Total Digital Economy	**£39,500**	**£23.38**

Source: LFS, calculated by Cambridge Econometrics.

- According to the Labour Force Survey, employees working in the UK Digital Economy have a median hourly wage of £23.40 (in current prices, as of 2024), and median annual earnings of £39,500.

- This is 50% higher than the median hourly wage of £15.60 for employees working in non-Digital Economy industries, and 60% higher than the median annual earnings of £24,500.

- As shown in Table 9, median hourly wages are 18% higher in Tier 1 than Tier 2 of the Digital Economy, though the difference is negligible in median annual earnings.

-         presents estimates of median annual earnings and hourly wages in the Digital Economy by occupation. Earnings and wages for those employed in managerial occupations is the highest, at £41,000 annually and £37 per hour, followed by professional occupations.

- On a per hour basis, wages for all occupation types in the Digital Economy exceed or at least match the average for that occupation in non-Digital Economy industries. The difference is largest (in order) for managerial, professional, associate professional, and skilled trades occupations.

SOC occupation group	Total Digital Economy		Non-Digital Economy	
	Median annual earnings (£, current prices)	Median hourly wage (£, current prices)	Median annual earnings (£, current prices)	Median hourly wage (£, current prices)
Managers, Directors And Senior Officials	£41,000	£37	£39,000	£24
Professional Occupations	£41,000	£27	£36,500	£22
Associate Professional Occupations	£39,000	£21	£27,500	£17
Administrative And Secretarial Occupations	£18,600	£15	£21,500	£14
Skilled Trades Occupations	£36,500	£19	£27,500	£14
Caring, Leisure And Other Service Occupations	Not available	£12	£16,200	£11
Sales And Customer Service Occupations	£24,500	£13	£13,800	£11
Process, Plant And Machine Operatives	Not available	£13	£25,200	£13
Elementary Occupations	£10,200	£11	£14,500	£11

Source: LFS, calculated by Cambridge Econometrics.

By Region

Table 11

- shows the distribution of UK Digital Economy employees by Tier and ITL region. London has the highest number of employees across both Tiers (1 million), representing 35% of total UK Digital Economy employees.

- For Tier 1, the South East has the next highest number of employees (250,400), while for Tier 2, the North West has the greatest number of employees outside London (199,300).

- Only the North West, South West, Wales and Yorkshire and The Humber have more of their Digital Economy employees working in Tier 2 enterprises.

- London's share of UK Digital Economy employees (35%) is almost double its share of UK total employee jobs (for all industries), which stood at 18% in 2024.[12]

- The South East is the only other region where it's share of UK Digital Economy employees (14%) was higher than its share of UK total employee jobs (13%).

---

[12] Region total employee shares from: Nomis - Query Tool - workforce jobs by industry (SIC 2007) - seasonally adjusted. Note values are for employee jobs, and are annual averages. Employee jobs are not directly comparable with the IDBR-derived employee estimates, and are included for reference only.

ITL Region	Employees			Share of UK Digital Economy	Share of UK Total*
	Tier 1	Tier 2	Tier 1 & 2		
East Midlands	57,868	54,180	**112,048**	4%	*7%*
East of England	104,993	96,071	**201,064**	7%	*9%*
London	552,349	475,621	**1,027,970**	35%	*18%*
North East	34,936	30,486	**65,422**	2%	*3%*
North West	119,923	199,347	**319,270**	11%	*11%*
Northern Ireland	30,606	28,072	**58,678**	2%	*3%*
Scotland	85,195	80,178	**165,373**	6%	*8%*
South East	250,363	148,479	**398,842**	14%	*13%*
South West	87,872	95,304	**183,176**	6%	*8%*
Wales	33,758	38,752	**72,510**	2%	*4%*
West Midlands	85,345	83,950	**169,295**	6%	*8%*
Yorkshire and The Humber	74,777	84,168	**158,945**	5%	*8%*

Table ... egion (2024)

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics. * this column shows the regions share of UK total employee jobs for all industries, for 2024. Note employee jobs are not directly comparable with the IDBR-derived employee estimates.

311,800   88%   25%

Business Demography

enterprises (2024)   **of enterprises employ less than 10 people** (2024)   **of enterprises have annual turnovers below £100,000** (2024)

By

Tab

Digital economy taxonomy	Enterprises	Share of UK Digital Economy
Frontier dedicated	25,563	8%
Established dedicated	175,283	56%
Total Tier 1	**200,846**	**64%**
Agile diversified	5,591	2%
Traditional diversified	105,359	34%
Total Tier 2	**110,950**	**36%**
Total Digital Economy	*311,796*	*-*

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics.

- As shown in Table 12, there are 311,800 enterprises active in the Digital Economy (as of 2024). In addition to this, there are 349,200 local sites associated with Digital Economy enterprises.

- This enterprise count equates to approximately 11% of total UK enterprises active during the same period (2024).[13]

- In terms of Tiers of the Digital Economy, in contrast to GVA and employees - which are more evenly distributed - Tier 1 accounts for a notably higher share of enterprises (200,800 enterprises - 64% of total enterprises).

By Region

- And at the taxonomy level, the distribution is much less balanced: for instance, only 10% of enterprises

Table 13
    are frontier dedicated or agile diversified, though such enterprises account for significantly higher shares of GVA and employees.

-         presents all Digital Economy enterprises in the UK by Tier and ITL region. London boasts the highest number, with 100,500 active Digital Economy enterprises (32% of UK total).

---

[13] Total UK enterprises from: Nomis - Query Tool - UK Business Counts - enterprises by industry and employment size band.

- This is followed by the South East, with 53,600 Digital Economy enterprises (17% of UK total), and the East of England with 28,600 enterprises (9%).

- This means the London and neighbouring regions (the South East and East of England), account for almost two-thirds (58%) of UK Digital Economy enterprises.

- As with GVA and employees, London's share of UK Digital Economy enterprises (32%) is significantly higher than its share of UK total enterprises (for all industries), which stood at 19% in 2024.[14]

- And the South East is the only other region where it's share of UK Digital Economy enterprises (17%) was higher than its share of UK total enterprises (15%).

Table

| ITL Region	Enterprises				
	Tier 1	Tier 2	Tier 1 & 2	Share of UK Digital Economy	Share of UK Total*
East Midlands	8,949	6,873	15,822	5%	7%
East of England	18,721	9,856	28,577	9%	10%
London	67,186	33,282	100,468	32%	19%
North East	3,327	2,371	5,698	2%	3%
North West	15,032	12,001	27,033	8%	10%
Northern Ireland	2,446	2,652	5,098	2%	3%
Scotland	8,987	5,674	14,661	5%	6%
South East	38,150	15,466	53,616	17%	15%
South West	14,390	8,333	22,723	7%	9%
Wales	4,698	3,491	8,189	3%	4%
West Midlands	13,123	7,663	20,786	7%	8%
Yorkshire and The Humber	9,402	6,758	16,160	5%	7%

By Employee and Turnover Size

Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics. * this column shows the regions share of UK total enterprises for all industries, for 2024.

Figure 1

- presents a breakdown of all Digital Economy enterprises in UK by annual turnover and employment size thresholds (as of 2024).

- A quarter (25%) of UK Digital Economy enterprises have an annual turnover below £100,000, which is significantly lower than the average of 34% for all UK industries during the same period.[15] Meanwhile 10% generate turnovers over £900,000.

