ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH

Office for National Statistics

Measuring the Green Economy: exploring new definitions and testing new methods

Will McDowall and Hamid Nejadghorban

ESCoE Discussion Paper 2026-11

August 2026

ISSN 2515-4664

DISCUSSION PAPER

Measuring the Green Economy: exploring new definitions and testing new methods
Will McDowall and Hamid Nejadghorban
ESCoE Discussion Paper No. 2026-11
August 2026

Abstract

As the world confronts the challenges of multiple environmental crises – particularly the climate emergency – it is important for policymakers to be able to track the scale of economic activity related to the achievement of environmental goals. We highlight the shortcomings of existing definitions and typologies for measuring the 'environmental goods and service sector' and related environmental activities. We first set out a new approach to defining green economic activity, which aims to overcome some of the shortcomings of existing approaches. We then explore the potential of new web-scraping tools for supporting the estimation of green economic activity in the UK. We test the alignment between lists of firms identified as involved in low-carbon and renewable energy by an existing ONS survey and by a web-scraping tool (TheDataCity). Our analysis highlights both the limitations and the advantages of web-scraping tools and advocates their use only when coupled with sufficient relevant sector-specific expertise. We also provide illustrative estimates of the scale of green economic activity in the UK, adjusted for our new set of definitions, showing that the true scale of such activities may be at least 70% greater than is captured in the current ONS Environmental Goods and Services Sector estimates.

Keywords: green economy, environmental accounting, web scraping, environmental goods and services sector, official statistics, economic measurement

JEL classification: C81, E01, Q01, Q56, Q58

Will McDowall, University College London
w.mcdowall@ucl.ac.uk

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

Measuring the Green Economy: exploring new definitions and testing new methods

Will McDowall & Hamid Nejadghorban

August 2026

Abstract

As the world confronts the challenges of multiple environmental crises – particularly the climate emergency – it is important for policymakers to be able to track the scale of economic activity related to the achievement of environmental goals. We highlight the shortcomings of existing definitions and typologies for measuring the 'environmental goods and service sector' and related environmental activities. We first set out a new approach to defining green economic activity, which aims to overcome some of the shortcomings of existing approaches. We then explore the potential of new web-scraping tools for supporting the estimation of green economic activity in the UK. We test the alignment between lists of firms identified as involved in low-carbon and renewable energy by an existing ONS survey and by a web-scraping tool (The Data City). Our analysis highlights both the limitations and the advantages of web-scraping tools and advocates their use only when coupled with sufficient relevant sector-specific expertise. We also provide illustrative estimates of the scale of green economic activity in the UK, adjusted for our new set of definitions, showing that the true scale of such activities may be at least 70% greater than is captured in the current ONS Environmental Goods and Services Sector estimates.

1 Introduction

The world faces a critical few decades in the fight against global heating and to prevent the irreversible loss of biodiversity. Developing policies that meet this challenge is difficult. Environmental policies sometimes come into conflict with economic objectives, while often the new industries associated with meeting environmental goals are a source of new economic growth opportunities. More than ever, policymakers need good evidence on the scale of economic activity related to achieving environmental goals. This report—funded by the Office for National Statistics as part of the work programme of the Economic Statistics Centre of Excellence—explores how measures of green economic activity in the UK can be improved.

We first set out a new approach to defining green economic activity, which aims to overcome some of the shortcomings of existing approaches. We then explore the potential of new web-scraping tools for supporting the estimation of green economic activity in the UK. Specifically, we apply the tool developed by The Data City, and test the alignment between their list of firms involved in Net Zero with the ONS dataset describing the Low Carbon and Renewable Energy Economy (LCREE)[1]. In section 4, we provide some new estimates of green economic activity in sectors that have been overlooked by previous UK green economy data. In doing so we show that these are large relative to the size of existing estimates of the Environmental Goods and Services Sector (EGSS) in the UK.

Our work shows that current estimates of green economic activity in the UK need improvement. Importantly, significant improvements can be made using existing data sources, by re-organising the way in which green economic activities are defined and reported. We also found clear value in web-scraping tools, in particular for informing sampling strategies for surveys—but we also found evidence of limitations in terms of their accuracy.

2 A revised definition of green economic activity

2.1 The need for a revised approach to defining green economic activity in the UK[2]

In this report, we focus on measurement of economic activities related to the environment, and specifically to definitions and measurement of the 'environmental goods and services sector' (EGSS). In the UK, EGSS accounts are produced drawing on a mix of existing data (where standard industrial classification (SIC) codes make this possible, as with waste management), from a dedicated low-carbon and renewable energy economy (LCREE) survey, and from sector-specific information such as shares of renewable electricity in total power generation.

As our previous report (McDowall and Fuchs 2025)[3] made clear, the most widely used definitions of the environmental goods and service sector (EGSS)—including those applied in the UK—are confused. EGSS definitions have been agreed internationally, and are embedded in the System of Environmental Economic Accounting – Central Framework. International comparability is of central importance to producing useful statistics. But the flaws with the current approach mean that this framework, and the way it is implemented in the UK, both need a rethink – not least because the problems with the framework itself undermine international comparability. We briefly summarise the major problems here.

The first is that current measurement and reporting practice conflates differing types of goods and services in ways that make it impossible to understand the economic or environmental significance of trends in total EGSS activity. For example, declines in electric vehicle manufacturing as a share of total vehicle output would be bad news both economically and environmentally. In contrast, declining pollution abatement activity is often good economic and environmental news, since such declines follow a switch away from the activities that were causing pollution in the first place. Aggregating such activities into total EGSS output obscures underlying trends.

---

[1] Note that the scope of these lists differs somewhat, a point that we address in some detail in section 3

[2] United Nations (2014) System of EnvironmentalEconomic Accounting 2012—
Central Framework. ISBN: 987-92-1-161563-0. https://seea.un.org/sites/seea.un.org/files/seea_cf_final_en.pdf

[3] McDowall and Fuchs (2025). Measuring the Green Economy. ESCoE Working Paper https://www.escoe.ac.uk/publications/measuring-the-green-economy/

Second, the SEEA definitions focus on the environmental *purpose* of goods and services, but guidance (from SEEA and Eurostat) and practice (in statistics agencies) are frequently inconsistent about the application of this idea.

In the SEEA-CF 2012, solar panels are given as an example of goods that "*directly serve an environmental protection or resource management purpose and that have no use except for environmental protection or resource management*." But solar panels clearly have a use other than environmental protection or resource management – they generate power, and people buy them regardless of their environmental or resource management concerns. Recent proposed changes to the SEEA-CF would help mitigate the problem, by explicitly incorporating the idea of 'secondary' purpose[4]. However, these revisions do not tackle the core conceptual problem, which is that for many goods and services that are environmentally-relevant there are different perspectives on whether they have an environmental purpose or not. There are two results from this confusion around 'purpose'. The first is inconsistent application of the definition, in which different statistical agencies, studies or countries will produce mutually incomparable outputs. This is partly because, as acknowledged in the recent consultation for the SEEA-CF, the 'purpose' criterion often cannot be inferred. The recommendation in guidance is that purpose is inferred from the 'technical nature' of the good or service, i.e. based on the characteristics of the good or service in terms of the environmental pressures generated. The result is a confusion: purpose lingers as an aspirational criterion, but in practice it is not consistently assessed or applied. Instead, statistical agencies increasingly rely on 'indicative lists' of environmental goods and services. These facilitate international comparability, but are not based on the consistent and clear application of a coherent definition. In short, a purpose-based definition is not directly actionable and it does not result in consistent treatment of relevant activities. Until recently, a second problem with a 'purpose-based' definition has been that it typically resulted in the omission of many goods and services that are green (relative to a relevant alternative) but that have existed for decades. Such goods and services are rarely seen as having an environmental purpose, and so were typically excluded from 'indicative lists' that are attempting to operationalise a definition based on purpose. A prime example is the bicycle, which has been long been excluded from EU and UK EGSS. It makes little sense to include solar panels in the green economy but not bicycles. Both are actively supported by public policy because of their environmental characteristics (in some countries but not in all), both are greener than a relevant alternative, and both have a main purpose that is not environmental (electricity generation, transport). However, the indicative lists published as part of the 2026 SEEA-CF consultation provide an improvement[5]: they include many such 'old-but-green' activities, including many that we propose in our estimates in section 4 of this report. A clearer set of definitions that remove the vagueness associated with 'secondary purpose' would help to ensure more consistent treatment of environmentally-relevant economic activities.

2.2 Distinguishing three types of environmental green activity

Here, we provide a set of coherent definitions for different types of green economic activity, which collectively contribute to the achievement of a green economy (i.e. an economy that is consistent with planetary boundaries for environmental sustainability). Acknowledging the importance of

---

[4] SEEA-CF 2026: Consultation guidance note issue C4: https://seea.un.org/content/consultation-guidance-note-issue-c4

[5] SEEA-CF 2026. Annex: indicative lists of environmental purposes. https://seea.un.org/content/consultation-guidance-note-issue-c4

different types of green economic activity avoids some of the challenges with an attempt to create a single 'green activity' definition. We distinguish between three core categories, each of which is explained in a sub-section below:

1. Main purpose environmental protection goods and services
2. Environmentally preferable goods, services and systems
3. Specialised enabling inputs for each of the above categories.

Estimates using these definitions can be aggregated (when care is taken to remove overlapping categories that would result in double-counting), and the aggregate would represent the total of 'green economic activities' (the output of EGSS). However, it will typically be useful to report *main purpose environmental protection activities* (such as pollution abatement) separately, since these activities are fundamentally different in character: the costs of such activities fall directly on businesses and households, while the benefits are entirely mediated through the economic and other benefits of a preserved/improved natural environment. In contrast, *environmentally preferable goods, services and systems* meet existing needs in the economy, while also improving environmental performance. Typically, environmentally preferable activities can be expected to grow (and we may see wholesale transitions to systems built around such activities), whereas many main purpose environmental protection goods and services may be expected to be maintained at a near steady state (such as waste-water treatment), or ultimately decline (such as pollution abatement technologies that become redundant). While the total aggregate across these activities is likely to be useful for political messaging around the scale of green economic activity, it is unlikely to be a useful input into decision-making, because the aggregate is comprised of activities with fundamentally different economic and environmental significance and implications.

It is important to note that our definition is aimed at measuring output of green activities. Some recent analysis has measured the 'green economy' by identifying sectors that have the capabilities required to deliver green goods and services[6]. They have done this by looking at comparative advantage in intermediate or final goods that can be used for green technologies (based on lists of goods agreed in green trade agreements). This measures the latent green capabilities of the economy, rather than the actual output of *main purpose environmental protection*, or *environmentally preferable activities*, or their *specialised inputs*. In this report, we focus on measuring the actual output of such activities, rather than estimating the capability of the UK economy.

