DEFINING AND MEASURING THE UK DIGITAL ECONOMY

A report prepared for the Department of Science, Innovation and Technology (DSIT)

May 2024

technopolis group

Authors: Thoraya El-Rayyes, Cristina Rosemberg, Alessandro Sica, Charlotte Glass, Adebisi Adewusi, Diane Coyle, Sergi Martorell, Tia J'Nae Murray

Acknowledgements: *The authors would like to thank Alex Bjorkegren and Nayim Kareem from Department for Science, Innovation and Technology for their valuable inputs and feedback throughout the development of this report. Additionally, we appreciate the contributions from multiple officials across DSIT, the Office for National Statistics, and Innovate UK. Our thanks also go to the participants of the two workshops held for this study whose insights have significantly enriched this work.*

Any enquiries regarding this publication should be sent to us at: digitaleconomydefinitionstudy@technopolis-group.com

Contents

1. Executive Summary

The degree of digitalisation in the economy has increased dramatically over the past decade, leading to an evolution in how the 'digital economy' is defined. Earlier conceptualisations of the digital economy employed a 'core-industries' approach, defining it as a specific set of economic activities that produce ICT goods and digital services. In contrast, more recent definitions tend to adopt a 'digital inputs' approach, which additionally includes economic activity enabled by the use of ICT goods and digital services, thus reflecting the spread of digitalisation across the economy (OECD, 2020).

The Department for Science, Innovation and Technology (DSIT) (and previously the Department for Digital, Culture, Media and Sport – DCMS), have historically used two definitions of the digital economy. As these are based on Standard Industrial Classification (SIC) codes, they do not provide a definition that is comprehensive or granular enough for current policy needs, as:

- SIC codes only capture the principal activity of a firm, so production of digital goods and services in non-ICT economic sectors such as engineering are excluded (see, for example, Nathan et al, 2013).
- The codes are not detailed enough to distinguish several key industries and technologies (e.g. AI or cyber-security).
- The classification frameworks are updated infrequently, so are not well-suited to assessing the economic impact of newer industries.

The ability to evaluate the impact of key emerging industries and technologies across the entire economy is increasingly key to DSIT's policy agenda, necessitating a revised approach to defining and measuring the digital economy.

In this context, DSIT has commissioned Technopolis to develop a revised definition of the UK's digital economy and a methodology to operationalise this definition into a series of metrics. Based on the key policy and data needs articulated by DSIT and other key stakeholders in the context of this study, the revised approach proposed in this study should be characterised by:

1. **An ability to measure the 'core' digital economy (i.e. Tier 1 in Box 1 below).**
2. **An ability to generate overarching measure of the digital economy (i.e. Tiers 1 and 2 in Box 1 below).**
3. **An ability to generate key measures of interest (including GVA, turnover and employment).**
4. **An ability to identify and measure strategically important subsectors and technologies.**

The revised definition of the digital economy presented in this report has been developed based on:

- A comparative analysis of key approaches to defining and measuring the digital economy that have been implemented in the UK and economies comparable to the UK over recent years.
- A consultation process aimed at identifying stakeholders' policy and data needs and soliciting their feedback on existing approaches to defining and measuring the digital economy.

The revised definition is comprised of two tiers (as presented in Box 1) and supplemented by a flexible approach to identifying individual subsectors and technologies.

Box 1: Revised Statistical Definition of the Digital Economy

All economic activity derived from the production of digital content, ICT goods or ICT services.

- **Tier 1** of the digital economy refers to economic activity derived from firms 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 firms 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 firms in the traditional economy.

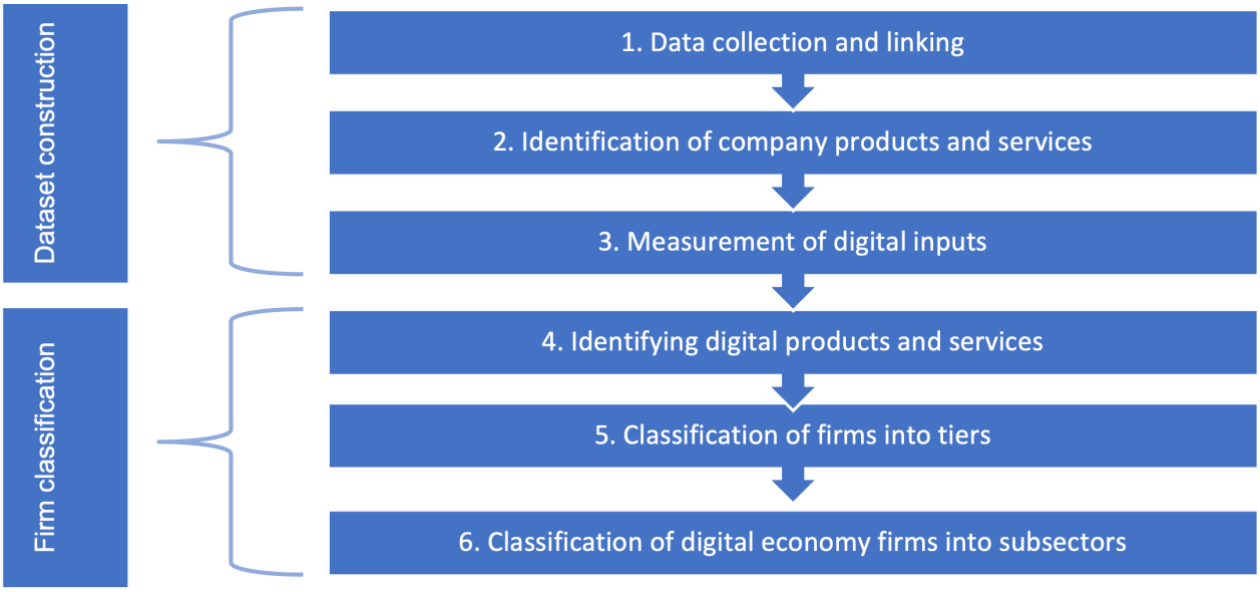
The two tiers include all firms comprising the UK digital economy. Firms within this digital economy boundary are then further classified into subsectors. A digital economy subsector is defined as a segment of the digital economy characterised by:

- The production of a set of goods and/or services that distinguishes it from other segments within the digital economy, and/or
- Development of a set of technologies that distinguishes it from other segments within the digital economy.

The sub-sectoral classification is not pre-determined. Rather, the identification of subsectors will be a dynamic empirical exercise combining data-driven approaches and expert knowledge. It will be dynamic in that the classification will be updated periodically (e.g. every 3-5 years) to capture emerging subsectors and potentially eliminate subsectors that lose economic relevance.

Finally, this report outlines an approach to operationalising the revised definition in order to measure the digital economy, as presented in Figure 1 below.

Figure 1: Approach to Measuring the UK Digital Economy



The content of this report lays the groundwork for DSIT to measure and monitor the digital economy in a way that more effectively supports the Department's strategy, setting the stage for better targeted and more impactful policy interventions.

2. Introduction

2.1 Context and objectives

The degree of digitalisation in the economy has increased dramatically over the past decade, leading to an evolution in how the 'digital economy' is defined. Earlier conceptualisations of the digital economy employed a 'core-industries' approach, defining it as a specific set of economic activities that produce ICT goods and digital services. In contrast, more recent definitions tend to adopt a 'digital inputs' approach, which additionally includes economic activity enabled by the use of ICT goods and digital services, thus reflecting the spread of digitalisation across the economy (OECD, 2020).

The Department for Science, Innovation and Technology (DSIT) (and previously the Department for Digital, Culture, Media and Sport – DCMS) have historically defined the digital sector in line with the OECD definition of the information economy. This definition was adopted by the department in 2017 and is based on Standard Industrial Classification (SIC) codes developed in 2007.[1] The department also has a wider definition of the digital economy. This includes all those employed in the digital sector and those employed in digital occupations[2] in non-digital sectors.

As these definitions are based on SIC codes, they do not provide a definition that is comprehensive or granular enough for current policy needs, as:

- SIC codes only capture the principal activity of a firm, so production of digital goods and services in non-ICT economic sectors such as engineering are excluded (see, for example, Nathan et al, 2013).
- SIC codes are not detailed enough to distinguish several key industries and technologies (e.g. AI or cyber-security).
- SIC classification frameworks are updated infrequently, so are not well-suited to assessing the economic impact of newer industries.

The ability to evaluate the impact of key emerging industries and technologies across the entire economy is increasingly key to DSIT's policy agenda, necessitating a revised approach to defining and measuring the digital economy.

In this context, DSIT has commissioned Technopolis to revise the definition of the UK's digital economy. Based on the key policy and data needs articulated by DSIT and other key stakeholders in the context of this study, the revised approach to defining and measuring the digital economy proposed in this study should be characterised by:

[1] The list of relevant SIC codes is available online [last accessed 19/04/2024].
[2] As defined in 2013 by eSkills UK using Standard Occupational Classification codes (eSkills UK, 2013).

1. **An ability to measure the 'core' digital economy.** This is conceptually comparable to the current DSIT definition of the 'digital sector' but should be measured using an approach that is more precise than reliance on ICT sector SIC codes.

2. **An ability to generate overarching measure of the digital economy**, which includes the digital economy activities of diversified companies that are involved in both digital and non-digital production.

3. **An ability to generate key measures of interest (including GVA, turnover and employment),** with the implication that it must be feasible to construct rich datasets using the definition.

4. **An ability to identify and measure strategically important subsectors,** with the implication that the definition should allow for flexible, fine-grained sub-sectoral disaggregation.

As an additional objective, this revision also aims to start building a common understanding of the definition of the digital economy across a variety of UK stakeholders beyond DSIT. Stakeholders considered international comparability of the digital economy definition to be a desirable but secondary feature.[3]

2.2 This report

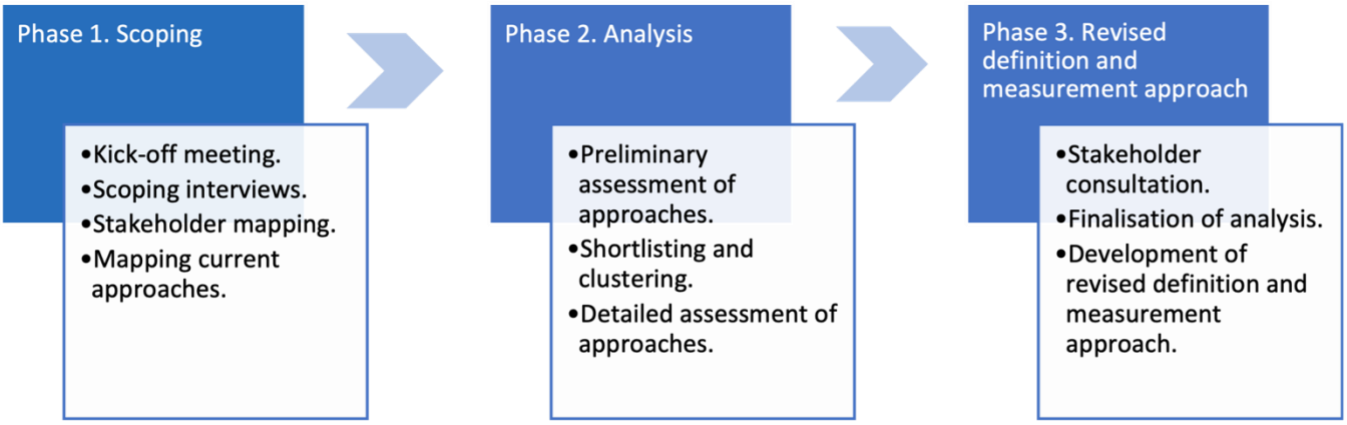
The remainder of this report is structured as follows. Section 3 presents an overview of the study methodology. Section 4 presents a description of the current DSIT approach to defining and measuring the digital economy. Section 5 presents our assessment of key recent approaches to the definition and measurement of the digital economy. Section 6 presents the proposed revised definition of the digital economy. Section 7 presents an approach to operationalising the proposed definition. Section 8 presents the conclusions and recommendations of this report. Section 9 contains references to the documents and articles consulted as part of this study.

[3] International comparability is best achieved through the application of the OECD's Digital Supply and Use Tables (DSUT) framework. Attempting to design a revised DSIT definition that is internationally comparable involves trade-offs that would undermine its ability to meet the other needs mentioned above. Additionally, the ONS is already engaged in work to apply the DSUT framework to UK national accounts. Therefore, the definition and measurement approach outlined here should be viewed as complementary to ONS' efforts.

3. Approach

This section of the report presents a summary of the study methodology (a detailed description of the methodology is presented in 0). The study involved three broad phases, as presented in Figure 2, below.

Figure 2: Phases of Defining and Measuring the UK Digital Economy Study



3.1 Phase 1: Scoping

The objective of the scoping phase was to develop a comprehensive understanding of the scope and policy context of the study. This involved:

- **Kick-off meeting and scoping interviews:** A study kick-off meeting and scoping interviews with DSIT officials to better understand their data and policy needs related to the digital economy.
- **Stakeholder mapping:** A participatory stakeholder mapping exercise conducted in collaboration with the DSIT Digital Economy and Data Policy Directorate (DEDP) to identify key stakeholders.
- **Mapping current approaches:** A mapping exercise to identify key approaches to defining and measuring the digital economy that have been implemented in the UK and internationally in economies comparable to that of the UK over recent years. This involved a literature review, complemented by consultation with DSIT officials and subject matter experts. The mapping exercise resulted in a longlist of 13 approaches (see Table 1 below).

Table 1: Longlist of Key Recent Approaches to Defining and Measuring the Digital Economy

No.	Name	Author(s)	Year
1	Global Industry Classification Standard (GICS)	MSCI / S&P Dow Jones[4]	2023
2	US Bureau of Economic Analysis approach to digital economy definition and measurement	US Bureau of Economic Analysis[5]	2019-2023
3	UN Statistics Division SNA approach to measuring economic digitalisation	UN Statistics Division[6]	2022
4	Proposed framework for updated DCMS definition of the digital sector and economy	UK Government[7]	2021
5	OECD digital economy definition and measurement toolkit	OECD[8]	2020-2023
6	DCMS digital sector definition and measurement	UK Government[9]	2016-present
7	National Institute of Economic and Social Research (NIESR) UK digital economy study	NIESR[10]	2013
8	BEIS proposed definition of the digital economy	UK Government[11]	Unknown
9	Defining and measuring the Yorkshire digital technology sector study	Glass.ai / Perspective Economics[12]	2023
Selected definition and measurement approaches for other relevant sectors			
10	Measuring the UK immersive economy study	Technopolis / Glass.ai[13]	Ongoing
11	UK artificial intelligence sector study	DSIT / Perspective Economics / Glass.ai / IPSOS[14]	2023

[4] MSCI (2023).
[5] United States Bureau of Economic Analysis (2021).
[6] United Nations Statistics Division (2023).
[7] DCMS (unpublished)
[8] OECD (2020; 2023).
[9] DCMS (2022).
[10] Nathan et al (2013).
[11] BEIS (unpublished).
[12] Glass.ai & Perspective Economics (unpublished).
[13] Technopolis & Glass.ai (forthcoming).
[14] Perspective Economics (2023).

12	UK cyber-security sectoral analysis	DSIT/ Perspective Economics/ Glass.ai/ IPSOS/ CSIT[15]	2023
13	NESTA dynamic mapping approach to economic sector measurement	NESTA[16]	2013-2015

Source: Technopolis, 2024

3.2 Phase 2: Analysis of existing approaches and stakeholder consultation

The objective of the analysis phase was to build a knowledge base about key recent approaches to the definition and measurement of the digital economy. This included systematic analysis of the advantages and disadvantages of these approaches in relation to the data and policy needs of DSIT and other relevant stakeholders. This phase involved:

1. **Preliminary assessment of longlisted approaches:** To conduct a preliminary assessment of the longlisted approaches, a critical assessment framework (CAF) was developed comprised of seven primary assessment criteria. The CAF was designed to capture whether the approaches in the longlist meet: 1) UK stakeholder needs, 2) relevant UK and international statistical quality criteria. It was developed based on consultation with the DSIT DEDP as well as relevant quality criteria described in the UK and European Statistics Codes of Practice. An overview of the assessment criteria is presented in Table 2.

   The preliminary assessment of longlisted approaches involved a high-level assessment of each approach according to each of the seven primary assessment criteria. Each criterion was assessed against a four-point scale. As a result of this assessment, two approaches that obtained very low scores were excluded from the detailed assessment as they were judged not to be relevant to DSIT's policy needs.[17] The full results of the preliminary assessment are presented in 0. The remaining ten approaches were then clustered into four groups for detailed assessment.

[15] Donaldson et al (2022).
[16] Bakhshi, et al. (2013); Spilsbury (2015).
[17] These were: 1) The Global Industry Classification Standard (GICS), which does not allow for comparisons over time, does not capture digitalisation across the economy and only includes stock-exchange listed companies; 2) The BEIS proposed definition of the digital economy, as only one component of this approach has been operationalised and it is therefore not possible to assess it against those framework criteria closely tied to data and measurement issues.

Table 2: Assessment Criteria

No.	Assessment criteria
1	Ability to generate an overarching measure of the value of the UK digital economy
2	Ability to generate key measures of interest
3	Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. quantum)
4	Translatability to Standard Industrial Classification
5	Ability to generate comparisons over time
6	Availability, timelines and statistical quality of relevant data sources (including both official and third-party data sources
7	Ability to capture firm-level data for individual subsectors (e.g. AI, cyber-security) and indicators of market competition (e.g. business entry and exit, market share).
8	Translatability to international classification systems.[18]

Source: Technopolis, 2024

2. **Critical assessment of key approaches:** To conduct the critical assessment of key approaches, a detailed version of the CAF was developed comprising an in-depth questionnaire that contains specific questions to be addressed in assessing each criterion (see 0 for the questionnaire). In addition, a secondary assessment criterion was added to the framework: translatability to international classification systems. The CAF questionnaire was developed in consultation with the DSIT DEDP to ensure that it reflects the department's needs. Each of the four key approaches to the definition and measurement of the digital economy (or relevant sectors) was assessed against this questionnaire.

   It is notable that some of the assessment criteria are interrelated.

   - Translatability to SIC codes (Criterion 4) has implications for Criterion 2 (ability to generate key measures of interest) and Criterion 5 (ability to generate comparisons over time). Specifically, if an approach is translatable to SIC codes, this can be used to generate the key measures of interest as well as comparisons over time. This means that when Criterion 4 scores highly, the scores for Criteria 2 and 5 are less consequential.

[18] This is not one of the seven primary assessment criteria used for the preliminary assessment. It is a secondary assessment criterion that was added to the framework for use in the detailed assessment.

  - There is overlap between translatability to SIC codes (Criterion 4) and translatability to international classification systems (Criterion 8). SIC codes comprise a key element of international classification systems and correspondence tables exist that allow for translatability between SIC codes and other international classification systems.

3.3 Phase 3: Development of a revised definition and measurement approach

The objectives of this phase were to:

- Iterate and refine the analysis conducted in Phase 2 through a process of stakeholder consultation.
- Develop a revised definition of the digital economy and approach to operationalising this definition in consultation with key stakeholders.

The key activities conducted for this phase were:

1. **Stakeholder workshop 1:** A workshop was held on the 22nd of February 2024 to introduce the study, collect stakeholder feedback on the assessment criteria and the extent to which different approaches to defining and measuring the digital economy meet their data and policy needs. The workshop was attended by 26 participants from nine organisations.

   A workshop synthesis note (presented in 0) summarised the key discussion points emerging from this first workshop. The output of this workshop was incorporated into a revised version of the critical assessment of approaches and informed the rationale for the development of a revised definition of the digital economy.

2. **Development of a proposed definition and high-level approach to measurement:** The study team developed a proposal for a revised definition that meets the key needs of DSIT and other key stakeholders, alongside a high-level description of how this definition can be operationalised.

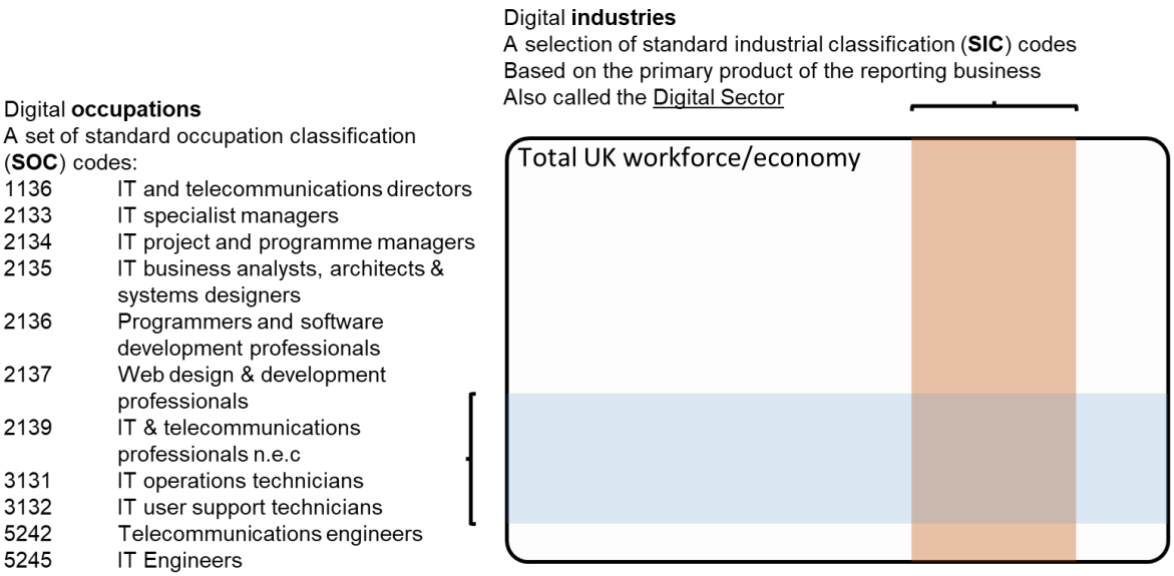
3. **Stakeholder workshop 2:** A workshop was held on the 21st of March 2024 to collect stakeholder feedback on the proposed revised digital economy definition and approach to measurement. The workshop was attended by 21 participants from eight organisations. A workshop synthesis note (presented in 0) summarised the key discussion points emerging from this second workshop. The output of the workshop was incorporated into this report.

4. **Finalisation of revised definition and approach to measurement:** The study team finalised the revised definition of the digital economy alongside a description of how this definition can be operationalised, as presented in this report.

4. Current DSIT Approach to Defining and Measuring the Digital Economy

DSIT has historically defined the digital economy as the digital sector (i.e. businesses whose main profit-generating activity is in a digital industry) in addition to individuals employed in a digital occupation outside the digital sector (see Figure 3 below).

Figure 3: DSIT Approach to Defining and Measuring the Digital Economy



Source: DEDP analytical team, DSIT

The **digital sector** is defined using SIC codes and is based on the OECD definition of the information economy (OECD, 2011), which is a combination of:

- **ICT sector:** Manufacturing and service industries "intended to fulfil or enable the function of information processing and communication by electronic means, including transmission and display."
- **Content and media sector:** Economic activities "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."

These are operationalised using a set of 4-digit SIC codes (see Table 3 below). In principle, SIC codes represent a firm's principal business activity – the activity that contributes most to its value added (ONS, 2007). More specifically SIC codes were designed to principally capture the nature of outputs ('the character of goods and services produced'), but also aggregate

information on inputs ('inputs, processes and technologies') and clients ('the uses to which these are put') of businesses in a sector.

In practice, assignment of a UK business to a SIC code (or multiple SIC codes) involves an informal process in which businesses select their own SIC code. When a business's SIC codes are used in official statistics, they are subjected to a sense check by ONS, but otherwise no further checks are conducted. In addition, a SIC classification may be correct but unintuitive because of phenomena such as contract manufacturing (for example, a company that designs but does not manufacture digital products may be classified as a wholesaler). It is therefore unsurprising that SIC codes may inaccurately represent a business' economic activity (Nathan et al, 2013).

Table 3: Standard Industrial Classification Codes Used in DSIT Digital Sector Estimates

SIC07 code	SIC07 description
2611	Manufacture of electronic components
2612	Manufacture of loaded electronic boards
2620	Manufacture of computers and peripheral equipment
2630	Manufacture of communication equipment
2640	Manufacture of consumer electronics
2680	Manufacture of magnetic and optical media
4651	Wholesale of computers, computer peripheral equipment and software
4652	Wholesale of electronic and telecommunications equipment and parts
5811	Book publishing
5812	Publishing of directories and mailing lists
5813	Publishing of newspapers
5814	Publishing of journals and periodicals
5819	Other publishing activities
5821	Publishing of computer games
5829	Other software publishing
5911	Motion picture, video and television programme production activities
5912	Motion picture, video and television programme post-production activities
5913	Motion picture, video and television programme distribution activities

5914	Motion picture projection activities
5920	Sound recording and music publishing activities
6010	Radio broadcasting
6020	Television programming and broadcasting activities
6110	Wired telecommunications activities
6120	Wireless telecommunications activities
6130	Satellite telecommunications activities
6190	Other telecommunications activities
6201	Computer programming activities
6202	Computer consultancy activities
6203	Computer facilities management activities
6209	Other information technology and computer service activities
6311	Data processing, hosting and related activities
6312	Web portals
6391	News agency activities
6399	Other information service activities n.e.c.
9511	Repair of computers and peripheral equipment
9512	Repair of communication equipment

Source: DSIT / DCMS

The definition of **digital occupations** used in DSIT's approach to definition and measurement of the digital economy was developed by eSkills UK (now TechSkills) in 2013. This definition represented industry consensus at the time, having been developed through a process of stakeholder consultation to identify the Standard Occupational Classification 2010 (SOC 2010) codes representing digital occupations. The definition primarily consists of IT specialists. Importantly, as early as 2015, research conducted by NESTA found that SOC 2010 classifications did not adequately capture emerging job roles in the information economy industries. This was particularly found to be the case in the area of data analytics, in roles such as Data Developer, Data Administrator and Data Scientist (Spilsbury, 2015). A new version of the Standard Occupational Classification (SOC 2020) was published in 2020, containing a more comprehensive list of occupations relevant to the digital economy. This has yet to be applied to the DSIT digital economy estimates.

The current DSIT approach has several limitations:

- First, it **does not capture production of digital goods and services across the whole economy.** For example, it does not capture economic activities such as the production of digital goods and services in economic sectors such as engineering (Nathan et al, 2013). It is therefore limited as a framework for measuring the overarching value of the digital economy.
- Second, **it does not allow for granular measurement of several key industries and technologies** (e.g. AI or cyber-security).
- Third, as both the sector and occupation definitions are implemented via the standard industry and occupation classifications, which are updated infrequently, **they are not well-suited to assessing the economic impact of newer industries.** Given the current remit of the department and the focus on innovative, emerging technologies, this represents a significant limitation.

5. Assessing Key Current Approaches to Digital Economy Definition and Measurement

5.1 Introduction

This section of the report presents a summary version of the critical assessment of the four key approaches to the definition and measurement of the digital economy. These four approaches are:

- Operationalisation of the OECD 2020 definition of the digital economy (OECD, 2020; 2023).
- Deep web reading approach to measuring economic sectors (Donaldson et al, 2023; Glass.ai and Perspective Economics, unpublished; Perspective Economics, 2023; Technopolis & Glass.ai, forthcoming).
- NIESR UK digital economy study (Nathan et al, 2013).
- NESTA dynamic mapping approach to economic sector measurement (Bakhshi et al, 2013; Spilsbury, 2015).

For each approach, this section of the report presents the findings of the detailed critical assessment in summarised form, in addition to an overview of its key advantages and disadvantages. The full version of the critical assessment is presented in 0. As discussed in section 0, it should be noted that these criteria are interrelated.

Our analysis found that none of the assessed approaches score highly on all of the assessment criteria. Put differently, none of the approaches meet all of DSIT's key data and policy needs. All four approaches can provide a measure of the overall size of the UK digital economy (albeit with varying levels of accuracy) and all are easily translatable to SIC codes. However, some approaches are less flexible in terms of providing a detailed snapshot of specific technologies and subsectors and capturing firm-level data (OECD and NESTA), while others are less well-suited to drawing comparisons over time and require more resources for the data collection process (Glass.ai and NIESR). All of the approaches assessed pose some challenges related to the availability of high-quality data. An overview of the assessment findings is presented in Table 4 below.

