Department for Science, Innovation and Technology
DSIT Research Paper Number 2024/001

Analytical Report
Identifying and describing UK Innovation clusters

alb@camecon.com
tt@camecon.com

Cambridge Econometrics
Cambridge, UK

www.camecon.com

Contents

Executive Summary 3

1 Cluster mapping project: Overview 7

1.1 Policy context and purpose of this project 7

1.2 How we define Innovation clusters 7

1.3 Our experimental method 9

2 How we identified different types of clusters 10

2.1 Definitions 10

2.2 Types of clusters 10

2.3 Overview of Data and Methodology 11

2.4 The Spatial Approach 14

2.5 The Network Approach 18

2.6 Summary of Results 21

3 Analysis of clusters 23

3.1 Introduction 23

3.2 Cluster composition 24

3.3 Cluster knowledge relationships 29

3.4 Cluster geography 33

3.5 Patterns of investment 39

4 Conclusion – How to use this new evidence and future research 43

References 45

Annex A – Data Analysis 47

Annex B – Sensitivity Analysis 52

Report written by Cambridge Econometrics:
www.camecon.com

Cambridge Econometrics Limited is owned by a charitable body, the Cambridge Trust for New Thinking in Economics:
www.neweconomicthinking.org

Executive Summary

Introduction

Strengthening research and innvoation across the UK is important to increasing UK productivity and prosperity. Many policy papers emphasise this, such as The Build Back Better Plan for Growth (2020), R&D Roadmap (2020), Innovation Strategy (2021), Levelling Up White Paper (2022) and Science and Technology Framework (2023). Place-based research, development, and innovation (RD&I) policy is also part of delivering the Levelling Up R&D mission and in strategic delivery plans across government departments and RD&I organisations such as the recent UKRI Five-year Strategy (2022). These plans recognise that the UK's competitive advantage derives from innovative places. The objective of delivering impact across the UK requires support to sustain and grow innovative places, as well as support to identify and unlock innovation potential in other places. These common themes have led to a strong focus on clusters across technologies and sectors.

Cluster mapping is an important step towards encouraging investment in RD&I from the private sector and international investors. With better mapped clusters government can direct public support and unblock the flow of private investment. Except for a few region-specific attempts, cluster mapping across the UK has not been attempted before. This was mainly because such an exercise requires a rigorous method to identify the location, technology and industry sector of firms.

The Department for Science, Innovation and Technology (DSIT) commissioned a consortium of leading data scientists and economists from Cambridge Econometrics, The Data City and Innovation Caucus to produce an **interactive mapping tool** of the UK's RD&I clusters**, accessible by businesses and members of public free at the point of use**. Readers can explore the tool at: www.innovationclusters.dsit.gov.uk.

This project applies innovative methodologies to identify and map the UK's RD&I clusters. The interactive tool and this accompanying report take a crucial step forward in building the evidence base in an important policy area. The tool showcases RD&I strengths and opportunities that exist across the UK for researchers, governments, businesses and potential investors. Although the tool is built on experimental data and has its limitations, it provides the most comprehensive picture of the RD&I clusters across the UK.

Defining and mapping clusters

Clusters are spatially concentrated groups of firms, research capabilities, skills, and support structures in related industries that benefit from spillovers associated with agglomeration.

Our approach sets four criteria to test whether aggregations of firms were innovation clusters. We used the best available data to find evidence that the groups of firms were:

1. RD&I-active,
2. Spatially co-located,
3. Engaged in related activities, for example within the same value chain or producing similar products,
4. Actively engaged in collaboration on public funded R&D projects, between organisations in the same group.

Typologies of clusters

For this project, we include only clusters which demonstrate evidence of meeting the criteria "RD&I-active" – thus, we refer to all clusters as innovation clusters. The innovation clusters are then labelled according to the following cluster types:

- "Diverse" innovation clusters – Co-located groups of firms that do not specialise in the same industrial sectors, and we were unable to identify solid evidence of collaboration within the cluster (meet criteria 2 but not 3 or 4).
- "Specialised" innovation clusters – Co-located groups of firms that specialise in the same industrial sectors, but we were unable to identify solid evidence of collaboration within the cluster (meet criteria 2 and 3).
- "R&D Collaborating" innovation clusters – Co-located groups of firms where solid evidence of collaboration within the cluster have been identified (meet criteria 2 and 4).
- "Dispersed" innovation communities – Groups of firms where evidence of collaboration within the cluster have been identified, but are spatially dispersed, i.e., not Co-located in a single place (meet criteria 4). These are best conceived of as collaboration "communities".