2.2.1 Main purpose environmental protection goods and services

Some goods and services are specifically designed and deployed with the main purpose of cleaning up environmental damage or preventing that damage. Examples include waste collection and landfill, waste water treatment, pollution abatement, habitat restoration, and carbon capture and storage. This aligns closely with the SEEA categories of 'environment-specific services', 'sole purpose environmental goods' and 'environmental technologies'.

Growth in such activities is not necessarily a predictor of better environmental performance of the economy. For example, catalytic converters are used to reduce air pollution from vehicles. A shift to electric vehicles will result in a decline in the production of catalytic converters, but will also result in

[6] IPPR 2024, Manufacturing Matters: https://www.ippr.org/articles/manufacturing-matters; Mealy & Teytelboym: https://doi.org/10.1016/j.respol.2020.103948

cleaner vehicles – and so the decline in this specific green economic activity will represent progress towards a greener economy.

These activities are well covered in the UK's existing green economy statistics (LCREE and EGSS datasets), with the notable exception of atmospheric greenhouse gas removals technologies other than CCS (such as enhanced rock weathering and biochar for carbon storage – though the scale of activity is currently very small).

2.2.2 Environmentally-preferable goods and services

The second major category of environmental goods and services is those that have a non-environmental main purpose, but which are environmentally preferable to relevant alternatives. This includes renewable energy, electric vehicles, recycling, and so on. Such activities have a main purpose (energy provision, mobility, and raw materials respectively) that is not focused on environmental protection, but they are all environmentally preferable to mainstream relevant alternatives (coal-fired power generation, internal combustion vehicles, and landfill/virgin material exploitation respectively). This definition is applied to economic production activities (i.e. they are environmentally preferable *production processes*, as with renewable energy generation, or they involve the production of final goods which are environmentally preferable in their consumption or end-of-life (electric vehicles, green packaging).

This category is close to the SEEA category of 'adapted goods', but it differs from the SEEA definition because it does not require such goods to be *'specifically modified to be more "environmentally friendly"'*[7].

Such goods must be environmentally preferable on at least one dimension of environmental harm (resource depletion, GHG emissions, biodiversity loss etc.) and should not be substantially worse performing on any dimensions to the point that the net environmental effect is negative. They should also avoid shifting environmental burdens to other life-cycle stages to the extent that the overall environmental performance is net negative (i.e. they should improve environmental performance on a life-cycle basis).

Many goods and services are only incrementally better environmentally than existing alternatives, and this can make measurement of such goods and services (often called 'adapted goods' in discussions of green economic measurement) more challenging. In order to be measurable, this definition is focused on activities that are *qualitatively different* from relevant (common and competing) alternatives and have a higher environmental performance. By 'qualitatively different', we refer to observable—and thus measurable—differences in production process or technology embodied in the good or service, which are sufficiently obvious and stable to allow clear recognition and measurement. For example, domestic clothes-drying equipment based on a heat pump is a recognisably different technology, and 'heat pump clothes dryers' is a coherent product category.

This approach inevitably excludes some economic activities that result in improved environmental performance. For example, incremental improvements in the energy efficiency of products are not captured, unless they are based on a recognisably different design that can be easily observed and for which data can be collected. Where incrementally environmentally-preferable goods and services

[7] UN (2014) System of Environmental-Economic Accounting – Central Framework. https://seea.un.org/sites/seea.un.org/files/seea_cf_final_en.pdf

are based on standardised certification such goods and services could also be included, as this enables measurement.

2.2.3 Specialised inputs and enablers

The manufacturing of gearboxes for wind turbines is not, under our definition, a green activity – it is not greener than the manufacture of gearboxes for other purposes. However, such activities are clearly of central policy interest, and capturing them is at the heart of attempts to measure green economic activity.

We therefore include a third type of economic activities within our definition of the green goods and services sector: 'specialised inputs', which are inputs that support environmentally preferable activities, and which are specialised to do so. In order to identify specialised inputs, we bring back the idea of a 'main purpose': where the main purpose of an activity is an input into either a *main-purpose environmental protection activity* or an *environmentally preferable good or service*, we include it within the scope of the environmental goods and services sector. This also includes services which are specialised for the delivery or retail of goods and services that are environmentally preferable (such as marketplaces for the re-use of goods), and which are thus specialised enablers of environmentally preferable consumption activities.

System perspective: enabling non-specialised infrastructure

A fundamental challenge of measuring green economic activity is that environmental outcomes are not necessarily associated with the life-cycle of specific individual goods and services, but rather with the systems in which they are embedded and to which they contribute. The infrastructure required for the transition to a clean energy system exemplifies this: electricity transmission lines are not specialised for a renewables-dominant grid, but the expansion of grid infrastructure is widely understood to be necessary for the achievement of net zero. Considerable economic activity around grid expansion is, in part, a consequence of the energy transition, and thus it is of interest to policymakers concerned with the economic and industrial consequences of pursuing and meeting environmental goals.

Supporting infrastructures like electricity grids are not straightforward to categorise, as they are not specialised to environmentally preferable systems, and do not themselves necessarily have a superior environmental performance compared to a relevant alternative. Since the power distribution infrastructure, or rail network, can just as easily be used to transfer dirty energy as clean, such infrastructures are not a core part of the green economy. However, their critical role in the energy transition means that it may be relevant to track and report activity in enabling system infrastructures alongside green economy measures. In the context of the transition to net zero, this could encompass the rail network and the electricity transmission and distribution system.

We therefore recommend that such infrastructures lie outside the core definition of 'green goods and services', but that they may be reported alongside green estimates. This could mean, for example, reporting on 'green goods and services and associated infrastructure'.

2.3 Towards greater granularity in data collection

A core challenge for measuring green economic activity is that the degree to which a given activity is 'environmentally preferable' is relative, and changes over time as new technologies are developed. For example, unleaded petrol was once seen as green, but the alternative of leaded petrol is no longer a relevant comparator, and a new relevant alternative is an electric vehicle which requires no petrol. To enable green economic activity statistics to be coherent over time, it is desirable to be able to remove such activities from historical data – otherwise the removal of an activity will appear to correspond to a reduction in the size of green economic activity, whereas it instead represents a reclassification.

In current UK practice, the LCREE survey has asked for rather broad categories. It would be desirable to collect micro-data around specific activities, to enable these activities to be removed or adapted over time. We suggest adapting the LCREE questionnaire so that it enables identification of more specific micro-data. Doing so would require the development of a more comprehensive list of relevant LCREE activities, which would likely be sub-categories of the existing 17 LCREE sectors, particularly for the 'energy efficient products' and 'Energy monitoring, saving and control systems' sectors in the existing LCREE structure. For example, the example activities set out in the LCREE sector descriptions could be made explicit as a set of more detailed options. Clearly this risks increasing the burden on respondents, and a larger survey is more costly, but a relatively small additional level of technological specificity would add considerable value in analysis of long-term trends.

2.4 Classification of environmental characteristics: a tagging system

'Environmentally preferable' is a multi-dimensional concept, and there are often trade-offs between different areas of environmental performance. Moreover, policymakers are frequently interested in specific environmental problems and the system transformations, rather than broadly construed sustainability. A system geared to produce a single headline for 'environmental goods and services' is thus less useful than one that can distinguish between the dimensions of environmental performance to which green activities are related.

The recently adopted Classification of Environmental Purposes (CEP)[8] provides a comprehensive structure for identifying relevant domains of environmental performance. The predecessor categorisations (the classification of environmental protection activities, and the classification of resource management activities) were applied as exclusive categories. This missed the point that many environmentally preferable activities can simultaneously have greater environmental performance across several dimensions.

We propose that the CEP should be implemented as a tagging system, not as a mutually-exclusive and collectively-exhaustive set of categories. For example, electric cars involve lower greenhouse gas emissions *and* lower air pollution emissions, as well as reduced overall energy consumption. It is meaningless to try to identify which of these is the 'main' purpose, and far more straightforward to use CEP categories as a tagging system. In the electric car example, relevant economic activities (manufacturing, charging infrastructure etc.) would be tagged as related to all three environmental purposes.

[8] https://seea.un.org/content/global-consultation-classification-environmental-functions

In a recent SEEA-CF discussion document, it is argued that a "*exclusive and exhaustive classification is critical, since it ensures consistency and avoids double counting. This is one of the core principles for statistical classifications endorsed by the United Nations Statistical Commission (Hoffmann and Chamie, 1999) and necessarily implies classifying under one and only purpose*."[9] However, global patent statistics are compiled in breach of this idea: they are based on a classification system that recognises that a single invention can relate to multiple areas of technology. Producing patent statistics that are useful depends precisely on this non-exclusive classification system. The environmental attributes of goods and services are inherently multi-dimensional, and it is a mistake to categorise them as if each environment-related activity had only one environmental dimension.

3 Testing new methods: web-scraped data and The Data City

3.1 The promise of web-scraped data

It is well known that the structure of industrial classifications is flawed in a changing economy[10]. Technological and economic structural changes mean that the importance of some sectors shrinks, while new sectors emerge. In today's economy, it is easy to find data on the economic activity of video rental shops—which few under the age of 35 will remember—but difficult to find data on the performance of companies that install heat pumps or manufacture, install and maintain wind turbines.

Industrial classifications do change, and indeed solar panel manufacture will soon be captured in internationally harmonised industrial classifications[11]. But at a time when rapid economic transformation towards a zero carbon economy is a necessity, the slow pace of change in economic statistics blinds policymakers and businesses to what is happening in the economy.

New data sources have sought to plug the gap. In many countries, including the UK, governments have established dedicated surveys to identify emerging low-carbon sectors. But surveys are slow, and the sector classifications that they use can also go out of date.

Web-scraped data sources have provided a new way of accessing up-to-date information on the activities of companies. The idea is that the text of a company's website provides good insight into the principal activities of the company, and that natural language processing and machine learning can be used to generate economic statistics from such data. In the UK, a leading proponent of this approach, and provider of such data, is The Data City.

[9] UN 2026, Guidance Note C4: Primary and Secondary Purpose. Global consultation on revisions to SEEA-CF; https://seea.un.org/content/consultation-guidance-note-issue-c4
[10] Dalziel et al. 2018: https://doi.org/10.1080/19488289.2017.1419319; Philips & Ormsby 2016: https://doi.org/10.1080/08963568.2015.1110229
[11] UN Stats, 2024. Introduction of the Standard Industrial Classification of All Economic Activities, Revision 5 (ISIC Rev.5). Report from the Task Team on ISIC. https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC5_Intro_11Mar2024.pdf

The data produced by The Data City has started to be used widely to inform policy audiences. In this section, we report on our assessment of the strengths and weaknesses of The Data City's platform in informing green economy estimates.