Table 4: Summary of critical assessment of approaches to defining and measuring the digital economy

	OECD	Deep web reading	NIESR	NESTA
Criterion 1: Ability to generate an overarching measure of the value of the UK digital economy	●●●●●	●●●●○	●●●●○	●●●○○
Criterion 2: Ability to generate key measures of interest	●●●●○	●●●●○	●●●○○	●●●●○
Criterion 3: Ability to capture the contribution of individual subsectors, including the identification and integration of future emerging technologies as subsectors	●●●○○	●●●●●	●●●●○	●●●○○
Criterion 4: Translatability to Standard Industrial Classification	●●●●●	●●●●○	●●●●●	●●●●●
Criterion 5: Ability to generate comparisons over time	●●●●○	●●●○○	●●●○○	●●●●○
Criterion 6: Availability, timeliness and statistical quality of relevant data sources	●●●○○	●●●○○	●●●○○	●●●○○
Criterion 7: Ability to capture firm-level data for individual subsectors and indicators of market competition	●●●○○	●●●●●	●●●●○	●●○○○
Criterion 8: Translatability to international classification systems	●●●●●	●●●○○	●●○○○	●●●●○

5.2.1 Operationalisation of the OECD 2020 definition of the digital economy

Overview of definition

The digital economy is defined by the OECD as "all economic activity reliant on, or significantly enhanced by the use of digital inputs, including digital technologies, digital infrastructure, digital services, and data; it refers to all producers and consumers, including government, that are utilising these digital inputs in their economic activities".

A tiered definitional framework is provided by **the OECD** to complement this conceptual definition, separating economic units into:

- The **'core'** digital economy. This is economic activity by producers of ICT goods, ICT services and digital content. It is also described as 'digitally enabling industries' and is similar to the current DSIT/DCMS definition of the digital sector.
- The **'narrow'** digital economy. This is economic activity from producers reliant on digital inputs.
- The **'broad'** digital economy. This is economic activity from producers significantly enhanced by the use of digital inputs.

A final tier, the **'digital society'**, includes all digital activities undertaken by individuals in a society that are not carried out for pay or profit (i.e. not included in the System of National Accounts production boundary), such as the use of free digital platforms (including public digital platforms). This tier is outside the scope of the current study and is not discussed further below.

Key advantages

- Ability to generate a wide range of key measures and indicators related to the digital economy, covering both the core ICT sector and firms reliant on ICT goods and services.
- Provides a flexible framework for measuring the digital economy that can potentially be adapted to include any relevant industry or product (e.g. AI, cyber-security, quantum).
- Aligned with international efforts to make the digital economy more visible in national accounts, facilitating international comparison.
- Most data required to operationalise this definition comes from existing official statistics (ONS surveys and other government statistics) and is consistent with national accounts. These data sources are readily available, non-proprietary, relatively timely and of high statistical quality.

Key disadvantages

- The framework includes elements whose measurement requires the development of new data sources. Thus, it is not possible to immediately measure all aspects of the framework (e.g. measurement of the value of data as a product).
- Deriving measures of digital industries and products from estimates in standard Supply and Use Tables (SUTs) requires making a number of assumptions, implementing corrections and balancing product lines.
- The addition of new lines to the SUTs in order to produce the Digital SUTs (DSUTs) is not necessarily backwards compatible.
- There is a very limited number of product codes in the UK, which makes it very difficult to define industries based on their mixture of inputs.

Overview of approach to measurement

This definition has been implemented in a number of countries using a national accounts approach (see Appendix 0 for examples). According to this approach, the first two tiers of the digital economy (core and narrow) are measured through the development of DSUTs. DSUTs are compiled by adding new rows and columns to traditional SUTs. These additions measure a range of digital products and services, digital industries, and the nature of economic transactions (i.e. digital/non-digital). DSUTs include:

- Seven additional columns for key digital industries defined by this approach: 1) digitally enabling industries, 2) digital intermediary platforms charging a fee, 3) data and advertising driven digital platforms, 4) firms dependent on intermediary platforms, 5) e-tailers, 6) digital-only firms providing financial and insurance services, 7) other producers only operating digitally.
- Six additional rows under each product (and total), separating transactions by whether they are digitally ordered or not digitally ordered, with digitally ordered transactions further broken down into ordered directly from the counterparty or ordered via a digital intermediation platform. A final breakdown disaggregates the products ordered via digital intermediary platforms between resident and non-resident platforms.
- Two additional columns showing the nature of the delivery of the service as either digitally delivered or not digitally delivered.
- Four additional rows, representing digital products of particular interest: 1) digital intermediation services. 2) cloud computing services, 3) ICT goods, 4) digital services.
- Three additional rows, representing data and digital service products that are currently outside the SNA production boundary (e.g. data).

DSUTs are either compiled through the reallocation of specific economic units that match one of the seven digital industries or through an aggregate reallocation based on indicators. The latter approach has been implemented using a variety of methods. For example, some countries have allocated units to the seven digital industries using business survey data, firm-level data (e.g. annual reports and tax fillings), external studies (e.g. on sharing economy data), expert judgements, and decision trees.

DSUTs contain information on output and consumption for different industries and goods, allowing for the generation of measures including GVA, employment, international trade, and expenditure disaggregated by the nature of the transaction (i.e. digital/ non-digital).

5.2.2 Deep web reading approach to measuring economic sectors

Overview of definition

The digital economy is defined as comprising all companies operating in a 'digital sector', as defined by a bespoke taxonomy that is developed through a combination of structured stakeholder consultation and deep reading of web content and official statistics about UK companies.

Companies may be classified into different categories, including developers of digital goods and services, diversified companies involved in both digital and non-digital production, and adopters of digital technologies (i.e. companies not engaged in the production of digital goods and services).

Key advantages

- Ability to generate a range of key measures and indicators related to the digital economy.
- Ability to capture the contribution of individual priority subsectors.
- Ability to capture comprehensive firm-level data.
- Ability to correct for misclassification of companies resulting from companies selecting their own SIC codes.
- Ability to impute SIC codes to companies that have not selected any.
- The approach does not require proprietary data (although such data sources may be used to enrich the dataset).

Key disadvantages

- Requires original data collection that may be costly and must be repeated on a periodic basis in order to update the dataset.
- Translation to international classification systems requires additional effort, and in some cases will not be comprehensive.

Overview of approach to measurement

Operationalising this approach involves:

- Developing a conceptual definition of the digital economy or a longlist of relevant companies in the sector (for a more data-driven approach).
- Building on the previous step, developing a taxonomy of the sector based on a combination of desk research and stakeholder consultation. This step may also draw on data-driven insights from a language model. Depending on the nature and size of the sector under study, developing the taxonomy may involve developing a list of subsectors and/or technologies related to the sector/subsectors. Development of these lists may involve workshop sessions with representatives from academia, industry, government and the study team.
- Producing a list of relevant keywords/ phrases associated with each taxonomic category. This is often an iterative process in which preliminary sample results are collected and shared with experts for feedback in order to refine and improve the language model used for taxonomic classification.

- Running a web crawl to read company websites and social media presence to construct a dataset of companies included in the sector. For example, the Glass.ai AI capability deep reads ~2.2 million UK organisation websites (including companies, partnerships, government, non-profits, and sole traders) and social media pages (e.g. LinkedIn), as well as potentially reading relevant news webpages and sector-specific sources to identify text that may suggest a company is active in the sector under study. In addition, the Glass.ai AI capability regularly crawls the websites of 35 million businesses globally and is therefore able to identify foreign companies with a presence in the UK.
- Link the list of companies identified in the previous step to other data sources to obtain additional information on companies' characteristics (e.g. employment, turnover, et cetera) including ONS statistics and the Companies House register data. It is also possible to include data from proprietary datasets such as the Fame, Beauhurst, or Crunchbase datasets in this data linking step.
- Where necessary, a final step may be implemented that involves fielding an original survey of companies to fill data gaps.

5.2.3 NIESR approach to measuring the UK digital economy

Overview of definition

The digital economy is defined as comprising companies that operate in a 'digital sector' **and** are involved in the production of 'digital goods' and/or 'digital services'.

Digital sectors are defined as comprising two sectoral groups: 'information and communications technology' (ICT) and 'digital content' as follows:

- ICT: Producers of ICT systems, hardware and software, in addition to producers of related services around these products (e.g. sales, installation and maintenance of the products). This sectoral group is characterised in terms of its outputs.
- Digital content: Firms in which the only or principal outputs are digital products or services.

Digital goods and services are defined as those commonly produced by firms classified under the SIC codes used in the UK Government's statistical definition of the digital economy.

Key advantages

- Ability to capture digital subsectors at a granular level.
- Ability to generate a range of key measures and indicators related to the digital economy, particularly if the approach is adapted to incorporate a greater number of data sources.
- Ability to correct for misclassification of companies resulting from the companies selecting their own SIC codes.
- Ability to impute SIC codes to companies that have not selected any.

Key disadvantages

- Requires extensive original data collection, involving both web-mining and data linking techniques.
- Lack of transparency regarding the quality assurance processes for evaluating the web-mined data's quality and representativeness, as well as the validity of its categorisations.

Overview of approach to measurement:

To define and measure the digital economy, this approach draws on a proprietary dataset comprised of all active companies in the UK registered at Companies House. To construct this dataset, the Companies House register data is matched to web-mined data and data from other public sources (including Companies House filings, patents, text from company websites and text from press coverage of the company). Quantitative text analysis techniques are used to classify companies according to a fine-grained, multi-dimensional classification system by categories including 'sector context' and 'product type' (henceforth NIESR sectors/products). These codings are probabilistic. The 'most likely' categories are used to allocate companies to sectors and products.

The approach is based on an assumption that the SIC codes used in the UK Government's statistical definition of the digital economy have some correspondence to the 'true' digital economy but do not fully capture it. It therefore uses these SIC codes as the starting point to develop a new definition and measurement of the sector. Importantly, however, it is not possible to operationalise this definition using SIC codes. This is because even the most detailed level of SIC codes does not provide adequate information on the nature of products and services within a sector that is needed to operationalise this definition.

The approach involves:

- Mapping all companies in the Companies House register with digital economy SIC codes to the corresponding NIESR sectors and products. This mapping provides an initial 'cut' of NIESR sectors and products that are relevant for the digital economy.
- Using a 'threshold rule' to exclude sparse sector and product groups from the analysis. This step provided an initial list of 21 NIESR sectors and 16 product groups that comprise the 'digital economy'.
- Applying manual filtering to correct for three types of potential misclassification resulting from: a) that resulting from companies selecting their own SIC codes; b) that resulting from lack of precise correspondence between SIC codes and NIESR categories; c) that resulting from probabilistic coding of NIESR sectors / products.
- Conducting a manual precision check. This involves cross-tabulating NIESR sector and product codes against 5-digit SIC codes to search for sets of companies that are not digital by the selection rules of this methodology (see Appendix 0 for further detail). These set of companies are removed from the definition.
- In order to be classified as digital, a company must be classified as both belonging to a digital sector and producing a digital product as defined in the previous steps. Thus, the final step is to construct a set of companies by 'sector-product', which consists of companies in digital sectors whose principal activity is also digital.

5.2.4 NESTA dynamic mapping approach to defining and measuring economic sectors

Overview of definition

The digital economy is defined as comprising companies in the ICT sector (by 4-digit SIC codes) in addition to other economic subsectors characterised by a comparable level of digital employment intensity.

Key advantages

- Ability to generate a wide range of key measures and indicators related to the digital economy.
- The data required to operationalise this definition is readily available, non-proprietary, timely and of high statistical quality.
- Easily translatable to SIC and SOC codes.

Key disadvantages

- Does not allow for measurement of several key strategic subsectors.
- Does not provide firm-level data or allow for measurement of market competition indicators, although the approach could be adapted to allow for this through a data linking exercise.
- Relies on SIC codes as self-selected by companies. This increases measurement error due to a substantial proportion of companies not classifying or misclassifying themselves.

Overview of approach to measurement

This approach was developed by NESTA in 2013 to define and measure the UK's creative industries (Bakhshi et al, 2013) and applied to the UK's information economy industries in 2015 (Spillsbury, 2015). For simplicity, it is described below as it could be applied to the digital economy.

The approach involves:

- Developing a list of 'digital occupations' based on the UK's Standard Occupational Classification (SOC). This involves developing a set of theoretically grounded criteria that describe the characteristics of digital employment. All SOC occupations are then scored against these criteria. The criteria are typically derived from a conceptual definition of the sector being measured.
- Using the list in Step 1 to calculate the 'digital intensity' of each economic subsector in the UK at the level of 4-digit SIC codes. The digital intensity of an economic subsector is defined as the proportion of workers employed in that sector who are in digital occupations.
- Defining a threshold for the level of digital intensity at which a subsector (4-digit SIC code) is included / excluded from the definition of the digital economy. Various methods for determining this threshold have been employed, some of which rely on the use of pre-existing UK Government sectoral definitions as a benchmark and others which combine this benchmarking with statistical techniques to determine the precise threshold.
- This approach relies entirely on employment intensity as a proxy measure for identifying digital sectors. The extent to which this introduces bias into the subsector selection process is an open empirical question, which may merit further examination. Research published by the OECD (Calvino et al, 2018) suggests that measuring the digital intensity of economic subsectors using

different proxy measures (e.g. investment in ICT equipment, investment in software, share of turnover from online sales, employment of ICT specialists) is likely to yield different results. Further research could explore the utility of updating the NESTA methodology by including additional proxy indicators in the subsector selection process. This would result in a more multidimensional approach to defining and measuring the digital economy.

5.3 Comparative assessment

The four approaches described above were assessed against the eight criteria described in section 3.2 of this report. The following tables present a summary of the assessment and a score against each criterion. 0 includes the full assessment for each definition.

Criterion 1: Ability to generate an overarching measure of the value of the UK digital economy		
OECD	DSUTs allow for estimation of the value of the core and narrow tiers of the digital economy, with the ability to disaggregate at the level of seven digital industries, including digital intermediary platforms as well as digital-only firms providing financial and insurance services. This approach also allows cross-industry measurement on transactions based on their nature, providing an alternative measure of the digital economy measure based on all economic activity that is digitally ordered and/or digitally delivered.	●●●●●
Deep web reading	The approach is able to generate an overarching measure of the value of the UK digital economy. It allows for a flexible, fine-grained classification of companies into subsectors and can therefore be designed to allow for the measurement of the economic value of digital-only service producers and digital intermediary platforms. If a tiered definition if adopted, this approach can accommodate the inclusion of companies where production has been significantly enhanced by digital technology (i.e. adopters of digital goods and services). The approach can also capture foreign digital companies active in the UK market.	●●●●○
NIESR	The findings of the study conducted using this approach demonstrate that it represents a more inclusive definition of the digital economy than the core digital sector definition used by the UK Government. The study found the digital economy to comprise almost 270,000 active companies in the UK (14.4% of all UK companies) compared to 167,000 companies (10.0%) when the government's conventional definition is used. In particular, the study finds that this approach includes a large number of companies in business and domestic software, architectural activities, engineering, and engineering-related scientific and technical consulting that are not captured by the conventional SIC-based definition of the digital economy. It is notable that this approach allows for a flexible, fine-grained classification of companies into subsectors. Thus, it could be adapted to allow for the measurement of the economic value of digital-only service producers and digital intermediary platforms. Finally, it is important to note that this approach is based on classifying companies according to their subsector and the goods/ services they produce. It does not involve the collection of data on digital inputs. Thus, it will not capture those economic units where production has been significantly enhanced by digital technology, but which are not in a digital sector or do not produce digital goods/ services.	●●●●○

NESTA	Such an approach would employ the existing UK Government definition of the digital sector as a benchmark for the level of digital intensity required for an economic subsector (4-digit SIC code) to be included/ excluded in the digital economy. Additionally, like the existing UK Government definition, this approach is based on 4-digit SIC codes. The resulting measure is therefore likely to be broadly comparable to the existing UK Government definition of the core digital sector. This approach would likely only capture the economic value of digital-only service producers where such producers are dominant in a particular economic subsector (4-digit SIC code). This is because each economic subsector (4-digit SIC code) is treated as a single unit, without regard for the variation between firms within the sector. For example, if an EdTech platform is categorised as falling under an education sector SIC code, it would be unlikely to be captured by this approach. This approach would likely only capture the economic value of digital-intermediary platforms where such platforms are dominant in a particular economic subsector (4-digit SIC code). For example, this would very likely exclude platforms such as Booking.com and OpenTable which are not classified under ICT or media SIC codes. It should be noted that this approach would not allow for disaggregation of either digital-only service producers or digital-intermediary platforms from the overarching measure of the digital economy. Finally, this approach would only capture economic units where production has been significantly enhanced by digital technology if these units belong to economic subsectors (4-digit SIC code) where digitalisation is dominant.	●●●○○

Criterion 2: Ability to generate key measures of interest		
OECD	Similarly to traditional SUTs, DSUTs are able to generate measures of GVA, employment, worker earnings and international trade, as well as the measurement of R&D (according to the Frascati definition[19]) in the different categories.	●●●●○
Deep web reading	Measures of interest that have been generated using this approach include gross value-added, employment, import and exports, business demography measures (including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites). It also allows for the calculation of estimates of investment in the sector disaggregated by stage (e.g. venture, seed).	●●●●○
NIESR	This approach allows for the measurement of a range of key measures and indicators related to the digital economy. It allows for the production of business demography measures, including number of businesses and regional distribution of business sites. An adapted version of this approach (linking the principal firm-level dataset with other datasets) would allow for the estimation of additional business demography measures, such as size of businesses by annual turnover and size of businesses by employees. The approach can also be used to measure the goods and services produced by individual firms and in subsectors. This approach does not allow for the measurement of the digital economy's gross value-added, employment or earnings. However, an adapted version of this approach (linking the principal firm-level dataset with other datasets and collecting employment data from web sources such as LinkedIn) would allow for the estimation of these measures.	●●●○○
NESTA	This approach allows for the measurement of the digital economy's gross value-added, employment, worker earnings and business demography (including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites). It also allows for the calculation of estimates for imports and exports of digital services and digital goods. More generally, this approach can be used to measure any indicator for which the ONS provides nationally representative data at the level of 4-digit SIC codes. In addition, it can be used to measure the number of digital workers across the economy (i.e. including digital workers employed outside the digital economy).	●●●●○

[19] A well-established methodology for collecting and using R&D statistics. See OECD (2015). Frascati Manual 2015. Guidelines for Collecting and Reporting Data on Research and Experimental Development.

Criterion 3: Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. Quantum)		
OECD	The standard DSUT framework highlights two products of interest: digital intermediary services and cloud computing services. A similar approach could be taken with other services such as artificial intelligence or cyber-security, using a mix of statistical sources and studies utilising alternative data sources (such as web data). The standard DSUT framework does include products outside the SNA production boundary, such as data. However, measurement of these is not foreseen in the near future as data on them is not captured in national accounts statistics (although an estimate for data by the sum of costs method is likely to be included in future).	●●●○○
Deep web reading	This approach involves a data-driven process that reads various non-proprietary sources to identify subsectors based on several sources of firm-level data. It therefore allows for the measurement of specific subsectors as it is possible to infer these subsectors from the firm-level data available. It further allows for the identification and incorporation of future emerging technologies as subsectors when needed.	●●●●●
NIESR	This approach allows for the measurement of the ICT trade subsector (including disaggregation into wholesale and retail trade) and the telecommunications subsector. More generally, this approach allows for a flexible, fine-grained classification of companies into subsectors. An adapted version of this approach would therefore allow for the measurement of additional subsectors, insofar as it is possible to infer these subsectors from the firm-level dataset used. Thus, while this approach does not currently allow for the measurement of key strategic subsectors (e.g. data, AI, cyber-security and digital services) it could be adapted to do so. Similarly, it could allow for the identification and incorporation of future emerging technologies as subsectors when needed.	●●●●○
NESTA	This approach allows for the measurement of the following key subsectors: Data processing, hosting and infrastructure (SIC code 6311) Software development (SIC code 6201) ICT manufacturing subsector (various SIC codes in SIC divisions 26 and 27) ICT trade subsector (various SIC codes in SIC divisions 46 and 47)	●●●○○

	Telecommunications subsector (SIC division 61) Other subsectors of interest that can be measured using this approach include 'web portals' (SIC code 6312) and 'computer consultancy activities' (SIC code 6202). In contrast, this approach does not allow for the individual measurement of the artificial intelligence subsector, the cyber-security subsector and digital services (e.g. FinTech, EdTech, HealthTech). It also does not allow for the identification and incorporation of future emerging technologies as subsectors.	

Criterion 4: Translatability to Standard Industrial Classification		
OECD	Many of the data sources used to compile SUTs use SIC codes and DSUTs are similarly expected to use them.	●●●●●
Deep web reading	The approach can assign individual companies in the dataset to a single industry which can be mapped back to SIC codes. Such mapping can be facilitated using automated models to match inferences from web data with official administrative data (e.g. the Companies House register).	●●●●○
NIESR	This approach is based on mapping digital economy SIC codes to the NIESR sectoral classification. The firm-level data is therefore easily translatable to SIC codes.	●●●●●
NESTA	This approach is based on SIC codes.	●●●●●

Criterion 5: Ability to generate comparisons over time		
OECD	The ability to generate comparisons over time depends on the measurement approach used for the different elements of the DSUTs. Most SNA component data are collected quarterly and allow for such comparisons. Elements of the DSUTs that are measured using new and/or alternative data sources (e.g. external studies, ad hoc measurement using digital tools) may be limited in the extent to which comparisons over time are possible.	●●●●○
Deep web reading	This approach allows for measurement of trends over time, although this would require a substantial data collection effort for each point in the time series. Importantly, however, collecting historical data would require web mining of archived firm-level text data and is unlikely to be feasible. In principle, this approach allows for measurement of time trends at any given level of granularity. However, it requires a substantial data collection effort for each point in a time series. It is therefore unlikely that it would be feasible to measure time trends more frequently than on an annual basis.	●●●○○
NIESR	This approach allows for measurement of trends over time, although this would require a substantial data collection effort for each point in every time series. Importantly, however, collecting historical data would require web mining of archived firm-level text data and is unlikely to be feasible. In principle, this approach allows for measurement of time trends at any given level of granularity. However, it requires a substantial data collection effort for each point in a time series. It is therefore unlikely that it would be feasible to measure time trends more frequently than on an annual basis.	●●●○○
NESTA	Because this approach is based on SIC codes, it can draw ONS statistics for measurement. Thus, it allows for measurement of trends over time, including historical comparison.	●●●●○

Criterion 6: Availability, timeliness and statistical quality of relevant data sources		
OECD	DSUTs involve reallocating economic activity already contained in national accounts and conventional SUTs in a way that provides insights into the digital economy. Thus, much of the data is already available from official national statistics. The statistical quality of this data is high. With regard to timeliness, SUTs are produced with some lag. At the time of writing, the most recent SUTs published for the UK are for the year 2021. There are data availability issues related to compiling the full set of supplementary DSUT columns and rows. For some elements of DSUTs, alternative sources of data are needed to calculate aggregate estimates associated with the production of units that cannot be identified using conventional SUTs (e.g. for firms dependent on intermediary platforms it might be necessary to use estimates from studies on the sharing economy). Supplementary data collection may be required, for example, through the addition of new questions to specialised surveys pertaining to the digital economy (in the case of the UK, the Digital Economy Survey).	●●●○○
Deep web reading	This approach employs datasets constructed using a combination of AI-driven web intelligence drawn from publicly available web data and linked firm-level data from numerous sources including ONS surveys, the Companies House register and non-proprietary datasets. It is possible to construct such a dataset without subscriptions to proprietary data sources. Data collected through deep web reading is as recent as the latest update to a given company website or social media page. The linked data sources used are typically updated on an annual basis. With regard to the representativeness of the data, the firm-level dataset draws primarily on web crawled data from company websites and social media pages. This affects the representativeness of the data in economic sectors where companies' online presence tends to be limited. In the case of digital economy companies, however, this data is highly likely to approximate a census. This would make it representative at the national, regional, sectoral and sub-sectoral levels. One caveat is that micro-sized firms are likely underrepresented in this dataset. If such an approach were to be adopted, it would be advisable to conduct an empirical analysis to validate the representativeness of the dataset.	●●●○○
NIESR	With regard to data availability, the Companies House data used to operationalise this definition is readily available. However, the construction of the firm-level dataset requires original data collection using web mining techniques and data linking. With regard to the timeliness of the data, the web-mined data would be as recent as the last update to a given company's website. The linked data sources used are typically updated on an annual basis. With regard to the representativeness of the data, the firm-level dataset used to operationalise this definition draws largely on web-mined data from company websites. This affects the representativeness of the data in economic sectors where companies' online presence tends to be	●●●○○

	limited. In the case of digital economy companies, however, this data is highly likely to approximate a census. This would make it representative at the national, regional, sectoral and sub-sectoral levels. One caveat is that micro-sized firms are likely underrepresented in this dataset. If such an approach were to be adopted, it would be advisable to conduct empirical analyses to validate the representativeness of the dataset.	
NESTA	The data required to operationalise the definition is largely available from ONS statistics. However, operationalisation of the definition requires an exercise to identify which SOC codes represent 'digital occupations'. Conducting this exercise in a rigorous manner would require the collection of expert judgements regarding whether individual SOC occupations fulfil the definition of a 'digital occupation'. Proprietary data sources are not required to operationalise this definition.	●●●○○

Criterion 7: Ability to capture firm-level data for individual subsectors and indicators of market competition		
OECD	Much of the data feeding into the framework for Digital SUTs is built from firm-level microdata. Indeed, some countries have used non-conventional sources of firm-level data (e.g. annual reports and tax fillings in Canada) to identify units to be allocated into the industry categories. In principle, accessing this microdata via the ONS Secure Research Service (SRS) could provide firm-level data for individual subsectors. In practice, however, the official firm-level data available through the ONS SRS provides very limited information that would allow for identification of the digital subsector to which a firm belongs. Firm-level data relevant for these purposes is often survey data (e.g. goods and services produced) rather than more comprehensive administrative or firm census data. This is likely to pose challenges for capturing firm-level data on individual digital subsectors because the sample sizes of relevant UK business surveys is often too small to obtain reliable estimates for smaller subsectors (e.g. AI or cybersecurity). In principle, it should also be possible to use the abovementioned sources of firm-level data to measures indicators of market competition such as market share concentration and business entry and exit. However, this is subject to the same limitations described above.	●●●○○
Deep web reading	This approach uses a firm-level dataset to measure the digital economy. It is designed to allow for a flexible, fine-grained classification of companies into subsectors and therefore allows for measurement of firm-level data for individual priority subsectors of interest to DSIT. With regard to indicators of market competition, this approach allows for the measurement of market share concentration as well as business entry and exit. For instance, business entry/exit can be measured through official sources (active/inactive in Companies House) and company websites (active / inactive).	●●●●●

NIESR	This approach uses a firm-level dataset to measure the digital economy. While the subsectoral classifications in this dataset are not comprehensive, the approach is designed to allow for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to allow for measurement of firm-level data for subsectors of priority to DSIT. With regard to indicators of market competition, this approach allows for the measurement of business entry and exit. It does not allow for the measurement of market share concentration. However, an adapted version of this approach (linking the primary firm-level dataset with other datasets) would allow for the estimation of market share concentration.	●●●●○
NESTA	This approach does not allow for the measurement of firm-level data. However, it could be adapted to allow for this through a data linking exercise. To ensure that such a data linking exercise produces a high-quality dataset would require: 1) A process to address misclassification of companies resulting from companies selecting their own SIC codes; 2) A process to impute SIC codes to companies for which this data is missing. To implement this to a high standard, collection of firm-level web data would be required. With regard to indicators of market competition: This approach allows for estimation of market share concentration in the digital economy, as well as selected subsectors (see response to criterion 3 for description of subsectors). High-level administrative statistics on companies' finances at the 4-digit SIC level are available from ONS and can be used to derive proxy measures of sectoral/ sub-sectoral market share concentration. This approach allows for measurement of business entry and exit in the digital economy, as well as selected subsectors (see response to criterion 3 for description of subsectors). This data is available from administrative statistics on business birth and death rates at the 4-digit SIC level available from ONS.	●●○○○

Criterion 8: Translatability to international classification systems		
OECD	This approach is mainly based on official national statistics and classifications (i.e. SIC and CPA). Correspondence tables to translate these classifications into other international classification systems such as SOC and the Harmonised System are publicly available.	●●●●●
Deep web reading	This approach is not designed to use international classification systems. However, the dataset includes firm-level web data on digital occupations, which can be mapped back to SOC codes. Similarly, information on company products in the dataset can be mapped to the Harmonised System. However, it is important to note that the correspondence between the web-crawled occupational and product categories will not correspond fully with these international classification systems and there will be a degree of measurement error in the resulting translation.	●●●○○
NIESR	This approach does not involve the collection of data on employment and is therefore not translatable to the Standard Occupational Classification. With regard to the Harmonised System, this approach does identify high-level product categories produced by individual companies. However, the classification system employed is driven by web-mined data collected on individual companies and does not map on to the Harmonised System. It would therefore be challenging to translate the product information collected by this approach to the Harmonised System, unless the approach to categorisation is redesigned to ensure that the product classification can be mapped on to these categories. It should be noted that such a change to the methodology could be technically challenging and may degrade the quality of data collected on goods produced by individual companies.	●●○○○
NESTA	This approach is based on SOC codes and is therefore easily translatable to the Standard Occupational Classification. With regard to the Harmonised System, this approach is based on 4-digit SIC codes and it is not possible to correlate the Harmonised System comprehensively to 4-digit SIC codes. The UK Standard Industrial Classification (SIC) 2007 is the national version of "The Statistical Classification of Economic Activities in the European Community (NACE) Rev. 2" and as such must be identical to NACE Rev 2, down to and including the fourth digit of the system. Correlations between NACE Rev 2 and other international classifications are produced and held by Eurostat. Due to fundamental differences between the two systems, Eurostat advises that it is not possible to correlate the Harmonised System comprehensively with NACE Rev 2.	●●●●○

6. Revised Definition of the Digital Economy

This section of the report presents a revised definition of the digital economy and describes the rationale informing this definition. As discussed in Section 5, each of the existing approaches to defining and measuring the digital economy assessed in this study has different advantages and disadvantages. Importantly, none of these approaches meet all of DSIT's key policy and data needs relating to measurement of the digital economy. Consequently, the proposal presented below combines elements from the four approaches to optimise alignment between the definition and these policy and data needs.