Given the best available data at the time of this report, collaboration is evidenced by joint work on government (research council) funded R&D projects. Supplementing the evidence with other types of collaboration such as patent applications is for future research.

Table 1 Four types of clusters and the criteria they satisfy

Innovation Clusters	RD&I-active	Co-located	Engaged in related activities	Internally Collaborative
Diverse	✓	✓	✗	✗
Specialised	✓	✓	✓	✗
R&D Collaborating	✓	✓	✓/✗*	✓
Dispersed	✓	✗	✓/✗*	✓

*While R&D Collaborating clusters and Dispersed communities may or may not meet the Specialised criterion, due to their cross-sector collaborations they warrant further exploration for their unique relationships.

Data and methodology

While taking this challenge on board, the research team acknowledges that data is never comprehensive and it is not possible to identify "all" innovation clusters with the current data. Instead, we took a pragmatic approach and identified as many clusters as possible with the available data. The clusters we present in the tool reflect data available in January 2023. We combined 5 different datasets to classify groups of organisations based on the four criteria described above:

- IDBR – 2018 Inter-Departmental Business Register dataset: 3.1 million business sites classified by 32 broad Standard Industrial Classification (SIC) code sectors.

- RTIC – 2023 Real-Time Industrial Classifications dataset:
  5 million firms classified by 46 emerging RTIC sectors identified by The Data City.
- IUK – Innovate UK dataset:
  46 thousand collaborators and 111 thousand project funding applications from innovative firms and other organisations from Sep 2016 to Jan 2023.
- UKRI – UK Research and Innovation dataset:
  24 thousand funded project from Sep 2016 to Jan 2023.
- MAKG – Microsoft Academic Knowledge Graph dataset:
  239 million scientific publications with co-authors.

All business sites are allocated an even share of the business's total estimated turnover and employee count, weighted by employee count on each site if the data is available. Funding received is allocated equally across all collaborators and then allocated equally across all sites unless specified by the dataset.

To utilise these datasets in a comprehensive manner, we chose to use two complementary approaches. The "Spatial approach" identifies spatial groupings of business sites with similar economic activities around the country using IDBR and RTIC data. The "Network approach" instead identifies network communities of collaborative firms using Innovate UK funding co-applications data.

In total, 3,443 innovation clusters were identified in the UK (of 10 or more firms). There are 429 R&D Collaborating clusters, 2,901 Specialised clusters, 7 Diverse clusters, and 106 Dispersed communities.

Cluster analysis and findings

The cluster mapping tool and database enable a wide range of comparative analysis not previously possible at this scale. Because the dataset is so vast, we highlight an indicative selection of insights in this report on:

- Composition – organisation and sectoral make up,
- Knowledge relationships – collaboration and overlaps,
- Geography – regional analysis of types of clusters,
- Patterns of Investment.

Cluster composition

- Cluster size is closely tied to factors such as the overall size of the industry (by number of business sites), and how tightly defined the industrial classification is.
- There is a low correlation between the number of clusters per sector and the average turnover or employment per sector. This suggests that sector size and cluster size are not suitable indicators of cluster success or potential.
- Clusters that collaborate on IUK R&D projects are correlated with higher average turnovers across a wide variety of sectors.

A broader range of indicators (such as level of collaboration) should be explored when identifying cluster potential for growth, and craft interventions based on industry-specific expertise.

Cluster knowledge relationships

- There are strong collaborations between clusters of the same sector. There is also a significant degree of interaction between clusters across related sectors, notably the Engineering sector.

- Strong collaboration networks span across long distances of at least 150km between research intensive universities, research organisations, and firms.
- Greater London emerges as the core knowledge “hub” in collaborative relationships. Other knowledge “hubs” include Hampshire, Bristol, Birmingham and Manchester.
- Outside of these research-intensive poles, places tend to cooperate more frequently with partners in their own broad geographic region.