The Data City platform enables users to create 'real time industrial classifications' (RTICs), which enable analysis of bespoke 'sectors' built of companies involved in a specific activity. One RTIC that has been widely used is focused on the "net zero supply chain", corresponding to firms involved in activities associated with the transition to a net-zero-carbon economy.

3.2 Using The Data City to expand the scope of green economic measurement in the UK

The Data City provides a way of identifying activities where no survey has been implemented. We have used The Data City to generate lists of firms involved in several areas of the green economy that are currently excluded by existing statistics. These are:

- **Alternative proteins**. Most of these firms are developing plant-based alternatives to animal proteins, which are typically *environmentally preferable* to animal-based proteins. Some are developing lab-grown animal proteins, which may also be environmentally preferable. Such products may facilitate a shift towards a diet that is less reliant on animal products, consistent with studies showing such diets are an effective strategy for meeting environmental goals.
- **Batteries** and **Energy transition materials.** These lists include companies involved in the development, manufacture and distribution of batteries, and in the supply chains for energy-related raw materials. While batteries are currently captured as part of the LCREE questionnaire, there is no way to disaggregate batteries specifically from the wider energy storage sector. These are *specialised inputs* into environmentally preferable goods and services (electrification replacing the use of fossil fuels).
- **Habitat management** and **Conservation tech**[12]. These firms offer either *main purpose environmental protection* services, or *specialist input* into such services, directly contributing to the management of natural capital and the protection of biodiversity. Their activities have typically been financed by other government agencies, or through the activities of charities. Their activities show up in EGSS statistics through the activities of government bodies and charities. Increasingly, however, such activities are likely to grow commercially, in part because of policies like biodiversity net gain requirements and environmental land management schemes.
- **Bicycle manufacture, retail and related services**. Bicycles are *environmentally preferable* relative to dominant road-transport options. Unlike in some other European countries—notably the Netherlands—the UK has no specific SIC codes for bicycle-related activities, whether manufacturing, distribution or retail. This list identifies firms involved in these activities.

[12] Conservation technology, sometimes known as 'naturetech' is a growing area of environment-related economic activity focused on technologies that measure, monitor and manage species distributions. Examples include satellite monitoring, drones, camera traps and acoustic monitoring.

- **Peer-to-peer online marketplaces for re-use**. This captures online marketplaces such as Vinted, Depop etc. that specialise in second-hand goods, for which no data is currently reported by ONS. While there is a dedicated SIC code for second-hand shops, this only relates to businesses with physical shops, and online platforms do not report their activities under this SIC code. The list provided here is focused on platforms that are wholly or very largely dedicated to second-hand rather than new goods. This means that it excludes some large players (including Ebay). It is also unable to capture platforms that are embedded within social media platforms (i.e. Facebook marketplace). It is thus an underestimate of this activity. These marketplaces are an instance of *specialised inputs and enablers*.

The process for producing each list follows the same core steps:

1. Identification of a 'training set' of firms that have the desired characteristics.
2. The Data City platform then uses this training set to predict similar firms, based on the similarity of website text to the companies in the training set.
3. The user then actively excludes companies that are not appropriate, which improves the training data.
4. The Data City platform then provides a revised list.

The process is repeated iteratively, adding firms to the training set and excluding inappropriate firms. The resulting list aims to capture as much relevant activity as possible, while minimising the inclusion of inappropriate firms.

The resulting lists of firms are appended to this report. The Data City itself has also produced a set of lists of firms under what it calls "real-time industrial classifications", or RTICs. We have also included RTICs of relevance to the green economy in our list of firms.

3.3 Assessing the validity of web-scraped data for green economy measurement

It is not straightforward to directly assess the validity of The Data City's approach to RTICs. One way of doing this would be to assess the extent to which The Data City is able to reproduce lists of firms that correspond to those identified under existing industrial classifications.

However, using existing industrial classifications can also be flawed. Companies have little incentive for accuracy in SIC reporting, and we have identified many companies with SIC codes that do not appear to be good descriptions of their core business. There are also examples of potential distortionary policies in which funding eligibility depends on SIC code[13], creating incentives for companies to choose SIC codes strategically.

Second, the key value in a web-scraping approach is the ability to create industrial classifications that cut across established sectors. In order to test this ability, it is necessary to compare a web-scraping approach to an established data source that measure the cross-sector activity.

In this work, we use the existence of the Low Carbon and Renewable Energy Economy (LCREE) survey, which directly asks firms whether they are involved in a set of 17 activities, to test The Data City platform. The LCREE dataset available to researchers in the ONS Secure Research Service

[13] For example, see [Withdrawn] Apply for the Energy Bills Discount Scheme support for Energy and Trade Intensive Industries (ETIIs) - GOV.UK

includes all firms that have been included in the LCREE sample (2014-2022), including those that report no activity in the specified low-carbon and renewable energy activities.

The comparison should be interpreted as a diagnostic assessment rather than a definitive test of whether either source is “correct”. LCREE has the advantages of being an official survey and of asking directly about LCREE activities, but it is affected by survey response, sampling-frame design, activity definitions and reference-year timing. DCNZ has the advantages of timeliness and flexible cross-sector coverage, but it is based on website text and therefore captures stated activities and capabilities rather than verified activity-specific turnover. We therefore treat LCREE-active status as the reference indicator for the diagnostic exercise, while interpreting apparent false positives and false negatives in light of differences in timing, scope and measurement design.

We set out to explore the following research questions:

1. To what extent does a company’s inclusion in The Data City’s net zero list predict whether it said “yes” to involvement in the LCREE?.
2. What proportion of companies that have said ‘yes’ to involvement in LCREE sectors are not included in The Data City’s net zero list? (i.e. to what extent does The Data City appear to miss low-carbon firms?).
3. What proportion of companies in the net zero list, when matched with companies sampled by LCREE, said that they were not involved in the 17 LCREE sectors? (i.e. to what extent does The Data City mis-classify firms as involved in low-carbon?)

We also explore the extent to which The Data City’s net zero (DCNZ) list reveals LCREE activities that would not be captured by the current LCREE survey, because they take place in firms outside the sampling frame used by the ONS. In order to balance the costs of running the survey with the benefits it provides, the LCREE is not randomly distributed across the economy, but targeted at sectors most likely to engage in LCREE activities. The Data City’s approach enables an assessment of whether there are areas of LCREE activity that are currently outside the sampling frame of the LCREE survey.

The DCNZ list and LCREE are covering similar topics. However, there are some differences in scope between these approaches. The DCNZ list is broader, including activities that are related to net zero, but which are not within the scope of the LCREE. For example, the net zero list includes GHG removal technologies such as enhanced weathering, which are not within the scope of the LCREE.

In order to align the scope of the DCNZ list with the LCREE, we used the ‘sector keywords’ provided by The Data City. We removed firms from the DCNZ list unless they had sector keywords that corresponded to the LCREE. This improved the share of matched firms that are LCREE-active, i.e. which both feature in the DCNZ and which, when sampled by LCREE survey, said that they are active in LCREE sectors.

3.4 Methods, data and assumptions

The Data City’s “net zero” list is continually updated. According to The Data City, it was developed in collaboration with experts and relevant stakeholders to inform both the training set and to review the list to improve it[14]. We extracted the net zero list of firms from The Data City in December 2025. For brevity, we refer to The Data City's net zero list of firms as the "DCNZ" list.

Our analysis treats The Data City as a human-in-the-loop, web-text-based machine-learning classification system. We do not audit or retrain the underlying algorithm. Instead, we conduct an external validation of the resulting DCNZ firm list by comparing it with LCREE survey responses and administrative business data. The evaluation is therefore conditional on four upstream processes: The Data City's matching of websites to company registration numbers, the construction and curation of the training set, the sector-keyword pruning used to align the DCNZ list with LCREE scope, and our subsequent matching of company registration numbers to enterprise references in ONS data.

A key strength of The Data City is its timeliness, since it provides data based on real-time application of the machine learning algorithm to website text, linked to data on companies. For our purposes, this strength creates methodological challenges, since the LCREE dataset available to the research team only runs to 2022 (more recent data is available[15], but was not yet in the ONS Secure Research Service when this research was undertaken). Each data source (The Data City and LCREE) provides a snapshot, but at different times. Changes to business activities between 2022 and 2024/2025 (such as companies going out of business, or diversifying into or out of green activities) are an unobserved source of discrepancy between the LCREE and The Data City approaches.

Our approach rests on our ability to match firm-level records in the LCREE dataset and in the DCNZ list. The Data City identifies firms using their Company Registration Number (CRN), as filed at Companies House. This is based on matching between the company's website and their CRN, a process that is subject to some error. We used the CRNs provided in the DCNZ list, and sought to match them to firms in the LCREE dataset and the Business Structure Database[16].

A business enterprise may comprise a complex structure of registered companies, each corresponding to different units, despite exhibiting coherent behaviour as a single organisational entity in practical business activity. For this reason, the UK Interdepartmental Business Register (IDBR) identifies firms through an identifier that aims to cover individual enterprises that may contain multiple CRNs – i.e. the enterprise reference (ENTREF).

We sought to match 32,734 firms in the DCNZ list with ENTREF records held in the IDBR, via the ENTREF-matching service provided by the IDBR team at the ONS. The IDBR team were able to match 27,542 CRNs to records held in the IDBR, a match rate of 84%. This means that our sample of firms was reduced substantially. The firms that were not matched were from a wide variety of SIC codes, including those likely to be relevant to low-carbon and renewable energy. The top 12 SIC codes represented in this unmatched group is shown in Table 1. It is not clear why it was not possible to match these firms to an ENTREF. This loss of firms from The Data City may have an influence on the results shown here.

Because the DCNZ firms are identified by their CRNs, whereas the IDBR uses a broader definition of the firm (i.e. a unique enterprise identified by an ENTREF may be made up of several registered companies each with a unique CRN), the total number of ENTREF firms is smaller: 25,962. We then matched this list of firms in the ONS Secure Research Service (SRS) to the Business Structure

[14] https://The Data City.com/blog/first-list-of-uk-net-zero-companies/
[15] https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/finalestimates/2024
[16] The Business Structure Database is a snapshot of the government's Inter-Departmental Business Register (IDBR) that is made available for researchers – it is a near comprehensive database of the UK's companies, providing information on their sector of activity, location, turnover, etc.

Database (BSD) and the LCREE dataset. Of the 25,962 firms for which we identified an ENTREF, around 2,526 were founded after 2022, and therefore could not be matched to that year's LCREE data (the most recent data for which we had access).