6.1 Revised definition

The revised definition of the digital economy comprises two tiers and it is complemented by a flexible approach to identifying individual digital subsectors. It draws elements from the four definitions assessed in the previous section, such as the tiered approach used in the OECD definition, and the concept of diversified firms from the 'deep web reading' approach.

Box 2: Revised Statistical Definition of the Digital Economy

The digital economy encompasses all economic activity derived from the production of digital content, ICT goods or ICT services. It includes two tiers:

- **Tier 1** includes the economic activity derived from firms that are primarily involved in the production of digital content and/or ICT goods and/or ICT services.
- **Tier 2** includes the economic activity derived from diversified firms 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 firms in the traditional economy.

These two tiers include all firms classified as comprising the digital economy. Firms within this digital economy boundary are then further classified into subsectors. A digital economy subsector is defined as a segment of the digital economy characterised by:

- The production of a set of goods and/or services that distinguishes it from other segments within the digital economy, and/or
- Development of a set of technologies that distinguishes it from other segments within the digital economy.

The sub-sectoral classification will not be pre-determined. Rather, the identification of subsectors will be a dynamic empirical exercise combining data-driven approaches and expert

knowledge. It will be dynamic in that the sub-sectoral classification will be updated in line with policy needs to capture emerging subsectors and potentially eliminate subsectors that lose economic relevance. This process is described further in sections 7.1 and 7.2 of this report. All subsectors will be mapped to SIC codes.

Broader definitions of the digital economy, such as the OECD definition, may additionally include economic activity derived from firms that are reliant on digital inputs but do not produce digital content, ICT goods and/or services. These firms, which can be described as ‘digital adopters’ are not included in the definition used in this report.

6.2 Rationale

The revised definition has three key characteristics. The rationale for each is presented below.

Product based classification

Tiered definition

Dynamic sub-sectoral classification

Product-based classification

In the revised definition, the principal basis for classifying firms is their product output, rather than the economic sector in which they are categorised.

In the revised definition, the principal basis for classifying firms is their product output, rather than the SIC code attributed to them. This alternative to a SIC-based definition is used to varying extents by the OECD, NIESR, and deep web reading approaches presented in the preceding section of this report and is a more precise approach to defining and measuring digital economy activity.

SIC codes represent economic sectors classified according to businesses' outputs ('the character of goods and services produced') in addition to aggregate information on their inputs ('inputs, processes and technologies') and clients ('the uses to which these are put'). In contrast, this approach will identify firms' product output based on a current, data-driven product classification. This will yield a more precise approach to measuring digital economy activity for the following reasons:

- It will be based on a current classification of products that reflects recent technological developments in the digital economy. Even with recent revisions of SIC codes (the latest being in 2022-2023, with ISIC Rev 5 due to be implemented in coming years), limitations remain regarding the representativeness of new and innovative sectors.
- It will allow for the recognition of digital firms operating within non-digital sector industries. For example, the NIESR UK digital economy study found that SIC-based definitions of the digital economy miss out a large number of digital firms in the engineering sector and engineering-related scientific and technical consulting sector (most notably, companies specialising in the development of custom software for the engineering sector, see Nathan et al, 2013).
- It will enable the identification of diversified firms that, while classified under non-digital sector industries, are significantly engaged in the production of digital goods and services.
- There is a large number of companies for which SIC codes are uninformative or missing. According to one study, 9.2 percent of companies who file at Companies House were found to have classified themselves under uninformative SICs ('Other business support service activities not elsewhere classified' or 'Other service activities not elsewhere classified') while around 20 percent were found to have provided no SIC information at all (Nathan et al, 2013).

Tiered definition

A tiered approach provides a solution to DSIT's sometimes divergent policy and data needs. To fulfil these needs a useful definition of the digital economy must both be comprehensive enough to generate measures of the overarching value of the digital economy and granular enough to generate measures of specialised subsectors.

The proposed tiering allows for:

- Measurement of the 'core' digital economy (Tier 1). This is conceptually similar to DSIT's existing definition of the digital sector but is more precise as it does not rely on sectoral classifications.
- Generation of overarching measures of the digital economy through the inclusion of Tier 2. Inclusion of this tier may lead to 'double counting' of economic sectors (i.e. when a diversified company is included both in the digital sector and a traditional one) and is therefore not appropriate in all contexts. However, the inclusion of Tier 2 companies can provide measures that reflect the spread of digital production across the economy. For contexts in which 'double counting' is inappropriate, users can revert to Tier 1.

This tiered approach to defining the digital economy draws on two of the approaches discussed in Section 5. First, it draws on the OECD approach, which recognises that firms participate in the digital economy to varying degrees and that this can be accommodated **in** defining the sector by using a tiered approach. Second, it draws on the deep web reading approach, which highlights the presence of diversified firms in the UK's digital economy and the importance of categorising these firms appropriately.

Dynamic sub-sectoral classification

Firms included in the digital economy boundary are further classified into subsectors through a process that combines data-driven approaches and expert knowledge.

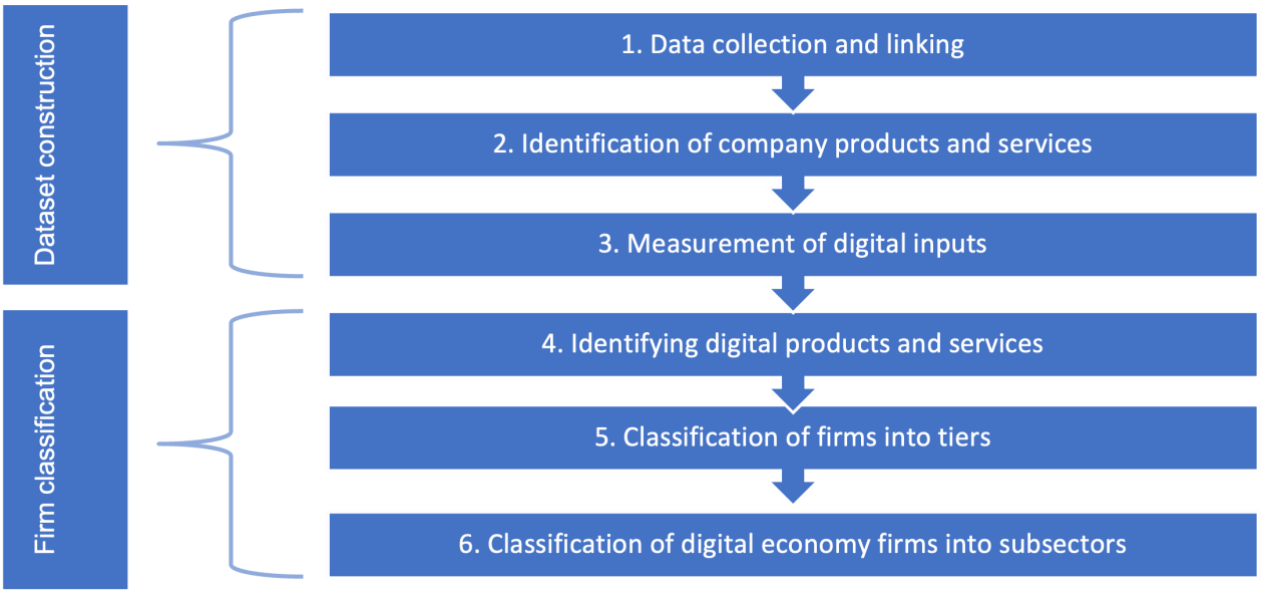
It is proposed that firms included in the digital economy boundary are further classified into subsectors through a process that combines data-driven approaches and expert knowledge, and that this process is repeated periodically in line with policy needs (and technological changes). The rationale for this is to ensure the availability of data on individual priority subsectors of policy relevance to DSIT. This flexible approach to taxonomic classification is adopted from the deep web reading approach to defining economic sectors.

7. Approach to Measurement

7.1 Approach to measurement

The process of operationalising the revised statistical definition of the digital economy involves six steps, as outlined in Figure 4 below.

Figure 4: Approach to measuring the UK digital economy



1. **Data collection and linking.** Operationalising this definition requires a rich, firm-level dataset. In the first step of measurement, a comprehensive dataset of UK companies will be constructed linking company data across various sources, including the Companies House register, a range of ONS surveys, HMRC administrative data, and public web crawled data from company websites and social media.[20] This approach is adopted from the deep web reading approach to defining economic sectors. The table below presents a list of key relevant data sources.

[20] All web crawled data used will be publicly available data to ensure that any collection, storage and dissemination of this data does not pose legal issues.

Table 5: Key data sources for measuring the UK digital economy (by alphabetical order)

	Source	Data type	Key company-level indicators of interest
Annual Business Survey	ONS (microdata accessible via Integrated Data Service)	Survey data	- GVA - Gross wages and salaries - Location - Total employment costs - Turnover - Number of employees
Annual Survey of Goods and Services (ASGS)	ONS (microdata accessible via Integrated Data Service)	Survey data	- Sales by product
Companies House register	ONS (microdata accessible via Integrated Data Service)	Administrative data	- Location - Status (active/ inactive) - Turnover - VAT number (unique identifier) - Self-assigned SIC code - Number of employees - Date of incorporation
Website data for individual companies	Requires collection using web crawling	Web data	- Company description - Products and services
Public LinkedIn webpage data	Requires collection using web crawling	Web data	- Employees by job title
HMRC firm-level data on earnings (PAYE real time information)	ONS (microdata accessible via Integrated Data Service)	Administrative data	- Employment costs - Number of employees - Wages and salaries
HMRC firm-level data on R&D tax credits	ONS (microdata accessible via Integrated Data Service)	Administrative data	- Proxy measure for R&D activity to distinguish digital adopters from digital developers

HMRC firm-level data on VAT	ONS (microdata accessible via Integrated Data Service)	Administrative data	- Proxy measure for turnover to fill data gaps
UK Manufacturers' Sales by Product Survey (PRODCOM)	ONS (microdata accessible via Integrated Data Service)	Survey data	- Sales by product

2. **Identification of company products and services.** In the second step of measurement, the products and services produced by individual UK companies will be inferred from the dataset using a language model. The primary data source used for this inference process is text data from company websites. This data source may be supplemented by public LinkedIn data about company employment and analysis of media coverage about companies.

   The above-described inference process will be calibrated through expert feedback. This is an iterative process in which preliminary sample results are collected and shared with experts for feedback in order to refine the criteria for inclusion and improve the quality of the results. This approach is adopted from the deep web reading approach to defining economic sectors.

   Survey data on the products and services sold by individual UK companies (including sales figures) for a small proportion of UK companies is available from the ONS' ASGS and PRODCOM surveys and can be used to supplement and validate the web crawled data.

3. **Measurement of digital inputs.** In the third step of measurement, proxy measures of digital inputs at the firm level will be calculated.

   a. The most well-established proxy measure of digital intensity is digital employment intensity (for example, Calvino et al, 2018; Spilsbury, 2015). This is defined as the proportion of workers in a firm employed in digital occupations. To calculate this, a list of digital occupations must first be developed. The occupations defined in the UK SOC 2020 can be used as the basis for developing this list. The list can be supplemented with additional digital occupations identified through web crawled data from company websites and LinkedIn.

      The numerator required for calculating digital intensity of employment (number of workers in a firm employed in digital occupations) can be obtained from company websites and public LinkedIn data. The denominator (total number of workers employed in a firm) is obtained from administrative data. This is because LinkedIn provides high quality data on digital workers but is not a reliable source

for workers employed in low-skill occupations. Thus, the total number of workers employed in a firm is best captured by administrative data on firms.

b. A second proxy measure that may be used to supplement digital employment intensity is a firm's level of R&D expenditure. This may improve the ability to distinguish between firms that are digital developers and those which are digital adopters. The data source for this is HMRC administrative data on R&D tax credits. This proxy measure is experimental and a decision regarding whether it will be used should be driven by empirical testing of its construct validity and properties.

c. A third potential proxy measure is a firm's ownership of ICT patents (number and value). The use of this proxy is generally not recommended. It is expected that this would be a weak indicator of whether a firm should be classified as falling within the boundary of the digital economy. This is because: 1) it is well established that patentable types of innovations are more common in certain digital subsectors, making this measure biased toward those subsectors (see, for example, Clancy, 2024; Lerner & Seru, 2022); 2) the speed of technological development in the digital economy tends to be faster than the patenting process (Singh et al, 2021), with the implication that this indicator will not capture recent digital innovation in firms. Nevertheless, it is important to recognise that patent data may offer valuable insights for specific subsectors, such as quantum computing.

4. **Identifying digital products and services.** The fourth step of measurement involves the development of a list of digital products (digital content, ICT goods, and ICT services). The development of this list can draw on:

    a. The list of goods and services collected through the web crawl.

    b. The UN Statistical Division Central Product Classification v2.1 "Products of the information economy" list (United Nations Statistical Division, 2015).

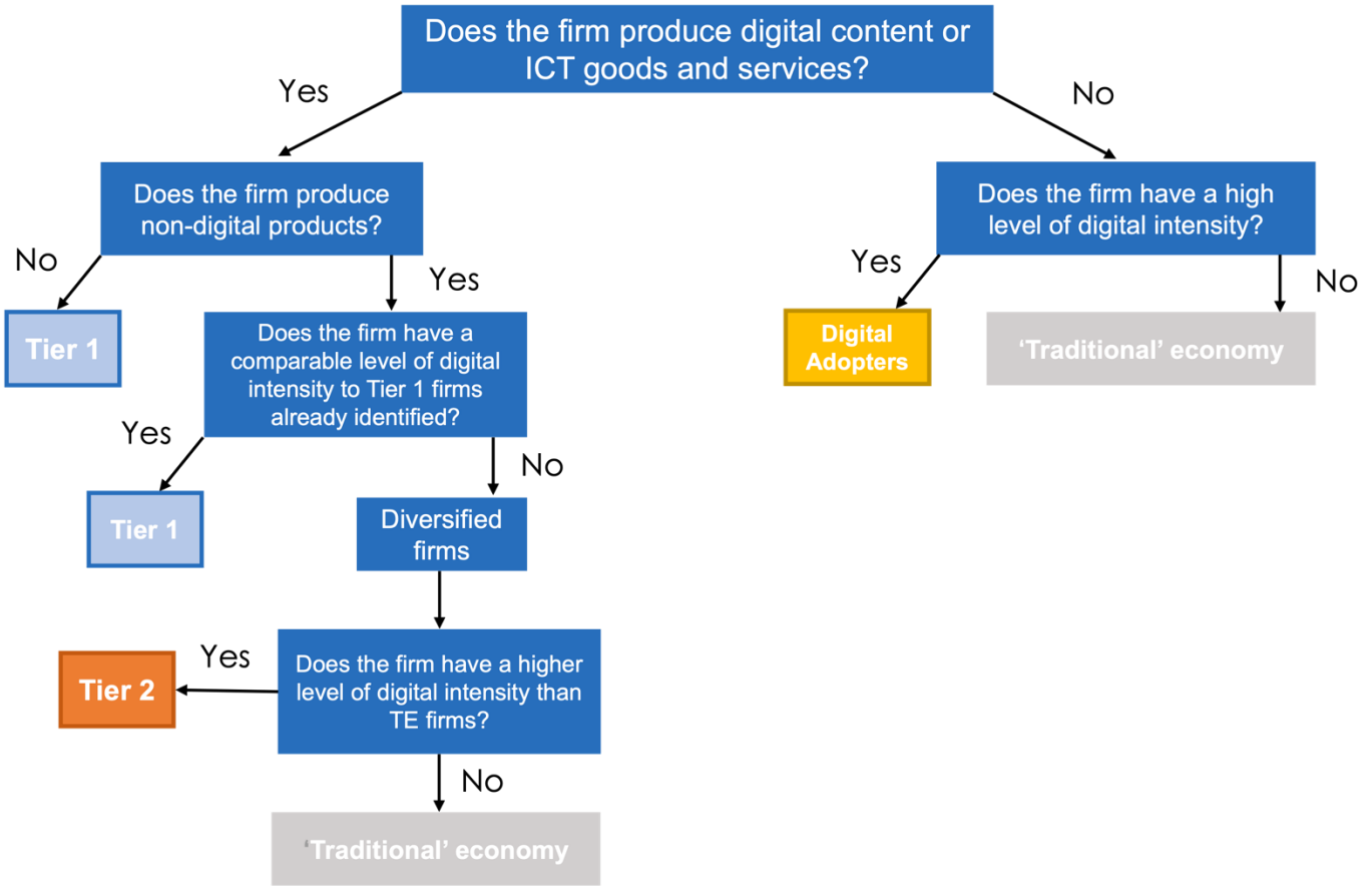
    c. The OECD's: a) list of digital products considered digitally deliverable, b) description of digital products and, c) description of digital industries (OECD, 2023).

    d. The United States Bureau of Economic Analysis list "BEA Items Included in the Digital Economy Estimates" (United States Bureau of Economic Analysis, 2021).

It is recommended that development of this list involves consultation across DSIT's digital and tech analysis teams. In addition, it is recommended that the list of digital products can be mapped to the CPA to ensure translatability to ONS data sources.

5. **Classification of firms into tiers.** In the fifth step of measurement, firms are classified as falling into Tier 1, Tier 2, digital adopters and the 'traditional' economy. An overview of the classification process is presented in the decision tree below.

Figure 5: Classification of firms into tiers



As illustrated in Figure 5, the primary criterion for classifying firms is their product output (digital / non-digital). In addition, measures of digital intensity are employed to determine the boundaries between Tier 1, Tier 2, digital adopters and the 'traditional' economy as follows:

a. For firms that produce both digital and non-digital products, it is necessary to draw a boundary between those belonging to Tier 1 (primarily digital) and those belonging to Tier 2 (diversified). Among this group of firms, those that have a similar level of digital intensity to firms that exclusively produce digital products will be classified as Tier 1. As a starting point, a similar level of digital intensity can be defined as one which is not distinguishable at conventional levels of statistical significance. However, it is recommended that this quantitative approach is used as a guideline. The final boundary can be determined according to expert judgement, based on analysis of the nature of firms observed at the guideline boundary.

b. For diversified firms, it is necessary to draw a boundary between those belonging to Tier 2 and those which fall outside the scope of the digital economy. Put differently, it is necessary to determine the lower boundary for inclusion in Tier 2. Among this group of firms, those with a similar level of digital intensity to firms in the 'traditional' economy will be classified as being outside the boundary of the digital economy. As a starting point, a similar level of digital intensity can be defined as one which is not distinguishable at conventional levels of statistical significance. However, it is recommended that this quantitative approach is used as a guideline. The final boundary can be determined according to expert judgement, based on analysis of the nature of firms observed at the guideline boundary.

c. For firms that do not produce digital products, it is necessary to draw a boundary between those which are digital adopters and those belonging to the traditional economy. This allows us to define the boundaries of the 'traditional' economy, which is used as benchmark in step b, above. To define this boundary, it is recommended that the overall distribution of digital intensity among this group of firms is first examined to determine whether there is a clear demarcation between digital adopters and 'traditional' economy firms (taking into consideration that this may change over time as a growing proportion of firms become digital adopters). If so, it is possible to use a data-driven approach to drawing this boundary and the observed demarcation can be used as a guideline for the threshold at which a firm is considered a digital adopter. This threshold can then be adjusted according to expert feedback based on analysis of the nature of firms observed at this boundary. If there is no clear demarcation between digital adopters and 'traditional' economy firms, an approach that is more reliant on expert judgement can be adopted. In such an approach, experts construct a list of digital adopter firms to act as a benchmark for developing the threshold.

d. Once the firms belonging to Tier 2 have been identified, it is necessary to measure the proportion of these companies that will be allocated to the digital economy. This would ideally require comprehensive information on Tier 2 firms' revenue by product type. However, this information is not available from existing sources of survey and administrative data or from the web crawled data described above. Following the extant practices of national statistical agencies in several OECD countries, partial data on Tier 2 firms' revenue by product type will be collected from the annual reports and financial statements of Public Limited Companies (PLCs). Given that these tend to be the largest companies, data collected using these reports is expected to represent a significant proportion of the value of Tier 2. Thus, this data can be used to arrive at a conservative estimate of the proportion of Tier 2 that belongs to the digital economy. For a more comprehensive estimate, this data can be supplemented by data from the ONS ASGS and PRODCOM surveys. In addition, it is possible to commission a targeted survey of Tier 2 firms to fill remaining data gaps.

The issue of boundary definition is addressed in various ways by existing approaches to defining and measuring the digital economy. The deep web reading approach to measuring digital and tech sectors relies on expert judgement, whereas the NESTA dynamic mapping approach relies on benchmarking economic units according to their digital employment intensity.[21] The methodology proposed here relies primarily on the benchmarking approach with a secondary role for expert judgement. This methodology has been proposed because the benchmarking element provides an objective, quantifiable framework that can be consistently applied to ensure reproducibility and comparability of results. The inclusion of an element of expert judgment allows for the flexibility needed to make nuanced decisions that account for emerging trends in the digital economy and resolve ambiguities that purely data-driven methods might overlook, leading to more accurate and meaningful assessments.

6. **Classification of digital economy firms into subsectors.** In the final step of measurement, digital economy firms are classified into subsectors (based on products and technologies) using a combination of data-driven approaches, stakeholder consultation and expert judgement. Language models can provide a first cut of subsectors, which can be refined and finalised through stakeholder consultation and expert judgement to develop a taxonomy that meets DSIT's policy needs. The inclusion of expert judgment and stakeholder consultation in this process will play a key role in demarcating the boundaries between subsectors, which may be challenging in instances where subsectors overlap. This approach is adopted from the deep web reading approach to defining economic sectors.

7.2 Limitations of measurement approach

This measurement approach has a number of limitations.

1. **Limitations of web data:** The accuracy and completeness of web-crawled company data can vary depending on the extent of a firm's digital presence. This is not expected to pose a significant limitation, as previous studies using a deep web reading approach for measuring digital sectors in the UK have found that a large majority of firms in this sector have a robust and accessible digital footprint (Donaldson et al, 2022; Glass.ai & Perspective Economics, unpublished; Perspective Economics, 2023; Technopolis & Glass.ai, forthcoming). Nevertheless, it is worth noting that the data available on micro

[21] The NIESR approach defines the digital boundary primarily using a decision rule which states that a firm should be **both** classified as operating in a digital sector **and** be involved in the production of 'digital goods' and/or 'digital services' in order to fall within the digital economy boundary. The rationale for this approach is not explicitly discussed in the study, but the resulting boundary is subject to adjustment based on expert judgement. In the case of the OECD approach, the method used to define the digital economy boundary varies between national statistical agencies but it generally relies on a mixture of: 1) Expert judgement to classify goods and services as digital / partly digital/ non-digital; 2) Decision rules to guide the reallocation of economic units from conventional SUTs to DSUTs.

and small-sized firms tends to be of lower quality than that for larger firms due to their typically less developed online presence.

In addition, measuring the UK operations of international companies using web-crawled company data presents a specific challenge. However, the Artificial Intelligence Sector Study commissioned by DSIT in 2023 demonstrates that it is possible to successfully address this challenge (Perspective Economics, 2023). Importantly, the abovementioned limitations highlight the need for robust data validation and triangulation processes to ensure that the insights derived from web-crawled company data are comprehensive and reliable.

2. **Limited data on diversified companies:** The digital economy activities of diversified companies are challenging to measure due to the limited data available on the proportion of digital versus non-digital products produced by these companies. In line with practices adopted by national statistical agencies in several OECD countries, the measurement approach outlined here involves collecting partial data on Tier 2 firms' revenue by product type from the annual reports and financial statements of PLCs. Since PLCs generally represent larger market players, the data derived from these sources will represent a significant portion of Tier 2's overall value. This approach allows us to derive a conservative estimate of the proportion of Tier 2 that falls under the digital economy. This data can be augmented with information from the ONS Annual Survey of Goods and Services (ASGS) and the PRODCOM surveys, which provide additional insights into company sales by product. However, these surveys only cover a small proportion of companies.
3. **Overlapping economic sectors:** Some firms or products span multiple economic sectors or digital subsectors. This results in two measurement issues.
    - The digital economy definition and measurement approach outlined in this report, partially allocates Tier 2 firm activity to the digital economy. It should be noted that these activities will also be included in the measurement of other sectors in national statistics. Hence, it is important to clearly communicate that this approach represents a ‘satellite account’ approach that operates alongside and complements standard national accounts. A satellite account extends the analytical capacity of national accounts by detailing additional aspects of economic activities. It incorporates sector-specific factors that are not usually included in primary accounts, offering a clearer picture of economic sectors that span traditional sector boundaries. By clearly communicating that this is a satellite account approach, one may better explain the rationale behind apparent ‘double counting’ of economic activities and illustrate how it enhances our understanding of the economic landscape rather than distorting it. This clarification ensures that stakeholders appreciate the nature of the data and the analytical benefits of this broader perspective.
    - In this approach to measuring digital economy subsectors, companies are allocated to the single subsector that most closely represents the firm-level data

available. This introduces some measurement error where a firm produces digital products and services that span multiple digital subsectors.

8. Conclusions and Recommendations

The ability to measure both the UK's digital economy as well as key emerging digital industries and technologies is increasingly crucial to DSIT's policy agenda. These evolving policy needs necessitate a revised approach to defining and measuring the digital economy, one that is both more comprehensive and more granular than the Department's current approach. This report has presented a revised statistical definition for the UK digital economy and a methodology to operationalise this definition into a series of metrics, laying the groundwork for DSIT to measure and monitor the digital economy in a manner that more effectively supports the Department's strategy and facilitates more targeted and impactful policy interventions.

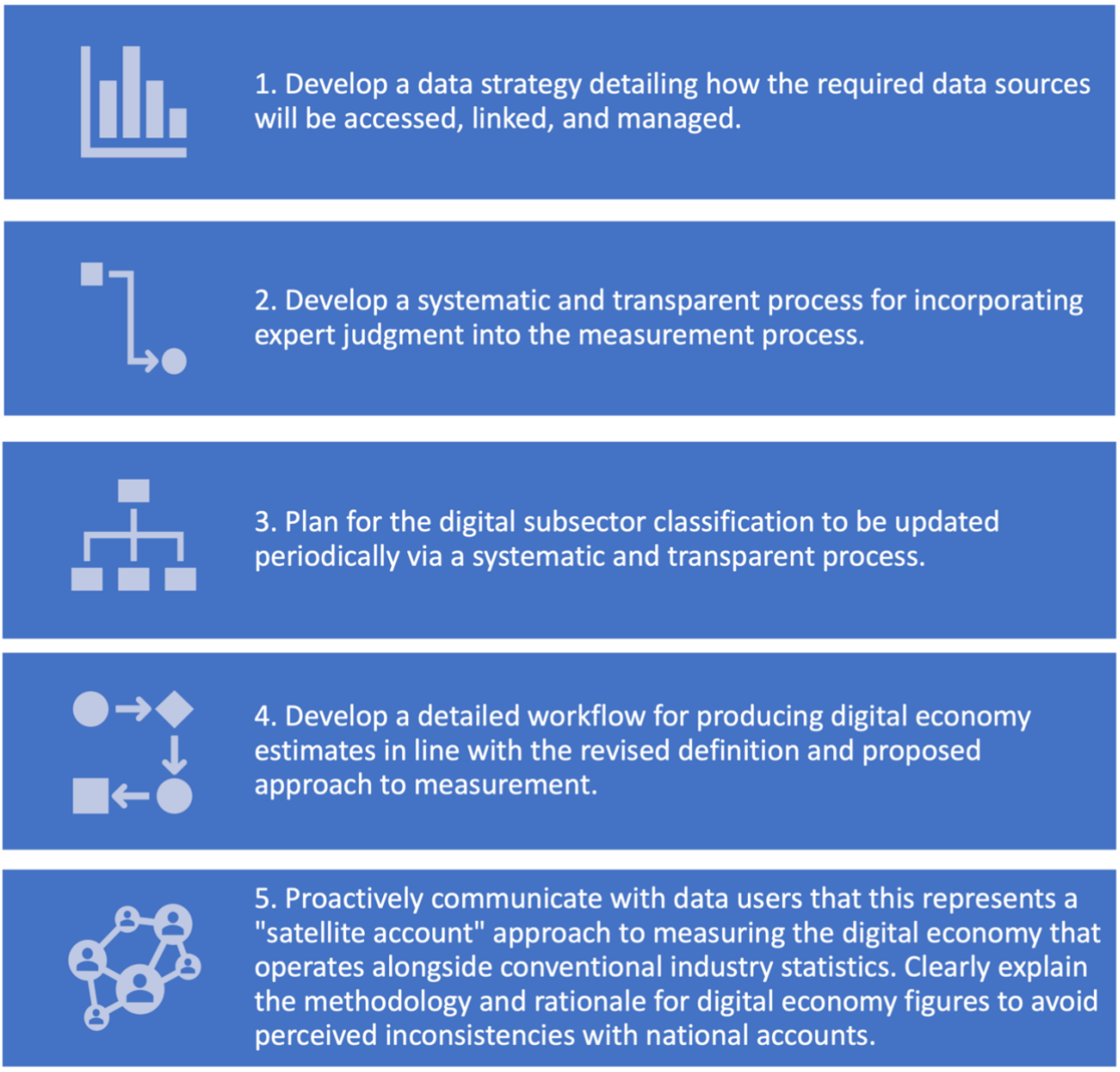
The proposed two-tier definition allows for measurement of both the core digital economy (Tier 1) and digital economy activities embedded within diversified firms across other sectors (Tier 2). This represents an improvement over the previous approach, which relies primarily on SIC codes and does not capture the production of digital goods and services across the whole economy.