While regional proximity and sectoral similarity promotes collaboration between clusters, firms are willing and able to span long distances to engage in research-intensive “hubs”. Thus, it is important to consider wider collaboration networks when developing policies to boost knowledge sharing across sectors/places.

Cluster geography

- Large cities are able to host dense clusters containing large numbers of firms. Across the UK, all regions host the top-3 largest cluster (in the number of firms) for at least one sector. For example, London hosts the largest cluster in 43 sectors; and the North West region hosts the top-3 largest clusters in 39 sectors.
- The type and location of clusters largely depend on the resource requirements (such as labour and land) of the sectors they belong to: resource-extracting sectors such as mining or offshore power generation mostly occur in limited geographies and cover wider territories; knowledge intensive activities (both services and non-services) tend to require access to large and diverse labour pools and therefore concentrate in denser urban environments and occupy a smaller spatial footprint.
- Industrial co-location generally occurs because sectors have resource needs that are satisfied by similar locations. Industries that need more space to operate are more likely to co-locate in suburban locations where office parks and manufacturing sites are more prevalent.

Regional support to satisfy resources requirements such as talent pool, land space and accessibility is crucial to build and sustain clusters.

Patterns of investment

- The unevenness of public R&D funding distribution across sectors reflects the different propensities and levels of competence across sectors in seeking innovation funding.
- Emerging sectors like Fintech received more private than public funding per firm, whereas the opposite is true for traditional sectors like Machinery.
- Sectors with greater levels of positive social externality (such as social work and care, arts and recreation, water and waste) receive relatively higher levels of public compared to private funding. Whereas in sectors where consumer surplus capture is easier (such as E-commerce and Advertising) receive relatively more private sector funding.

The tool that we have produced provides a good indication of where cores of clustered activities are located and estimates across a variety of indicators. It is particularly useful for identifying industry hotspots and places with growth potential. It will help places understand their strengths – some of which might come as a surprise – and be a useful tool for developing innovation strategies that will increase and leverage potential synergies. This data driven analysis needs to be supplemented with local intelligence and qualitative research to explore enablers and barriers to cluster growth and productivity.

1 Cluster mapping project: Overview

1.1 Policy context and purpose of this project

The Build Back Better Plan for Growth (2020), R&D Roadmap (2020), Innovation Strategy (2021), Levelling Up White Paper (2022), and Science and Technology Framework (2023) all emphasise the importance of strengthening research and innovation as mechanisms for the UK to increase productivity and prosperity. These expectations are further developed in strategic delivery plans across government departments and RD&I organisations such as the recent UKRI Five-year Strategy (2022), which recognises that the UK's competitive advantage derives from innovative places and that supporting the objective of delivering impact across the UK requires support to sustain and grow innovative places, as well as identify and unlock innovation potential in other places. These common themes have led to a strong focus on clusters.

Clusters are spatially concentrated groups of firms, research capabilities, skills, and support structures in related industries that benefit from spillovers associated with agglomeration. These dynamics have been found to boost local and regional innovation and productivity which are thought to be crucial to initiatives to improve regional productivity growth and smooth spatial inequalities (Sunley et al. 2022). Clusters are also key to increasing regional resilience and speed of rebound from shocks and downturns (Delgado and Porter 2021). A critical mass of specialised expertise with high potential for innovation is an important signal in competitive markets and can help attract talent, foreign investment, and public funding. In short, clusters are seen as mechanisms that can catalyse, and maintain, virtuous cycles of growth that can generate important place-specific and economy-wide benefits.

However, the first step to support clusters is to identify and describe them. To date, there have been few attempts to map innovation clusters across a wide variety of industrial sectors and across the entire geography of the UK.[1] This project is a large-scale effort to apply innovative methodologies and triangulate cutting-edge data to map the UK's Research, Development and Innovation (RD&I) clusters and provide a tool for investment, research, and policy communities to better understand cluster growth potential. This report and the accompany interactive mapping tool (explore at www.innovationclusters.dsit.gov.uk) represents a crucial asset to progressing place-based growth strategies, contributing to the levelling up agenda, and raising the international profile of the UK's clusters.