The empirical workflow consisted of six steps. First, we extracted The Data City's net zero list in December 2025. Second, we restricted the DCNZ list to firms with sector keywords corresponding to LCREE activities, in order to reduce scope differences between the two sources. Third, we matched company registration numbers in the DCNZ list to enterprise references in the IDBR. Fourth, we linked the matched enterprises to the Business Structure Database and to LCREE survey records in the ONS Secure Research Service. Fifth, we constructed mutually exclusive comparison groups: LCREE-active firms included in DCNZ; LCREE-active firms not included in DCNZ; LCREE-inactive firms included in DCNZ; LCREE-inactive firms not included in DCNZ; and DCNZ firms not sampled by LCREE. Sixth, we assessed the comparison using both firm-count and turnover-weighted metrics, while separately examining firms outside the LCREE sampling frame.

Table 1. Sectoral distribution of DCNZ firms that could not be matched to an ENTREF

SIC	No. Firms	Share of unmatched	SIC description
43210	267	5.1%	Electrical installation
64209	237	4.6%	Activities of other holding companies n.e.c.
70229	187	3.6%	Management consultancy activities other than financial management
35110	167	3.2%	Production of electricity
82990	167	3.2%	Other business support service activities n.e.c.
43220	158	3.0%	Plumbing, heat and air-conditioning installation
68209	151	2.9%	Other letting and operating of own or leased real estate
74901	147	2.8%	Environmental consulting activities
38110	144	2.8%	Collection of non-hazardous waste
96090	133	2.6%	Other personal service activities n.e.c.
99999	123	2.4%	Dormant Company
74990	114	2.2%	Non-trading company

3.5 Results and discussion

In describing the results, we use a Euler diagram showing the overlapping sets of firms involved, in which the area of the figure is proportional to the numbers shown. The Data City provides broad coverage of web-visible active UK firms and is designed to identify firms whose website text resembles firms in a curated training set. It should therefore be interpreted as a web-text-based discovery and classification tool, not as a census of LCREE activity. However, in coverage it is a near-census of relevant firms, since most active firms involved in LCREE activities can be expected to have a website. We would expect that most firms that are "LCREE-active" (i.e. that have been surveyed by the LCREE and said that they are involved in LCREE activities) feature in the DCNZ list (though it is important to recall that 18% of DCNZ firms could not be matched to an ENTREF, and hence were excluded from the analysis). In other words, looking at Figure 1, we would expect A to be a large share of A+B+C. B should be small (i.e. firms that have been sampled by LCREE, that have indicated that they are LCREE-active, but which are not on the DCNZ list), as should C (firms that have responded to the LCREE survey that they are not involved in LCREE-related activities, yet they are included in the DCNZ list).

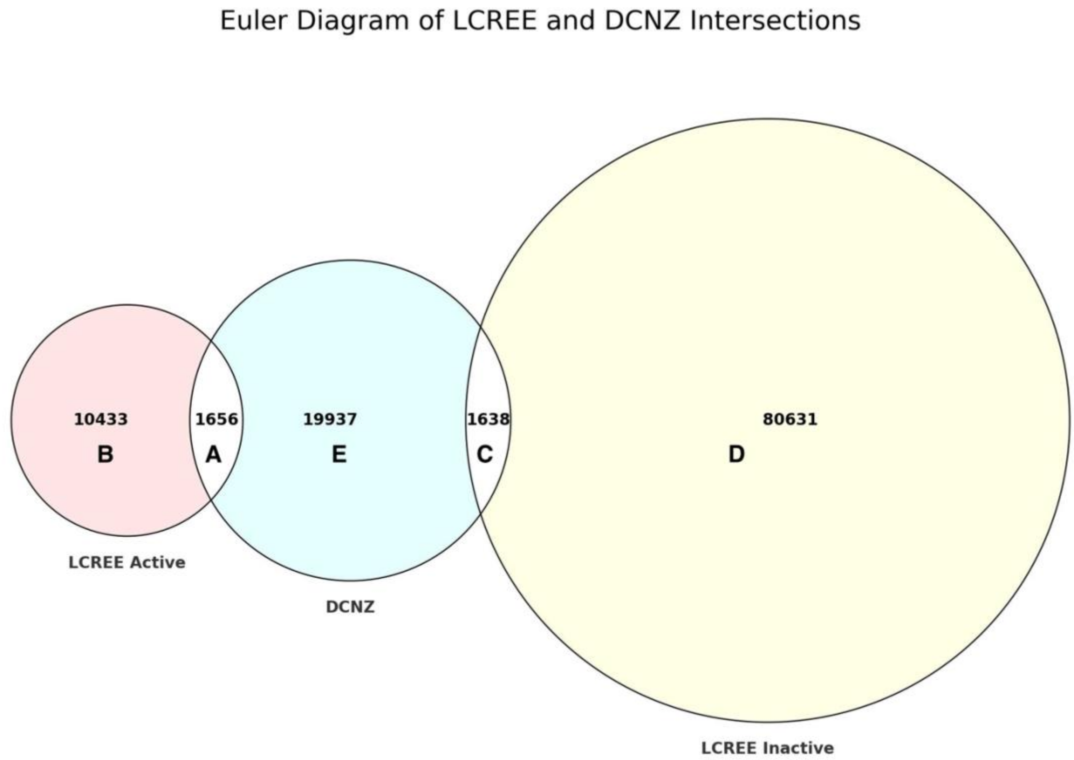

Figure 1. Euler diagram showing the sets of firms involved in the analysis. Areas A+C represent firms that have been sampled by the LCREE survey, and that appear in the DCNZ list. Numbers show the results for LCREE returns in 2022 (A-D). LCREE-active firms are those have been sampled by the LCREE survey and indicated that they are involved in LCREE activities. LCREE-sampled but inactive are those that were sampled but said that they are not involved in LCREE activities. Firms in region E are in the DCNZ, but were not sampled by the LCREE survey. Areas are proportional to the numbers shown. Figure produced using eulerr.co

The Euler diagram can also be read as a diagnostic classification table. Taking LCREE-active status as the reference indicator, area A represents true positives, area B apparent false negatives, area C apparent false positives, and area D true negatives. We report four core metrics: recall, A/(A+B), which measures the share of LCREE-active firms identified by DCNZ; precision, A/(A+C), which measures the share of DCNZ-positive firms that are LCREE-active within the LCREE sample; specificity, D/(C+D), which measures the share of LCREE-inactive firms excluded from DCNZ; and the F1 score, which summarises the balance between precision and recall. We do not emphasise overall accuracy because the comparison population is highly imbalanced: LCREE-inactive firms substantially outnumber LCREE-active firms.

Table 2. Firm-count and turnover-weighted diagnostic metrics for DCNZ relative to LCREE within the LCREE-sampled comparison set

Metric	Formula	General Interpretation	Firm-count Results	Turnover-weighted Results
Recall / Sensitivity	A/(A+B)	Measures the share of LCREE-active firms or turnover captured by DCNZ. Higher values indicate better coverage of the LCREE-active reference group.	13.7%	51.0%
Miss Rate	B/(A+B)	Measures the share of LCREE-active firms or turnover not captured by DCNZ. Lower values indicate fewer apparent false negatives relative to LCREE.	86.3%	49.0%
Precision / Positive Predictive Value	A/(A+C)	Measures the share of DCNZ-positive firms or turnover that is also LCREE-active. Higher values indicate that DCNZ-positive classifications are more likely to correspond to LCREE activity.	50.3%	60.0%
False Discovery Rate	C/(A+C)	Measures the share of DCNZ-positive firms or turnover that is LCREE-inactive. Lower values indicate fewer apparent false positives relative to LCREE.	49.7%	40.1%
Specificity	D/(C+D)	Measures the share of LCREE-inactive firms or turnover excluded from DCNZ. Higher values indicate better exclusion of firms that do not report LCREE activity.	98.0%	88.5%
False Positive Rate	C/(C+D)	Measures the share of LCREE-inactive firms or turnover included in DCNZ. Lower values indicate less over-inclusion relative to the LCREE reference indicator.	2.0%	11.5%
F1 Score	2A/(2A+B+C)	Combines precision and recall into a single summary measure. Higher values indicate a better balance between capturing LCREE-active firms and avoiding apparent false positives.	21.5%	55.1%

On a firm-count basis, DCNZ has low recall relative to LCREE. Using the counts in Figure 1, Table 2 presents the firm-count diagnostic metrics for DCNZ relative to LCREE within the LCREE-sampled comparison set. It shows DCNZ identifies 13.7% of firms that report LCREE activity, implying an apparent miss rate of 86.3%. This is not, by itself, evidence that the web-scraped approach has no value: the comparison is affected by timing, scope, firm matching and the fact that LCREE is a survey-based reference indicator rather than a perfect ground truth. However, it does show that the DCNZ list should not be used as a direct firm-count substitute for LCREE without further validation or survey calibration.

Precision is higher than recall but still limited. Within the LCREE-sampled comparison set, 50.3% of DCNZ-positive firms report LCREE activity. Specificity is high, at 98.0%, meaning that most LCREE-inactive firms are not included in DCNZ. Overall accuracy is 87.2%, but this figure is not informative because LCREE-inactive firms dominate the comparison population. The more meaningful interpretation is therefore that DCNZ is selective and highly specific, but has low firm-count coverage of LCREE-active firms.

When matching firms that have been surveyed by the LCREE with the firms on the DCNZ list, we would expect to find matches mostly in firms that are LCREE-active (i.e. that responded to the LCREE survey by saying that they are involved LCREE activities). We do find more LCREE-active firms (A) in the DCNZ than LCREE-inactive firms (C), but the difference is not large.