Another 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. This also represents an improvement over the previous approach which is based on standard industry and occupation classifications that are updated infrequently and are not well-suited to assessing the economic impact of newer industries.

It is important to note that there are also limitations to this revised approach:

- This approach makes extensive use of web data. Importantly, the quality of web data varies with higher quality data available for firms with a more substantial online presence. For this reason, implementation of this approach requires thorough data validation procedures.
- Data on digital activities within diversified Tier 2 firms is limited. Revenue data collected from annual reports will provide partial coverage of these activities, with the implication that the reported size of Tier 2 will represent a conservative estimate.
- The approach involves some double counting of economic activity, as the Tier 2 portion of diversified firms is included in both digital economy estimates as well as the national accounts for their primary industry. This is an intended effect reflecting the economy-wide spread of digitalisation. However, this must be communicated to stakeholders and data users effectively.

To effectively implement this revised definition and methodology, it is recommended that DSIT take the following steps:



9. References

Bakhshi, H., Freeman, A., & Higgs, P. (2013). A Dynamic Mapping of the UK's Creative Industries. Available online [last accessed 27/03/2024].

Bukht, R. and Heeks, R. (2017). “Defining, Conceptualising and Measuring the Digital Economy.” Development Informatics Working Paper No. 68. Available online [last accessed 27/03/2024].

Calvino, F., Criscuolo, C., Marcolin, L., & Squicciarini, M. (2018). “A Taxonomy of Digital Intensive Sectors.” OECD Science, Technology and Industry Working Papers, No. 2018/14. Available online [last accessed 27/03/2024].

Clancy, M. (2024, March 29) Can We Learn About Innovation from Patent Data? Available online [last accessed 19/04/2024].

Department for Business, Energy and Industrial Strategy (unpublished). Proposed Definition of the Digital Economy.

Department for Digital, Culture, Media & Sport (unpublished). Proposed framework for updated DCMS definition of the digital sector and economy.

Department for Digital, Culture, Media & Sport (2022). DCMS Sector Economic Estimates Methodology. Available online [last accessed 19/04/2024].

Donaldson, S., Crozier, D., Martorell, S., McLaren, I., Douglas, J., & Navin Shah, J. (2022). UK Cyber Security Sectoral Analysis 2022. Department for Digital, Culture, Media and Sport. Available online [last accessed 27/03/2024].

eSkills UK (2013). Information Economy: Economic Estimates 2013. Available online [last accessed 19/04/2024].

Glass.ai and Perspective Economics (unpublished). Defining and Measuring the Yorkshire Digital Technology Sector Study.

Lerner, J. and Seru, A. (2022) “The Use and Misuse of Patent Data: Issues for Finance and Beyond.” The Review of Financial Studies, 35(6), pp. 2667–2704.

MSCI (2023). Global Industry Classification Standard (GICS) Methodology. Available online [last accessed 19/04/2024].

Nathan, M., Rosso, A., Gatten, T., Majmudar, P., & Mitchell, A. (2013). Measuring the UK's Digital Economy with Big Data. National Institute of Economic and Social Research. Available online [last accessed 27/03/2024].

OECD. (2011) OECD Guide to Measuring the Information Society. OECD. Available online [last accessed 27/03/2024].

OECD (2020). A Roadmap Toward a Common Framework for Measuring the Digital Economy. Report for the G20 Digital Economy Task Force. Available online [last accessed 27/03/2024].

OECD (2023). OECD Handbook on Compiling Digital Supply and Use Tables. OECD Publishing, Paris. Available online [last accessed 27/03/2024].

ONS (2007). UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007): Structure and Explanatory Notes. London: Palgrave MacMillan.

ONS (2022). UK Digital Economy Research: 2019. Available online [last accessed 27/03/2024].

Perspective Economics (2023). Artificial Intelligence Sector Study: Research report for the Department for Science, Innovation & Technology (DSIT). Department for Science, Innovation and Technology. Available online [last accessed 27/03/2024].

Singh, A., Triulzi, G., & Magee, C. L. (2021). "Technological Improvement Rate Predictions for All Technologies: Use of patent data and an extended domain description." Research Policy, 50(9), 104294.

Spilsbury, M. (2015). Dynamic Mapping of the Information Economy Industries. Nesta/Tech UK. Available online [last accessed 27/03/2024].

Statistics Canada (2021). Measuring the digital economy: The Canadian digital supply and use tables 2017-2019. Available online [last accessed 27/03/2024].

Statistics Netherlands (CBS) (2021). Supply-Use tables for the digital economy. Experimental research. Available online [last accessed 27/03/2024].

Technopolis and Glass.ai (forthcoming). Measuring the UK Immersive Economy. Innovate UK.

United Nations Statistical Division (2023). Chapter 22 (2025 SNA)/Chapter 16 (BPM7): Digitalisation: Version circulated for AEG review. Available online [last accessed 19/04/2024].

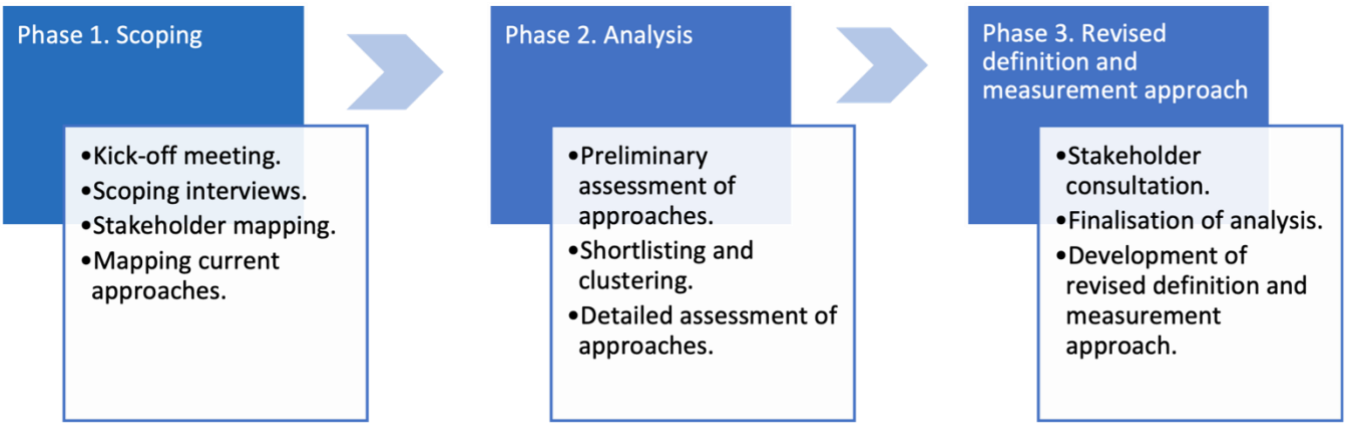
United Nations Statistical Division (2015). Central Product Classification (CPC) Version 2.1. Available online [last accessed 28/03/2024].

United States Bureau of Economic Analysis (USBEA) (2021). Updated Digital Economy Estimates: June 2021. Available online [last accessed 28/03/2024].

Appendix A: Study Methodology

This section of the report describes the study methodology. The study involved three broad phases, as presented in Figure 6, below.

Figure 6: Phases of Defining and Measuring the UK Digital Economy Study



A1 Phase 1: Scoping

The objective of the scoping phase was to develop a comprehensive understanding of the scope and policy context of the study. This involved:

1. **Kick-off meeting and scoping interviews:** A study kick-off meeting and scoping interviews with DSIT officials to better understand their data and policy needs related to the digital economy.

2. **Stakeholder mapping:** A participatory stakeholder mapping exercise conducted in collaboration with the DSIT DEDP to identify key stakeholders. This provided a basis for stakeholder engagement activities conducted in Phase 3 of the study, as well as providing preliminary insights on the range of needs that the revised definition should endeavour to meet.

3. **Mapping current approaches:** A mapping exercise to identify key approaches to defining and measuring the digital economy that have been implemented in the UK and internationally in economies comparable to that of the UK over recent years. This involved a literature review, complemented by consultation with DSIT officials and subject matter experts. The mapping exercise resulted in a longlist of 11 approaches (see Table 6 below). Eight of these approaches involve definition and measurement of the digital economy, while the remaining three involve definition and measurement of economic sectors other than the digital economy. The latter group was included because these approaches present promising methods for the measurement of economic sectors that do not easily fall into standard classification systems, and which could be applied to the digital economy.

Table 6: Longlist of Key Recent Approaches to Defining and Measuring the Digital Economy

No.	Name	Author(s)	Year
1	Global Industry Classification Standard (GICS)	MSCI/ S&P Dow Jones	2023
2	US Bureau of Economic Analysis approach to digital economy definition and measurement	US Bureau of Economic Analysis	2019-2023
3	UN Statistics Division SNA approach to measuring economic digitalisation	UN Statistics Division	2022
4	DCMS definition of the digital sector and economy	UK Government	2016
5	OECD digital economy definition and measurement toolkit	OECD	2020-2023
6	Proposed framework for updated DCMS definition of the digital sector and economy	UK Government	2021
7	National Institute of Economic and Social Research (NIESR) UK digital economy study	Nathan, M., Rosso, A., Gatten, T., Majmudar, P. & Mitchell, A.	2013
8	BEIS proposed definition of the digital economy	UK Government	Unknown
9	Defining and measuring the Yorkshire digital technology sector study	Glass.ai	2023
Selected definition and measurement approaches for other sectors			
10	Measuring the UK immersive economy study	Technopolis/ Glass.ai	Ongoing
11	UK artificial intelligence sector study	DSIT/ Perspective Economics/ Glass.ai/ IPSOS	2023
12	UK Cyber Security sectoral analysis	DSIT/ Perspective Economics/ Glass.ai/ IPSOS/ CSIT	2023

13	NESTA dynamic mapping approach to economic sector measurement	NESTA	2013-2015

Source: Technopolis, 2024

A2 Phase 2: Analysis of existing approaches and stakeholder consultation

The objective of the analysis phase was to build a knowledge base about key recent approaches to the definition and measurement of the digital economy. This included systematic analysis of the advantages and disadvantages of these approaches in relation to the data and policy needs of DSIT and other relevant stakeholders. This phase involved:

Preliminary assessment of longlisted approaches: To conduct a preliminary assessment of the longlisted approaches, a critical assessment framework (CAF) was developed comprising of seven primary assessment criteria. The CAF was designed to capture whether the approaches in the longlist meet: 1) UK stakeholder needs, 2) relevant UK and international statistical quality criteria. It was developed based on consultation with the DSIT DEDP as well as relevant quality criteria described in the UK and European Statistic Codes of Practice. An overview of the assessment criteria is presented in Table 7.

Table 7: Primary Assessment Criteria

No.	Principal assessment criteria
1	Ability to generate an overarching measure of the value of the UK digital economy
2	Ability to generate key measures of interest
3	Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. quantum)
4	Translatability to Standard Industrial Classification
5	Ability to generate comparisons over time
6	Availability, timelines and statistical quality of relevant data sources (including both official and third-party data sources

7	Ability to capture firm-level data for individual subsectors (e.g. AI, cyber-security) and indicators of market competition (e.g. business entry and exit, market share).

Source: Technopolis, 2024

The preliminary assessment of the longlisted approaches involved a high-level assessment of each approach according to each of the seven primary assessment criteria. Each criterion was assessed against a four-point scale, as shown below.

1.	2. Does not meet the criterion at all
3.	4. Meets the criterion to some extent
5.	6. Meets the criterion to a great extent
7.	8. Fully meets the criterion

As a result of this assessment, two approaches which obtained very low scores were excluded from the detailed assessment as they were judged not to be relevant to DSIT's policy needs.[22] The full results of the preliminary assessment are presented in 0. The remaining nine approaches were then clustered into four groups for detailed assessment. The four clusters selected for detailed assessment are:

1. Operationalisation of the OECD 2020 definition of the digital economy. This includes the following approaches:
    - US Bureau of Economic Analysis approach to digital economy definition and measurement.
    - UN Statistics Division SNA approach to measuring economic digitalisation.
    - Proposed framework for updated DCMS definition of the digital sector and economy
    - OECD digital economy definition and measurement toolkit.
2. Deep web reading approach to measuring economic sectors. This includes following approaches:
    - Defining and measuring the Yorkshire digital technology sector study
    - Measuring the UK immersive economy study
    - UK Artificial Intelligence sector study
    - UK Cyber Security sectoral analysis

[22] These were: 1) The Global Industry Classification Standard (GICS), which does not allow for comparisons over time, does not capture digitalisation across the economy and only includes stock-exchange listed companies; 2) The BEIS proposed definition of the digital economy, as only one component of this approach has been operationalised and it is therefore not possible to assess it against those framework criteria closely tied to data and measurement issues.

3. NIESR UK digital economy study.
4. NESTA dynamic mapping approach to economic sector measurement.

Critical assessment of key approaches: To conduct the critical assessment of key approaches, a detailed version of the CAF was developed comprising an in-depth questionnaire that contains specific sub-questions to be addressed in the assessment of each criterion (see 0 for the full questionnaire). In addition, a secondary assessment criterion was added to the framework: translatability to international classification systems. The CAF questionnaire was developed in consultation with the DSIT DEDP to ensure that it reflects the department's needs.

Each of the four key approaches to the definition and measurement of the digital economy (or comparable sectors) was assessed against this questionnaire. The findings for each approach were synthesised into an overview of its key advantages and disadvantages.

A3 Phase 3: Development of a revised definition and measurement approach

The objectives of this phase were to:

1. Iterate and refine the analysis conducted in Phase 2 through a process of stakeholder consultation. This purpose of the consultation is two-fold. First, it allowed the study team to solicit feedback on the analysis conducted from specialist policy practitioners with expertise in various areas of the digital economy. Second, it allowed the study team to develop a better understand of stakeholders' data and policy needs relating to the digital economy.
2. Develop a revised definition of the digital economy and approach to operationalising this definition in consultation with key stakeholders.

This phase involved:

1. **Stakeholder workshop 1:** A two-hour workshop was held on the 22nd of February 2024 to introduce the study, collect stakeholder feedback on the assessment criteria and the extent to which different approaches to defining and measuring the digital economy meet their data and policy needs. The workshop was attended by 26 people from nine organisations.

   Prior to the workshop, attendees received preliminary information containing an overview of the study and the policy context, an overview of the preliminary assessment of the longlisted approaches and an overview of the shortlisted approaches. During the workshop, participants were invited to provide their reflections and assessment of

existing approaches as well as a set of preliminary proposed definitions through open discussion and online polls managed through Mentimeter. A Workshop Synthesis Note (presented in 0) summarised the key discussion points emerging from this first workshop. The output of this workshop was incorporated into a revised version of the critical assessment of approaches and informed the rationale for the development of the proposed definition.

2. **Collecting of further stakeholder feedback:** Following the workshop, the Workshop Synthesis Note was shared with workshop participants and invitees who had been unable to attend the workshop. This provided an opportunity to obtain both an overview of the discussion and to share further reflections and feedback. Further feedback collected at this stage was collated, synthesised and incorporated into the analysis and design of the proposed definition set out below.

3. **Development of a proposed definition and high-level approach to measurement:** The study team developed a proposal for a revised definition that meets the key needs of DSIT and other key stakeholders, alongside a high-level description of how this definition can be operationalised. This was presented in a Draft Methodology Note.

4. **Stakeholder workshop 2:** A two-hour workshop was held on the 21st of March 2024 to collect stakeholder feedback on the Draft Methodology Note. The workshop was attended by 21 people from eight organisations.

   Prior to the workshop, attendees received preliminary information containing an overview of the Draft Methodology Note. During the workshop, participants were invited to provide their reflections and assessment of the content. A Workshop Synthesis Note (presented in 0) summarised the key discussion points emerging from this second workshop. The output of this workshop was incorporated into a revised version of the methodology note.

5. **Finalisation of revised definition and approach to measurement**: The study team finalised the revised definition of the digital economy, alongside a description of how this definition can be operationalised, as presented in this methodology note.

Appendix B: Preliminary Assessment of Definitions

9. 1	10. Does not meet the criterion at all
11. 2	12. Meets the criterion to some extent
13. 3	14. Meets the criterion to a great extent
15. 4	16. Fully meets the criterion
17. NO	18. Not operationalised

	Ability to generate overarching measure	Ability to generate key measures of interest	Ability to capture contribution of individual subsectors, including future emerging technologies	Translatability to Standard Industrial Classification	Ability to generate comparisons over time	Availability, timeliness and statistical quality of relevant data sources	Ability to capture firm-level data for individual subsectors and indicators of market competition
Current DCMS digital sector definition and measurement (benchmark)	2	3	2	4	4	4	2
Global Industry Classification Standard (GICS)	2	2	3	1	2	2	2
US Bureau of Economic Analysis approach to digital economy definition and measurement	4	3	2	4	4	4	2

Defining and Measuring the UK Digital Economy

UN Statistics Division SNA approach to measuring economic digitalisation	4	3	2	4	4	4	2
Proposed framework for updated DCMS definition of the digital sector and economy	4	2	2	2	4	NO	NO
OECD Digital Economy Definition	4	3	3	4	4	3	2
Growth Intelligence / NIESR Definition of the UK Digital Economy	4	2	4	3	2	2	3
BEIS Proposed Definition of the Digital Economy	4	NO	3	2	NO	NO	NO
Defining and measuring the Yorkshire digital technology sector study	3	3	4	1	2	2	4
Measuring the UK Immersive Economy	3	3	4	3	2	2	4
UK Artificial Intelligence Sector Study	3	3	4	3	2	2	4
UK Cyber Security sectoral analysis	3	3	4	3	2	2	4
NESTA dynamic mapping approach	3	3	2	4	4	3	1

As a result of this assessment, two approaches which obtained very low scores were excluded from the detailed assessment as they were judged not to be relevant to DSIT's policy needs. These were:

- The Global Industry Classification Standard (GICS), which does not allow for comparisons over time, does not capture digitalisation across the economy and only includes stock-exchange listed companies.
- The BEIS proposed definition of the digital economy, as only one component of this approach has been operationalised and it is therefore not possible to assess it against those framework criteria closely tied to data and measurement issues.

The remaining ten approaches were then clustered into four groups for detailed assessment. The four clusters selected for detailed assessment are:

1. Operationalisation of the OECD 2020 definition of the digital economy. This includes the following approaches:
    a. US Bureau of Economic Analysis approach to digital economy definition and measurement.
    b. UN Statistics Division SNA approach to measuring economic digitalisation.
    c. Proposed framework for updated DCMS definition of the digital sector and economy
    d. OECD digital economy definition and measurement toolkit.
2. Deep web reading approach to measuring economic sectors. This includes following approaches:
    a. Defining and measuring the Yorkshire digital technology sector study
    b. Measuring the UK immersive economy study
    c. UK Artificial Intelligence sector study
    d. UK Cyber Security sectoral analysis
3. NIESR UK digital economy study.
4. NESTA dynamic mapping approach to economic sector measurement.

Appendix C: Critical Assessment Framework Questionnaire

Primary assessment criteria

1. Ability to generate an overarching measure of the value of the UK digital economy.
    a. To what extent does it allow for measurement of the economic value of the core digital sector (OECD definition as used by the UK Government)?
    b. To what extent does it allow for the measurement of the economic value of digital-only service producers (e.g. digital-only banks, e-payment service providers, online only media outlets, streaming services, edtech platforms, healthtech platforms)?
    c. To what extent does it allow for measurement of the economic value of digital intermediary platforms, including both those which charge a fee those which are data/advertising driven?
    d. To what extent does it allow for measurement of wider digitalisation: economic units where production has been significantly enhanced by digital technology?
    e. To what extent does it allow for disaggregation of the economic activities listed in questions 1a-1d from the overarching measure of the digital economy?

2. Ability to generate key measures of interest.
    a. To what extent does it allow for the generation the following measures?
        i. Gross value-added.
        ii. Employment.
        iii. Worker earnings.
        iv. International trade measures, including imports and exports of services and goods.

v. Business demography measures, including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites.

b. What other measures of interest can be generated from this definition (e.g. valuation of tech companies, number of unicorns, VC investment in the tech sector, FDI in the tech sector, growth indicators beyond revenues and employees)?

3. Ability to capture the contribution of individual subsectors of high priority to the UK Government,[23] including the identification and integration of future emerging technologies as subsectors.

   a. To what extent does it allow for measurement of the following subsectors?

      i. Data subsector (including data processing, hosting and infrastructure).

      ii. Artificial intelligence subsector.

      iii. Cyber-security subsector.

      iv. Software development.

      v. ICT manufacturing (including semiconductors).

      vi. ICT trade.

      vii. Telecommunications.

      viii. Digital services (e.g. FinTech, EdTech, HealthTech)

   b. To what extent does it allow for the measurement of other subsectors not listed above, and at what level of granularity?

   c. To what extent does it allow for the identification and incorporation of future emerging technologies as subsectors when needed (e.g. Quantum)?

4. Translatability to Standard Industrial Classification.

   a. To what extent is it translatable to SIC codes?

[23] As identified by the DSIT Science and Tech Framework and through consultation with the DSIT DEDP.

   b. How easily is it translatable to SIC codes (i.e. what is required to translate the definition to SIC codes)?

5. Ability to generate comparisons over time.

   a. To what extent does it allow for measurement of trends over time, including historical comparison?

   b. At what level of temporal granularity does it allow for measurement of time trends?

6. Availability, timeliness and statistical quality of relevant data sources.

   a. To what extent is the data required to operationalise the definition currently, available? If it is not available, can it be collected?

   b. To what extent is the data required to operationalise the definition publicly available or available from ONS statistics? How easily accessible are any ONS statistics required to operationalise the definition?

   c. To what extent does the data required to operationalise the definition require subscriptions to proprietary data sources?

   d. How recent is the available data required to operationalise the definition?

   e. To what extent is the data required to operationalise the definition representative at the following levels: nationally, regionally, sectoral, sub-sectoral?

7. Ability to capture firm-level data for individual subsectors and indicators of market competition.

   a. To what extent does it allow for measurement of firm-level data for the individual subsectors listed in question 3a?

   b. To what extent does it allow for measurement of the following indicators of market competition?

      i. Market share concentration.

      ii. Business entry and exit.

Secondary assessment criterion

8. Translatability to international classification systems.
    a. To what extent is it translatable to the Standard Occupational Classification?
    b. To what extent is it translatable to the Harmonised System.

Appendix D: Stakeholder Workshop 1 Synthesis Note

D1 Introduction

The degree of digitalisation in the economy has increased dramatically over the past decade, leading to an evolution in how the 'digital economy' is defined. Earlier conceptualisations of the digital economy employed a 'bottom-up' approach, defining it as a specific set of economic activities that produce ICT goods and digital services. In contrast, more recent definitions tend to adopt a 'top-down' approach, which additionally includes economic activity enabled by the use of ICT goods and digital services, thus reflecting the spread of digitalisation across the economy. In the UK, DCMS – and more recently DSIT – historically defined the digital sector using a bottom-up definition based on Standard Industrial Classification codes, which does not capture digitalisation across the whole economy nor allow for granular measurement of key industries/ technologies (e.g. AI or cyber-security). This approach no longer meets DSIT's policy needs.

In this context, DSIT has commissioned Technopolis to conduct a study on defining and measuring the UK digital economy. The objective of this study is to develop: 1) a revised definition of the digital economy which meets DSIT's policy needs; 2) a methodology to operationalise this definition. The study involves the following key elements:

1. **Review and assessment of existing definitions of the digital economy** to capture the extent to which they meet UK stakeholder needs and international standards.

2. **Stakeholder consultation** to develop a rigorous, comprehensive and measurable definition that meets stakeholder and user needs. This involves holding two stakeholder workshops. The first workshop will introduce the study to stakeholders and solicit their feedback on the extent to which different approaches to defining and measuring the digital economy meet their policy needs and those of DSIT. The second workshop will be used to solicit feedback on the study findings and recommendations.

3. **Design of a measurement methodology** setting out how the new definition of the digital economy will be operationalised.

This workshop synthesis note presents a summary of the discussion from the first workshop. The workshop was held virtually on 22 February 2024, and was attended

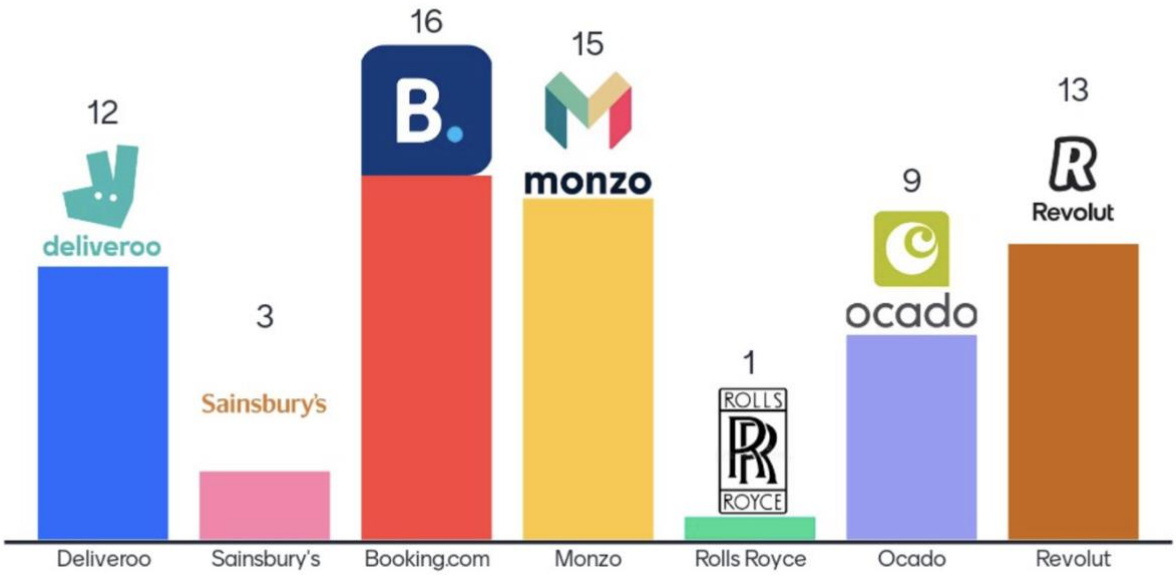
by 26 people from nine organisations. This note presents the feedback gathered during the workshop discussion and through the online chat function, in addition to the live online polling conducted during the workshop.

D2 Overview of Workshop Discussion

D2.1 Icebreaker Exercise

To kick-off discussion in the workshop and highlight the challenge of setting a definition, the study team ran a brief ice-breaker exercise. Workshop participants were asked to identify which of a subset of companies would be classed as being part of the 'Digital Economy'. The figure below shows the output of a poll of participants.

Figure 7: Workshop poll "Which of these companies is part of the 'digital economy'?"



- This poll and results triggered much disagreement amongst workshop participants, with some noting "all of them" and others noting "none of them".
- A member of the study team also highlighted that you get very different definitions and results depending on where one draws the boundaries, so it is important to focus on what the needs are, as this will determine the definition.
- One participant noted that digital activities could be a proportion of the firm's activities, whilst one of the study team noted it was interesting that only one person selected Rolls Royce, which gains more than half its revenue and profits from digitally enabled services. Another participant highlighted that

Rolls Royce are a world leader in digital simulation for avionics, manufacturing, and digital twins, again emphasising the challenges in attributing organisations to single sectors. Another participant also raised the question about how this might reflect or not the change in a company over time, noting Rolls Royce it has been trading for longer than 'digital' has been a concept.

- To highlight the limitations of SIC codes, the study team noted that according to the SIC codes captured under Companies House, only Ocado and Revolut would be classed as being part of the Digital Economy (see table below).