- Some 88% of UK Digital Economy enterprises employed fewer than 10 people in 2024, similar to the UK average for all industries of 89%[16], while 2% have more than 50 employees.

---

[14] Region total enterprise shares from: Nomis - Query Tool - UK Business Counts - enterprises by industry and employment size band.

[15] UK all industry turnover size shares from: Nomis - Query Tool - UK Business Counts - enterprises by industry and turnover size band

[16] UK all industry employment size shares from: Nomis - Query Tool - UK Business Counts - enterprises by industry and employment size band
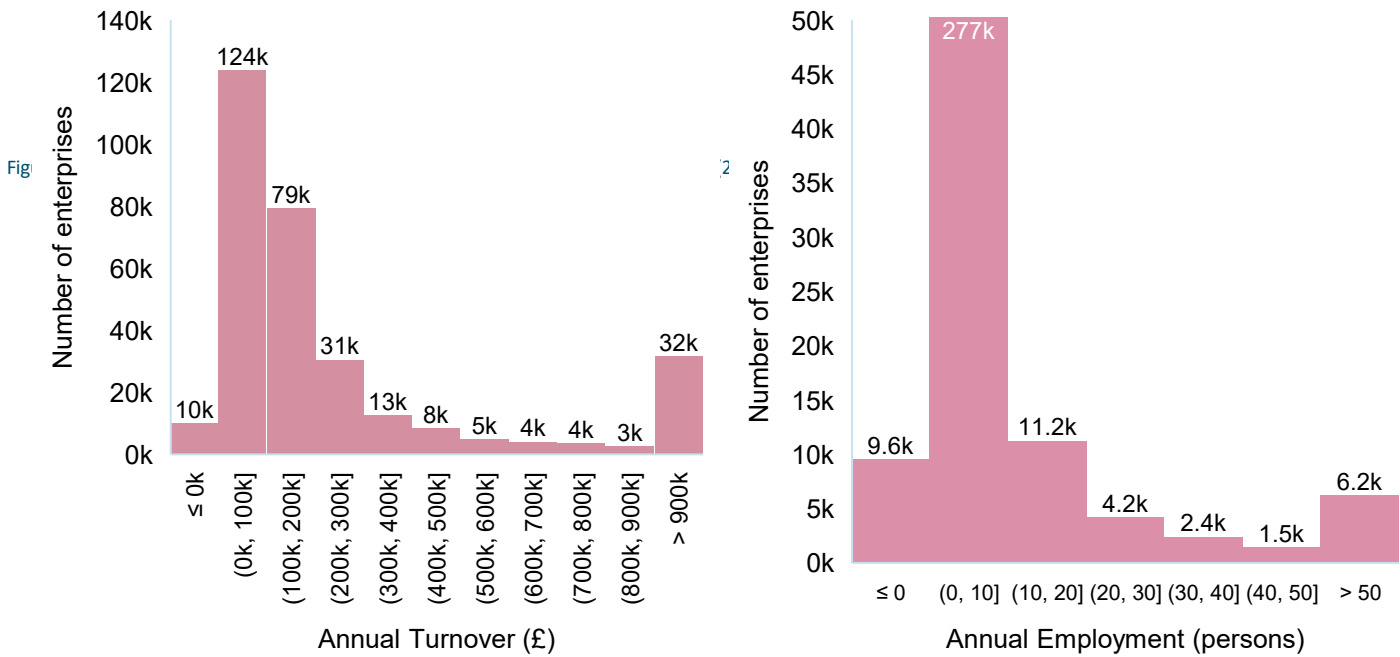
Source: IDBR Q4 extract, ONS, calculated by Cambridge Econometrics.

£72m   £4,065m

Fu   e)

**of UKRI funding
received per annum
(2000 - 2024)**

**of VC funding
received per annum
(2010 - 2024)**

By

Tab

Digital economy taxonomy	UKRI funding (£m, current prices)	Share of UK Digital Economy	Average per annum (£, current prices)
Frontier dedicated	593	34%	25
Established dedicated	543	31%	23
Total Tier 1	**1,136**	**66%**	**47**
Agile diversified	536	31%	22
Traditional diversified	54	3%	2
Total Tier 2	**590**	**34%**	**25**
Total Digital Economy	*1,726*	*-*	*72*

Source: IDBR Q4 extract, UKRI, calculated by Cambridge Econometrics.

Table 14

- shows the total UKRI funding received by Digital Economy enterprises between January 2000 and January 2025. This can act as a proxy for public funding to innovative enterprises.
- Digital economy enterprises have received an estimated £1,726 million of UKRI funding to date, equivalent to £72 million of funding per annum (in current prices).

Table 15: VC funding by Digital Economy taxonomy (2010 to 2025)

- The majority - £1,136 million (66%) - of this was allocated to enterprises within Tier 1, and £590 million (34%) to those in Tier 2.

Digital economy taxonomy	VC funding (£m, current prices)	Share of UK Digital Economy	Average per annum (£m, current prices)
Frontier dedicated	31,257	55%	2,233
Established dedicated	22,761	40%	1,626
Total Tier 1	**54,018**	**95%**	**3,858**
Agile diversified	1,044	2%	75

Traditional diversified	1,850	3%	132
Total Tier 2	**2,894**	**5%**	**207**
Total Digital Economy	*56,912*	*-*	*4,065*

Source: IDBR Q4 extract, Dealroom (via The Data City Data Explorer), calculated by Cambridge Econometrics.

- Table 15 shows the total VC funding received by Digital Economy enterprises between January 2010 and January 2025. This can act as a proxy for private funding to innovative enterprises.

- Digital economy enterprises have received an estimated £59,912 million of VC funding to date, equating to an average of £4,065 per annum (in current prices).

- The vast majority of VC funding between 2010 and 2025 - £54,018 million (95%) – was attributed to Tier 1 enterprises, with only £2,894 million (5%) to Tier 2.

- In terms of the Digital Economy taxonomy subsectors, more than half (55%) of VC funding is concentrated in frontier dedicated enterprises, despite representing only 8% of enterprises.

By Region

- Table 16 shows the distribution of UKRI funding to Digital Economy enterprises by ITL Region. London accounts for more UKRI funding in the Digital Economy than any other ITL region (48% of UK total).

- This is followed by the South East (13%), which in contrast to London has the majority of it's of UKRI funding associated with Tier 1 of the Digital Economy.

Table 16: UKRI funding in the Digital Economy by ITL Region (2000 to 2025)

ITL Region	UKRI funding (£m, current prices)			
	In Tier 1	In Tier 2	In Tier 1 & 2	Share of UK Digital Economy
East Midlands	35	15	**50**	3%
East of England	88	6	**94**	5%
London	341	488	**829**	48%
North East	41	22	**63**	4%
North West	67	7	**75**	4%
Northern Ireland	67	4	**71**	4%
Scotland	45	4	**49**	3%
South East	207	14	**220**	13%
South West	87	6	**93**	5%
Wales	26	4	**30**	2%
West Midlands	85	10	**95**	5%

Yorkshire and The Humber	46	10	**56**	3%

Source: IDBR Q4 extract, UKRI, calculated by Cambridge Econometrics.

The following summary highlights the economic contribution of the UK Digital Economy - using the new framework and definition adopted for the Phase 2 work - across key measures such as gross value added, productivity, employees (and their characteristics), and funding (both public and private).