1.2 How we define Innovation clusters

Following the definition in Nelles et al. 2022, a cluster is a "group of firms and intermediary organisations involved in related activities and that derive individual and collective benefits from co-location with each other such as through access to shared knowledge bases, labour markets, specialised services, infrastructure, support services, training and other industry-specific pooled resources".

The literature related to clusters stresses that while quantitative approaches to defining clusters are frequently sought, it is critical to

[1] Notable exceptions tend to focus specifically on mapping clusters or hotspots in urban areas (e.g., Centre for Cities 2023, 2014), focus only on specific industries (e.g., Cottineau 2020), or focus on qualitative case studies (e.g., Nelles et al. 2022, 2023).

understand that clusters are more than the sum of their statistical parts. A cluster is more than just a critical mass of “the right ingredients” – such as firms and assets – but is also a function of networked relationships and flows and exchange of resources between firms and other localised assets. In other words, agglomeration alone is not a guarantee that the particular types of benefits associated with clusters will materialise or be optimised. This is among the reasons that data-driven approaches to cluster identification are so difficult. The presence of agglomeration of something (an industrial sector, for example) can be relatively easy to detect. The degree to which that agglomeration is resulting in the clustering dynamics that enhance innovation outcomes is much trickier to determine. Even more difficult is determining quantitatively which agglomerations may have the *potential* to develop those dynamics or are earlier in that evolutionary process.

Our approach sets four criteria to test whether aggregations of firms were innovation clusters. We used the best available data to find evidence that the groups of firms were:

1. RD&I-active,
2. Spatially co-located,
3. Engaged in related activities, for example within the same value chain or producing similar products,
4. Actively engaged in collaboration on public funded R&D projects, between organisations in the same group.

For this project, we include only clusters which demonstrate evidence of meeting the criteria “RD&I-active” – thus, we refer to all clusters as innovation clusters. The innovation clusters are then labelled according to the following cluster types:

- “Diverse” innovation clusters – Co-located groups of firms that do not specialise in the same industrial sectors, and we were unable to identify solid evidence of collaboration within the cluster (meet criteria 2 but not 3 or 4).
- “Specialised” innovation clusters – Co-located groups of firms that specialise in the same industrial sectors, but we were unable to identify solid evidence of collaboration within the cluster (meet criteria 2 and 3).
- “R&D Collaborating” innovation clusters – Co-located groups of firms where solid evidence of collaboration within the cluster have been identified (meet criteria 2 and 4).
- “Dispersed” innovation communities – Groups of firms where evidence of collaboration within the cluster have been identified, but are spatially dispersed, i.e., not Co-located in a single place (meet criteria 4). These are best conceived of as collaboration “communities”.

Given the best available data at the time of this report, collaboration is evidenced by joint work on government (research council) funded R&D projects. Supplementing the evidence with other types of collaboration such as patent applications is for future research.

1.3 Our experimental method

The datasets and tools developed in this study are one of many representations of reality, each with their own unique advantages and limitations. We urge users interacting with and interpreting these outputs to proceed with the following considerations in mind:

- **Judgements and Thresholds**: Mapping clusters requires a combination of art and science. While we have designed and applied a rigorous methodology consistently (elaborated below), every stage of the process required decisions that could have been made differently. Altering any part of that process would have delivered slightly different results. We made every effort to ensure that those decisions were made based on the state of the art and to combine different processes to give a more holistic view. However, it is important to recognise that judgement plays an important role and that, at times, it can involve more art than science. One of the most prominent examples of the judgment calls described in the previous point is in setting thresholds – of distances between firms, of minimum number of participants in clusters to be considered, of minimum strengths of ties, to name a few. Different thresholds give different results, which is why we felt it was important to include all types of clusters. We have included the sensitivity analysis of these thresholds in Annex B.
- **Limitations of Data:** We can confidently say that our mapping tool shows where clustering activities are taking place. However, we cannot claim to have included every firm or that the lines on the map tell the sum total of any place or industrial sector's story. We were systematically looking for evidence of clusters, but *absence of evidence is not evidence of absence*. Data is an imperfect sample of reality, and there is some data loss in converting between data sets, and each source has its own unique advantages and limitations. Adding new data sources could improve the robustness of the results and this is a route for future improvements to this project as it evolves. An example would be the addition of patent applications to supplement the evidence of collaboration. See Annex A for the detailed discussion of methodology and caveats during data cleaning, merging, and analysis.