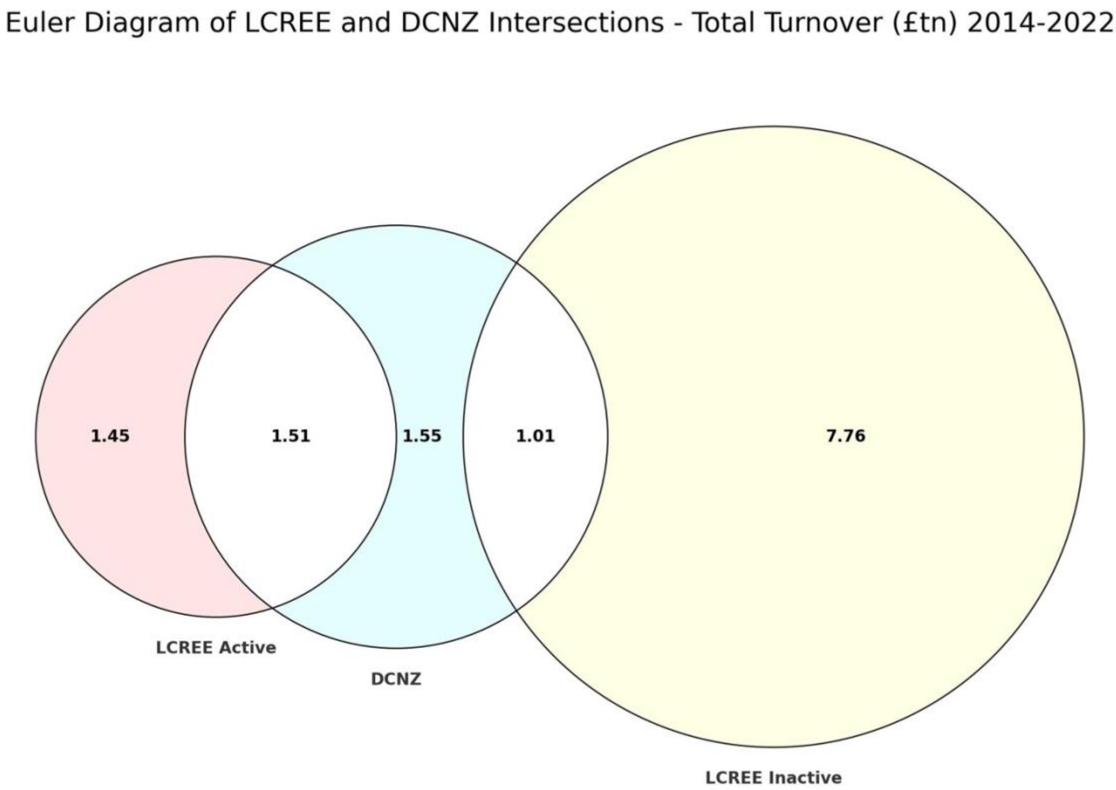

Figure 2. An area-proportional Euler diagram showing LCREE-active, LCREE-inactive and DCNZ firms in terms of turnover summed across the period. Data is in £tnGBP. Note that the figure shows total turnover, not the turnover that is specific to LCREE activities.

The Euler diagrams show that the DCNZ matches the estimates of the LCREE more closely in terms of turnover than in terms of the number of firms. Firms in area E on the figures—i.e. those firms that are included in the DCNZ but which have never been surveyed by the LCREE—are on average much smaller than those in the matched data (mean 2022 turnover in E = £17m; in A+C = £245m, and median for E = £0.2m compared to £2.2m for A). Many of the unmatched firms in the DCNZ are small firms. Furthermore, a large proportion (~65%) of the turnover in area E is from firms that are outside the LCREE sampling frame, i.e. which would not ever receive the LCREE survey (as explored further in section 3.5.3). This suggests that the additional firms in the DCNZ that the LCREE has not sampled and that are within the sampling frame are largely small firms that would not make a substantial impact on estimates of UK LCREE activities.

As can be seen in Table 2, the turnover-weighted results give a more favourable picture than the firm-count results. Using the turnover values shown in Figure 2, turnover-weighted recall is approximately 51%, compared with 13.7% on a firm-count basis. Turnover-weighted precision is approximately 60%, and the turnover-weighted F1 score is approximately 55%. This indicates that many of the LCREE-active firms missed by DCNZ are relatively small in turnover terms. The result is important for economic measurement because missing many small firms has a different implication from missing a few very large firms.

This interpretation should nevertheless be treated cautiously. Figure 2 uses total firm turnover, not turnover specifically attributable to LCREE activities. The turnover-weighted metrics therefore measure the economic scale of firms in each overlap category, not the value of LCREE-specific activity captured or missed by DCNZ.

3.5.1 Explanations for divergence between DCNZ and LCREE

We have identified several possible explanations for the observed differences. These explanations should be read together rather than as mutually exclusive alternatives. The observed divergence between DCNZ and LCREE can arise because the two sources measure different reference periods, use different activity definitions, identify firms using different units, and capture different aspects of firm behaviour. LCREE asks firms to report activity and turnover in defined low-carbon and renewable energy categories. DCNZ uses current website text to identify firms whose activities or capabilities resemble those of firms in a curated net zero training set. Divergence is therefore expected even if both sources are performing well within their own design.

1. Differences in time horizon

The Data City platform provides "real-time" industrial classifications based on indexed website text. It provides a snapshot of today's economy. In contrast, the most recent LCREE dataset available to us is based on a survey that recorded the activities of firms in 2022. Moreover, many of the firms in the LCREE dataset are not sampled every year, and their most recent inclusion in the LCREE dataset will be much earlier than 2022. That discrepancy in time creates opportunities for discrepancy between the DCNZ and LCREE characterisations of individual firms.

Notably, firms may have entered or exited low-carbon activities during the intervening period since they were sampled by LCREE. It is possible that firms that indicated in LCREE surveys in 2022 that they are not involved in LCREE activities have subsequently diversified into such activities, and are captured in the DCNZ list. The reverse is also true: firms that were involved in LCREE activities when surveyed may have abandoned such activities by 2024-2025, the period for which we collected data from The Data City.

We have sought to understand the significance of this effect by looking for trends over time in the matched data (i.e. firms that are both in the DCNZ list *and* were sampled by the LCREE survey). If firm entry and/or exit from LCREE activities is a significant source of mismatch between DCNZ and LCREE-active lists of firms, we would expect the share of LCREE-active firms in the matched data (A+C) to be highest in the most recent years, and lowest for the earliest years of the LCREE survey. If we assume a constant rate of diversification into/out of LCREE activities, firms surveyed in 2014 are most likely to diversified into new activities by 2022. We do observe this in our data (Figure 3) – the DCNZ is much more accurate in its classification of firms in the most recent year for which LCREE data is available. Only around 35% of firms that are in the DCNZ list and that were surveyed by the LCREE in 2014 said at that time they were active in LCREE activities, but that rises to 55% for firms surveyed in 2022.

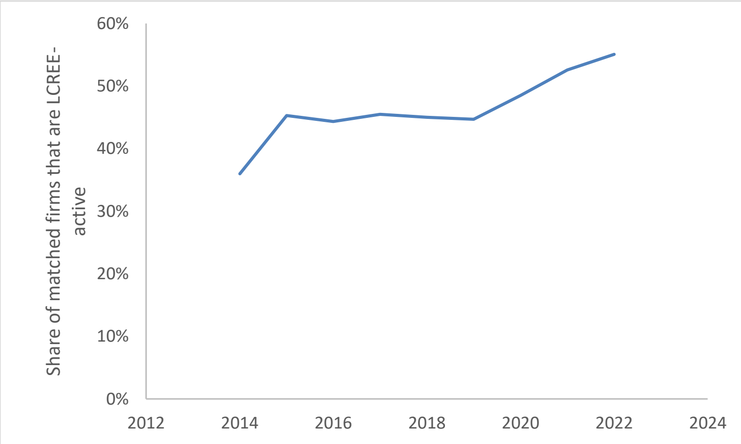

Figure 3. Share of firms in the matched DCNZ-LCREE dataset that are LCREE-active, i.e. A/(A+C). Years show the year of response to the LCREE survey.

To further explore the time-horizon question, we have examined the extent of switching in/out of LCREE activities in firms that have been repeatedly included in the LCREE sample. Of firms that have been surveyed more than once in the observation period, around 80% were LCREE active in all years in which they responded. Around 10% of firms diversified into LCREE activities (i.e. when first surveyed they reported no LCREE activity, and when last surveyed they reported being LCREE active), while around 10% of firms appear to have moved out of LCREE activities. Entry/exit from LCREE activities does occur, and this may underpin some of the discrepancy between DCNZ and LCREE lists.

Finally, some firms sampled by the LCREE will have gone out of business since they completed the survey. Where this is the case, they would not appear in the DCNZ, and such firms would appear in area B in our Euler diagrams, inflating the apparent failure of DCNZ to capture the green economy. We restricted our analysis to the firms that were still active in 2024 (the final year for which BSD data was available). It is possible that some firms went out of business between 2024 and 2025, but it is likely that this accounts for only a small portion of the discrepancy.

2. Discrepancies driven by differences in definitions of ‘low-carbon’ and ‘renewable’.

The definitions and guidance in the LCREE survey were established more than 15 years ago, and do not always align with current thinking on what constitutes activities relevant for net zero as reflected in the stakeholder-informed DCNZ. The two data sources are attempting to measure a similar idea (especially once we have pruned firms with non-LCREE-related keywords from the DCNZ list), but choices at the margins of the LCREE definitions lead to different outcomes. This clearly accounts for some of the discrepancy: for example, the LCREE includes activities related to hybrid petrol and diesel cars (i.e. all hybrids, not just plug-in hybrids). The LCREE definitions and guidance were established prior to the adoption of the UK’s net zero target, at a time when petrol and diesel hybrids were seen as low carbon technologies. The manufacture of such vehicles is excluded from the definitions underpinning the DCNZ. In contrast, the DCNZ appears to be rather conservative with respect to the inclusion of automotive firms. For example, the UK's largest producer of electric vehicles was not featured in the DCNZ. This difference has a large impact on the apparent 'miss' by the DCNZ in terms of turnover. The largest single SIC division in area 'B' relates to automotive, accounting for nearly a quarter of the total turnover in B. Other divergent scoping choices are revealed in the comparison of the most prominent SIC codes and divisions: higher education organisations hardly feature in the DCNZ, but organisations with a 85421 SIC code (*First-degree level higher education*) accounts for around 5% of the turnover in B. This source of discrepancy highlights that the definitions used by stakeholders in constructing the DCNZ, and that underpin the LCREE, differ in important ways.

3. Misclassification in the DCNZ

It is also possible that DCNZ mis-classifies firms, identifying them as involved in low carbon or renewable energy activities when they have reported via the LCREE survey that they are not active in such activities. In order to better understand this possibility, we examined the sectoral distribution of firms in C (in the DCNZ, but LC-inactive according to the LCREE). The sectors of most firms in C are similar to those in A: in both, though in different rank order, the top five SIC divisions with the most firms are:

- 35, i.e. Electricity, gas, steam and air conditioning supply
- 43, i.e. Specialised construction activities
- 71, i.e. Architectural and engineering activities; technical testing and analysis
- 38, i.e. Waste collection, treatment and disposal activities; materials recovery
- 46, i.e. Wholesale trade, except of motor vehicles and motorcycles

In other words, while the DCNZ list is identifying some firms that have responded negatively to the LCREE, it is largely identifying firms that operate in the same sectors as firms that are involved in LCREE. A similar pattern is observed for the most common 5-digit SIC codes: 7 of the top 10 SIC codes by number of firms are the same in A and C. The similarity of the sectoral coverage of DCNZ firms that are, and are not, LCREE-active according to the LCREE survey may be due to similarity of the website text of such firms. The machine learning algorithm that underpins The Data City is seeking firms similar to those in the training set. This may lead to misclassification in some cases. For example, if the training set includes firms involved in 'electrical installation SIC 43210' (e.g. electricians that install EV chargers, solar panels etc.), it may also catch electricians that do not provide these specific services because of similarities in website text shared across all electricians. This highlights that the results produced from The Data City depend very strongly on the specificity of the training set and the user-driven approach to building and testing that training set. It is perhaps worth noting that this means that the mis-classification tends to include firms that have closely related capabilities to the low-carbon firms. One way of interpreting this is that DCNZ is overestimating the size of the low-carbon economy by including 'latent' low carbon firms—those that could relatively easily become low carbon active firms because they are involved in similar activities.