Summary

Table ... for the UK Digital Economy (2024)

Digital economy taxonomy	GVA (£m, current prices)	GVA per employee (£, current prices)	Employees	Enterprises
Frontier dedicated	60,518	89,007	679,927	25,563
Established dedicated	74,629	89,050	838,058	175,283
Total Tier 1	**135,147**	**89,031**	**1,517,985**	**200,846**
Agile diversified	40,938	111,997	365,528	5,591
Traditional diversified	110,240	105,082	1,049,080	105,359
Total Tier 2	**151,178**	**106,869**	**1,414,608**	**110,950**
Total Digital Economy	*286,325*	*97,636*	*2,932,593*	*311,796*
DSIT Digital Sector Economic Estimates*	*153,483 (2023)*	*Not available*	*1,882,622 (2022)*	*Not available*

Source: Department for Science, Innovation & Technology, Cambridge Econometrics, Innovation and Research Caucus, The Data City. Note: * DSITs current Digital Sector Economic Estimates are included here for reference. Due to the differences in underlying methods, definitions, data, and timeframes, these will be different to the economic estimates derived as part of the Phase 2 work and are only broadly comparable with Tier 1 estimates.

Gross Combined (GVA) ... produced by

- The UK Digital Economy is estimated to have total annual GVA of £286,325 million (in current prices, as of 2024). This equates to approximately 13% of total annual UK national accounts GVA.

- Output (GVA) per employee in the UK Digital Economy is estimated at £97,600 as of 2024. This compares to a whole economy value of £62,300 (2023 estimate, the latest comparable value available).

Employees and employee characteristics

- This means that the average Digital Economy employee is 57% more productive than average UK employee.

- Across the UK, it is estimated there are a total of 2.93 million employees working in the Digital Economy (as of 2024).

- Two-thirds (66%) of Digital Economy employees in the UK work in professional and associate professional occupations (as of 2024), compared to 39% in non-Digital Economy industries.

- Employees working in the UK Digital Economy are estimated to have a median hourly wage of £23.40 (in current prices, as of 2024) and median annual earnings of £39,500.

- This is 50% higher than the median hourly wage of £15.60 for employees working in non-Digital Economy industries, and 60% higher than the median annual earnings of £24,500.

Business demography and funding

- There are 311,800 enterprises currently active in the Digital Economy (as of 2024). In addition to this, there are 349,200 local sites associated with Digital Economy enterprises.

- A quarter (25%) of UK Digital Economy enterprises have an annual turnover below £100,000, which is significantly lower than the average of 34% for all UK industries during the same period. Meanwhile 10% generate turnovers over £900,000.

- Some 88% of UK Digital Economy enterprises employed fewer than 10 people in 2024, similar to the UK average for all industries of 89%, while 2% have more than 50 employees.

- Digital economy enterprises in the UK have received an estimated £72 million of UKRI funding per annum (2000-2025), and £4,065 million of VC funding per annum (2010-2025).

FAQs

Definitions

- Can a single enterprise have different Company Reference Numbers? If so, why?
    - Yes, many enterprises in the IDBR dataset have multiple CRNs which do not correspond in any way to individual sites of that enterprise. This is because an enterprise can be a group of legal units under common ownership, whereas a CRN is based on a registered business entity.
    - When linking datasets to the IDBR dataset based on CRNs, we adopt an approach that prioritises completeness, ensuring that the activities of all companies (CRNs) are associated with the correct enterprise.
- Can a single enterprise have different RTICs? Is the enterprise counted in all the RTICs assigned to it?
    - Yes, a single enterprise can have multiple CRNs and therefore multiple companies. Each company CRN can be assigned multiple RTICs.
    - Yes, for completeness, we ensure that the enterprise's activities are included in each associated RTIC sector as a whole.
- If an enterprise is included in all the RTICs assigned to it, how is double counting avoided?
    - When aggregating data across multiple RTIC sectors, each enterprise is counted only once within the grouping. This prevents double counting when an enterprise is assigned to multiple RTIC sectors. As a result, the total for a group covering multiple RTICs is lower than the sum of individual RTICs calculated separately.
- Can a single enterprise appear in different Tiers simultaneously?
    - No, each enterprise has one primary SIC code which is uniquely identified to be in Tier 1, Tier 2, or not in Tiers. We check the enterprises not in Tiers if they have secondary SICs in Tiers and the result is that they do not.
- Are there any limitations to using a SIC code classification for defining Tiers 1 and 2?

Data Vintage

    - SIC codes are defined based on traditional sectors which may or may not represent the UK Digital Economy completely. Based on Phase 1 of the project and DSIT's investigation, we adopted the method to extend beyond the Tier 1 and 2 definition, using the subsector taxonomy.

- What years are the employment and turnover data on?
    - The IDBR dataset include data available on active enterprises as of January 2025, with enterprise records covering the Quarter 4 period (September to December) of 2024. Employment and turnover are annual figures.
- Is UKRI and VC funding data for a single year or cumulative through the years?
    - UKRI project funding application dataset includes all funding received since January 2000 up

until January 2025. VC funding data is sourced from Dealroom and is also cumulative, covering the period January 2010 to January 2025. We do not analyse funding received per year trends because funding in emerging sectors may be diluted and volatile given their relative recency.

Data Analysis

- How is business performance measured across time?

  o Because the IDBR dataset is a snapshot of all enterprises in the economy, we do not analyse the business performances across time. Instead, we present the latest levels and focus on estimating the size of the Digital Economy as of January 2025.

    However, our methodology has been designed in a way that, if applied consistently, could allow for performance to be measured over time.

- Why is it relevant to account for enterprise local sites?

  o Local sites of enterprises offer a geospatial understanding of the Digital Economy. Employment figures at the local site level offer a way to measure activity by region.

- How are median annual earnings and hourly wage calculated from LFS?

  o LFS has hourly wage data for 51 thousand employees in 2023 and 2024. LFS has salary data categorised into weekly, monthly, and annual salary bands. We compute the average salary within each band and standardise all values to annual salaries. This allows us to derive 'average annual salaries,' which we then summarise by occupation, SIC sector, and gender.

Economic Estimates

- Why are the estimates different to those presented in related publications on the UK Digital Economy?

  o Due to the differences in underlying methods, definitions, data, and timeframes, the Phase 2 estimates will be different to those in related Digital Economy publications (such as DSIT's Digital Sector Economic Estimates).

  o For instance, at the time this analysis was completed, DSIT's Digital Sector Economic Estimates uses different SIC codes (with no Tier distinction), were last updated as a 2022/23 vintage, and do not use an enterprise-level (IDBR-based) methodology. However, our QA and validation checks showed when our database was restricted to the current Digital Sector SIC codes, estimates had a >95% match.

- Where can I find estimates for specific Digital subsectors/technologies, such as AI, or Quantum?

  o DSIT continues to prioritise the estimates from its dedicated publications on Digital subsectors/technologies of interest (including its annual Digital Sector Economic Estimates, as well as the standalone [Artificial Intelligence sector study](#) and [Cyber security sectoral analysis)](#).

  o Though the Phase 2 database provides RTIC-level estimates (as required for the taxonomy classifications), there are no plans to publish these as standalone estimates, though they are available for DSIT to consider and interrogate further. The RTICs used in the Phase 2 work include those representing Digital industries and technologies of significant interest to DSIT, including Artificial Intelligence (AI), Cyber, Data and Software.