2 How we identified different types of clusters

2.1 Definitions

The project aims to identify innovation clusters – groups of related organisations (such as firms, universities, or Catapults). Four criteria were selected to identify types of clusters:

- RD&I-active – Groups of firms that undertake Research, Development and Innovation
- Co-located – Office sites of firms that form a single spatial grouping
- Specialised – Majority of firms from the same industrial sectors
- Internally Collaborative – Evidence of collaboration between organisations in the same cluster

Given the best available data at the time of this report, internal collaboration is evidenced by joint work on government (research council) funded R&D projects. Note that the absence of evidence for collaboration is not evidence of absence in collaboration.

2.2 Types of clusters

For this project, we include only clusters which demonstrate evidence of meeting the criteria “RD&I-active” – thus, we refer to all clusters as innovation clusters. The innovation clusters are then labelled according to the following cluster types:

- Diverse clusters – Co-located groups of firms that do not specialise in the same industrial sectors, and we were unable to identify solid evidence of collaboration within the cluster.
- Specialised clusters – Co-located groups of firms that specialise in the same industrial sectors, but we were unable to identify solid evidence of collaboration within the cluster.
- R&D Collaborating clusters – Co-located groups of firms where evidence of collaboration within the cluster have been identified.
- Dispersed communities – Groups of firms where evidence of collaboration within the cluster have been identified, but are spatially dispersed, i.e., not Co-located in a single place. These are best conceived of as collaboration “communities”.

Table 2.1 Four types of clusters and the criteria they satisfy

Innovation Clusters	RD&I-active	Co-located	Specialised	Internally Collaborative
Diverse	✓	✓	✗	✗
Specialised	✓	✓	✓	✗
R&D Collaborating	✓	✓	✓/✗*	✓
Dispersed	✓	✗	✓/✗*	✓

*While R&D Collaborating clusters and Dispersed communities may or may not meet the Specialised criterion, due to their cross-sector collaborations they warrant further exploration for their unique relationships.

2.3 Overview of Data and Methodology

There is no definitive way to identify Innovation clusters; different approaches of equal conceptual validity may provide a very different set of answers. Our approach was informed by the fact that the data is never fully comprehensive and there is no way to identify "every" innovation cluster in the UK. We have identified as many clusters as the best data allows in as consistent a manner as possible.

All data in this research used in identifying clusters were extracted in January 2023. Updates to organisations, applications, and projects beyond this date are not included in the analysis. Further details on the data and methodology can be found in Annex A.

We combined 5 different datasets to classify groups of organisations based on the four criteria:

- IDBR – 2018 Inter-Departmental Business Register dataset:
  3.1 million business sites classified by 32 broad Standard Industrial Classification (SIC) code sectors.
- RTIC – 2023 Real-Time Industrial Classifications dataset:
  5 million firms classified by 46 emerging RTIC sectors identified by The Data City.
- IUK – Innovate UK dataset:
  46 thousand collaborators and 111 thousand project funding applications from innovative firms and other organisations from Sep 2016 to Jan 2023.
- UKRI – UK Research and Innovation dataset:
  24 thousand funded project from Sep 2016 to Jan 2023.
- MAKG – Microsoft Academic Knowledge Graph dataset:
  239 million scientific publications with co-authors.

In the IDBR dataset, there are 2.3 million headquarters and 0.8 million local unit, both of which are treated as office sites. See Table 2.2 and Table 2.3 for the number of sites in each broad SIC code and RTIC sector. All sites are allocated an even share of the business's total estimated turnover/employee count, weighted by employee count on each site if the data is available (see Annex A). Suppose a firm has 10 million turnover, its two sites of 70 and 30 employees would be allocated 7 and 3 million turnover, respectively.

To utilise these datasets in a comprehensive manner, we chose to use two complementary approaches. **The "Spatial approach" identifies spatial groupings of sites with similar economic activities around the country using IDBR and RTIC data**, and then testing these groupings for evidence of a) innovation and b) internal collaboration. Note that a firm can appear in more than one cluster due to having several sites, allowing for clusters with shared members. This Spatial approach provided us with the majority of our final list of R&D Collaborating clusters, as well as the long list of Specialised clusters. We also repeat this spatial procedure for IUK applicants without pre-categorising by sector – this allowed us to identify Diverse clusters.