Table 8: SIC codes for subset of companies that could be part of the digital economy

Deliveroo	01110 - Growing of cereals (except rice), leguminous crops and oil seeds
Sainsbury's	47110 - Retail sale in non-specialised stores with food, beverages or tobacco predominating
Booking.com	82990 - Other business support service activities not elsewhere classified
Monzo	64191 - Banks 64999 - Financial intermediation not elsewhere classified
Rolls Royce	25300 - Manufacture of steam generators, except central heating hot water boilers 25620 - Machining 28110 - Manufacture of engines and turbines, except aircraft, vehicle and cycle engines
Ocado	47110 - Retail sale in non-specialised stores with food, beverages or tobacco predominating 62012 - Business and domestic software development 62090 - Other information technology service activities

- Participants thought that this exercise highlighted the limitations of SIC codes, as Monzo very much positions itself as a digital economy company.
- A participant noted that the SIC codes were defined in 2007, the same year that iPhone 1 was released and that the whole economy has changed dramatically since the SIC list was defined'
- A participant from UKRI notes that using Company House data is a challenge, as these codes are based on the company's self-assessment which are not validated, although a representative from DSIT noted that the Interdepartmental Business Register does sense check those assessments.
- As one participant noted in the chat of the workshop, “It's almost like we need 'core digital economy' and then 'digitally intensive' for the next raft out?”

D2.2 Assessment Criteria

To capture feedback on the assessment criteria, the study team presented an overview of the assessment criteria developed during the scoping phase of the study (see Box 1 below).

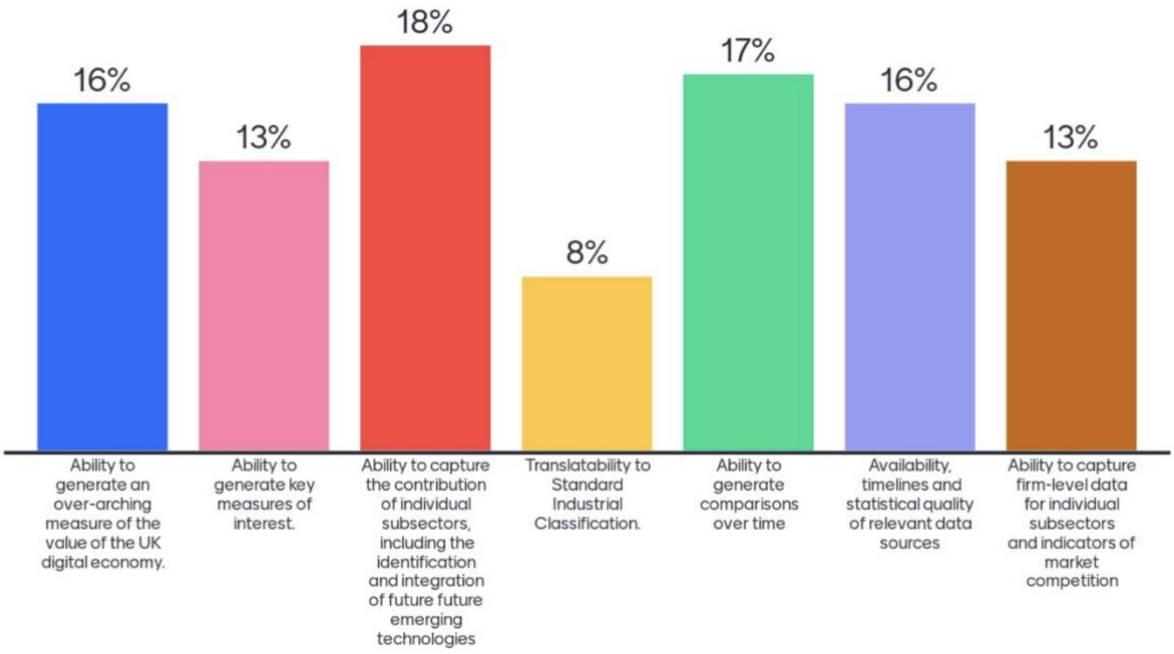
Box 2: Assessment criteria

1. Ability to generate an overarching measure of the value of the UK digital economy.
2. Ability to generate key measures of interest.
3. Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. quantum).
4. Translatability to Standard Industrial Classification.
5. Ability to generate comparisons over time.
6. Availability, timelines and statistical quality of relevant data sources (including both official and third-party data sources).
7. Ability to capture firm-level data for individual subsectors (e.g. AI, cyber-security) and indicators of market competition (e.g. business entry and exit, market share).

Summary of feedback and discussion points:

- Workshop participants did not identify or highlight any criteria that were missing at this stage.
- Workshop participants were asked to identify and prioritise those criteria that were the most important for them (see Figure 3 below).

Figure 8: Workshop poll: "Which criteria are essential for you to be able to apply this in your work?"



- A member of the study team asked workshop participants for more information or reflections on the criteria ‘Available, timeliness and statistical quality of relevant data sources’.
- DSIT participants highlighted the importance of data in terms of timeliness because subsectors grow rapidly, availability, because it determines the types of analysis one can undertake and quality, dictates the quality and usefulness of the analysis one is able to conduct.
- One participant noted the challenges / issues arising from the databases used to analyse firms and the need to have an understanding of how these databases have been generated and the source data they use. This ensures analysts are able to understand how comparable different databases are and how best to use them together. The risk of this being that different databases

could create several sources of 'the truth', as they have been developed using different methods or based on different source data.

- A member of the study asked about the participant's view on proprietary data sources
- A representative from DSIT noted that usefulness of the definition is in part contingent on the availability of the data. "If the data aren't available, then we can have the best definition in the world. But we can't actually produce measures to inform policy."
- A representative from DSIT's Analysis Directorate noted they're doing a data discovery exercise, but also that there is an opportunity to explore other sources of information. However, we also want to make sure these data sources are accessible to others (across DSIT), which has implications for signing agreements to access proprietary data sources.
- DSIT participants noted that the approach selected needs to be futureproof and accommodate a degree of flexibility because the digital economy is rapidly evolving.

D2.3 Preliminary Assessment of Definitions

The table below presents the preliminary assessment of definitions against the assessment criteria to reach a shortlist of definitions / approaches to be assessed in more detail.

Table 9: Preliminary assessment of longlisted definitions

	Ability to generate overarching measure	Ability to generate key measures of interest	Ability to capture contribution of individual subsectors, including future emerging technologies	Translatability to Standard Industrial Classification	Ability to generate comparisons over time	Availability, timeliness and statistical quality of relevant data sources	Ability to capture firm-level data for individual subsectors and indicators of market competition
Current DCMS digital sector definition and measurement (benchmark)	2	3	2	4	4	4	2
Global Industry Classification Standard (GICS)	2	2	3	1	2	2	2
US Bureau of Economic Analysis approach to digital economy definition and measurement	4	3	2	4	4	4	2
UN Statistics Division SNA approach to measuring economic digitalisation	4	3	2	4	4	4	2
Proposed framework for updated DCMS definition of the digital sector and economy	4	2	2	2	4	NO	NO
OECD Digital Economy Definition	4	3	3	4	4	3	2
Growth Intelligence / NIESR Definition of the UK Digital Economy	4	2	4	3	2	2	3
BEIS Proposed Definition of the Digital Economy	4	NO	3	2	NO	NO	NO
Defining and measuring the Yorkshire digital technology sector study	3	3	4	1	2	2	4
Measuring the UK Immersive Economy	3	3	4	3	2	2	4
UK Artificial Intelligence Sector Study	3	3	4	3	2	2	4
UK Cyber Security sectoral analysis	3	3	4	3	2	2	4
NESTA dynamic mapping approach	3	3	2	4	4	3	1

- Workshop participants did not have any questions or reflections on this preliminary analysis during the workshop.

D2.4 Shortlisted Definitions / Approaches

To capture participants' more specific feedback on the four shortlisted definitions, the study team presented each of the approaches in more detail alongside an assessment of the pros and cons of each. Workshop participants were then asked for their reflections on the approaches and whether they had anything further to add. The detailed presentation of each of the four shortlisted approaches is presented in the PowerPoint slides for the workshop. A summary of the discussion around each approach is presented in the table below.

Notably, the study team highlighted that each of these shortlisted 'definitions' are more akin to groupings of approaches which have each been applied in different ways. As a result, the data sources, parameters and outputs of these approaches will be partly shaped by the availability of data and the priorities and interests of those applying them.

Table 10: Discussion of shortlisted approaches

Approach	Discussion / reflections
OECD/G20 Tiered Definition	Participant asked if this approach also drew on SOC codes, study team clarified that it is a flexible framework which can incorporate the use of SOC codes. Workshop participant asked the study team if they have a view on the extent to the OECD definition overlaps with other sectoral definitions. Noting that the limitations of the SIC code definition captures both false negatives and false positives. Study team noted that the definition is quite flexible and can be adapted to include different elements, and that the extent to which overlaps with other sectors can be avoided depend on the quality and granularity of the data collected, allowing for instance to distinguish between digital and non-digital elements of specific sectors. The study team also noted that the overlaps may be attributed to the use of digital supply and use tables, which can overlap with traditional sectoral supply and use tables. Representative from TechUk noted that whilst this approach has limitations on what it can capture, there are other surveys and data sources that can identify and capture firm level data. Based on this, the OECD could be complemented with other data collected by other courses? Member of the study team noted that this is an important point and that the study might also seek to highlight some of these alternative data sources and the extent to which they're complementary as part of the final presentation of the approaches.

NIESR/ Growth Intelligence 'Big Data' Approach	Members of the study team noted this approach is interesting because one of one of the key problems is that the SIC codes are not validated and 1/3 of the SIC codes are classified as 'other'. This approach aims to solve this problem to some extent by looking at additional information the is publicly available (and non-proprietary) from the web to see how and where the SIC codes can be augmented or reclassified, so aim is not to replace SIC but to enhance and provide additional information. Participant from DSIT noted that, under this approach, as you adjust the starting point, the definition involves the subjective judgement about what is included or excluded, for example, the proportional share of the firm whose output is generated in the digital economy. Part of that subjective judgement will be down to the person doing that judgement to determine the parameters. If we then apply this approach to reach a measure, e.g. employment or output in terms of the value of the industry, instead of a simple count of the number of companies, how do we make sure this approach ensures it is not overestimating that measure? I.e. not capturing those firms which, though supporting some aspect of digital, is actually a lower proportion of their total output than other companies? Member of the study team noted that under this approach, companies haven't been partially allocated to the digital economy. To be included, they have to be classified under the digital sector and producing digital products and services. Representative from UKRI also noted that they also saw / experienced this in their analytical work, where there will be overlaps in analysis for pieces of their internal work looking at the immersive sector vs the AI sector, for example, but also with an acknowledgement that there are grey areas and these sectors do 'bleed into' one another. The subjectivity of these analysis means that if they're added up, they sum to being far greater than the UK economy. So if we do choose to take this approach, there will be a consistent methodology and guidance for how to conduct this analysis that is agreed upon and standardised to ensure a degree of robustness within the analysis. A member of study team also noted that Technopolis and Glass.ai are currently working on a study with Innovate UK that has applied this approach to measure the immersive economy in the UK which follows a Deep Web Reading Approach. DSIT participant asked if this was a 'Satellite Account' approach, study team noted that it was not within the framework of national accounts, it was a single one off study so in that sense, yes, is a satellite account approach.

Deep Web Reading Approach	A member of the study team shared three further reflections on this approach: Even though we call it deep web reading approach, which means reading millions of websites, social media, news sources, the data is matched to the official data sets. So it's not just web data, it is matched to SIC codes whenever possible. So it shares some common commonality with the previous approach in the sense that it augments the official data sources. I want to make it so it can differentiate between core developers, wider ecosystem and adoption, which I believe it's important if we want to measure the digital economy and it depends on what definition we agree on. But are we talking about measuring the adopters? I think what is important is it doesn't rely on proprietary sources even though we do it at glass AI, we read the web, we bring the data that produce data set is owned by the government and it is matched to the official data set. So it doesn't rely on company databases. Discussion amongst the study team and participants noted that this approach assigns values or proportions of the extent of 'digital' within individual companies based on the proportion of employees and the products and services offered. This can then be used to calculate other measures such as GVA and the number of employees for the sector as a whole. Representative from Glass.ai noted that this is the approach that they're taking on a current study in collaboration with Prospective Economics to map the AI Sector in the UK as well as other sector mapping studies in the UK.
NESTA Dynamic Mapping Approach	Representative from UKRI noted that this approach (or something very similar) had been applied to a study delivered in 2013 relating to the creative industries and found that there were some issues as it did not highlight known hotspots for the sector. Participant acknowledged the approach and tools improved since then, seemed on first glance to work quite well, but when exploring the details in wasn't capturing aspects of the sector they knew to be true. Representative from DSIT noted that in their experience, this approach doesn't capture the digital occupations as well as they would like, particularly for the very new or innovative sectors of interest as part of their emerging technologies work. Member of the study team noted there's almost a systematic bias against capturing the most innovative parts of the sector because those are the ones that are most likely to be missed out in the SOC codes. This is

	particularly important to consider for the definition given the need to ensure policy relevance of the outputs. Representative from DSIT noted that using the dynamic mapping approach also assumes that your inputs to production can be adequately captured by employment, which if you consider the franchising out or you're contracting out of key parts of your production process may not be the case. From their initial analysis didn't show it as particularly useful for the digital sector, but that was using the SoC 2010 codes, so perhaps if ran it again it would be better. Member of the study team noted that this is where and why it's interesting to look at that OECD research on other proxy measures you could add to employment in terms of identifying which sectors are digitally intensive. Representative from DSIT noted the importance to ensure any statistics produced using this definition are compliant with the code of practise for statistics, which means there are certain requirements around transparency and reproducibility. A representative from DSIT noted that as part of the dynamic mapping approach, they had explored using indicators against both SIC and SOC codes (e.g. Tech professional SOC % of workforce), which the representative from TechSkills UK found to be interesting and of use.

D2.5 Preliminary Definitions

To capture workshop participant's feedback on a preliminary set of definitions, the study team presented four draft approaches, each of which are derived or inspired by one of the shortlisted definitions (see box below).

Box 3: Preliminary definitions

1. Derived from NESTA approach.

Companies in the ICT sector (by SIC codes) plus companies in other economic subsectors characterised by a high level of digital intensity (e.g. by employment).

2. Derived from OECD approach

Economic activity reliant on, or significantly enhanced by, the use of digital inputs, including digital technologies, digital infrastructure, digital services and data. This is measured by enhancing its visibility in national accounts through the use of digital supply-use tables.

3. Derived from NIESR approach.

Companies that operate in a 'digital sector' and are involved in the production of 'digital goods' and/or 'digital services'. The digital sectors, products and services used are identified and defined through analysis of text data on all UK companies included in the UK Government defined 'digital

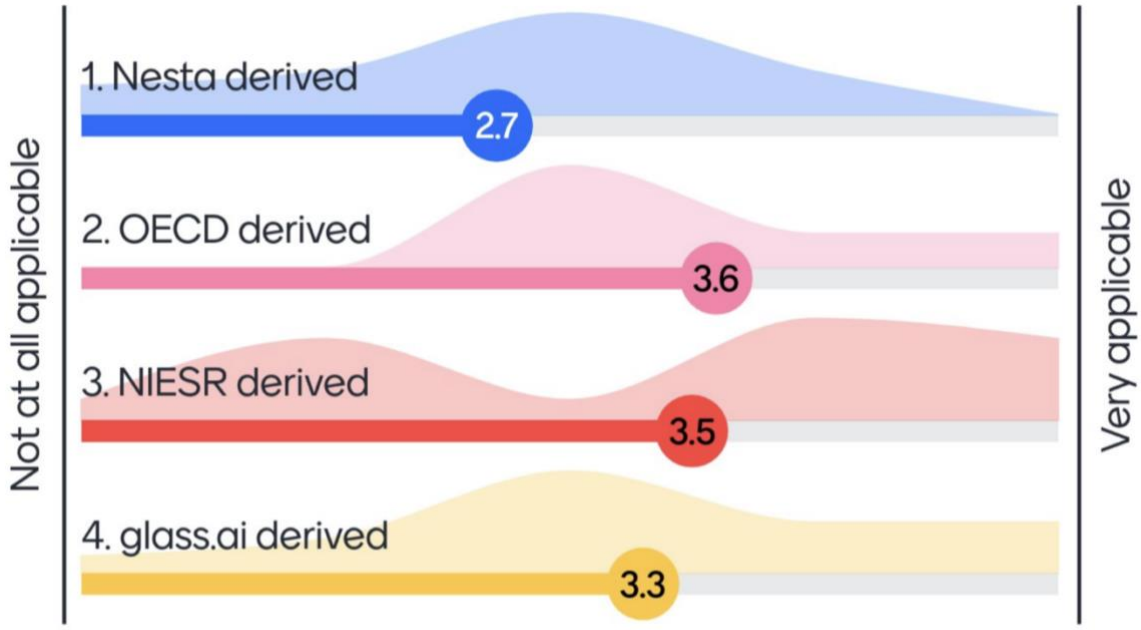
sector' (including company website descriptions, patent filings, Companies House register data and online media coverage).

4. Derived from glass.ai approach.

Companies operating in a 'digital sector', as defined by a bespoke taxonomy developed through a combination of structured stakeholder consultation and deep reading of web content on UK companies. Companies are allocated to a sector based on deep web crawling of data on company information.

Prior to discussion these definitions in detail, workshop participants were asked to score each of the definitions against their potential application to their work. The outputs of this is presented in the figure below.

Figure 9: Workshop Poll: "To what extent are these draft approaches applicable for your work and needs?"



- One representative DSIT noted that there might be interest and appetite to explore two approaches in parallel, one that supports reporting one set of statistics that talks about the total number of firms, IE accounts, and another one that then talks more about their value. Key to the value or success of such an approach would then be to have mechanisms to compare between these approaches.
- One participant from DSIT asked how, in the OECD derived definition, the focus is on 'significantly enhanced by, the use of digital inputs' and how we

identify this level. The study team noted that the current example mainly focus on the core and the narrow, while the broader one still needs to be explored. The study team also noted that different countries have used different methods for applying the OECD derived definition, for example the Netherlands and Canada have used different methods depending on the data for which they have access to.

D2.6 Other notes and discussion points that relate to the overall study and framing

- Overall reflection throughout the discussion on the important of being able to capture employment and skills.
- Representative from TechSkills noted the value / importance of being clear when discussion SIC as to whether that also relates to SOC, as this isn't always clear from the presentation and discussion but will have important implications on how applicable this will be to their work in future. Furthermore, that half of the tech workforce is outside the industry vertical of digital SIC codes.
- A member of the study team noted this is particularly important to factor into the discussion of the approaches as one starts to look at those organisations that are outside of the 'core digital sector' (i.e. with reference to taking a tiered approach).
- A member of the study team noted that, when approaching the more detailed analysis, it may also be possible to explore how the different approaches have been applied by others and highlighting the extent to which they make use of different data sources and address different key issues, including employment.
- DSIT participant noted from a trade perspective, HMG prefers to say goods and services - 'products' is not a classification recognised by the WTO, and that we only have the GAT (goods) and GATTS (services) agreements at WTO level. A Member of the study noted that it's getting harder to distinguish the two as the more physical goods embed IP and software and are bundled with services, though a 'Product' would in theory sit outside this.
- Throughout the workshop, multiple points were raised about the importance of adopting a methodology / approach that doesn't fall into the same trap as SIC / SOC codes in that it becomes outdated and unfit for purpose in a relatively short period of time.
- Participants noted that the approach selected needs to be futureproof and accommodate a degree of flexibility because the digital economy in particular

is changing and rapidly evolving and highly dynamic. In practical terms, this will mean trying to avoid locking the definition into a particular methodology with a set of ingredients that doesn't change.

- Representative from DSIT also noted that where we reference SIC codes in the presentation, it is a shorthand for a range of measures that are used in standard government datasets (e.g. HN8 codes, CPA, SOC etc.). This is done to get consistent measures across ONS datasets in particular for all the economic measures of interest, including employment.
- Throughout the workshop, participants also noted that it was unlikely we would be able to reach a perfect definition that works well for all purposes. Instead, the study (and wider stakeholders) would be seeking to identify complementary approaches that can be used together or in parallel to provide the picture and body of evidence needed.
- In this vein, participants often highlighted the need to ensure there was some guidance provided to support consistent application of any approaches put forward and providing assurance of the rigor of the analysis. In particular, this would be important for minimising contractions in analysis provided by different departments or teams, to mitigate against any reputational damage.
- There were some questions and discussion from participants around global comparisons, and the approaches of other nations.
- A member of the study team noted that the OECD approach is one that is being used more often in other countries' national statistical agencies as it is proposed by the UN Statistics Division. Important to note that other countries might also have their different definitions within ministries that reflect their national policies and priorities. This is not something covered in detail within this study and currently have a limited view of these other definitions, however the study team are soliciting feedback from a couple of international organisations including the US Government.
- Member of the study team noted that translatability to SIC codes is important as this is what allows for international comparisons.
- A member of the study team noted that the SNA revision due by next year is expanding the scope for Digital SUTs, so it has some international momentum.
- A representative from DSIT expressed interest in sense checking some of these definitions or approaches, for example to see whether these methodologies would rule in or out particular firms - e.g. the firms the participants were asked about in the beginning. Participants and the study team agreed that these approaches will needed to be sense checked and

validated. The study team noted at this point that this work would aften require the full implementation of these methods, which unfortunately would be outside the scope and timeframe of the current study, however is something that will be reflected in the final output, ideally alongside some recommendations for any such implementations.

- Throughout the workshop, participants also shared other references or work:
    - https://unece.org/sites/default/files/2021-04/ECE-CES-GE20-2021-10-EN.pdf
    - https://www.cbs.nl/en-gb/background/2021/49/research-on-supply-use-tables-for-the-digital-economy-in-the-netherlands
    - https://www.productivity.ac.uk/news/professor-raquel-ortega-argiles-to-head-productivity-lab/
    - https://innovationgraph.github.com/economies/gb
    - https://waifinder.iuk.ktn-uk.org/explorer

D3 Workshop Participants

The workshop was attended by 26 people from the following nine organisations:

- Department of Science, Innovation and Technology (10 participants)
- Glass.ai (1 participant)
- Perspective Economics (1 participant)
- Technopolis (6 participants)
- TechSkills (1 participant)
- Tech UK (1 participant)
- UKRI - Innovate UK (2 participants)
- University of Cambridge (1 participant)
- Welsh Government (2 participants)

Appendix E: Critical Assessment of Key Current Approaches

E1 Operationalisation of the OECD 2020 definition of the digital economy

Building on work by Bukht and Heeks,[24] the following definition of the digital economy was put forward by the OECD in the G20 digital economy task force report[25]: "The digital economy incorporates all economic activity reliant on, or significantly enhanced by, the use of digital inputs, including digital technologies, digital infrastructure, digital services and data. It refers to all producers and consumers, including government, that are utilising these digital inputs in their economic activities."

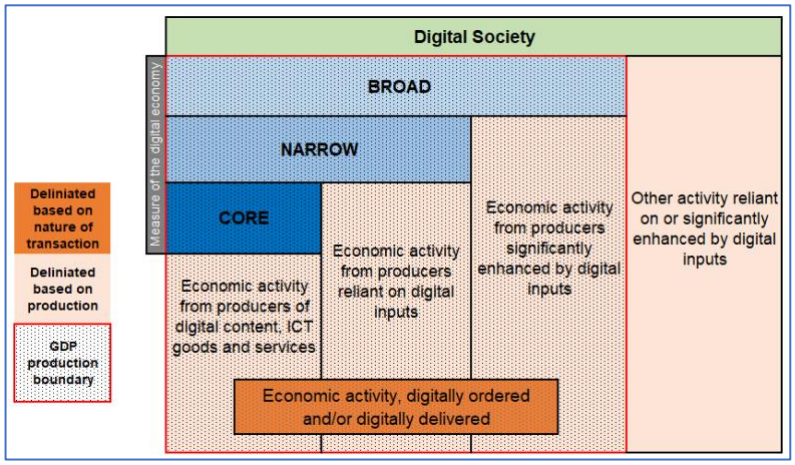
This definition, which was designed to align with the concepts used in the System of National Accounts (SNA), was translated into an approach to measurement built around four tiers (or measures) of the digital economy, moving from the centre towards broader definitions:

- The ‘Core’ tier includes all economic activities related to the production of ICT goods such as semiconductors and processors, computers, smartphones, software and algorithms, as well as digital services such as internet and telecom networks, as defined in ISIC. Rev. 4.
- The ‘Narrow’ tier includes all core digital activities, in addition to economic units that rely entirely on digital technology and data to operate, such as mobile payment platform, e-commerce or digital labour platforms. These economic units do not directly produce ICT goods and digital services as defined by ISIC rev. 4.
- The ‘Broad’ tier includes the first two tiers in addition to economic activity from firms significantly enhanced by digital technologies and data.
- The final tier, the Digital Society, includes all digital activities undertaken by individuals in a society that are not carried out for pay or profit (i.e. not included in the SNA production boundary), such as the use of free digital platforms (including public digital platforms).

[24] Bukht, R. and Heeks, R. (2017). Defining, Conceptualising and Measuring the Digital Economy. Development Informatics Working Paper no. 68.
[25] OECD (2020). A Roadmap Toward a Common Framework for Measuring the Digital Economy. Report for the G20 Digital Economy Task Force.

This approach also presents an additional definition: economic activity digitally ordered and/or digitally delivered. This can be considered an alternative perspective on the digital economy, as it is delineated based on the nature of the transaction, rather than based on firms' output or production methods. This measure focuses on the method used for ordering or delivery, regardless of the final product or how it is produced.

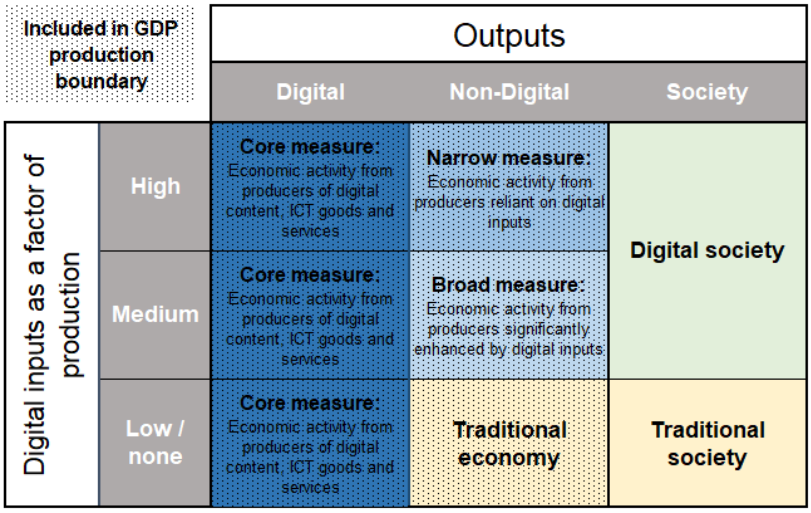
Figure 10: Tiered definition of the Digital Economy



Source: OECD, A Roadmap Toward a Common Framework for Measuring the Digital Economy

These tiers can further be classified (as shown in Figure 11) by the type of production (digital, non-digital, society) and the level of digital inputs used in the production process (high, medium, low/none). All output deemed digital is regarded as part of the Core tier, regardless of the level of digital input. In practice, this includes all the economic activity generated by the ICT sector. Economic activities included in the Narrow and Broad tiers include output that is not digital in nature, so this is classified according to digital inputs: high for the Narrow tier (reliant on digital inputs), medium for the Broad tier (significantly enhanced by digital inputs), and low/none for the traditional economy. Finally, the Digital Society includes non-market outputs delivered with high or medium digital inputs, while non-market outputs with low/no digital inputs are the traditional household sector activities.

Figure 11: Measures of the Digital Economy within the traditional economy



Source: OECD, A Roadmap Toward a Common Framework for Measuring the Digital Economy

As mentioned, an alternative approach is defined using two concepts:

- 'Digitally ordered goods and services' is equivalent to the current e-commerce definition already broadly in use, which comprises "the sale or purchase of goods or services, conducted over computer networks by methods specifically designed for the purpose of receiving or placing orders".
- 'Digitally delivered services' comprise "all services that are delivered remotely in an electronic format, using computer networks specifically designed for the purpose". This category includes services for which a computer network is necessary for the service to be administered, as well as other intangible goods and services.