Appendices

SIC-Tier definition

This section summarises research carried out by DSIT prior to phase 1 and the economic project

Tier 1 SICs

Within the Phase 1 framework, Tier 1 of the Digital Economy is economic activity derived from enterprises that are primarily involved in the production of:

1. Digital content,
2. ICT goods, and/or
3. ICT services.

There is no set definition of Digital content at the international level. The UK Digital Markets, Competition and Consumers Act (2024) defines Digital content as "data which is produced and supplied in Digital form".[17] The OECD defines the content and media component of the Information Industries, a conceptual pre-cursor to the Digital Economy, as "the group of economic activities that are primarily "engaged in the production, publishing and/or the distribution of content (information, cultural and entertainment products), where content corresponds to an organised message intended for human beings."[18]

The Phase 2 consortium proposes combining these approaches to define Digital content as all content that is inherently Digital or delivered Digitally, where "content" is taken to mean written, audio or visual media that is prepared for a plurality of consumers i.e. personal services such as individual artistic commissions should be excluded where viable. The definition therefore includes the following 4-digit SIC codes:

- 58.14: Access to learned and other professional journals is more typically via an institutional login rather than a printed copy
- 58.21: The vast majority (89.5%) of computer games sold in 2023 were Digital downloads rather than physical copies (source) and computer games are an inherently Digital product
- 58.29: Software is, by definition, Digital content. Where possible, reproduction of this software (18.20/3) should also be included in the definition.
- Division 59: all television has been distributed Digitally in the UK since 2012 (source); as of 2017, 98% of the world's cinema screens were Digital (source); and as of 2021 the majority of those listening to the radio in the UK did so Digitally (source), and in 2023, Digital sales accounted for 86% of music revenue in the UK (source).
- 63.91: There is likely to be substantial crossover between "News Agency activities" and other activities (newspaper, TV and radio broadcasting) that are considered to be within the Digital content definition

---

[17] See: UK Digital Markets, Competition and Consumers Act (2024)

[18] See: OECD (2020)

The definition excludes:

- 58.11: The majority of book sales in the UK in 2024 were physical, rather than Digital copies ([source](#)).

- 58.12: It is unknown what proportion of revenue of producers of directories and mailing lists are Digital rather than paper, so this class is excluded due to its proximity/similarity to classes 58.11 and 58.13. It is also unclear whether a directory or mailing list can really be described as "content". This class will not be included in the next revision of the NACE classification.

- 58.13: In 2021 around two thirds of newspaper revenue in the UK was from paper sales rather than Digital ([source](#)). The proportion of newspaper revenue accounted for by Digital income is growing so this position should be kept under review.

- 58.19: The [description](#) of this class suggests that the majority of revenue will be derived offline.

- 63.99: some of the activities in this class are computer based but it is not clear what proportion they make up of the whole.

- Section R (excluding 92.00): this section includes activities that produce entertainment e.g. 90.01 "Performing arts" however the vast majority of these are, due to the nature of the activity, delivered in person rather than Digitally.

The definition initially excluded 63.12 Web portals, although following QA and validation checks (detailed under *QA and data validation* below), this was added to the Tier 1 definitions list. SIC divisions 60 to 62 are not listed above as they are considered part of the "IT services" component of the Tier 1 Digital Economy definition. ICT goods are defined per the UNCTAD classification for international trade in ICT goods.[19] ICT services are defined per the UNCTAD technical note on international trade in ICT services and ICT-enabled services.[20]

Taken together these give the Phase 1 SIC definition for Digital content, ICT goods and services. Companies with these SIC codes can be considered to belong to Tier 1, i.e. they are primarily involved in the production of Digital content and/or ICT goods and/or ICT services.

Tier 2 SICs

The outputs of Tier 1 company activity can also be considered inputs to further production. Input-output tables can be used to identify SIC codes where these activities form a non-zero minority of outputs, and supply-use tables identify which SIC codes have a similar Digital intensity of inputs to enterprises in Tier 1. The intersection of these two groups are the Tier 2 enterprises.

The input-output tables are not available at the necessary granularity to support the Tier 1 SICs and must instead be approximated as the sum of SIC divisions. This accuracy of this approach can be assessed by estimating the proportion of economic activity in each division that is accounted for by activity in the listed classes. ONS does not publish GVA data at 4-digit SIC level, however it does publish approximate GVA (aGVA) estimates at the 4-digit SIC level for the UK based on annual business survey data.

These show that Digital activities formed a substantial minority of divisions 26 (38%) and 58 (45%), but accounted for a moderate majority of aGVA in divisions 63 (56%) and 95 (60%) in 2022. Inclusion of the operation of web portals and search engines as a Digital service increases the proportion of aGVA for division 63 to 91%. These proportions are not static (see Figure A.1 below), for example, software publishing is

---

[19] See: [UNCTAD classification for international trade in ICT goods](#)

[20] See: [UNCTAD technical note on international trade in ICT services and ICT-enabled services](#)

accounting for an increasing proportion of publishing sector aGVA.

Proportion of division aGVA accounted for by digital goods and services
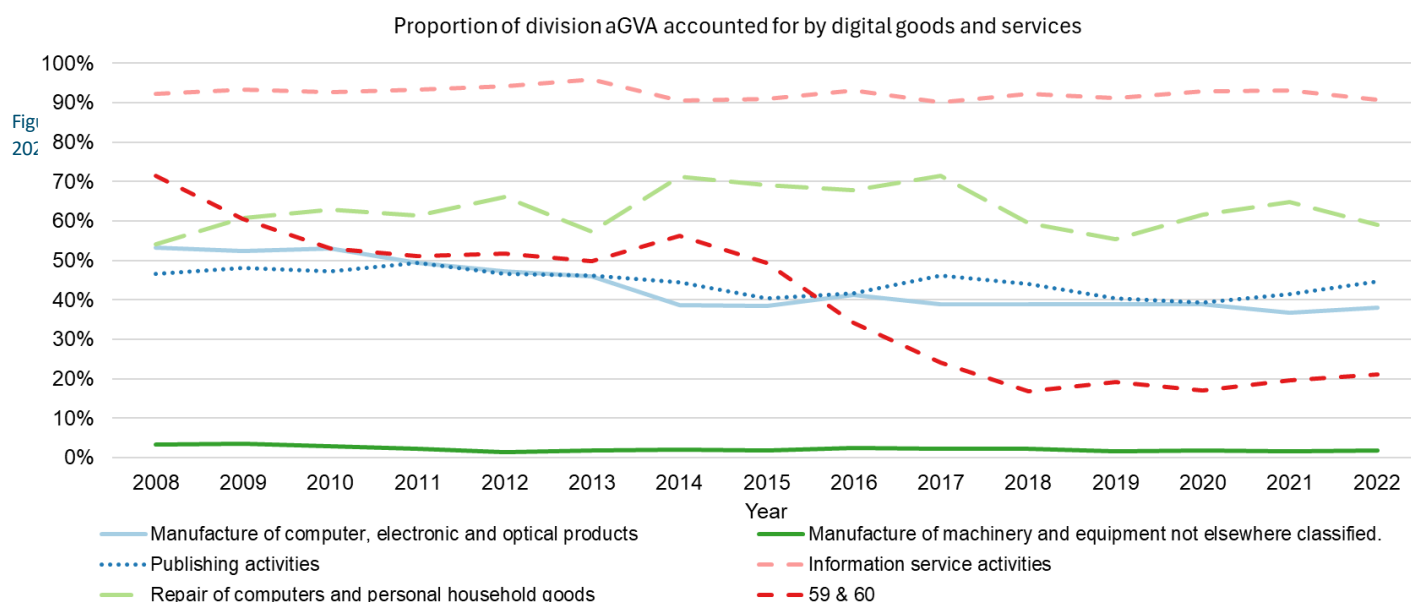
Source: ONS Annual Business Survey (see: Non-financial business economy, UK: Sections A to S - Office for National Statistics).