Looking at top ten SIC codes by total turnover, we do identify some surprises – i.e. sectors that do not feature in the LCREE data. Oil refining and oil extraction firms both make it into the top 10 5-digit SIC codes, accounting for more than a third of the total revenue in C. One potential explanation for the inclusion of such firms relates to a specific risk for misclassification in the green economy: greenwashing. Firms have an incentive to talk about sustainability, because of broad societal concern about the importance of climate change and the environment. We identified a handful of firms that appeared to be included in the DCNZ list because of generic language on their websites about their commitments to renewable energy, energy efficiency and so on. This might drive some of the apparent misclassification of firms. Since we are unable to conduct this analysis outside the ONS secure research environment, it is not possible to assess the magnitude of this effect without a detailed firm-by-firm manual assessment of the 'true' LCREE nature of each of the >20,000 firms in the DCNZ list.

There are two further explanations for potential mis-classification. One is that firms may advertise low-carbon or renewable energy services, but in any given year they may have zero turnover in those activities (for example, engineering services firms that operate in a range of sectors). Such firms would show up in the DCNZ, but unless they are surveyed each year, a portion of such firms will appear LCREE-inactive in the LCREE results. Again this reflects the idea that DCNZ assesses the latent capacity for delivery of low-carbon and renewable goods and services even when those are not being provided.

A final reason for misclassification is that DCNZ may miss LCREE activities when they form a very small share of activities of a large firm, with the result that website text largely focuses on those other activities. This could result in The Data City failing to identify these firms in the DCNZ. The LCREE survey asks companies for their turnover in LCREE activities. In order to look at this possibility, we looked at whether the share of LCREE activities in A is higher than in B. We found that firms in B are typically smaller than those in A, suggesting that DCNZ tends to miss smaller LCREE-active firms (mean total turnover 2014-2022 of firms in B was £74m compared to £360m for firms in A). We also found that median *low carbon* turnover is much lower in B compared to A, suggesting that many of the firms missed by DC have very low turnover related to LCREE activities.

3.5.2 Identifying the share of a company's activity that is green

One of the weaknesses of The Data City approach is that it is only able to identify whether a firm is involved in a specific activity. It cannot determine whether the activity is the firm's main area of business. This means that firms that are involved in multiple activities, and for whom LCREE activities are only a very small area of activity, will still be captured by the DCNZ.

However, the machine learning approach used generates a score for each firm that reflects the relative similarity between the website text of the firm, and that of the training set. We tested whether the score might help to identify firms that have a larger share of LCREE-related activity. In order to do this, we generated a list of firms involved in wind energy within The Data City platform, and examined the correlation between each firm's similarity score and their share of LCREE activity as reported to the LCREE survey. This was based on the hypothesis that firms with a greater share of wind-related activity would be likely to show stronger similarity to the training set (which are firms specialised in wind).

However, we did not find a clear correlation between each company's share of activity in wind-related activities and the similarity score. This leads us to expect that the degree of similarity of website text to firms that specialise in a given activity cannot be used to approximate the share of a company's turnover in that activity. This finding rules out a tempting shortcut for green economy estimation. The Data City similarity scores may be useful for ranking or prioritising firms for review, but our results suggest that they should not be used as activity weights. In particular, a high similarity score should not be interpreted as evidence that most of a firm's turnover is generated by the relevant green activity, and a lower score should not be interpreted as evidence of only marginal involvement. For official statistics, web-text classification can help identify candidate firms, but activity-specific turnover shares still require survey responses, administrative data, manual validation or another source of activity-level allocation.

3.5.3 Insights into the LCREE sampling frame

The LCREE survey is distributed to firms operating in specific SIC divisions, and excludes others[17]. The DCNZ list includes firms in 30 five-digit SIC codes that are not within this LCREE sampling frame. These are firms that would never be included in estimates of LCREE activity based on the LCREE survey, because they would never be sampled. This is shown as the shaded area in Figure 4, which divides E into sections within and outside the LCREE sampling frame.

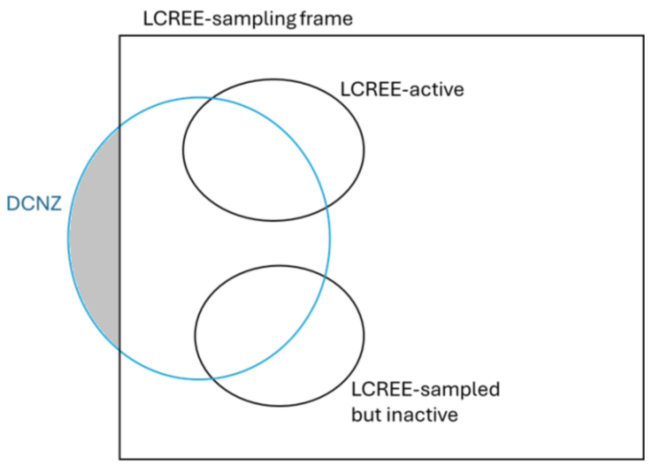

Figure 4. The sets of firms that are included in the DCNZ, in the LCREE sampling frame, and that have been sampled by the LCREE. The shaded area shows firms that are identified as LCREE-related by The Data City, but which are not in sectors that are sampled by the LCREE survey.

Summary statistics for DCNZ firms outside the LCREE sampling frame are presented in Table 3 below. As can be seen from the table, the DCNZ list suggests that a large amount of economic activity takes place in SIC divisions that are not sampled by the LCREE. This could mean that the LCREE sampling frame is unnecessarily narrow, and is missing a sizable volume of relevant activity – £60-100bn. Note that this accounts for much of area E in figures 1 and 2.

This is one of the clearest use cases for web-scraped ML in official statistics. The result does not imply that all turnover in these firms should be counted as LCREE activity. Rather, it shows that DCNZ can identify economically significant firms and SIC divisions that are plausibly relevant to LCREE but are not reached by the current survey frame. The policy implication is therefore about sampling-

17
https://www.ons.gov.uk/economy/environmentalaccounts/methodologies/lowcarbonandrenewableenergyeconomylcreesurveyqmi

frame design, not immediate point estimation. Web-scraped classifications can be used to identify candidate SIC divisions, firms or activities for targeted survey expansion, after which survey responses can determine whether and how much activity should be counted.

Table 3. Summary statistics for the turnover firms in the DCNZ but outside the sectors sampled by the LCREE survey. St. dev. = standard deviation, turnover figures in £1000s of GBP.

Year	Number of firms	Mean	Standard deviation	Median	Total
2019	319	273,173	4,633,981	300	87,142,084
2020	343	177,745	2,989,242	350	60,966,616
2021	352	173,237	2,950,514	327	60,979,595
2022	381	267,320	4,855,439	303	101,848,977

However, this result is driven by a relatively small number of very large firms, with a long tail of firms that have zero or very small turnover: the median firm had an annual turnover of less than £300k. The SIC divisions that feature in the DCNZ list, but which are outside the LCREE sampling frame, are:

- 63: Information service activities
- 65: Insurance, reinsurance and pension funding, except compulsory social security
- 66: Activities auxiliary to financial services and insurance activities
- 84: Public administration and defence; compulsory social security
- 86: Human health activities
- 87: Residential care activities
- 90: Creative, arts and entertainment activities
- 91: Libraries, archives, museums and other cultural activities
- 93: Sports activities and amusement and recreation activities
- 95: Repair of computers and personal and household goods

The share of activity is highly skewed across these divisions: around 99% of the turnover in DCNZ firms found in these divisions is accounted for by a small number of firms in divisions 63, 65 and 66.

We identified four reasons for net zero firms appearing in these SIC divisions. We conducted some work outside the SRS using only data provided by The Data City to explore these reasons. The limitation of this work is that the analysis is reliant on the completeness of The Data City's records, which link to Companies House, but which may be less complete than the Business Structure Database.

1. Genuine LCREE activity can occur in these divisions, particularly in 63-66 (specialised financial and consulting services focused on low-carbon and renewables). We identified firms within these SIC divisions that we believe to be genuinely involved in LCREE activities.
2. Public bodies are captured by the DCNZ platform where they are registered as companies, and in some cases we identified public bodies with involvement in the energy system.
3. Some companies have made errors in registering their SIC codes. We identified several companies that did not appear to be active in the area of business described by their SIC code.

4. Errors in classification by The Data City. We identified a handful of companies for which there did not appear to be genuine LCREE activity.

Overall, our view is that the majority of turnover captured outside the sampling frame—most of which is in division 65, is genuine LCREE activity. We suggest that there would be value in expanding the scope of the LCREE survey to include firms involved in SIC division 65. This division includes firms that provide specialist insurance services for the renewable energy sector.

3.6 Conclusions on the use of web-scraped data in green economy estimates

This analysis has shown both the strengths and weaknesses of The Data City—and by extension similar tools based on web-scraping and machine learning. These tools are a hugely valuable complement to existing national statistical approaches. For some areas of the green economy, such as alternative proteins, The Data City and similar tools are an essential approach, given the lack of existing survey data.

Web-scraped data is timely, flexible, and enables very specific sectoral categorisation with relatively low effort. This is transformational for economic data, particularly for areas like the green economy that cut across traditional sectoral classifications.

A particularly notable strength, from the perspective of the ONS, is that such tools can quicky and cheaply identify areas of the economy that contain particular activities, which can then be targeted for surveys. In the work presented here, we have found that considerable LCREE activity takes place outside the sectors to which the LCREE survey is distributed. For other areas of the green economy, The Data City and similar tools could provide a very valuable way of informing the development of a sampling frame.

The analysis also highlights the weaknesses of web-scraping tools. One is that these tools can identify whether a company is engaged in a specific area of activity, but it cannot provide insight into the extent of the company's involvement. Our analysis suggests that the degree web-text similarity to specialist clean energy firms is not enough to predict the share of clean energy activity within a firm that is engaged in multiple activities.