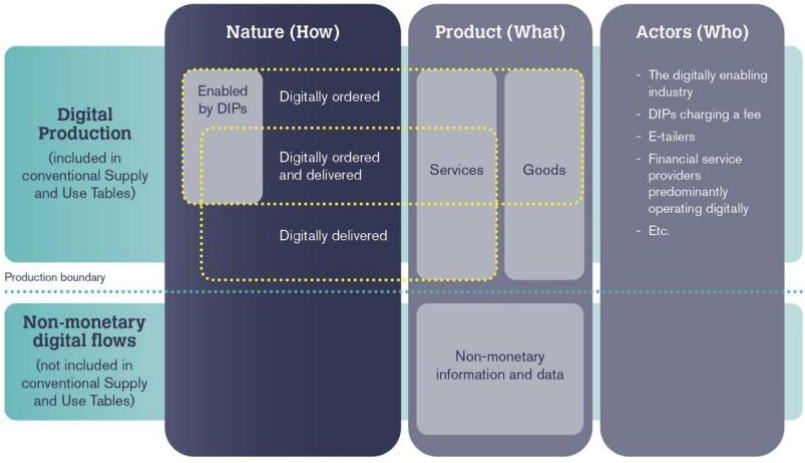
The OECD report provides examples of international definitions for key components of the digital economy, and which are needed to operationalise the high-level conceptual definition described above. This includes work from the UNCTAD, the IMF-OECD-WTO Handbook on Measuring Digital Trade as well as Eurostat and national statistical offices, including ONS (see more below). Moreover, related work is being undertaken internationally for the formal revision process of the SNA, due to complete in 2025, which aims to combine the different elements described above in a chapter on digitalisation. In this revision process, the use of three analytical tools is being suggested to make the digital economy more visible in the national statistics:

1. A thematic account on the digital economy
2. Digital Supply-Use Tables (SUTs)
3. An extended account on free services of online platforms.

These tools complement each other, with the thematic account bringing visibility to accounts already included in the standard classifications, while the SUTs enable a decomposition of the sources of supply and use, and the extended account allows for measurement of an estimate of the value of free services. Thus, when this process will be completed, new measurements will be possible, such as that of the value of data as a product.

To date, the most widespread approach to operationalising this definition is the work in the area of Digital SUTs and digital trade. A framework for this was set out in the OECD Handbook on Measuring Digital Trade,[26] categorising transactions according to the 'where' (in the accounts are these transactions recorded), the 'how' (are digital transactions defined), the 'what' (types of products are included) and the 'who' (are the buyers and sellers).

Figure 12: Proposed framework for Digital SUTs



Source: OECD Handbook on Compiling Digital Supply and Use Tables

To create Digital SUTs, the standard OECD supply-use tables have been modified to produce additional information on the digital economy. The modifications include:

[26] OECD (2023). OECD Handbook on Compiling Digital Supply and Use Tables. OECD Publishing, Paris.

- Seven additional columns for digital industries that are considered worth presenting separately: Digitally enabling industries, Digital intermediary platforms charging a fee, Data and advertising driven digital platforms, Firms dependent on intermediary platforms, E-Tailers, Digital only firms providing financial and insurance services, and Other producers only operating digitally.
- Six additional rows under each product (and total), separating transactions by whether they are: digitally ordered or not digitally ordered, with digitally ordered transactions further broken down into ordered directly from the counterparty or ordered via a digital intermediation platform (DIP), with a final breakdown splitting the products ordered via DIPs between resident and non-resident platforms.
- Two additional columns showing the nature of the delivery of the service as either digitally delivered or not digitally delivered.
- Four additional rows, representing digital product of particular interest: digital intermediation services (DIS) and cloud computing services (CCS), as well as total Information and Communication Technology (ICT) goods and digital services that fall within the SNA production boundary.
- Three additional rows, representing data and digital service products that are currently outside the SNA production boundary (along with some other intangibles).

The digital industries proposed within the Digital SUTs framework broadly align with the tiers outlined in the OECD definition. The Core measure corresponds to the “Digitally enabling industries”, while all the other industries would include firms classified within the Narrow tier as they are reliant on digital inputs. The Broad tier is not directly visible in the digital industries, but potentially it could be measured using the alternative measure and assigning output of a product to either the Narrow or the Broad tier based on the level of digital ordering and/or delivery taking place or the amount of digital goods and services used as an input into production, both of which are outputs of the Digital SUTs.

Table 11: Definition of the seven industries in the digital SUT framework

Industry	Definition	Explanation	Examples
Digitally enabling industries	Businesses engaging in production that enables the function of information processing and communication by electronic means including transmission and display; explicitly it is those industries defined in the ICT sector list in ISIC Rev. 4	It includes Internet service providers, telecommunications companies, providers and developers of software, Computer manufacturers, and website developers While excluding free and priced digital media providers, social media providers, digital platforms directly or intermediately providing goods and services not included in the defined ICT sector list for ISIC Rev.4	Amazon Web Services, BSNL, Dell, Indosat, Ooredoo, Orange, Verizon
Digital intermediary platforms charging a fee	Businesses that operate online interfaces that facilitate, for a fee, the direct interaction between multiple buyers and multiple sellers, without the platform taking economic ownership of the goods or services that are being sold (intermediated)	It includes food delivery companies, travel booking portals, platforms facilitating online auctions or marketplaces that assume no ownership of stock While excluding digital platforms that sell their own goods or services, platforms that do not receive an explicit monetary fee from either the producer or consumer	Airbnb, Booking.com, Deliveroo, Didi, Mercado Libre, OLA, Trivago, Uber
Data and advertising driven digital platforms	Businesses that are operating exclusively online that predominately generate revenue via selling data or advertising space	It includes search engines, social media platforms, developers of zero-priced phone applications and information sharing platforms While excluding business that sell goods or service (excluding data or advertising space) for a monetary price, subscription-based services providers, priced phone applications and information sharing platforms	Citymapper, Facebook, Google, Tik Tok, Twitch, Youku

Firms dependent on intermediary platforms	Businesses that always or a significant majority of the time transact with consumers via an independently owner third party digital platform	It includes independent service providers who source work from digital platforms, business who sell via a third-party digital platform While excluding business who sell predominately digitally but do so via their own website/digital platform	Bicycle couriers, Ghost kitchens, Uber drivers
E-Tailers	Retail and wholesale businesses engaged in purchasing and reselling goods or services who receive a majority of their orders digitally	It includes businesses receiving orders digitally that sell their own inventory and/or have set contracts with producers and suppliers While excluding; businesses that carry no ownership of the purchased good or service, businesses who contribute no additional value added to the consumed good or service	ASOS, JD.com, Sarenza, Yesstyle, Zalando
Digital only firms providing financial and insurance services	Businesses providing financial and insurance services that are operating exclusively digitally, with no interaction with consumers physically	It includes online only banks and other financial service providers, online only payment system providers While excluding; banks and other financial service providers that include consumer-facing locations, platforms solely acting as intermediaries between lender and borrower (i.e. crowd funding websites)	Ally financial, Directline, Fidor bank, Open bank, Paypal, Seven bank, Transferwise
Other producers only operating digitally	Businesses that produce their own services for sale but operate exclusively digitally	It includes priced digital media providers, subscription based service providers (assuming the service is delivered digitally) While excluding; business who do not deliver their good or service digitally regardless of how they receive orders	Bet365, The Independent newspaper, Netflix, Showmax, Spotify, Starz Play

Source: OECD Handbook on Compiling Digital Supply and Use Tables

As explained above, the other dimensions of the Digital SUTs concern the nature of the transaction (the 'how') and the goods and services produced (the 'what'). The former focuses on how products and services are ordered and delivered. The latter allows for aggregation and presentation of ICT goods and digital services, as well as products considered to be of policy interest such as digital intermediation services (DIS) and cloud computing services (CCS). In the framework, Digital SUTs include also rows that provide for estimates related to data, zero-priced digital services provided by enterprises and zero-priced digital services provided by communities, but their operationalisation is still under debate (as part of the Digital Economy Satellite Account work).

Once the Digital SUTs are populated, an array of indicators can be derived from these tables. The OECD indicates as priority the following:

1. Expenditure disaggregated by the nature of the transaction, including household final consumption expenditure digitally ordered, imports and exports digitally ordered.
2. Output and/or intermediate consumption of DIS, CCS and total ICT goods and digital services.
3. Digital industries' output, GVA and its components.

Clearly, there are other indicators that individual countries could pursue, reflecting their policy needs and data availability. Examples include additional information on research and development, the growth and level of investment in digital products, as well as labour-related indicators such as hours worked or occupations in the digital industries.

Digital SUTs aims to reallocate production already contained in the national accounts and conventional SUTs in a way that provides insights into the digital economy. The compilation methods for deriving the high priority indicators associated with digital industries fall into two categories:

1. Reallocation of specific units: where specific units are identified as matching the criteria of the new digital industry and estimates of output, intermediate consumption and value added associated with these units are moved to the new industry.
2. Aggregate reallocation based on indicators: where specific units cannot be identified, aggregated estimates associated with the production of these units is calculated using alternative indicators. The aggregate amounts can then be deducted from existing industry classes and moved to the new industry.

Data collection for the other two dimensions of the Digital SUTs (the products and the nature of transactions) mainly depend on existing sources, for example by adding questions to the existing business surveys (as already done by ONS with the Digital Economy Survey[27]), but also household surveys.

[27] See https://www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/2021digitaleconomysurveysurveyquestions

The following table shows examples of approaches followed by other countries. Statistics Canada[28] has been able to produce estimates of the output of digital industries, the share of digitally ordered products and services compared to the total supply, the share of digitally delivered exports, the share of digital products compared to total use, and a figure on the contribution of the Digital Economy to total GDP and jobs nationally and at regional level. Statistics Netherlands[29] was able to estimate the output and the GVA of the Digital Economy as a whole and of the individual industries, supply and use of goods and services (including intermediation services and cloud computing), and the estimate of the transaction methods.

Table 12: Examples of digital SUT compilations

Element	Canada	Netherlands
Digitally enabling industries	ICT industries (IOIC codes) mapped from standard SUTs, split into hardware, software, telecommunications and other services	Identified businesses part of this category using their NACE class. Some business units in the scope of the NACE classes of the enabling industries are included in other digital industries (e.g. intermediary platforms). According to the 'Digital industry decision tree', these are not included in the enabling industries. In cases where ICT industries make up only part of an industry as available in the standard SUTs compiled by Statistics Netherlands, the Structural Business Statistics (SBS)survey data is used to calculate the shares of supply, use and value added to be re-allocated from each standard industry to the digitally enabling industry. Since the SBS is a sample survey, the sampled businesses need to incremented/weighted.
Digital intermediary platforms charging a fee	Most of the activity is dominated by a few large firms in taxi, delivery and short-term accommodation. Firm level data was used to construct the industry estimates for domestic platforms, including publicly available annual reports and tax filings.	From a survey by Statistics Netherlands on resident digital intermediary platforms, they retrieved the subsection of businesses which operate an online platform (with fifty percent or more of their employees working specifically for the platform), list further extended with large platforms based on expert knowledge. To identify those charging a fee they used

[28] Statistics Canada (2021). Measuring the digital economy: The Canadian digital supply and use tables 2017-2019.
[29] CBS (2021). Supply-Use tables for the digital economy. Experimental research.

		answers on survey questions based on the sources of revenue for the business in question or on the (non)requirement of payment for both users and providers on the platform. Only a subsample of the identified fee-driven platforms is present on SBS but represent the vast majority of the total revenue (based on VAT turnover data). Data on this subsample was used to make estimates for the rows and columns of the SUTs for the entire industry.
Data and advertising driven digital platforms	Estimates are based on a number of firms classified to the advertising industry that reported mostly online revenues. A number of firms that would qualify for this heading are currently classified to the ICT industries under “Data processing and hosting services”, so no special effort was made to reallocate them as they would still be captured under the general digital economy aggregates. Data driven digital platforms are dominated by large global firms. No evidence was found of Canadian firms involved in selling user data as their main source of revenues.	Equivalent to the method for platforms charging a fee, platforms are identified through surveys, and the industry and product values present on SBS are scaled up based on the total revenue of the survey population compared to the complete population.
Firms dependent on intermediary platforms	Firm-level tax data was used to split the taxi and local messenger and delivery industries into their traditional and platform-dependent portions. Secondary output was residually derived as the difference between resident platform revenues and the fees generated from producers whose main revenues rely on platforms. Estimates in the accommodation sector relied on information from private sector data providers.	ICT usage survey with questions on the sales made through digital intermediary platforms was used for business with 2 or more employees. Additional sources were used for particular cases. For drivers on Uber and similar platforms, and home cleaners on Helpling and similar platforms, external reports were used to estimate their number and average gross monthly earnings. The estimates on holiday rentals are based on research by Statistics Netherlands (using number of nights and average price of rental).
E-Tailers	Estimates for these industries are available from the Annual Wholesale	For retail, the values for the E-Tailers present in the SBS was scaled to national

	Trade Survey, captured under the "Business-to-business electronic markets" industry and the Annual Retail Trade Survey, under "Electronic shopping and mail-order houses". However, further research is needed to determine the extent to which online-only retailers may still be classified under the brick-and-mortar retailing industries.	population totals. Wholesalers using mainly online channels had to be identifies through the ICT usage survey.
Digital only firms providing financial and insurance services	Firm level data from regulatory and tax filings were used to derive the estimates for this industry.	As no other source was available, for now this digital industry is limited to a small share (0.8%) of NACE section 'K - Financial and insurance activities', based on the ICT usage survey (answering that all sales are made online).
Other producers only operating digitally	For practical reasons, the units classified here are not required to generate 100% of their revenues from online activities but rather a large majority of their revenues is deemed a sufficient condition. Some goods producers that reported a large proportion of e-commerce sales were excluded from this category. These are units that were not deemed to be truly reliant on e- commerce for their activities and their inclusion would not have provided analytically useful results for the relative role of digitalization in the economy.	After testing three alternative approaches to identify businesses that meet the criteria of 'Other producers only operating digitally', they were unable to identify any substantial amount of relevant business units for this digital industry.
Digitally ordered	Digitally ordered products were split based on the source, i.e. from a counterparty, via a resident digital intermediary, via a non-resident one, and via a resident retailer or wholesaler. For the first one, outputs were derived mainly from e-commerce revenues captured on annual establishment-based business surveys. These estimates were confronted and complemented by estimates from the 2019 Survey of Digital Technology and Internet Use, which were also trended back to 2018 and 2017	For intermediate consumption, data from the ICT usage survey is employed, using information on the (NACE) industry in which each business is active. This allows to derive fractions for the proportion of digitally ordered sales and purchases for each industry that is present in the ICT usage survey. For household final consumption, they decided to use the SBS retail questionnaire's revenue, taking the total sales revenue of a certain product multiplied by the online fraction of sales, and divide this by the total sales

	based on general observed trends in the annual surveys. Digitally ordered imports of goods were based on a combination of merchandise trade data on small-valued couriered items as well as the value-added tax remittances by some know large non-resident retail platforms. Information on digitally delivered products from the Business Survey on Exports of Commercial Services were used to model estimates for both the imports and exports of digitally ordered services. E-commerce modules from the wholesale and retail industry surveys are used to derive the proportion of digitally ordered retail and wholesale margin output in the SUTs. Margin output values are subsequently grossed up by the survey sales-to-margin ratio to derive the gross value of e-commerce sales in the SUTs. Total gross sales by product are allocated to demand based on the valuation tables and to domestic output and imports based on the relative shares of these two sources of supply.	revenue of that product to obtain an online fraction of sales. For exports, the share of online exports in total turnover for the business units of ten or more persons employed in the ICT usage survey sample were obtained by combining the share of digital sales and the share of non-resident customers in the digital sales. The share of digitally ordered exports in total turnover and the turnover from the SBS data (for small units not included in SBS the turnover from the ICT usage survey was used) result in a value of digitally ordered export for the units in the sample of the ICT survey. These units were merged with the international trade data (ITSS and ITGS).
Digitally delivered	The proportion by product of digitally delivered services from the Business Survey on Exports of Commercial Services are applied to both exports and imports in the SUTs. Digitally delivered exports are proportionally allocated to producing industries to derive the associated outputs. Similarly, digitally delivered imports are proportionally allocated to domestic uses to derive the equivalent demand. The trade-based estimates are complemented by an approach that relies on the characteristics of the products in the SUTs.	Previous research by Statistics Netherlands on modes of supply (MoS) in the international trade of services was used. For each service in the BPM6 classification in Dutch trade, the MoS was estimated.
ICT goods	Disaggregation of market transactions already covered in the current SUTs.	Any ICT goods in the OECD list of ICT products that coincide one-to-one with a product in the working level of the Dutch national accounts are shifted in their

		entirety to ICT goods. Products are identified by coupling the CPA 2.1 codes that correspond to the CPC 2.1 codes in the OECD guidelines for the DSUTs. When codes include non-ICT goods, PRODCOM is used.
Digital services		The approach for the domestic output is largely similar to the approach for ICT goods.
Cloud computing services		From SBS questions on the provision of cloud computing, they estimate for some categories a fraction of all sales regarding software. This fraction is put on the total production. For the product data processing, hosting and related services they include 100% of production as a cloud service, ‘Computer programming, consultancy and related services’ (CPA 2.1, 62) that are produced by businesses in ‘Data processing, hosting and related activities; web portals’ (NACE Rev. 2, 63.1) in the business registry are included as cloud computing services as well.
Digital intermediation services		From the subset of businesses identified to be resident digital intermediary platforms in the digital industry section, they first take the subsections that are fee platforms. They then retrieve the percentage of employees working specifically for each platform as an indicator for the share of the production that is to be seen as a priced digital intermediary service. These shares are applied to the SBS data on the output of any business units that are at least partially a fee driven platform. Finally, using a correspondence table between product classifications in the SBS survey and product classifications in the final SUTs, they can calculate the fraction of (fee-driven) platform services in the total production of each industry and product

		combination. These fractions are re-allocated from their origin product to priced digital intermediary services.

Although not based on the latest framework for Digital SUTs, ONS has also published work on the digital economy,[30] mainly focused on defining digital products inside the SNA 2008 production boundary (excluding digital intermediary services) and non-digital products significantly affected by digitalisation (also referred to as digitally affected products; examples include accommodation services, financial and insurance services, education services). The digital products are approximated through the definition and delineation of the ICT sector in the Complimentary Class of ICT products in the Central Product Classification (CPA) version 2.1. The definition of digital is drawn from the OECD Guidelines for Supply-Use, which include 10 categories of products. ONS then used data from UK Manufacturers' Sales by Product (PRODCOM) and the Annual Survey of Goods and Services (ASGS) to expand the product categories in the SUTs supply table into their respective digital products (including ICT goods, digital services, and cloud services), digitally affected products, and non-digital components. These estimates of output by product are then translated into output by industry using the domestic output table to give an estimate of output (by industry) of digital products, digitally affected products, and non-digital output. Finally, digital and digitally affected output (by industry) is translated into gross value added (GVA) using the ratio of output to GVA for each industry from the intermediate consumption table of the SUTs. Here, the ratio of output to GVA for each industry is assumed to be the same for each industry's digital and non-digital component, which is a reasonable assumption but may be invalid if the production structure of the two differs significantly. This approaches still use only some elements of the Digital SUTs framework, and the ONS recognises that more work is needed to estimate the remaining elements.

Criterion 1: Ability to generate an overarching measure of the value of the UK digital economy.

To what extent does it allow for measurement of the economic value of the core digital sector (OECD definition as used by the UK Government)?

As shown in the examples from Canada and the Netherlands, with Digital SUTs is possible to estimate the output and GVA from the core and narrow tiers of the digital economy. This is mainly based on the measurement of the 7 digital industries identified in the Digital SUTs framework, and the measurement of each industry can vary based on the approach decided and the data sources (apart from the Digitally enabling industry that is based on the ISIC definition of the ICT sector).

[30] ONS (2022). UK Digital Economy Research: 2019.

To what extent does it allow for the measurement of the economic value of digital-only service producers (e.g. digital-only banks, e-payment service providers, online only media outlets, streaming services, edtech platforms, healthtech platforms)?

There are different ways in which Digital SUTs could allow for the measurement of digital-only services. In terms of products, the framework suggests including a row specifically for digital services, based on the CPC classification. Considering transaction type, the SUTs allow to split the output of services in digitally and non-digitally ordered and delivered, providing a measure of the level of output for digital services. Finally, some of the industries included in the framework could be considered part of this measure such as Digital only firms providing financial and insurance services.

To what extent does it allow for measurement of the economic value of digital intermediary platforms, including both those which charge a fee those which are data/advertising driven?

Such categories are explicitly included in the Digital SUTs (Digital intermediary platforms charging a fee and Data and advertising driven digital platforms in the industries, Digital intermediary services in the products). For their measurement, a combination of the ONS Annual Business Survey and the new Digital Economy Survey (once operative) could be suggested.

To what extent does it allow for measurement of wider digitalisation of the economy: economic units where production has been significantly enhanced by digital technology?

Even if the definition includes a broader tier of the digital economy, attempts to measure it have not yet been attempted. Potentially, it could be observed in the digital SUTs based on indicators such as the level of digital ordering and/or delivery taking place or the amount of digital goods and services used as an input into production, both of which are outputs of the digital SUTs.

To what extent does it allow for disaggregation of the economic activities listed in the above questions from the overarching measure of the digital economy?

The disaggregation of the different activities should be automatically made visible in the Digital SUTs.

Criterion 2: Ability to generate key measures of interest.

To what extent does it allow for the generation the following measures?

- Gross value-added.
- Employment.
- Worker earnings.

International trade measures, including imports and exports of services and goods.

Business demography measures, including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites.

Measures a. to d. have all been generated in the digital SUTs for Canada and the Netherlands. Digital SUTs would also allow measurement of R&D (on the Frascati definition) in the different categories. Specific business

demography measures have not been included in the examples, and the feasiblity of including them would mainly depend on the data available.

What other measures of interest can be generated from this definition (e.g. valuation of tech companies, number of unicorns, VC investment in the tech sector, FDI in the tech sector, growth indicators beyond revenues and employees)?

Similarly to the previous question, the feasibility of generating other measures of interest would mainly depend on the data available.

Criterion 3: Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. Quantum).

To what extent does it allow for measurement of the following subsectors?

- Data subsector (including data processing, hosting and infrastructure).
- Artificial intelligence subsector.
- Cyber-security subsector.
- Software development.
- ICT manufacturing (including semiconductors)
- ICT trade.
- Telecommunications.
- Digital services (e.g. FinTech, EdTech, HealthTech)

The framework for Digital SUTs decided to make more visible two products of interest, Digital intermediary services and Cloud computing services. Something similar could be done with other subsectors such as artificial intelligence or cyber-security, using a mix of statistical sources and studies using innovative methods (or combining the approaches of other definitions, such as deep web reading). Additional rows are also included for products outside the SNA production boundary, such as data, but measurement for many of these is not foreseen in the near future, as information on these aspects is not captured in national accounts statistics (although an estimate for data by the sum of costs method is likely to be included in future).

To what extent does it allow for the measurement of other subsectors not listed above, and at what level of granularity?

As mentioned, digital intermediary services and cloud computing services are included in the aspects to be measured, with some examples shown in Canada and the Netherlands, measuring their value.

To what extent does it allow for the identification and incorporation of future emerging technologies as subsectors when needed (e.g. Quantum)?

The definition is flexible enough to allow countries to measure different aspects of interest.

Criterion 4: Translatability to Standard Industrial Classification.

To what extent is it translatable to SIC codes?

The definition and framework for digital SUTs include SIC codes by default.

How easily is it translatable to SIC codes (i.e. what is required to translate the definition to SIC codes)?

The definition and framework for digital SUTs include SIC codes by default.

Criterion 5: Ability to generate comparisons over time.

To what extent does it allow for measurement of trends over time, including historical comparison?

The comparability over time depends on the measurement approach for the different elements of the Digital SUTs, and the flexibility offered (e.g. in what to include under the new industries) could limit to some extend the measurement of trends.

At what level of temporal granularity does it allow for measurement of time trends?

The granularity completely depends on the data sources selected; most SNA component data are collected quarterly.

Criterion 6: Availability, timeliness and statistical quality of relevant data sources.

To what extent is the data required to operationalise the definition currently, available? If it is not available, can it be collected?

The Digital SUTs are about reallocating production that is already contained in the national accounts and conventional SUTs in a way that provides insights into the digital economy. Thus, the majority of data should already be available, or its collection would require some changes in the surveys, and additional work requiring resourcing, by ONS. In some case, alternative indicators might be used to calculate the aggregated estimated associated with the production of units which cannot be identified using conventional SUTs (e.g. for Firms dependent on intermediary platforms it might be necessary to use estimates from studies on the sharing economy).

To what extent is the data required to operationalise the definition available from ONS statistics?

As explained in the previous question, most data should be available from ONS statistics, but this depends on the methods used for estimating industries and products.

To what extent does the data required to operationalise the definition require subscriptions to proprietary data sources?

No proprietary data sources should be required.

How recent is the available data required to operationalise the definition?

This mainly depends on the sources employed, but their update would mainly depend on decisions from ONS.

To what extent is the data required to operationalise the definition nationally representative, regionally representative, representative at the sectoral level, representative at the sub-sectoral level?

In the example of Canada, indicators from the Digital SUTs have been presented both at national and regional levels, although the UK has fewer regional statistics than many other comparator economies. As already expected, different sectors and subsectors can be made visible in the tables, depending on the underlying data.

Criterion 7: Ability to capture firm-level data for individual subsectors and indicators of market competition.

To what extent does it allow for measurement of firm-level data for the individual subsectors listed in question 3.1?

Much of the data feeding the framework for Digital SUTs is built from firm-level micro data and accessing this via the ONS Secure Research Service (SRS) would allow subsectors to be analysed.

To what extent does it allow for measurement of the following indicators of market competition?

- Market share concentration.
- Business entry and exit.

No examples of this included in the framework, but in the Canada and the Netherlands cases firm-level data (e.g. VAT) has been used to facilitate estimates for entire groups starting from data from a subsample, and as mentioned in the previous question these indicators can be studied in SRS at a sufficiently aggregated level that the results are non-disclosive.

Criterion 8: Translatability to international classification systems.
To what extent is it translatable to the Standard Occupational Classification? *To a great extent.* To what extent is it translatable to the Harmonised System. *To a great extent.*

E2 Deep web reading approach to measuring economic sectors

This group of data-driven approaches of defining and measuring economic sectors has been applied to the measurement of the Yorkshire digital technology sector and the UK's AI and cyber-security sectors.[31] In addition, it is currently in the process of being applied to the measurement of the UK's immersive economy.[32] These studies have different aims and there are some differences in the specific methodology employed by each study, however, all three use an AI capability (in these case, the glass.ai capability) that deep reads the web and official sources (e.g. Companies House) to track the activity of millions of companies and analyse information on them. Definition and measurement of the digital economy using this approach would involve identifying the companies that operate in the digital economy, as defined by a bespoke taxonomy. This taxonomy would be developed through a combination of structured stakeholder consultation, deep reading of web content on UK companies and data linking to relevant ONS statistics.

Implementing this approach involves the following:

1. Development of either a conceptual definition of the digital economy or, for a more data driven approach, a longlist of relevant companies in the sector.
2. Based on this definition, a taxonomy of the sector is developed based on a combination of desk research and stakeholder consultation, as well as potentially using data-driven insights from a language model. Depending on the nature and size of the sector under study, this step may involve developing a list of subsectors and/or technologies related to the sector/ subsectors. Development of these lists may involve workshop sessions with representatives from academia, industry, government and the study team.
3. A list of relevant keywords/ phrases associated with each taxonomic category is developed. This is often an iterative process in which preliminary sample results are

[31] Donaldson et al (2023); Perspective Economics (2023).
[32] Technopolis & Glass.ai (forthcoming) UK Immersive Economy Report 2023.

collected and shared with experts for feedback in order to refine the language model and improve the quality of taxonomic classification.

4. Once the model is finalised, a web crawl is run to construct a dataset of companies included in the sector. For example, the glass.ai AI capability deep reads ~2.2 million UK organisation websites (including companies, partnerships, government, non-profits, and sole traders) and social media pages (e.g. LinkedIn), as well as potentially reading relevant news webpages and sector-specific sources to identify text that may suggest a company is active in the sector under study. In addition, the glass.ai AI capability regularly crawls the websites of 35 million businesses globally and is therefore able to identify foreign companies that have presence in the UK market.
5. Additional information on the characteristics of companies in the dataset (e.g. employment, turnover, etc.) are collected by matching the crawled dataset to other UK business datasets including ONS statistics, Companies House register data. It is also possible to include data from proprietary datasets such as the Fame, Beauhurst or Crunchbase datasets in this data linking step.
6. Where necessary, an original survey can also be fielded to fill data gaps.

Criterion 1: Ability to generate an overarching measure of the value of the UK digital economy.

To what extent does it allow for measurement of the economic value of the core digital sector (OECD definition as used by the UK Government)?

Based on the AI sector and Cyber Security sector studies, this approach has allowed the measurement of the size of the sector, revenues from the firms, GVA and employment.

To what extent does it allow for the measurement of the economic value of digital-only service producers (e.g. digital-only banks, e-payment service providers, online only media outlets, streaming services, edtech platforms, healthtech platforms)?

This approach allows for a flexible, fine-grained classification of companies into subsectors and could be adapted to allow for the measurement of the economic value of digital-only service producers.

To what extent does it allow for measurement of the economic value of digital intermediary platforms, including both those which charge a fee those which are data/advertising driven?

This approach allows for a flexible, fine-grained classification of companies into subsectors and could be adapted to allow for the measurement of the economic value of digital intermediary platforms.