The appropriate averaging period for each division was determined by visual inspection of the trends in Figure A.1 and is recorded in Table A.1 below, along with the proportion of the division used in the subsequent input-output analysis.

Table A.1: Estimated proportion of each aGVA for each SIC division accounted for by production of Digital goods and services

SIC Division Code	SIC Division Description	% aGVA	Averaging period
18	Printing and reproduction of recorded media	Unknown – no data	N/A
26	Manufacture of computer, electronic and optical products	39%	2014 to 2022
28	Manufacture of machinery and equipment not elsewhere classified.	2% (set to 0%)	2012 to 2022
58	Publishing activities	43%	2014 to 2022
59	Motion picture, video and television programme production, sound recording and music publishing activities	100%	N/A *
60	Programming and broadcasting activities	100%	N/A *
61	Telecommunications	100%	N/A

62	Computer programming, consultancy and related activities	100%	N/A
63	Information service activities	92%	2014 to 2022
95	Repair of computers and personal and household goods	62%	2020 to 2022

Source: ONS Annual Business Survey (* Division 60 is only reported in the input output tables in combination with division 59).

The Digital intensity of an industry is estimated by summing the weighted Digital goods and services inputs over each industry division and dividing by total intermediate consumption.

In 2021, Digital intensity was highest for the information and communication division (J, 30%) and lowest for electricity, gas, steam and air conditioning supply (D, 1%), excluding activities of households (T) and extraterritorial bodies (see Figure A.2 below). Overall UK average Digital intensity of inputs was 7% for the same year.

Figure A.2

Digital goods and services as % of intermediate consumption

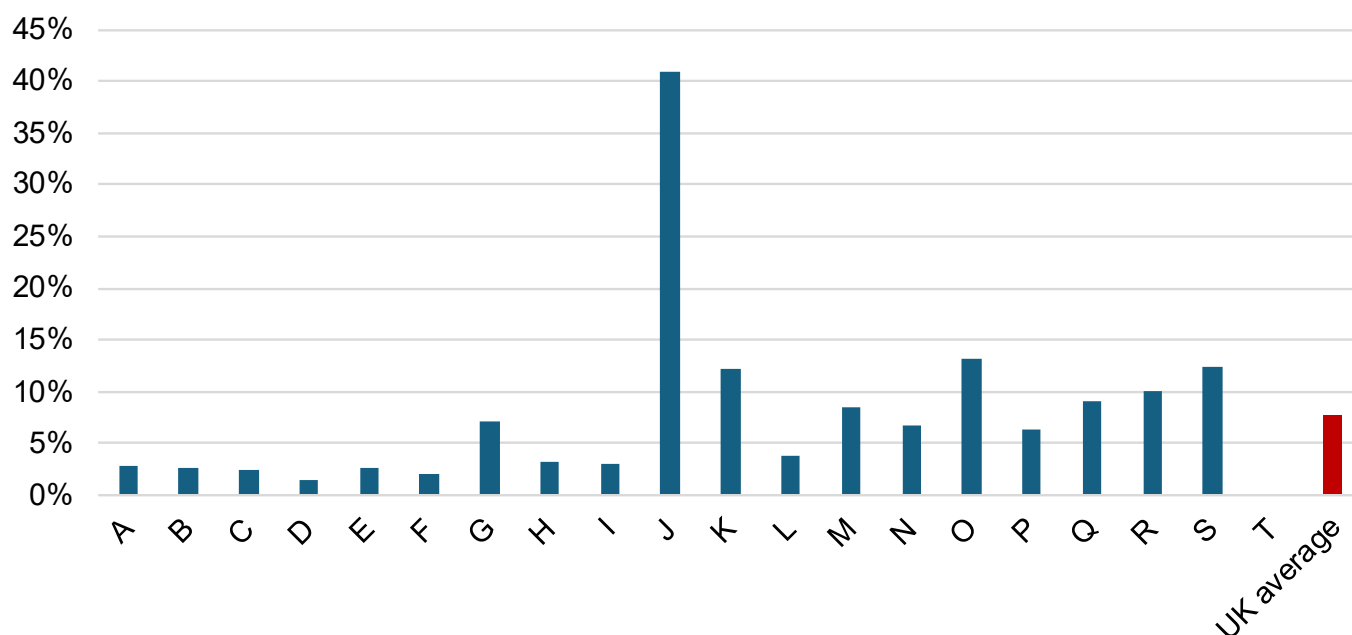Source: ONS Annual Business Survey.

Table A.2: Digital intensity of intermediate product consumption for SIC divisions which produce Digital goods and services, and the percentage of each SIC division's aGVA accounted for by production of Digital goods and services

There is considerable variability within divisions (see Table A.2 below) that produce Digital goods and services. Average Digital intensity of Tier 1 was calculated by weighting the Digital intensity of each division by the proportion of its aGVA that derived from production of Digital goods or services.

SIC Division Code	SIC Division Description	% aGVA due to Digital production	Digital intensity of inputs, 2021

18	Printing and reproduction of recorded media	Unknown	N/A
26	Manufacture of computer, electronic and optical products	39%	15%
58	Publishing activities	20%	23%
59	Motion picture, video and television programme production, sound recording and music publishing activities	100%	52%
60	Programming and broadcasting activities	100%	52% *
61	Telecommunications	100%	39%
62	Computer programming, consultancy and related activities	100%	42%
63	Information service activities	73%	22%
95	Repair of computers and personal and household goods	62%	38%

Source: ONS Annual Business Survey (* Division 60 is only reported in the input output tables in combination with division 59).

The average Digital intensity of the Tier 1 Digital Economy was 36% in 2021. There are no divisions outside Tier 1 that have a Digital intensity this high, however there are several that have a Digital intensity above the lowest ranked Tier 1 Digital industry (see Table A.3 below).

Table A.3: SIC divisions with above-average Digital intensity of inputs, 2021 (Tier 1 Digital Economy divisions are shaded in blue)

| Division	Description	Digital Intensity (2021)
J59 & J60	Motion picture, video and television programme production, sound recording and music publishing activities; Programming and broadcasting activities	* 52%
J62	Computer programming, consultancy and related activities	42%
J61	Telecommunications	39%
S95	Repair of computers and personal and household goods	38%
J58	Publishing activities	23%
K66	Activities auxiliary to financial services and insurance activities	22%
J63	Information service activities	22%
M692	Accounting, bookkeeping and auditing activities; tax consultancy	19%
M691	Legal activities	19%
R92	Gambling and betting activities	17%
C26	Manufacture of computer, electronic and optical products	15%
L683	Real estate activities on a fee or contract basis	13%
O84	Public administration and defence; compulsory social security	13%
N82	Office administrative, office support and other business support activities	12%
S94	Activities of membership organisations	11%
M70	Activities of head offices; management consultancy activities	11%

K64	Financial service activities, except insurance and pension funding	10%
R91	Libraries, archives, museums and other cultural activities	10%
R90	Creative, arts and entertainment activities	10%
Q86	Human health activities	10%
N79	Travel agency, tour operator and other reservation service and related activities	9%
G47	Retail trade, except of motor vehicles and motorcycles	8%
	UK total intermediate demand	8%

Source: ONS Annual Business Survey (* Division 60 is only reported in the input output tables in combination with division 59).