More generally, web-scraping tools are only as good as the work underpinning their use. The logic of the tools (similarity of website text) is straightforward, but developing, testing and refining the training set requires careful work and analysis by people with detailed understanding of the sectors and activities that they are seeking to capture.

While the discrepancies between The Data City and LCREE (show in in figure 1 and 2) look large, the analysis suggests that much of this relates to differences in the time horizon, and different inclusion/exclusion choices for what constitute low-carbon and renewable energy activities in the datasets. Where there is misclassification by The Data City, at least some of that is capturing firms that are very similar to LCREE-active firms, and hence The Data City is likely to be capturing latent green economy capabilities. We did, however, find some evidence of misclassification driven by 'greenwashing', in which company text misleads The Data City algorithm into thinking that the company is engaged in renewable or low-carbon activities when it is not.

The great strength of web-scraped tools is that they can be used flexibly, and analysts can adapt the training set for the machine-learning algorithm, and set more stringent inclusion criteria to improve accuracy. Any list generated by such tools will strike a balance between comprehensiveness (including all relevant firms) and accuracy (excluding irrelevant firms). The challenge for users of the tool is that they may be unable to identify how comprehensive the resulting list is, and may have a low tolerance of irrelevant firms, skewing the list towards a high 'miss-rate'.

The comparison has also five important limitations. First, LCREE and DCNZ refer to different points in time: LCREE data available for this analysis run to 2022, whereas the DCNZ extraction reflects the web-visible economy in 2025. Second, LCREE and DCNZ differ in scope, even after sector-keyword pruning. Third, matching company registration numbers to enterprise references is imperfect, and unmatched DCNZ firms are excluded from parts of the analysis. Fourth, LCREE-active status is treated as a reference indicator for diagnostic purposes, but it is not a perfect ground truth. Fifth, The Data City identifies firm involvement, not activity-specific turnover shares. These limitations mean that the results should be interpreted as evidence on the strengths and weaknesses of web-scraped classification for discovery and sampling design, rather than as a definitive estimate of low-carbon activity.

4 Revising the UK's green economy estimates: illustrative data on key sectors

This section provides evidence on the scale of green activities currently omitted from the ONS EGSS account. These are consistent with the definitions set out above. The section illustrates that a shift in definitions can involve substantial revisions to the apparent scale of the green economy in the UK.

Here, using a combination of evidence from The Data City, from existing ONS surveys, and from other sources, we identify:

- Economic activities central to the energy transition and clean transportation that are currently excluded from LCREE and EGSS estimates
- Circular economy activities currently not captured in the UK's EGSS estimates

4.1 Clean transport and energy: technologies, systems and infrastructure.

The UK already collects considerable data on low carbon and renewable energy, including low emissions vehicles, through the Low-Carbon and Renewable Energy Economy (LCREE) survey. However, our review of this survey and the definitions used suggest that it is missing several important areas of activity that contribute to the achievement of the UK's clean energy transition.

4.1.1 Cleaner transportation: the cycling economy, electric transport services, and the rail system

Existing EGSS statistics take a limited view of what constitutes clean transportation. The only economic activities currently included are the manufacture of electric vehicles and the production of biofuels.

This omits large areas of transport-related economic activity that are cleaner than the dominant petrol-based road transport, meeting the criteria set out above. This section highlights three areas of activity that are consistent with the green economy definitions set out above.

Electric public transport services

The EGSS and LCREE accounts provide data on the manufacture of electric vehicles (including buses), biofuels production, and charging infrastructure. However, they do not capture rail-related output, despite electric rail being environmental preferable to both private road transport and diesel-powered rail. We have estimated the manufacturing of electric rolling stock as having a turnover of £2,992m. This is based on the turnover of SIC 302 from the Business Structure Database, multiplied by the average share of electric rolling stock in orders made between 2016-2019[18].

Public transport services that are electric, rather than running on fossil fuels, are also examples of environmentally preferable transport activities. In Table 4, we provide an estimate of turnover in electric public transport services.

Table 4. Electric public transport services

Activity	Turnover estimate (£m, 2021)	Data source
Electric bus services	393	BDS for SIC 49319, multiplied by the share of bus services that are electric (DfT)[19]
Electric rail services	7,154	Total turnover of SIC 4910, 4920 and 49311 from BDS, multiplied by share of rail that is electric from DfT (table ENV0102).
Electric taxi services	53	Total turnover of SIC 49320, multiplied by the share of battery electric vehicles in the taxi and private hire vehicle fleet (from DfT taxi statistics)[20].
Total	**7,600**	

The cycling economy

Cycling is poorly represented in existing economic statistics. There is no distinct SIC code for cycling-related retail and service activities (such as bicycle repair and maintenance), and the SIC code for manufacturing aggregates bicycles with "invalid carriages". The most comprehensive data on the cycling market in the UK is collected by industry group the Bicycle Association. In the below, we report their data in the context of other relevant independent data sources.

[18] Data on orders of electric and diesel rolling stock is available from The Rail Engineer: https://www.railengineer.co.uk/a-rolling-stock-retrospective-2004-to-2023/

[19] https://www.gov.uk/government/statistical-data-sets/bus-statistics-data-tables

[20] https://www.gov.uk/government/statistics/taxi-and-private-hire-vehicle-statistics-england-2024/taxi-and-private-hire-vehicle-statistics-england-2024#fuel-type-and-age-of-taxis-and-phvs

Manufacturing of bicycles in the UK is rather small. PRODCOM data for 2021 show sales from manufacturing of bicycles at around £63m in 2021. However, this neglects manufacturing of components, spare parts, and relevant accessories (such as helmets, locks and lights). The Bicycle Association suggests that total bicycle-related manufacturing in the UK was around £314m in 2021[21].

Data from the industry group the Bicycle Association suggests that wholesale and retail—and associated services including repair and maintenance—were around £3.4bn in 2021[22]. We used The Data City to identify bicycle-related manufacture, retail and wholesale businesses in the UK, and these businesses had a total turnover in 2021 of around £3bn. However, around half of total turnover in the businesses identified by The Data City was accounted for by three retailers that are not cycling specialists, but which sell bicycles alongside other goods, and it is not possible to disaggregate those figures to arrive at an independent estimate.

Estimates of cycling economic activities *as a green activity* are complicated by the fact that cycling is both a mode of transport and a leisure activity. While cycling is certainly a green mode of transport, it is not necessarily a green leisure activity. Including cycling within green economy estimates ideally requires identification of the share of cycling-related economic activity that is associated with transport, rather than leisure. One way of estimating the share of cycling-related purchases that are for the purpose of transport is to look at data from the government's cycle to work scheme. This scheme provides a tax incentive for bikes bought for community. Data provided by the Cycle to Work Alliance suggests that around one in four bikes purchased in the UK are bought through the scheme[23].

Drawing on the Bicycle Association data, and conservatively assuming that only one in four bicycles and associated accessories are primarily purchased for transport, we estimate around £1.5bn in economic activity related to cycling as a transport mode. This includes manufacturing, wholesale and retail (including repair services), but excludes bicycle infrastructure. This is a rough estimate for use in illustrating the relative scale of different areas of the green economy that are currently excluded from existing green economy data published by ONS.

4.1.2 Enabling systems for clean energy and transport: the electricity grid and rail infrastructure

In our revised definitions of green economic activity, we noted that it may be desirable to record the scale of activity in enabling systems, which by themselves are not necessarily green, but which are core enablers of a transition to a greener economy. Two are particularly relevant: rail infrastructure and the electricity grid.

Based on the turnover data from the Business Structure Database, the manufacture of electricity transmission and distribution equipment had a 2021 turnover of around £4.8bn, while the

[21] Hopkinson 2024. https://transportforqualityoflife.com/wp-content/uploads/2024/01/the-uk-cycle-industry-economic-and-employment-benefits.pdf
[22] Hopkinson 2024. https://transportforqualityoflife.com/wp-content/uploads/2024/01/the-uk-cycle-industry-economic-and-employment-benefits.pdf
[23] This is an underestimate, as the 'cycle to work alliance' data covers only around 80% of shops selling through the scheme. Nevertheless, it provides a conservative basis on which to make the estimate.

construction of railways was £2.75bn. Operation of the transmission and distribution grids was considerably larger, with a turnover of £63.4bn[24].

4.2 Circular economy

The UK's current EGSS accounts do not capture a large body of activity associated with activities associated with the achievement of a circular economy. Currently the EGSS accounts capture waste management activities, including landfill, energy-from-waste, and the sorting and processing activities that take place to turn waste streams into secondary raw materials.

Landfill is a waste management activity, and has a clear environmental purpose (preventing the leakage of waste into the environment). However, it has no place in an attempt to measure the circular economy, and for most materials is environmentally less preferable than circular material flows. In the estimates presented here, we exclude landfill, and we expand the range of activities beyond those captured in existing EGSS accounts. This includes:

- repair, leasing and rental activities (following the European monitoring framework for the circular economy)
- the share of manufacturing output of core materials that has been derived from secondary materials.

4.2.1 Re-use and repair

We have used both The Data City and the Business Structure Database to provide estimates of re-use and repair activities. Rental and leasing activities are typically recorded as part of a circular economy within 're-use', because they typically enable more frequent re-use, and a higher usage intensity than purchases. These are summarised in Table 5 below. The largest single category is the rental and leasing of machinery, equipment and tangible goods, which comprises leasing of agricultural equipment, office equipment, and other equipment often used as capital goods by other sectors[25].

Table 5. Estimates of turnover in 2021 in re-use and repair activities, in £m.

Route	Activity	Turnover	Data source
Re-use	Second-hand shops	3,198	Turnover from the Business Structure Database for companies with SIC 4779
	Online second-hand platforms	650	Estimate from list produced using The Data City
	Rental & leasing: household goods	1,818	Turnover from the Business Structure Database for companies with SIC 772
	Rental & leasing: machinery, equipment and tangible goods	19,023	Turnover from the Business Structure Database for companies with SIC 773
Repair	Personal computers and household goods	8205	Turnover from the Business Structure Database for companies with SIC 95

[24] These figures are based on the four-digit SIC codes: 3512, 3513, 4212 and the three digit code 271.

[25] These rental and leasing sectors have been included following the European monitoring framework for a circular economy – but further work to examine specific activities would be valuable, since it is not necessarily always the case that leasing models generate lower throughput of materials.