To what extent does it allow for measurement of wider digitalisation of the economy: economic units where production has been significantly enhanced by digital technology?

Similarly to the approach for the AI sector and Cyber Security sector studies, where firms have been split between dedicated and diversified AI businesses (based on the role that AI technologies played in their activities) a similar approach could be taken for the digital economy, with different degrees of utilisation of digital technology. The approach also captures foreign digital companies active in the UK market.

To what extent does it allow for disaggregation of the economic activities listed in the above questions from the overarching measure of the digital economy?

The approach would allow for disaggregation of digital-only service producers and digital intermediary platforms from the overarching measure of the digital economy.

Criterion 2: Ability to generate key measures of interest.

To what extent does it allow for the generation the following measures?

- Gross value-added.
- Employment.
- Worker earnings.
- International trade measures, including imports and exports of services and goods.
- Business demography measures, including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites.

Based on the AI sector and Cyber Security sector studies, this approach has allowed the measurement of GVA, employment, import and exports, and business demography measures.

What other measures of interest can be generated from this definition (e.g. valuation of tech companies, number of unicorns, VC investment in the tech sector, FDI in the tech sector, growth indicators beyond revenues and employees)?

In the AI sector study, this approach allowed for estimation of investment in the sector disaggregated by stage (e.g. venture, seed, et cetera).

Criterion 3: Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. Quantum).

To what extent does it allow for measurement of the following subsectors?

- Data subsector (including data processing, hosting and infrastructure).
- Artificial intelligence subsector.
- Cyber-security subsector.
- Software development.
- ICT manufacturing (including semiconductors).

- ICT trade.
- Telecommunications.
- Digital services (e.g. FinTech, EdTech, HealthTech)

This approach involves a data-driven process that reads various non-proprietary sources to identify subsectors based on several sources of firm-level data. It therefore allows for the measurement of specific subsectors as it is possible to infer these subsectors from the firm-level data available.

To what extent does it allow for the measurement of other subsectors not listed above, and at what level of granularity?

This approach involves a data-driven process to identify subsectors based on several sources of firm-level data. It therefore allows for the measurement of subsectors not listed above insofar as it is possible to infer these subsectors from the firm-level data available.

To what extent does it allow for the identification and incorporation of future emerging technologies as subsectors when needed (e.g. Quantum)?

This approach allows for the identification and incorporation of future emerging technologies as subsectors when needed.

Criterion 4: Translatability to Standard Industrial Classification.

To what extent is it translatable to SIC codes?

As implemented in the AI sector study, each company identified can be assigned to a single industry which is derived from and can be mapped back to SIC codes.

How easily is it translatable to SIC codes (i.e. what is required to translate the definition to SIC codes)?

The mapping to SIC codes can be done using automated models to match the web results with official data.

Criterion 5: Ability to generate comparisons over time.

To what extent does it allow for measurement of trends over time, including historical comparison?

This approach allows for measurement of trends over time, although this would require a substantial data collection effort for each point in the time series. Importantly, however, collecting historical data would require web mining of archived firm-level text data and is unlikely to be feasible.

At what level of temporal granularity does it allow for measurement of time trends?

In principle, this approach allows for measurement of time trends at any given level of granularity. However, it requires a substantial data collection effort for each point in a time series. It is therefore unlikely that it would be feasible to measure time trends more frequently than on an annual basis.

Criterion 6: Availability, timeliness and statistical quality of relevant data sources.

To what extent is the data required to operationalise the definition currently, available? If it is not available, can it be collected?

The approach employs a combination of AI driven web intelligence, and collation of company data from numerous open sources including Companies House.

To what extent is the data required to operationalise the definition available from ONS statistics?

The Companies House data used to operationalise this definition is available through ONS.

To what extent does the data required to operationalise the definition require subscriptions to proprietary data sources?

The data required to operationalise the definition does not necessarily require subscriptions to proprietary data sources.

How recent is the available data required to operationalise the definition?

Data for the deep web reading would be as recent as the last update to the company websites, while the other sources have different update frequencies.

To what extent is the data required to operationalise the definition nationally representative, regionally representative, representative at the sectoral level, representative at the sub-sectoral level?

The approach could have a low representativeness in those economic sectors where companies' online presence tends to be limited. In the case of digital economy companies, however, this data is highly likely to approximate a census. This would make it representative at the national, regional, sectoral and sub-sectoral levels.

Criterion 7: Ability to capture firm-level data for individual subsectors and indicators of market competition.

To what extent does it allow for measurement of firm-level data for the individual subsectors listed in question 3.1?

Firm-level data can be captured for the individual subsectors.

To what extent does it allow for measurement of the following indicators of market competition?

- Market share concentration.
- Business entry and exit.

Indicators of market competition can be measured using this approach. For instance, business entry/exit can be measured through the companies' websites (active/inactive) and official sources (active/inactive in Companies House).

Criterion 8: Translatability to international classification systems.

To what extent is it translatable to the Standard Occupational Classification?

In the AI sector and Cyber Security sector studies, the approach was able identify AI and Cyber Security job roles. A similar approach could be implemented for the digital economy, with these roles mapped back to SOC codes similarly to the way this is possible for SIC codes.

To what extent is it translatable to the Harmonised System.

This approach identifies the categories of goods produced by individual companies. However, the classification system does not map on to the Harmonised System.

E3 NIESR approach to measuring the UK digital economy using big data

This approach was developed by the National Institute of Economic and Social Research (NIESR) and Growth Intelligence in 2013 to define and measure the UK's digital economy.[33] As a starting point, it takes the UK Government's definition of the digital economy as comprising two sectoral groups: 'information and communications technology' (ICT) and 'digital content'. In defining each of these sectoral groups:

- It defines the ICT sector as comprising producers of ICT systems, ICT hardware, ICT software, and producers of related services around these products (e.g. sales, installation and maintenance of the products). This sector is characterised in terms of its outputs.
- With regard to the digital content sector, this approach defines it as comprising firms in which the only or principal outputs are digital products or services. For example, this definition excludes large parts of the architecture sector but includes firms specialising in CAD and technical drawing. Similarly, the definition excludes supermarkets but includes retailers whose principal offering is digital (such as retailers of digital music or audio-visual content).

Importantly, it is not possible to operationalise this definition using Standard Industrial Classification (SIC) codes. Even the most detailed level of SIC codes does not provide

[33] Nathan, M., Rosso, A., Gatten, T., Majmudar, P., & Mitchell, A. (2013). *Measuring the UK's Digital Economy with Big Data*. London: National Institute of Economic and Social Research.

adequate information on the nature of products, services, production techniques or distribution systems within the sector that is needed to operationalise such a definition.

To define and measure the digital economy, this approach draws on a proprietary Growth Intelligence dataset (it should be noted that the company that produced this dataset has since been dissolved and the dataset is no longer available). This dataset is comprised of all active companies in the UK registered at Companies House. The Companies House register dataset provides information on companies' start dates, directors and shareholders, full registered addresses and balance sheets. To construct the Growth Intelligence dataset, the Companies House register dataset is matched to web-mined data and data from other public resources (including Companies House filings, patents, text from company websites and text from press coverage of the company). Quantitative text analysis techniques are used to classify companies according to a fine-grained, multi-dimensional classification system by categories including 'sector context', 'product type', 'sales process' and 'client type'. These codings are probabilistic. In the NIESR/ Growth Intelligence UK digital economy study, the 'most likely' categories are used to allocate companies to sectors and products.

The NIESR/ Growth Intelligence approach is based on an assumption that the SIC codes used in the UK Government's statistical definition of the digital economy have some correspondence to the 'true' digital economy but do not fully capture it. It therefore uses these SIC codes as the starting point to develop a new definition and measurement of the sector. The approach involves the following process:

1. All companies in the Companies House register that have digital economy SIC codes are mapped to their corresponding Growth Intelligence sector and product. This mapping provides an initial 'cut' of Growth Intelligence sectors and products that are relevant for the digital economy.

2. A 'threshold rule' is used to exclude sparse sector and product groups from the analysis. Sparse groups are defined as those with a less an 0.2% share of the Growth Intelligence sector/product within the overall digital economy as conventionally defined. This step provides an initial list of 21 Growth Intelligence sectors and 16 product groups that comprise the 'digital economy'.

3. The Growth Intelligence digital economy sector list is manually edited. This step is necessary for two reasons. The first is the misclassification observed when companies select their own SIC codes. The second is the fact that 4-digit SIC codes are a more detailed classification system than Growth Intelligence sector codes. Thus, some SIC categories are coded into broader Growth Intelligence sectors which are irrelevant to the digital economy. This step follows a number of rules:

4. Sector groups where the only or principal output is non-digital were excluded. For instance, the following were excluded: 'oil and energy', 'apparel and fashion', 'non-profit organization management', 'utilities' and 'wholesale'.

5. Sectors with companies that offer both digital and non-digital outputs, such as 'architecture' and 'defence/ space' are retained in the list and parsed later (in step 6).

6. Sectors that were not included in the initial shortlist, but where the principal output is likely digital, like ‘online publishing’, ‘e-learning’ or ‘computer network security’ are added to the sector list.

7. The product list is manually edited in the same manner. From the 16 products selected on the base of the threshold, several are excluded such as ‘care or maintenance’ and ‘clothing’. As in the previous step, some products/services are added that were not in the shortlist but describe digital outputs, like ‘digital media’, ‘media delivery’ and ‘software desktop or server’.

8. A manual precision check is conducted using 5-digit SIC codes. This involves cross-tabulating Growth Intelligence sector and product codes against 5-digit SIC codes. This process revealed two sets of companies that are not clearly digital by the selection rules of this methodology: ‘Activities of head offices’ (SIC 70100) and ‘Financial intermediation not elsewhere classified’ (SIC 64999). These are removed from the definition.

9. In order to be classified as digital, a company must be classified as both belonging to a digital sector and producing a digital product as defined in the previous steps. Thus, the final step is to construct a set of companies by ‘sector-product’, which consists of companies in digital sectors whose principal activity is also digital.

It is noteworthy that the manual filtering steps outlined above (steps 3-5) have a substantial effect on the definition and measurement of the digital economy. Without these steps 25.5 percent of the companies in the sample would be classified as digital, compared to 14.4 percent when the manual filtering steps are implemented.

When applied to data from 2012, this approach finds the digital economy to be substantially larger than the UK Government’s conventional SIC-based definition would suggest. The study found the digital economy to comprise almost 270,000 active companies in the UK (14.4% of all UK companies) compared to 167,000 companies (10.0%) when the Government’s conventional definition is used. In particular, the study finds that the purely SIC-based definition of the digital economy missed out a large number of companies in business and domestic software, architectural activities, engineering, and engineering-related scientific and technical consulting.

Criterion 1: Ability to generate an overarching measure of the value of the UK digital economy.

To what extent does it allow for measurement of the economic value of the core digital sector (OECD definition as used by the UK Government)?

The findings of the study conducted using this approach demonstrate that it represents a more comprehensive definition of the digital economy than the core digital sector definition used by the UK Government. The study found the digital economy to comprise almost 270,000 active companies in the UK (14.4% of all UK companies) compared to 167,000 companies (10.0%) when the government's conventional definition is used.

To what extent does it allow for the measurement of the economic value of digital-only service producers (e.g. digital-only banks, e-payment service providers, online only media outlets, streaming services, edtech platforms, healthtech platforms)?

This approach allows for a flexible, fine-grained classification of companies into subsectors and could be adapted to allow for the measurement of the economic value of digital-only service producers.

To what extent does it allow for measurement of the economic value of digital intermediary platforms, including both those which charge a fee those which are data/advertising driven?

This approach allows for a flexible, fine-grained classification of companies into subsectors and could be adapted to allow for the measurement of the economic value of digital intermediary platforms.

To what extent does it allow for measurement of wider digitalisation of the economy: economic units where production has been significantly enhanced by digital technology?

This approach is based on classifying companies according to their subsector and the goods/ services they produce. It does not involve the collection of data on digital inputs. Thus, it is unlikely to capture those economic units where production has been significantly enhanced by digital technology, but which are not in a digital sector or do not produce digital goods/ services.

To what extent does it allow for disaggregation of the economic activities listed in the above questions from the overarching measure of the digital economy?

An adapted version of this approach would allow for disaggregation of digital-only service producers and digital intermediary platforms from the overarching measure of the digital economy.

Criterion 2: Ability to generate key measures of interest.

To what extent does it allow for the generation of the following measures?

- Gross value-added.

This approach does not allow for the measurement of gross value-added. However, an adapted version of this approach (linking the principal firm-level dataset with other datasets) could allow for the estimation of gross value-added.

- Employment.

This approach does not allow for the measurement of employment. However, an adapted version of this approach (linking the principal firm-level dataset with other datasets) could allow for the estimation of employment.

- Worker earnings.

This approach does not allow for the measurement of worker earnings. However, an adapted version of this approach (linking the principal firm-level dataset with other datasets) could allow for the estimation of worker earnings.

- International trade measures, including imports and exports of services and goods.

This approach does not allow for the measurement of international trade measures.

- Business demography measures, including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites.

This approach allows for the measurement of business demography measures, including number of businesses and regional distribution of business sites. An adapted version of this approach (linking the principal firm-level dataset with other datasets) could allow for the estimation of other business demography measures, such as size of businesses by annual turnover and size of businesses by employees.

What other measures of interest can be generated from this definition (e.g. valuation of tech companies, number of unicorns, VC investment in the tech sector, FDI in the tech sector, growth indicators beyond revenues and employees)?

This approach can be used to measure the goods and services produced by individual firms and in subsectors.

Criterion 3: Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. Quantum).

To what extent does it allow for measurement of the following subsectors?

- Data subsector (including data processing, hosting and infrastructure).

This approach does not allow for the measurement of the data subsector. However, it allows for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to do so.

- Artificial intelligence subsector.

This approach does not allow for the measurement of the artificial intelligence subsector. However, it allows for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to do so.

- Cyber-security subsector.

This approach does not allow for the measurement of the cyber-security subsector. However, it allows for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to do so.

- Software development.

This approach does not allow for the measurement of the software development subsector. However, it allows for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to do so.

- ICT manufacturing (including semiconductors).

This approach does not allow for the measurement of the ICT manufacturing subsector. However, it allows for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to do so.

- ICT trade.

This approach allows for the measurement of the ICT trade subsector, including disaggregation into wholesale and retail trade.

- Telecommunications.

This approach allows for the measurement of the telecommunications subsector.

- Digital services (e.g. FinTech, EdTech, HealthTech)

This approach does not allow for the measurement of the digital services subsector. However, it allows for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to do so.

To what extent does it allow for the measurement of other subsectors not listed above, and at what level of granularity?

This approach involves a data-driven process to identify subsectors based on several sources of firm-level data. It therefore allows for the measurement of subsectors not listed above insofar as it is possible to infer these subsectors from the firm-level data available.

To what extent does it allow for the identification and incorporation of future emerging technologies as subsectors when needed (e.g. Quantum)?

This approach allows for the identification and incorporation of future emerging technologies as subsectors when needed.

Criterion 4: Translatability to Standard Industrial Classification.

To what extent is it translatable to SIC codes?

This approach is fully mappable to SIC codes.

How easily is it translatable to SIC codes (i.e. what is required to translate the definition to SIC codes)?

This approach is based on mapping digital economy SIC codes to the Growth Intelligence sectoral classification. Thus, translating the firm-level data to SIC codes does not require additional effort.

Criterion 5: Ability to generate comparisons over time.

To what extent does it allow for measurement of trends over time, including historical comparison?

This approach allows for measurement of trends over time, although this would require a substantial data collection effort for each point in the time series. Importantly, however, collecting historical data would require web mining of archived firm-level text data and is unlikely to be feasible.

At what level of temporal granularity does it allow for measurement of time trends?

In principle, this approach allows for measurement of time trends at any given level of granularity. However, it requires a substantial data collection effort for each point in a time series. It is therefore unlikely that it would be feasible to measure time trends more frequently than on an annual basis.

Criterion 6: Availability, timeliness and statistical quality of relevant data sources.

To what extent is the data required to operationalise the definition currently, available? If it is not available, can it be collected?

The Companies House data used to operationalise this definition is readily available. However, the construction of the firm-level dataset requires original data collection using web mining techniques and data linking.

To what extent is the data required to operationalise the definition available from ONS statistics?

The Companies House data used to operationalise this definition is available through ONS.

To what extent does the data required to operationalise the definition require subscriptions to proprietary data sources?

The data required to operationalise the definition does not necessarily require subscriptions to proprietary data sources.

How recent is the available data required to operationalise the definition?

The data mined from the web would typically be several months old, given that the data collection process requires several months. However, it should be noted that some of the company website text that is collected will be significantly older than this. The linked data sources used are typically updated on an annual basis.

To what extent is the data required to operationalise the definition nationally representative, regionally representative, representative at the sectoral level, representative at the sub-sectoral level?

The comprehensive firm-level dataset used to operationalise this definition draws largely on web-mined data from company websites. This affects the representativeness of the data in those economic sectors where companies' online presence tends to be limited. In the case of digital economy companies, however, this data is highly likely to approximate a census. This would make it representative at the national, regional, sectoral and sub-sectoral levels. One caveat is that micro-sized firms are likely underrepresented in this dataset. If such an approach were to be adopted, it would be advisable to conduct empirical analyses to validate the representativeness of the dataset.

Criterion 7: Ability to capture firm-level data for individual subsectors and indicators of market competition.

To what extent does it allow for measurement of firm-level data for the individual subsectors listed in question 3.1?

This approach uses a firm-level dataset to measure the digital economy. The subsectoral classifications in this dataset do not include all of the individual subsectors listed in question 3.1. However, this

approach is designed to allow for a flexible, fine-grained classification of companies into subsectors and therefore could be adapted to allow for measurement of firm-level data for these subsectors.

To what extent does it allow for measurement of the following indicators of market competition?

- Market share concentration.

This approach does not allow for the measurement of market share concentration. However, an adapted version of this approach (linking the principal firm-level dataset with other datasets) could allow for the estimation of market share concentration.

- Business entry and exit.

This approach allows for the measurement of business entry and exit.

Criterion 8: Translatability to international classification systems.

To what extent is it translatable to the Standard Occupational Classification?

This approach is not translatable to the Standard Occupational Classification.

To what extent is it translatable to the Harmonised System.

This approach identifies the categories of goods produced by individual companies. However, the classification system employed is driven by the web-mined data collected on individual companies and does not map on to the Harmonised System. It would therefore be challenging to translate the product information collected by this approach to the Harmonised System, unless the approach to categorisation is redesigned to make the product classification map on to these categories. It should be noted that this change to the methodology could be technically challenging and may degrade the quality of data collected on goods produced by individual companies.

E4 Dynamic mapping approach to defining and measuring economic sectors

This approach was developed by NESTA in 2013 to define and measure the UK's creative industries[34] and applied to the UK's information economy industries in 2015.[35] For simplicity, it is described below as it could be applied to the digital economy. The approach involves the following:

1. A list of 'digital occupations' is developed based on the UK's Standard Occupational Classification (SOC). This involves developing a set of theoretically grounded criteria that describe the characteristics of digital employment. All SOC occupations are then

[34] Bakhshi, H., Freeman, A., & Higgs, P. (2013). A dynamic mapping of the UK's creative industries. London: NESTA/Arc Centre of Excellence for Creative Industries and Innovation.
[35] Spilsbury, M. (2015). Dynamic Mapping of the Information Economy Industries. London: NESTA/Tech UK.

scored against these criteria. The criteria are typically derived from a conceptual definition of the sector being measured.

2. The above-described list of occupations is used to calculate the 'digital intensity' of each economic subsector in the UK at the level of 4-digit Standard Industrial Classification (SIC) codes. The digital intensity of an economic subsector is defined as the proportion of workers employed in that sector who are in digital occupations.

3. A threshold is defined for the level of digital intensity at which a subsector (4-digit SIC code) is included/excluded from the definition of the digital economy. Various methods for determining this threshold have been employed, some of which are data-driven[36] and others which rely on the use of pre-existing UK Government sectoral definitions as a benchmark.[37]

Importantly, a key limitation of the 2015 NESTA/ Tech UK study that applied this approach to the UK Information Economy was that the SOC 2010 (the version in use at the time of the study) failed to capture many digital occupations, particularly in the field of data analytics.[38] The current version of the UK's SOC (SOC 2020) better captures digital occupations and could provide the basis for an improved application of this approach to the digital economy.

It is also important to note that this approach relies entirely on employment intensity as a proxy measure for identifying digital sectors. The extent to which this introduces bias into the subsector selection process is an open empirical question, which may merit further examination. Research published by the OECD[39] suggests that measuring the digital intensity of economic sectors using different proxy measures (e.g. investment in ICT equipment, investment in software, share of turnover from online sales, employment of ICT specialists) is likely to yield different results. Further research could explore the utility of updating the NESTA methodology by using additional proxy indicators in the subsector selection process. This would result in a more multidimensional approach to defining and measuring the digital economy.

Criterion 1: Ability to generate an overarching measure of the value of the UK digital economy.

To what extent does it allow for measurement of the economic value of the core digital sector (OECD definition as used by the UK Government)?

Such an approach would employ the existing UK Government definition of the digital sector as a benchmark for the level of digital intensity required for an economic subsector (4-digit SIC code) to be included/ excluded in the digital economy. Additionally, like the existing UK Government definition, this approach is based on 4-digit SIC codes. The resulting measure is therefore likely to be broadly

[36] Bakhshi, H., Freeman, A., & Higgs, P. (2013). A dynamic mapping of the UK's creative industries. London: NESTA/Arc Centre of Excellence for Creative Industries and Innovation.
[37] Spilsbury, M. (2015). Dynamic Mapping of the Information Economy Industries. London: NESTA/Tech UK.
[38] *Ibid.*
[39] Calvino, F., Criscuolo, C., Marcolin, L., & Squicciarini, M. (2018). A taxonomy of digital intensive sectors. Paris: OECD.

comparable to the existing UK Government definition of the core digital sector, both conceptually and in terms of measurement.

To what extent does it allow for the measurement of the economic value of digital-only service producers (e.g. digital-only banks, e-payment service providers, online only media outlets, streaming services, EdTech platforms, HealthTech platforms)?

This approach would likely only capture the economic value of digital-only service producers where such producers are dominant in a particular economic subsector (4-digit SIC code). This is because each economic subsector (4-digit SIC code) is treated as a single unit, without regard for the variation between firms within the sector. For example, if a EdTech or HealthTech platform are respectively categorised as falling under an educational or healthcare services SIC code, they would be unlikely to be captured by this approach.

To what extent does it allow for measurement of the economic value of digital intermediary platforms, including both those which charge a fee those which are data/advertising driven?

This approach would likely only capture the economic value of digital-intermediary platforms where such platforms are dominant in a particular economic subsector (4-digit SIC code). For example, this would very likely exclude platforms such as Booking.com and OpenTable which are not classified under ICT or media SIC codes.

To what extent does it allow for measurement of wider digitalisation of the economy: economic units where production has been significantly enhanced by digital technology?

This approach would only capture economic units where production has been significantly enhanced by digital technology if these units belong to economic subsectors (4-digit SIC code) where digitalisation is dominant.

To what extent does it allow for disaggregation of the economic activities listed in the above questions from the overarching measure of the digital economy?

This approach does not allow for disaggregation of the economic activities listed in the above questions from the overarching measure of the digital economy.

Criterion 2: Ability to generate key measures of interest.

To what extent does it allow for the generation the following measures?

- Gross value-added.

This approach allows for the measurement of gross value-added.

- Employment.

This approach allows for the measurement of employment.

- Worker earnings.

This approach allows for the measurement of worker earnings.

- International trade measures, including imports and exports of services and goods.

This approach allows for the calculation of estimates for imports and exports of digital services and digital goods.

- Business demography measures, including number of businesses, size of businesses by annual turnover, size of businesses by employees, regional distribution of business sites.

This approach allows for the measurement of business demography.

What other measures of interest can be generated from this definition (e.g. valuation of tech companies, number of unicorns, VC investment in the tech sector, FDI in the tech sector, growth indicators beyond revenues and employees)?

This approach can be used to measure any indicator for which the ONS provides nationally representative data at the level of 4-digit SIC codes. In addition, it can measure the number of digital workers across the economy, including digital workers employed outside the digital sector.

Criterion 3: Ability to capture the contribution of individual subsectors (e.g. AI, cyber-security), including the identification and integration of future emerging technologies as subsectors (e.g. Quantum).

To what extent does it allow for measurement of the following subsectors?

- Data subsector (including data processing, hosting and infrastructure).

This approach allows for the measurement of the data processing, hosting and infrastructure subsector (SIC code 6311).

- Artificial intelligence subsector.

This approach does not allow for the measurement of the artificial intelligence subsector.

- Cyber-security subsector.

This approach does not allow for the measurement of the cyber-security subsector.

- Software development.

This approach allows for the measurement of the software development subsector (SIC code 6201).

- ICT manufacturing (including semiconductors).

This approach allows for the measurement of the ICT manufacturing subsector (various SIC codes in SIC divisions 26 and 27).

- ICT trade.

This approach allows for the measurement of the ICT trade subsector (various SIC codes in SIC divisions 46 and 47).

- Telecommunications.

This approach allows for the measurement of the telecommunications subsector (SIC division 61).

- Digital services (e.g. FinTech, EdTech, HealthTech).

This approach does not allow for the measurement of the digital services as subsectors.

To what extent does it allow for the measurement of other subsectors not listed above, and at what level of granularity?

Other sectors of interest that can be measured using this approach include 'web portals' and 'computer consultancy activities'.

To what extent does it allow for the identification and incorporation of future emerging technologies as subsectors when needed (e.g. Quantum)?

This approach does not allow for the identification and incorporation of future emerging technologies as subsectors when needed.

Criterion 4: Translatability to Standard Industrial Classification.

To what extent is it translatable to SIC codes?

This approach is based on SIC codes.

How easily is it translatable to SIC codes (i.e. what is required to translate the definition to SIC codes)?

This approach is based on SIC codes.

Criterion 5: Ability to generate comparisons over time.

To what extent does it allow for measurement of trends over time, including historical comparison?

This approach is based on SIC codes and can therefore employ ONS statistics for measurement. Thus, it allows for measurement of trends over time, including historical comparison.

At what level of temporal granularity does it allow for measurement of time trends?

Most of the relevant data sources are collected quarterly.

Criterion 6: Availability, timeliness and statistical quality of relevant data sources.

To what extent is the data required to operationalise the definition currently, available? If it is not available, can it be collected?

The data required to operationalise the definition is largely available from ONS statistics. However, operationalisation of the definition requires an exercise to identify which SOC codes represent 'digital occupations'. Conducting this exercise in a rigorous manner would require the collection of expert judgements regarding whether individual SOC occupations fulfil the definition of a 'digital occupation'.

To what extent is the data required to operationalise the definition available from ONS statistics?

This approach is designed to be operationalised using ONS statistics.

To what extent does the data required to operationalise the definition require subscriptions to proprietary data sources?

Proprietary data sources are not required to operationalise this definition.

How recent is the available data required to operationalise the definition?

Most of the relevant data sources are published annually.

To what extent is the data required to operationalise the definition nationally representative, regionally representative, representative at the sectoral level, representative at the sub-sectoral level?

The data required to operationalise the definition is nationally, regionally and sectorally representative. At the sub-sectoral level, representativeness depends on the size of the sector. Data for smaller subsectors, such as AI, is unlikely to be representative.

Criterion 7: Ability to capture firm-level data for individual subsectors and indicators of market competition.

To what extent does it allow for measurement of firm-level data for the individual subsectors listed in question 3a?

This approach does not allow for the measurement of firm-level data.

To what extent does it allow for measurement of the following indicators of market competition?

- Market share concentration.

This approach allows for estimation of market share concentration in the digital economy, as well as selected subsectors (see response to question 3a for description of subsectors). High-level administrative statistics on companies' finances at the 4-digit SIC level are available from ONS and can be used to derive proxy measures of sectoral/ subsectoral market share concentration.

- Business entry and exit.

This approach allows for measurement of business entry and exit in the digital economy, as well as selected subsectors (see response to question 3a for description of subsectors). This data is available from administrative statistics on business birth and death rates at the 4-digit SIC level available from ONS.

Criterion 8: Translatability to international classification systems.

To what extent is it translatable to the Standard Occupational Classification?

This approach is based on SOC codes.

To what extent is it translatable to the Harmonised System.

This approach is based on 4-digit SIC codes. The UK Standard Industrial Classification (SIC) 2007 is the national version of "The Statistical Classification of Economic Activities in the European Community (NACE) Rev. 2" and as such must be identical to NACE Rev 2, down to and including the fourth digit of the system. Correlations between NACE Rev 2 and other international classifications are produced and held by Eurostat. Due to fundamental differences between the two systems, Eurostat advises that it is not possible to correlate HS comprehensively with NACE Rev 2.