Assessment of total approximate GVA produced by Digital intensity of the producing industry (Figure A.3) over multiple years does not show a stable distribution; as expected, Digital intensity of production has tended to increase over time, with a step change during the pandemic.

A potential point of distinction between low and higher Digital intensity industries (the trough at ca. 0.15 or 15% Digital intensity) coincides with the lowest Digital intensity of a Tier 1 industry. For the purpose of the Phase 2 work, Tier 2 can therefore be defined as businesses in all SIC codes that do not have a Digital good or service as their primary product, and with a Digital intensity of inputs that is at least as high as the lowest intensity Tier 1 division.

Figure A.3: Total average aGVA produced by the Digital intensity of the producing industry
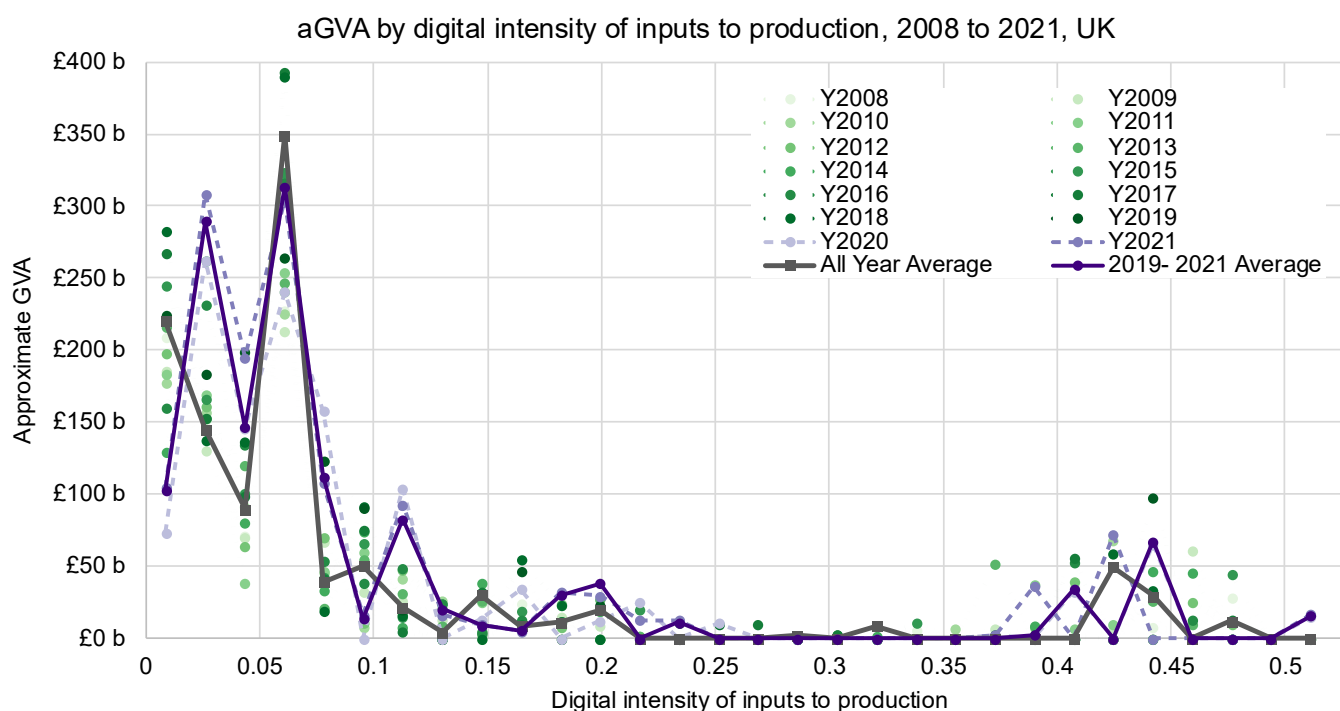
Source: ONS Annual Business Survey. Note: aGVA was not estimated for the financial services industry (divisions 64 to 66), which include a division of high Digital intensity. See the notes for the ONS publication: Non-financial business economy (Annual Business Survey) 2022 results for further information.

QA and data validation

The process to identify and adopt the Tier 1 and 2 SICs uses a clear and robust methodology underpinned by

established, official data sources. Naturally, given the nature of the selection process, there were some 'edge case' SIC codes that fall close to the required thresholds for inclusion or warranted further investigation due to disclosive or limited official data.

These 'edge cases' underwent additional QA and validation by the Phase 2 consortium, with the close involvement of DSIT. This was done by appraising 'edge case' SIC codes against broader measures of Digital intensity, using alternative, up to date (to December 2024) company-level data from The Data City's Data Explorer platform. Specifically, it appraised 'edge case' SICs in terms of:

- The proportion of companies with 'Digital-intense' RTICs: a series of RTICs were identified which had companies overrepresented with Tier 1 SICs.[21] For each 'edge case' SIC, the proportion of companies with one or more of these RTICs was calculated.

- The proportion of company job postings 'Digital': using Lightcast job posting data from 2020 onwards, the proportion of job postings in 'Digital occupations' (SOC-based, as defined in the Phase 1 report[22]) within each 'edge case' SIC was calculated.

- Manual company-list checks: company lists for each 'edge case' SIC were manually reviewed to check for the omission of key UK Digital Economy companies, while potential false positives were also identified.

Table A.4: 'Edge case' SIC codes additional Digital intensity checks (data to December 2024)

Edge case' 4-digit SICs	Proportion of companies with a 'Digital-intense' RTIC	Proportion of company job postings in 'Digital occupations'	RTIC company count*
58.11: Book publishing	6%	16%	679
58.12: Publishing of directories and mailing lists	0%	2%	37
58.13: Publishing of newspapers	7%	3%	60
58.19: Other publishing activities	10%	17%	1,347
63.12: Web portals	26%	13%	306
63.99: Other information service activities n.e.c.	32%	11%	2,371
Section R: Arts, entertainment and recreation (excl. 92.00)	7%	4%	5,551
Tier 1 SIC average	**41%**	**27%**	**52,246**
All SIC average	*13%*	*9%*	*287,188*

Source: The Data City Data Explorer (December 2024), calculated by Cambridge Econometrics. Note: * company counts are sourced from Companies House and

[21] These RTICs – which had more than half of companies in their company lists self-reporting Tier 1 SIC codes – are (in order of intensity): Cloud Computing; Software Development; Data Infrastructure; Data Intermediaries; Cyber; Artificial Intelligence Technologies and Applications; Artificial Intelligence Ecosystem; Gaming; Software as a Service (SaaS); Telecommunications; Future Telecoms Supply Chain.

[22] See Figure 3 here: [Defining and Measuring the UK Digital Economy](#)

therefore are not comparable to the IDBR-derived estimates presented elsewhere in this report. The above data should only be referred to in a QA/validation context.