	Fabricated metal products, machinery and equipment	16579	Turnover from the Business Structure Database for companies with SIC 331
Total		**49,473**	

4.2.2 Accounting for the conversion of secondary raw materials into goods

The re-use of materials in a 'circular economy' is widely recognised to be environmentally preferable to an economy in which materials are extracted, used and then disposed of. However, traditional EGSS statistics have omitted many activities that are central to the re-use of materials. Current EGSS estimates include the activities captured in SIC 38.3 ('materials recovery'), but these activities are only one stage of the process required to turn end-of-life products into new products. For example, in the context of paper recycling, activities captured under SIC 38.3 relate to the sorting and separation of paper, and the bundling and baling of paper for shipment to a paper mill. The actual conversion of used paper into a new product—i.e. the step at the heart of a circular material flow—is the conversion of that used paper into new paper in the pulping process. The activity of the pulping process is not captured in SIC 38.3, but in SIC division 17 (the manufacturing of paper and paper products). The result is that current green economy statistics capture the activity of sorting and separating material for recycling, but the conversion of the material into a new product is excluded.

However, the processes of converting secondary raw materials into new goods are typically different, at least in some respects, from the processes for converting primary raw materials into intermediate goods. For example, steel-making with scrap is a different set of processes than steel-making from iron ore. Manufacturing from secondary materials is qualitatively different, observable, measurable and in the majority of cases has a higher environmental performance than production from virgin materials. We thus propose that such processes should be captured in estimates of green economic activity.

In Table 6 we provide an overview of the relative scale of economic activity captured in a 'circular economy'. This can compared to the 2021 turnover associated with SIC 38.3, which was just under £10bn. The process of estimation involves using the turnover of the relevant sector, multipled by the share of that sector's outputs that are recycled content. As can be seen, the current approach to measuring the green economy excludes £13bn of economic activity that is at the heart of a circular material system. A similar approach could be used for other sectors, such as aluminium.

There are weaknesses in the estimation approach used here – notably that the value of goods derived from secondary materials can be lower. The value-added may differ across them, and assuming that they are the same will lead to overestimation. However, it is straightforward way of providing an initial estimate.

Table 6. Turnover in 2021 associated with the conversion of secondary raw materials into new goods and services, in £m

	Estimation approach	2021 Turnover (£m)	Data
Steel	Share of scrap in total steel output	3,198	Total steel output and scrap steel input from UK Steel[26]; turnover from BDS SIC 241

[26] UK Steel Statistics 2022: https://www.uksteel.org/reports-and-publications

Glass bottles	Share of recycled content in glass bottle output	414	Share of recycled content in UK glass bottle manufacturing from Eunomia[27]; turnover from BDS SIC 23130
Paper and cardboard	Share of recycled content in feedstock	9,713	Share of recycled content in UK paper and card output from the Confederation of Paper Industries[28]; turnover from BDS SIC 17
Total		**13,325**	

4.3 Comparison with EGSS accounts

In this section, we provide some simple comparisons to illustrate how changes in the way that environmental goods and services are recorded could substantially change perceptions of the scale of such activities in the economy.

In 2021, the EGSS sector as reported by the ONS had a total output of around £105bn, of which the single largest component was the production of renewable energy at £26bn. The additional activities identified above, if included, would suggest a considerable increase in the scale of environmental goods and services. For that year, inclusion of the three categories pictured in Figure 5 below leads to an estimate of EGSS that is 70% bigger than the published estimate, at £180bn.

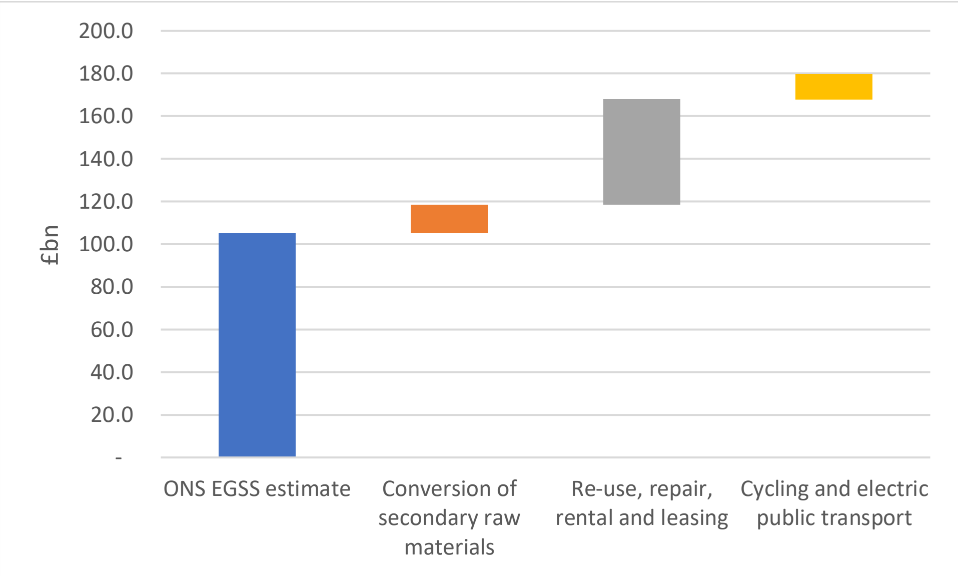

Figure 5. Scale of green activities previously excluded from UK EGSS estimates. Data for 2021. Data sources as described throughout the report section.

All three categories are made up of either 'environmentally preferable activities' or 'specialist inputs to environmentally preferable activities'. These results suggest that the consistent application of the revised definition presented in this report could substantially change the apparent scale of green economic activity within the UK.

[27] Eunomia 2022, "How circular is glass? A report on the circularity of single-use glass packaging, using Germany, France, the UK and the USA as case studies" https://eunomia.eco/reports/how-circular-is-glass/
[28] https://thecpi.org.uk/library/PDF/Public/Publications/Annual%20Reviews/CPIAnnualReview2022-23.pdf

The estimates presented here also suggest the need for revisions in the ONS Green Jobs estimates, for those based on 'green industries'[29]. It is worth noting that the ONS Green Jobs estimates demonstrate a broader approach to measuring green economic activity, including not only green activities (the 'green industries' estimates), but also firm-level measures (based on employment in firms with low greenhouse gas emissions) and occupation-level measures (based on a typology of green occupations).

5 Conclusions: new approaches to measuring green economic activity in the UK

This report has shown that the measured size and composition of the green economy depend materially on definitional and methodological choices. We propose a revised activity-based typology that distinguishes between main-purpose environmental protection activities, environmentally preferable goods and services, and specialised inputs and enablers. Applying this typology to selected activities currently excluded from UK EGSS estimates suggests that the 2021 estimate of green economic activity would rise from around £105bn to around £180bn, an increase of approximately 70%. This is not presented as a final official estimate, but it demonstrates that current measurement choices are likely to omit economically significant areas of activity.

Our analysis also shows both the usefulness and limitations of web-scraping tools, based on our assessment of The Data City platform. The tool clearly creates a cost-effective way of exploring activities that cut across existing sectors, and it is undoubtedly useful in identifying areas of activity that can inform survey strategy. But the analysis also highlights that the tool can result in mis-classification of firms in ways that are likely to influence aggregate estimates. The central finding is that using web-scraping tools well requires detailed work by sector experts to develop and test the training set(s) on which new industrial classifications are based. With sufficient expert input alongside the tool, The Data City (and similar tools) can be an accurate approach for identifying activities that are poorly represented by existing sectoral classifications.

The LCREE survey and DCNZ list have both become well established approaches widely used in the policy debate, and which measure closely related areas (i.e. activities related to the UK's transition to net zero through low-carbon and renewable energy). It is rather striking that they overlap rather little in terms of the firms identified, which illustrates the scale of the challenge in accurately measuring green economic activity in the UK. However, while the DCNZ list was informed by the LCREE sectors, it is not a bottom-up reproduction based exactly on LCREE. The lack of granularity in the LCREE made it difficult to implement a bespoke LCREE list within the DCNZ, but such efforts in future could result in a stronger alignment of the methods. In any case, we recommend that the LCREE survey be revised to capture greater granularity, and that it be updated to reflect changes in the relative environmental performance of key activities (notably, hybrid vehicles, compact fluorescent lighting, and condensing gas boilers).

[29] https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/experimentalestimatesofgreenjobsuk/july2025#jobs-in-green-industries

The practical implication is that ONS should treat web-scraped ML tools as part of a measurement system rather than as a replacement for existing statistical infrastructure. Their comparative advantage lies in discovery: identifying candidate firms, activities and SIC divisions that are poorly captured by existing classifications. Official estimates should continue to rely on survey and administrative data where activity-specific turnover, employment and value added are required. The most promising route is therefore a hybrid approach in which web-scraped classifications inform survey-frame design and targeted data collection, while official statistics provide calibrated estimates of economic scale.

We conclude this report by suggesting the following key actions for ONS:

1. **Revise existing publications and definitions in ways that are consistent with existing international practice**. While this report argues the need for changes to core definitions (in particular making a clear distinction between 'main purpose', 'environmentally preferable' and 'specialised inputs'), it is also recognised that the desire for international comparability may require continued publication of statistics produced on the basis of SEEA-CF guidance and definitions. There are several changes that ONS can and should make to existing statistics that are consistent with these definitions. There are relatively straightforward changes without major cost implications. This includes:
    a. Removal from LCREE-survey of activities that are no-longer environmentally preferable, and that can no longer be seen as having an environmental purpose (e.g. condensing gas boilers)
    b. Adjustment of EGSS accounts to include the activities discussed in section 4, which is consistent with the proposed indicative list recently published in the SEEA-CF 2026 consultation. It is also recommended that 'primary' and 'secondary' purpose activities are published separately. This would be consistent with SEEA-CF guidance, but would capture at least some of the key definitional distinction made in section 2.
2. **Improve the granularity of data collection via the LCREE and other surveys**. We recognise that a substantial increase in the granularity of data collection for surveys is likely too costly, both for the ONS to run, and in terms of the burden on respondents. However, we see value in a small increase in the granularity of data collection within the LCREE survey to capture the 3-4 main components of 'energy efficient products' – since without granularity it is not currently possible to interpret changes over time in this category.
3. **Use web-scraping tools to explore the scale of green economy activities** for which no other data sources currently exist, and for which surveys are not yet produced (such as alternative proteins).
4. **It would be desirable to work with international partners at the SEEA to revise international guidance, in two key ways:**
    a. Formalise the distinction between 'main purpose', 'environmentally preferable' and 'specialised input' forms of environmentally preferable goods and services. This would result in statistics that are more meaningful for policymakers, since the current aggregation of these different types of activities neglects key differences in their economic characteristics and relevance.
    b. Shift from a 'mutually exclusive' approach to categorising environmental purposes to a tagging system in which a given activity is recognised to have multiple environmental dimensions.

6 Acknowledgement

This research was funded by the Office for National Statistics as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). The authors are grateful to the ONS and the ESCOE anonymous reviewer for comments on drafts. They are also grateful to The Data City team for their support in using and interpreting their data.