Appendix F Stakeholder Workshop 2 Synthesis Note

F1 Introduction

The degree of digitalisation in the economy has increased dramatically over the past decade, leading to an evolution in how the 'digital economy' is defined. Earlier conceptualisations of the digital economy employed a 'bottom-up' approach, defining it as a specific set of economic activities that produce ICT goods and digital services. In contrast, more recent definitions tend to adopt a 'top-down' approach, which additionally includes economic activity enabled by the use of ICT goods and digital services, thus reflecting the spread of digitalisation across the economy. In the UK, DCMS – and more recently DSIT – historically defined the digital sector using a bottom-up definition based on Standard Industrial Classification codes, which does not capture digitalisation across the whole economy nor allow for granular measurement of key industries/ technologies (e.g. AI or cyber-security). This approach no longer meets DSIT's policy needs.

In this context, DSIT has commissioned Technopolis to conduct a study on defining and measuring the UK digital economy. The objective of this study is to develop: 1) a revised definition of the digital economy which meets DSIT's policy needs; 2) a methodology to operationalise this definition. The study involves the following key elements:

1. Review and assessment of existing definitions of the digital economy to capture the extent to which they meet UK stakeholder needs and international standards.
2. Stakeholder consultation to develop a rigorous, comprehensive and measurable definition that meets stakeholder and user needs. This involves holding two stakeholder workshops. The first workshop will introduce the study to stakeholders and solicit their feedback on the extent to which different approaches to defining and measuring the digital economy meet their policy needs and those of DSIT. The second workshop will be used to solicit feedback on the study findings and recommendations.
3. Design of a measurement methodology setting out how the new definition of the digital economy will be operationalised.

This workshop synthesis note presents a summary of the discussion from the second workshop.

The workshop was held virtually on 21 March 2024, and was attended by 21 people from eight organisations. The objective of the second workshop was to obtain stakeholder feedback on the proposed approach to defining and measuring the digital economy, with a view to the finalisation of the methodology note. This note presents the feedback gathered during the workshop discussion and through the online chat function, in addition to the live online polling conducted during the workshop.

F2 Overview of workshop discussion

F2.1 Proposed conceptual definition of the digital economy

The workshop began with a recap of the policy context in which the study is being conducted and a presentation of key takeaways from the discussion during the first workshop. The study team then presented the proposed definition of the digital economy (see Box Box 4 below).

Box 4: Proposed definition of the digital economy

All economic activity derived from the production of digital content, ICT goods, or ICT services.

- **Tier 1:** Economic activity derived from firms that are primarily involved in the production of digital content and/or ICT goods and/or ICT services.
- **Tier 2:** Economic activity derived from diversified firms producing digital content and/or ICT goods and/or ICT services, as part of a broader diversified product or service offer.

The study team highlighted that the proposed definition excludes 'digital adopters', such as online retailers of physical goods (see Box 5 below).

Box 5: Definition of 'digital adopters'

Digital adopters: Economic activity derived from firms reliant on digital inputs that do not produce digital content, ICT goods or ICT services.

- One workshop participant asked if the definitions would take account of market failures. The study team responded that the scope of work did not include measuring market failures. Rather, the objective is to measure the size of the digital economy and key indicators of business demography, including key subsectors.
- A DSIT representative noted that the exclusion of digital adopters from the definition excludes some ecommerce segments from the scope of the digital economy. The study team noted that a decision had been taken to exclude online retail of non-digital goods from the definition following initial consultation with the DSIT DEDP The rationale for this is that ecommerce falls outside the scope of DSIT's remit.
- A representative from DSIT asked about the motivation for measuring the digital economy. A member of the study team stated that the digital economy was an important part of DSIT's remit, and that DSIT aims to measure the digital economy in a way that aligns with DSIT's policy needs and priorities. In particular, the existing definition does not allow for granular analysis of subsectors. A representative from DSIT highlighted that the primary need was for a definition that reflects current DSIT policy priorities and a methodology that can identify those areas of policy interest within the broader digital

economy. If DSIT only wanted to measure the overall size of the digital economy then the OECD SUT approach would be as good as any, but the Department wants to be able to assess cyber, data, telecoms, software, AI, et cetera, on a comparable basis and to know how much they contribute to the economy overall, without double counting.

- A representative from Innovate UK noted the need to get away from double counting necessarily being considered a bad thing. Double counting may be necessary in this space, as some products will sit in more than one area.
- A member of the study team noted that double counting is an issue with regard to national accounting (e.g. how big the digital economy is in the context of GDP) but not necessarily for other purposes. The appropriate digital economy definition will be at least partly determined by purpose.

F2.2 Key characteristics of the proposed definition

The study team presented the three key characteristics of the proposed definition and the rationale underlying each of these (see Figure 13 below).

Figure 13: Key characteristics of proposed definition.

Product based classification

Tiered definition

Dynamic sub-sectoral classification

F2.2.1 Rationale for product-based classification

The rationale for a product-based classification is presented in the boxes below.

In this proposed definition, the principal basis for classifying firms is their product output, rather than the economic sector in which they are categorised.
Allows for identification of digital firms operating within non-digital sector industries (e.g companies specialising in the development of custom software of the engineering sector).
Allows for identification of diversified firms that are classified under non-digital sector industries but are significantly engaged in digital activities or the production of digital goods and services (e.g Rolls Royce).
SIC codes are not available for a large number of companies.

F2.2.2 Rationale for tiered definition

The rationale for the tier definition is presented in the boxes below.

A tiered approach provides a solution to DSIT's policy and data needs. To fulfil these needs, the definition must both be comprehensive enough to generate measures of the overarching value of the digital economy and targeted enough to generate measures of specialised subsectors.
Tier 1 measures the 'core' digital economy. Conceptually similar to DSIT's existing definition of the digital sector but more precise as it does not rely on sectoral classifications.
Tier 2 allows for generation of overarching measures of the digital economy.
Distinguishing between digital developers (Tiers 1 and 2) and digital adopters. The latter are excluded from the proposed digital economy definition.

Proposed tiers

The proposed tiers are presented in the box below.

Tier 1: Economic activity derived from firms that are primarily involved in the production of digital content, ICT goods and/ or services. Conceptually maps to 'core; digital economy (OECD)/ digital sector (DSIT) **Tier 2: Economic activity derived from diversified firms producing digital content, ICT goods and. Or services as part of a broader diversified product or service offer.** Allows for generation of overarching measures of digital economy.

The study team highlighted that digital adopters were excluded from the boundary of the digital economy.

> **Digital adopters:** Economic activity derived from firms reliant on digital inputs that do not produce digital content, ICT goods and/or services.
>
> Represents adopters (as opposed to developers) of digital content/technologies.
>
> Can capture key sectors of policy interest (FinTech, HealthTech, e-tailers) without overstretching boundary of digital economy

The study team presented firm-level examples to:

1. Illustrate the scope of each proposed tier (see Figure 14 below).
2. Highlight how the proposed definition of the digital economy captures the evolution of firms as they move up the value chain in terms of their involvement in the digital economy (see

3. Figure 15 below).

Figure 14: Firm-level examples of proposed tiers

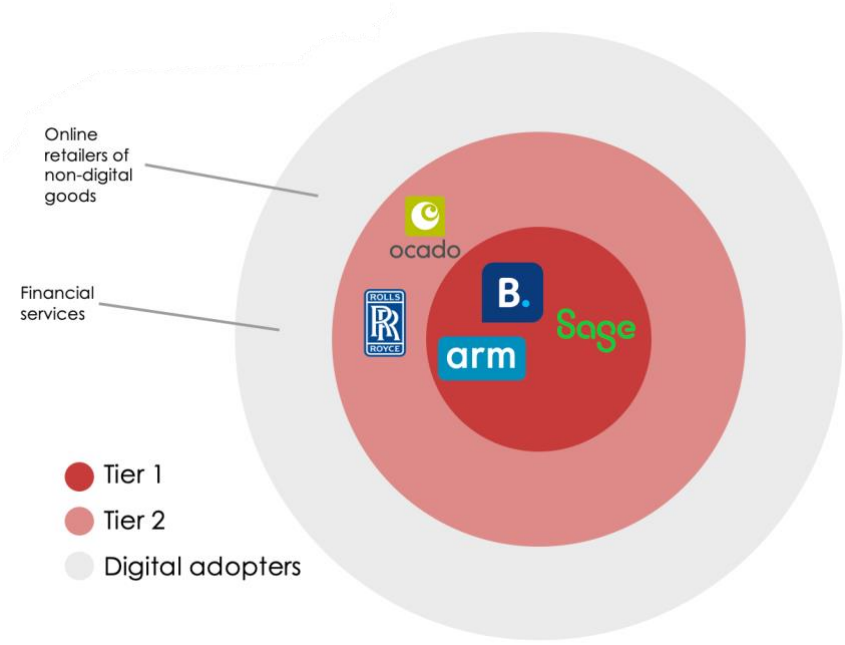
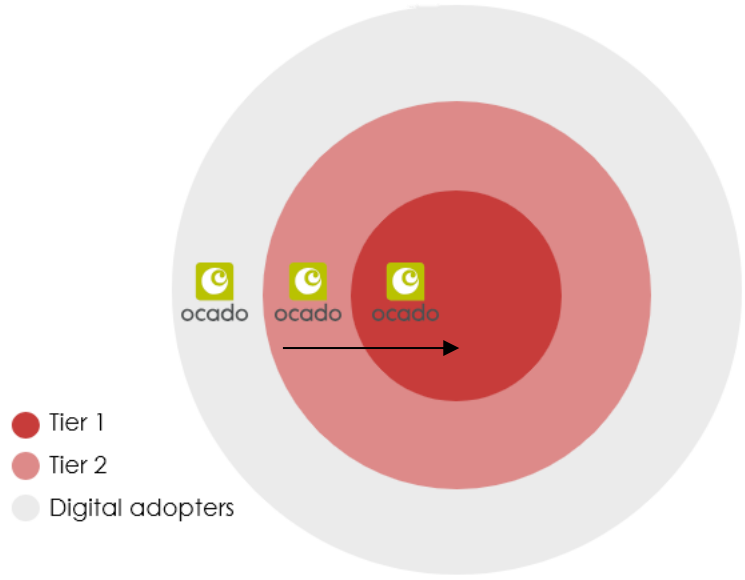


Figure 15: Capturing evolution of firms



- A DSIT representative highlighted that while Ocado is moving from food retail to warehouse fulfilment, Ocado was a digital adopter. A member of the study team highlighted that Ocado is a started as an adopter and then increased their automation and investment in AI and robotics. Therefore, it could be argued Ocado moved into Tier 2. Another member of the study team further noted that Ocado's activities have moved into processing the data from physical infrastructure and has built a team of data scientists to analyse the data for decision making.
- Several workshop participants noted that Ocado was an interesting example because it highlighted the challenge in setting a threshold, as many companies have increased their degree of the diversification.
- An Innovate UK representative asked if there would be any future work looking at tiering digital adopters? He noted that some are much closer to Tier 2 than others especially when looking at potential high-growth tech disrupters and that there is a lot of value in monitoring / measuring this category. A representative from DSIT responded that there are no specific plans / timelines on this as of right now, as the focus will be on operationalising this work (i.e. Tier 1 and 2). However, it is an area of interest for future work.
- A DSIT representative asked at what point a firm moves between tiers, noting that this would be quite nebulous and a matter of debate. A member of the study team noted that this would be discussed further in the next session of the workshop.
- A DSIT representative asked what tier neo-banks (i.e. digital-first financial companies that offer banking services) would fall under. A member of the study team noted that the OECD approach to defining the digital economy puts neo banks into a special category of the digital sector, which is measured and included in the digital economy. However, this was not the approach taken here. According to the proposed approach, neo-banks

would fall under digital adopters, but this could be discussed further during the next session of the workshop. Another member of the study team noted that neo-banks are essentially banks that use a lot of technology and therefore should be categorised as adopters.

- A DSIT representative noted a general trend towards more use of data and AI, with firms and sectors becoming more digitalised. In future, retailers like Tesco will become more data driven and perhaps move into tier two or be classified as digital adopters. The economy is changing and there is a risk that over time this approach will become unsuitable. The study team noted that the proposed definition may not be able to accommodate the evolving nature of the digital economy in the long term, but rather has been conceptualised for use in the short-to-medium term. This will be useful until there is generally a better understanding of the degree of digitalisation that will take place across the economy in future.

F2.2.3 Rationale for classification into subsectors

The rationale for the classification into subsectors is presented in the boxes below.

Firms included in the digital economy boundary are further classified into subsectors through a process that combines data-driven approaches and expert knowledge.
Ensures the availability of data on individual priority subsectors of policy relevance to DSIT.
This process can be repeated periodically (e.g every 3-5 years) to capture and measure emerging subsectors.

- The study team highlighted that to keep up with the evolution in the digital economy, the tiers in the proposed approach capture which firms are relevant, while a separate element of the definition captures subsectors. This is analogous to the approach of the Consumer Price Index, which measures prices paid by consumers based on a representative basket of goods and services that changes over time as consumer behaviour changes. Similarly, the subsectors identified as falling within the boundary of the digital economy will be updated periodically to reflect the digital economy's changing characteristics.
- A representative from Innovate UK noted that updating the subsectoral taxonomy every 3-5 years is not frequent enough to capture sectors at the frontier of digital technology.

Subsector definition

A digital economy subsector is defined as a segment of the digital economy characterised by:

- The production of a set of goods and/or services that distinguishes it from other clusters within the digital economy, and/or
- Utilisation of a set of technologies that distinguishes it from other clusters within the digital economy
- An Innovate UK representative noted that 'utilisation of' should be 'development of' if we are focusing on product developers rather than adopters. A DSIT representative noted that utilisation can bleed into adoption.
- A DSIT representative noted that the subsector definition should refer to producers developing the set of technologies rather than utilising them. A member of the study team stated that the wording of the definition would be modified to clarify this.

F3 Proposed Approach to Measurement

The study team presented the approach to operationalising the proposed definition and solicited feedback from the workshop participants (see Figure 16 below).

Figure 16: Process

The following section highlights the comments of the workshop participants on the proposed approach to measurement.

F3.1 Data collection and linking

- The study team highlighted that the three key data sources were web crawled data, the Companies House register and LinkedIn data. LinkedIn data involves collecting information on the nature of job roles in any given company and enables the identification of digital occupations. This helps to identify the level of digital intensity of any given firm.

F3.2 Identification of company products and services

- The study team highlighted that the key data sources for this would be web crawled data. It was noted that the ONS PRODCOM and Annual Survey of Goods and Services surveys could potentially be used as secondary data sources to identify the goods and services produced by firms. However, the study team noted that the sample size is too small for this to be a key source of data for measurement.
- A workshop participant from ONS asked if any preliminary analysis had been done on the aforementioned two sources of data and what triangulation had been conducted between these data sources and web crawled data. A member of the study team responded that data analysis was not within the scope of the current project.

F3.3 Measurement of digital inputs

- The study team highlighted that firm R&D spending might be an interesting indicator to cross reference against some of the others. Another member of the study team further mentioned that there was a growing consensus in the US towards measuring AI and other frontier technologies by tracking the key researchers in these fields and their employment.
- A DSIT representative highlighted that a lot of firms have negative profit and this could cause problems for producing meaningful GVA.
- Another DSIT representative noted that patents are only obtained in certain types of industries and that this metric is biased against small firms. They highlighted the need to be aware of these biases and/or to use multiple metrics with diverging coverage.
- A representative from Innovate UK highlighted that patent data will be a weak indicator for measuring the digital economy because the speed of technological development is faster than the patenting process.

F3.4 Identifying digital products and services

- A DSIT representative asked about how to disentangle digital products that are embedded with services such as mobile phones. Members of the study team acknowledged the challenges related to bundled products. They highlighted that to

understand the products and services that companies offered, the proposed approach is deep reading of multiple data sources including firms' websites, media coverage and employment data. The language model used in this process infer the products and services using all of these data sources.

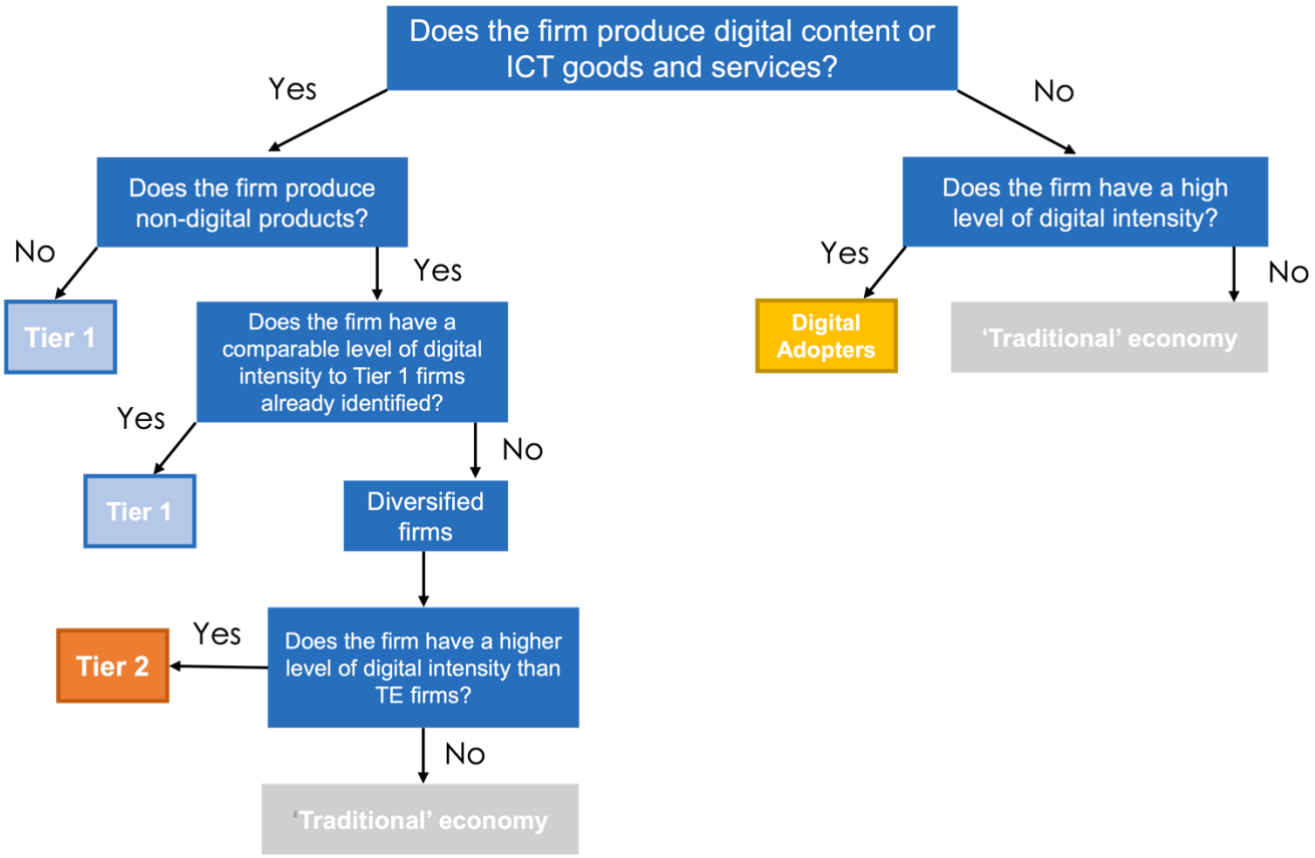
- A member of the study team noted that there is controversy around how to attribute the value from bundled products like tractors/software/data or phone/data services/apps. There are no easy answers. This issue is why looking at technology employment intensity is useful.
- A DSIT representative stated that in using web data "You always have some in that you want out and some out that you want in. It is not exact, but nor were the SIC. But we have the law of large numbers on our side."

F3.5 Classification of firms into tiers

The study team presented a decision tree for the classification of firms according to the proposed definition (see).

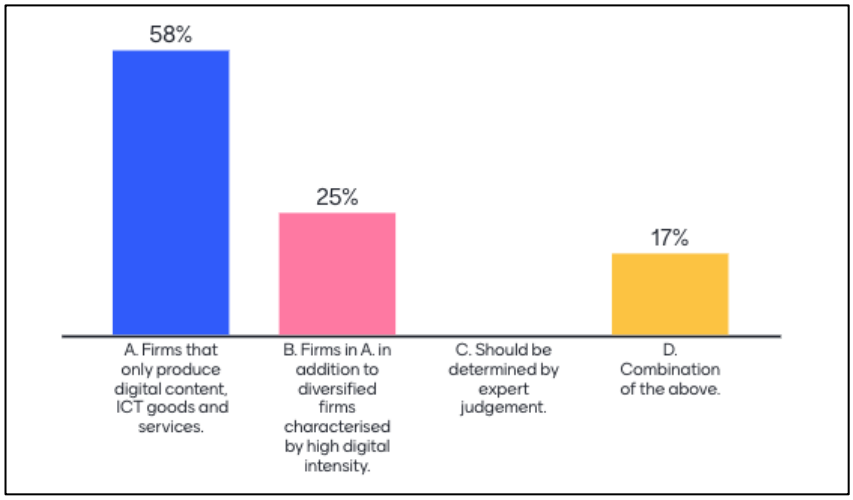
Figure 17 below).

Figure 17: Classification of firms into tiers



The study team conducted a poll asking workshop participants to select the most appropriate boundary for Tier 1 (see Figure 18 below for the poll results).

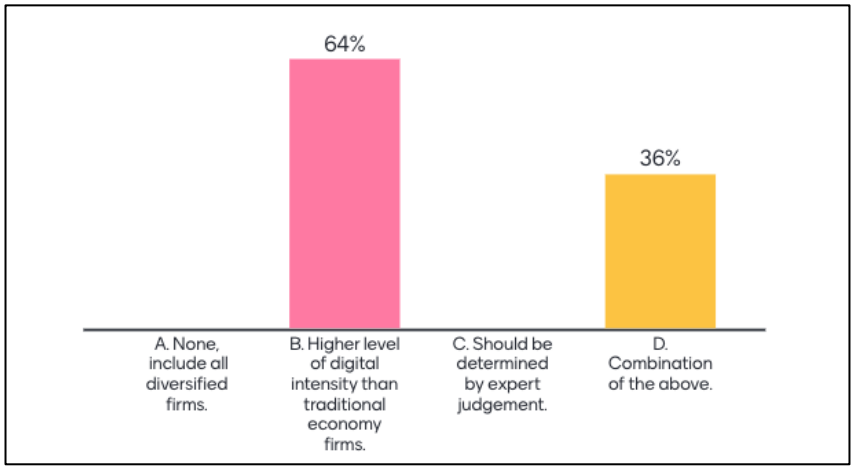
Figure 18: Workshop poll: "What is the most appropriate boundary for Tier 1?"



- An Innovate UK representative asked if the 'T' part of ICT was still relevant to digital economy, or should it be separated into a standalone telecoms definition? A member of the study team noted that it is included in all of the extant definitions reviewed for this study, including recent ones. Another member of the study team noted that it is an integral part of digital infrastructure without which the digital economy would not be possible and therefore should be included.

Workshop participants were asked "What is the most appropriate boundary for Tier 2?" (see Figure 19 below for the poll results).

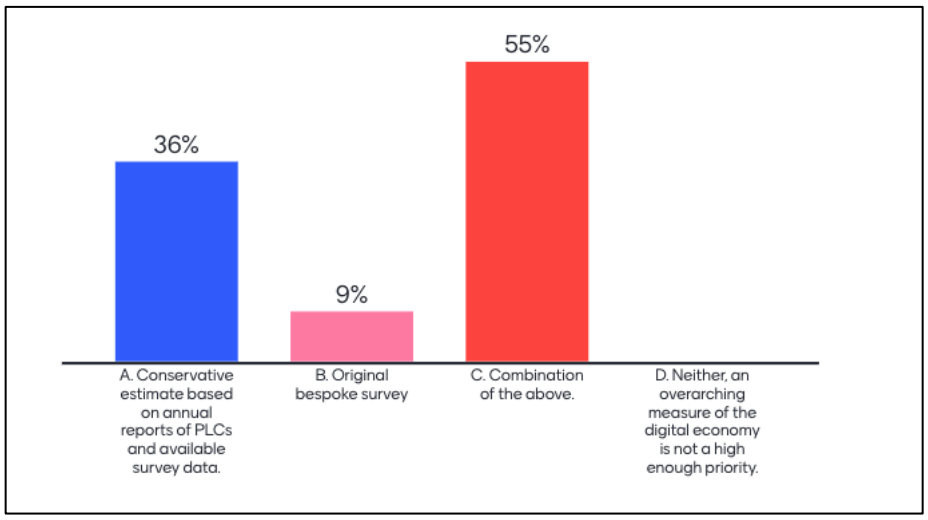
Figure 19: Workshop poll: "What is the most appropriate lower boundary for Tier 2?"



Workshop participants were asked “On what basis should a proportion of the Tier 2 companies be allocated to the digital economy?” (see

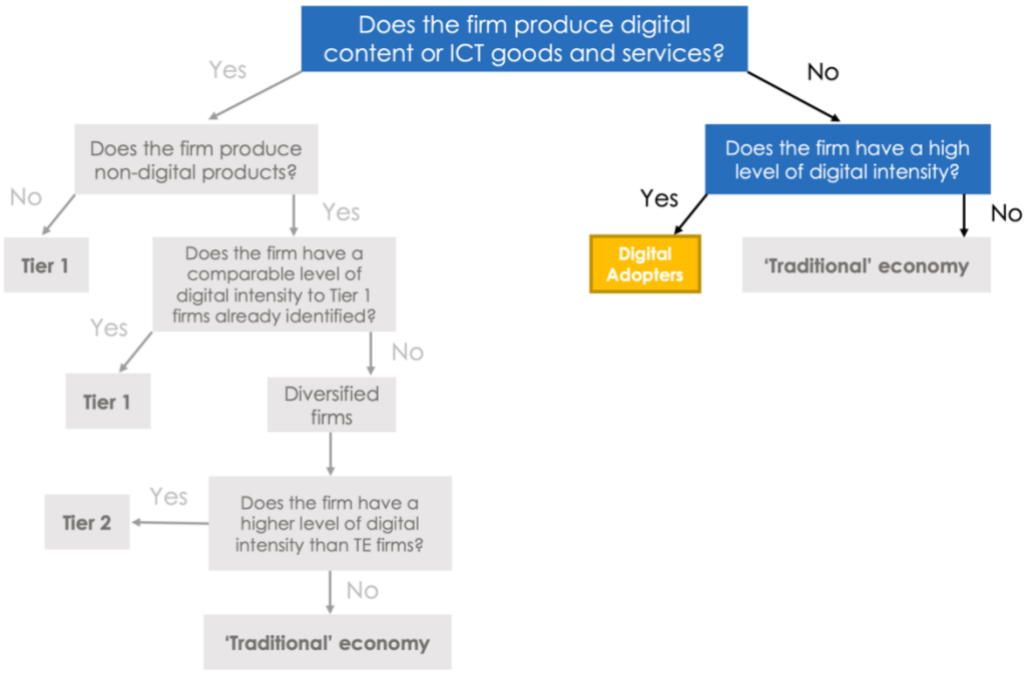
Figure 20 below for the poll results).

Figure 20: Workshop poll: “On what basis should a proportion of Tier 2 Companies be allocated to the digital economy?”



Workshop participants were asked about the level of digital intensity that should be required for inclusion as a digital adopter and how this should be measured (see the box in the lower right corner of Figure 21 below).

Figure 21: Classification of firms into tiers



F3.6 Classification of digital economy firms into subsectors

The study team noted that the classification of the digital economy firms into subsectors involves the development of a taxonomy, typically using a combination of data-driven approaches, stakeholder consultation and expert judgement. They further explained that language models can provide a first cut of subsectors, which can be refined and finalised through stakeholder consultation and expert judgement. A member of the study team noted the importance of effective stakeholder engagement in producing a taxonomy that meets user needs.

F4 Summary of workshop feedback and discussion points

- There is a need to effectively communicate the objectives of this study in the methodology, both in terms of what it is for and what it is not for.
- The recommendations should be clear on how methodological decisions was reached.
- There is a consensus that the product-based approach is good and a consensus on the use of employment data to measure digital intensity across tiers, combined with some information on innovation.
- The polls results suggest that stakeholders prefer a reasonably high threshold for defining Tier 2 that does not dilute the definition of the sector overall.
- There was a consensus that the operationalisation of the definition needs to be data driven, but with a human in the loop.
- A member of the study team shared a reference showing an example of a new approach to measuring the AI sector: https://iris.isr.umich.edu/tip/

F5 Workshop participants

The workshop was attended by 21 people from the following eight organisations:

- Department of Science, Innovation and Technology (9 participants)
- Glass.ai (1 participant)
- Office for National Statistics (1 participant)
- Technopolis (5 participants)
- TechSkills (1 participant)
- UKRI - Innovate UK (2 participants)
- University of Cambridge (1 participant)
- Welsh Government (1 participant)

This publication is available from: www.gov.uk/dsit

If you need a version of this document in a more accessible format, please email alt.formats@dsit.gov.uk. Please tell us what format you need. It will help us if you say what assistive technology you use.