Following these checks (presented in Table A.4 above), the 63.12: Web portals SIC was recommended by the Phase 2 consortium for inclusion. Relative to other 'edge cases', 63.12 exhibited a Digital intensity profile closer to the Tier 1 SIC average, but most importantly (in contrast to other 'edge case' SICs) it included a high count of key UK Digital Economy companies, and a lower count of false positives.

Such 'edge case' SICs will continue to be monitored and reviewed as the Phase 2 framework evolves and develops in future, particularly where new data is made available or existing data is updated. This could include analysis of The Data City's recently developed Real-Time Standard Industrial Classifications (RSICs) to observe what extent 'edge-case' SICs include misclassified companies who instead should be classified with Digital SICs.[23]

RTIC methodology

Real-Time Industrial Classifications, or RTICs, are datasets that represent the emergent economy. They are built with The Data City's machine learning and website scraping technology. In short, RTICs group companies that describe their activity similarly in their website text, allowing the user to make custom sectoral lists. This approach has been combined with expert input and review to create RTICs. There are currently more than 60 RTICs defined on The Data City's Data Explorer platform, listed in Table A.5 below, and which have been used to inform the Phase 2 program of work.

Taxonomies

The first thing needed to build an RTIC is a sector taxonomy. The taxonomy represents the structure of the sector and is a key resource that informs the machine learning classification of companies. Specifically, in this context, a taxonomy is a framework that identifies industry verticals (subsectors) within a sector. The aggregated view of all companies captured in each industry vertical builds the resulting RTIC.

Industry verticals are defined by keywords and key phrases that represent their activity, specialised technologies or role in a supply chain. The resulting framework and language guide the selection of company websites used to train the machine learning algorithm and subsequent iterations of the exercise.

There are four steps in the process of taxonomy creation.

1. Desk research: existing literature regularly gives a good insight into current developments and major verticals of activity in industrial sectors. It also contributes to the identification of key technical language.

2. Website discourse analysis: diving into a few company websites is important. Usually, academic literature or expert reports describe processes in a different way than companies when addressing a wide audience. Understanding how companies use sector-specific language is crucial to identify the most representative keywords for an economic sector.

3. Choosing an approach: When it comes to categorising sectors, there are various methods to choose from. Here are four main approaches The Data City have identified.

    a. Technology Approach: the verticals are defined according to the enabling technologies

---

[23] See here for more information: Real-Time SIC Codes (RSICs) - Product - The Data City

companies use.

b. Applications Approach: Classification of verticals according to the market applications of products, technology, and/or services.

c. Functional Approach: Define verticals based on a company's activity within a sector's supply chain.

d. Services Approach: the verticals are defined according to the services companies provide to the sector.

4. Selecting company websites for the training sets: this stage involves the selection of the first 10/20 company websites to train the algorithm. These can be provided by experts, existing databases and/or found through a combination of keyword searches.

Production

The Data City create one machine learning list for each vertical identified in the taxonomy. This is an iterative process that requires:

- Updating the training data: the algorithm needs different training data at different stages. Mostly, this refers to the addition of company websites that do not represent the targeted industry vertical of activity to remove false positives.

- Expert review: the different iterations' outputs are reviewed by an industry expert to confirm the data is representative. The expert review can lead to the application of changes to the training data.

- Publication of the data: once the machine learning lists have been produced and signed off, they are published on the platform as an RTIC. The raw data is downloadable in .xlsx, JSON and .csv formats. In addition, the platform also provides summary results and processed insights of the raw data, also downloadable in .csv format.

QA and data validation

The Data City's quality assurance processes are applied to ensure data integrity and robustness. The Data City implements a manual quality assurance process before the publication of RTICs.

The process consists of manually checking 30 random companies from each subcategory identified in the taxonomy. The manual check includes a review of:

- Accuracy of the URL-company match: The Data City manually checks whether the URL assigned to a particular company is correct. This is done by comparing the company name, company postcodes and director details.

- Accuracy of the machine learning classification method at the RTIC level: The Data City manually checks whether the company that has been classified in an RTIC provides products and/or services for the sector. In short, it implies checking whether the captured companies work in the sector. This is corroborated by the companies' website information.

- Accuracy of the machine learning classification method at the vertical level: manual check of the activity of the companies captured in each vertical.

Through this process, The Data City can estimate the accuracy of its datasets. The Data City's published data has a minimum confidence level of 90%, based on the random sampling described above. Updates to RTICs are done annually, with certain key ones refreshed bi-annually.

Caveats of RTIC methodology

- RTICs are a novel data product enabled by supervised machine learning technology. Several aspects of the RTIC methodology set limitations as to which companies are captured and how they are classified. The main elements to consider regarding RTIC caveats are:

  - RTICs classify only companies with a website match. The algorithm needs website text data to classify companies because language is the main criterion for classification. Therefore, companies with no URL match cannot be captured in an RTIC. However, it is important to note that RTICs tend to focus on new sectors where websites are a market need for companies.

  - There may be up to 10% of false positives in the data. Supervised machine learning technology creates sophisticated classifications that consider large numbers of keywords and key phrases. It enables finding companies in the same field even if they do not use the exact same language, overcoming the limitations of simple keyword filtering. This sometimes leads to the inclusion of edge cases in the final list, which could be referred to as false positives. The QA process limits the inclusion of this type of company in the list and user feedback helps to improve the quality of the lists published on the platform over time.

  - Companies may change their website text. The Data City re-scrapes companies' websites quarterly. The classification of companies in one RTIC could potentially be affected by changing website text.

Table A.5: Currently defined RTICs (as of December 2024)

RTICs (as of December 2024)		
AdTech	E-Commerce	Media and Publishing
Advanced Manufacturing	EdTech	MedTech
Advanced Materials	Electronics Manufacturing	Modular Construction
Agency Market	Energy Generation	Net Zero
AgriTech	Energy Management	Neurotechnology
Artificial Intelligence Ecosystem	Energy Storage	Omics
Artificial Intelligence Technologies and Applications	Engineering Biology Application	Pharma
Autonomy and Robotics	Engineering Biology Supply Chain	Photonics
B2B Services	ESG	Quantum Economy
Battery Supply Chain	FemTech	Quantum Technology
Biopharmaceutical	FinTech	Rehabilitation
Business Support Services	FoodTech	Research and Consulting - Physical Sciences and Engineering
CleanTech	Future Telecoms Supply Chain	Robotics and Autonomous systems
Cloud Computing	Gaming	Sensors
Computer Hardware	Geospatial Economy	Software as a Service (SaaS)
Cryptocurrency Economy	Immersive Technologies	Software Development
Cyber	In-Orbit Servicing and Manufacturing	Space Economy
Data Infrastructure	Internet of Things	Space Energy

Data Intermediaries	Land Remediation	Streaming Economy
Defence	Legal Services	Supply Chain Logistics
Design and Modelling Technologies	Life Sciences	Telecommunications
Digital Creative Industries	Marine and Maritime	Wearables and Quantified Self

Source: The Data City. See here for latest RTIC lists and accompanying descriptions: Real-Time Industry & Sector Data - RTICS - The Data City

Estimating Digital Economy GVA and productivity

Table A.6 presents output (GVA) per job values at the 1-digit SIC section level, which are used to estimate enterprise-level GVA for the Digital Economy (an identical methodology used in DSITs Data Driven Markets work). One limitation of this approach is these high-level estimates do not fully capture variation in productivity estimates at the 4-digit SIC or subsector taxonomy level.

Having cross-validated our estimates with related estimates (notably, DSIT's Digital Sector Economic Estimates, of which there is a 95% match in values when SICs are aligned), we recommend assessing Digital Economy productivity at the aggregate level (i.e., across the entire Digital Economy), though we also provide Tier and enterprise-level estimates for reference. In future, it is recommended the DMDE work looks to develop more granular output per job estimates to better capture productivity variations within Digital sectors.

Table A.6: Output (GVA) per job by SIC section (2023, the latest year available)

SIC Code	Description	Tier	SIC Section	GVA per job (£)
18.20/3	Reproduction of computer media	1	C	85,976
26.1 - 26.4	Manufacture Of Computer, Electronic and communication equipment	1	C	85,976
58.14 & 58.2	Publishing of Journals, Periodicals and Software	1	J	90,017
59	Motion picture, video and television programme production, sound recording and music publishing activities	1	J	90,017
60	Programming and broadcasting activities	1	J	90,017
61	Telecommunications	1	J	90,017
62	Computer Programming, Consultancy And Related Activities	1	J	90,017
63.11 - 63.91	Information Service Activities	1	J	90,017
95.11	Repair Of Computers and Peripheral Equipment	1	S and T	44,696
66	Activities Auxiliary To Financial Services And Insurance Activities	2	K	188,986
69	Legal and accounting activities	2	M	66,325
92	Gambling and betting activities	2	R	32,274

Combining IDBR data on enterprises and local sites

Source: ONS, Out per job (See: Output per job, UK - Office for National Statistics, Table_12)

The IDBR dataset contains in 3 files:

- `ents.csv` – enterprises by enterprise reference number (entref)
- `lus.csv` – local sites/unit reference number (luref) by entref
- `crns.csv` – Company Registration Number (CRN) by entref

The first two files contain details of the sites of the enterprises. Enterprises contain name, postcode, employment, turnover and SIC codes of the enterprise, whereas local sites do not have turnover and only one SIC code.

Enterprises with only one local site and no head site (i.e., present in `lus.csv` but not in `ents.csv`) lack turnover data. In such cases, the single local site is treated as the head site.

For enterprises with both a head site and local sites (i.e., present in both `ents.csv` and `lus.csv`), their data is combined and adjusted accordingly as follows:

1. If the enterprise has a single local site with zero employment:
   - Same postcode as the head site: Disregard the local site.
   - Different postcode from the head site: Treat the local site as a separate entity with zero turnover.

2. If the enterprise has a single local site with positive employment:
   - Same employment and postcode as the head site: Disregard the local site.
   - Different employment or postcode from the head site: Treat the head site and local site as separate entities.

3. If the enterprise has multiple local sites:
   - Total employment across local sites equals head site employment, and at least one local site shares the same postcode as the head site: Disregard the head site and designate the local site with the shared postcode as the new head site.
   - Total employment across local sites does not match head site employment, or no local site shares the head site postcode: Treat the head site and local sites as separate entities.

Allocating IDBR enterprise turnover to local sites

Since turnover data is missing in the local sites file (`lus.csv`), the enterprise's total turnover is distributed based on employment shares across its sites.

When the head site and local sites are treated as separate sites:

1. Subtract the total employment of local sites from the enterprise's total employment to determine the head site's employment (which may be zero).

2. Allocate the enterprise's total turnover proportionally based on the employment shares of all sites.

When a local site is designated as the head site:

1. Distribute the enterprise's total turnover based on the employment shares of all sites.

2. Assign one of the local sites that share a postcode with the enterprise as the new head site.

Companies House data is used to assign a Company Registration Number (CRN) to enterprises in the IDBR dataset that lack one. This matching process is as follows:

1. Using the Company House mapping to find enterprises with exact name matches and assign the corresponding CRN.

Identifying enterprise CRNs using Company House data

2. For unmatched enterprises, perform a fuzzy name match within the set of companies sharing the same first three digits of their postcode. Approve matches only if they meet a high-quality threshold.[24]

The below quality assurance (QA) log summarises the QA and data validation checks undertaken during the development of the Phase 2 database and economic estimates. For information on the QA and validation of

Quality Assurance log

the framework and SIC definition, please refer to the *SIC-based definition* section earlier in the Appendices.

Date	Check(s) undertaken	Observation(s)	Action(s), if required
24-Feb-25	Review raw IDBR extract from DSIT IDBR sponsor	nameline1 and postcode column values missing	Missing column values requested from DSIT IDBR sponsor
27-Feb-25	Review revised raw IDBR extract	Missing column values included	
27-Feb-25	Review implementation of data cleaning scripts	Removal of negative employment and turnover estimates	
27-Feb-25	Review implementation of fuzzy company name scripts	Missing CRNs added to dataset through fuzzy name match	
04-Mar-25	Extract and review first SIC and RTIC estimates	Total number of enterprises by SICs and by RTICs (linking by CRNs) differ significantly because some enterprises have more than one CRN	Reordered script to create SIC summary first before merging with CRN and RTICs. Fixed script to ensure secondary SIC codes not counted
05-Mar-25	Extract and review revised SIC and RTIC estimates and cross-validate with BRES, ABS, TDC estimates	SIC enterprise counts notably lower than cross-validated estimates. Aggregate employment and turnover estimates within 10% of cross-validated datasets	Fixed errors and added enterprises that ignored its single empty local site back in. Improved fuzzy match algorithm by allowing more fuzzy postcodes
05-Mar-25	Extract and review revised SIC and RTIC estimates and cross-validate with BRES, ABS, TDC estimates	All aggregate estimates within 10% of cross-validated datasets	
06-Mar-25	Extract and review revised SIC and RTIC company lists	Number of key UK Digital Economy companies excluded by omission of SIC 63.12	Added SIC 63.12 to database

---

24 The RapidFuzz package provides `token_ratio` and `token_set_ratio` scorers to evaluate match quality between company names. A match is approved only if it meets both of the following criteria: `token_set_ratio` score of at least 90%, `token_ratio` score of at least 80%. This ensures a high-confidence match when linking company names to CRNs.

06-Mar-25	Extract and review revised SIC and RTIC estimates and cross-validate with revised BRES, ABS, TDC estimates (reflecting inclusion of SIC 63.12)	All aggregate estimates remain within 10% of revised cross-validated datasets	
13-Mar-25	Extract and review appended LFS and UKRI data	Distribution of LFS SIC/SOC earnings estimates significantly different to cross-validated estimates	Improved LFS cleaning script
13-Mar-25	Extract and review revised LFS data	Distribution of LFS SIC/SOC earnings are normal	
18-Mar-25	Review appended GVA estimates, restrict SICs to cross-validate with DSIT Digital Sector Economic Estimates	Aggregate GVA estimates are within 5% of Digital Sector Economic Estimates	
19-Mar-25	Extract and review taxonomy estimates	Aggregate taxonomy estimates are higher than aggregate database estimates, as CRN RTIC values double-counted	Controlled for CRN-RTIC double counting in script
19-Mar-25	Extract and review revised taxonomy estimates. Check taxonomy company lists to ensure key companies grouped correctly	Double-counting removed, taxonomy estimates now mutually exclusive, and collectively exhaustive	
26-Mar-25	Extract and review ITL regions estimates	Aggregate ITL region estimates significantly lower than aggregate database estimates	Fixed region mapping to 3-digit postcode
26-Mar-25	Extract and review revised ITL regions estimates	Aggregate ITL region estimates normal	