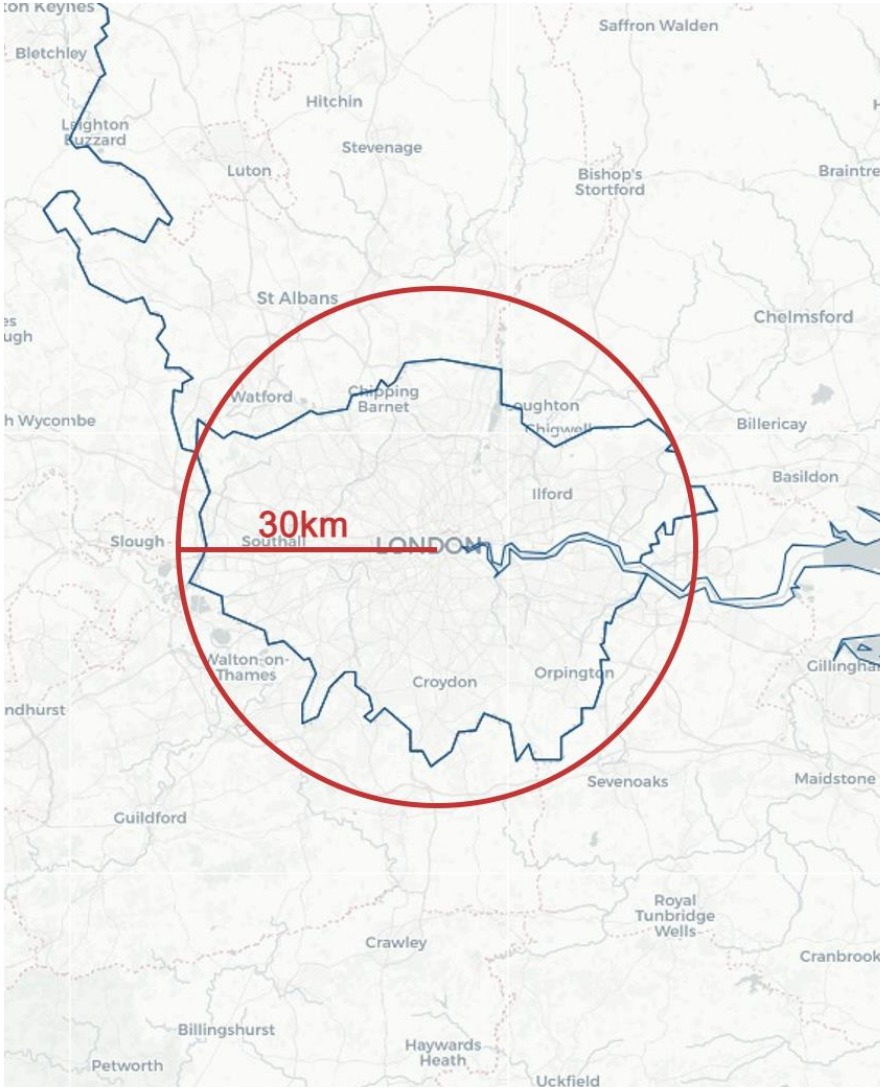
The "Network approach" instead identifies network communities of collaborative firms using Innovate UK funding co-applications data. We then tested these communities for spatial co-location and industrial relatedness. This didn't identify many R&D Collaborating clusters, as the vast majority of collaboration communities were dispersed across space, but did allow us to identify the Dispersed communities. The approach essentially reverses the steps in the spatial approach to locate collaborative clusters first before filtering by other criteria. Details of both approaches follow in the sections below.

The technical definitions of each criteria used are as follows:

- RD&I-active – Clusters identified from RTIC, IUK and MAKG are assumed to be RD&I-active, due to the nature of their activity in emerging sectors and participation in research and innovation. IDBR clusters are RD&I-active if at least 50 sites (or 20%) in the cluster have a non-zero Innovation Score[2], or at least 20 sites (or 10%) have been active in applying for IUK funding.
- Specialised – Clusters identified from IDBR and RTIC datasets consist of firms pre-classified by broad sectors and are therefore Specialised in related sectors. Other clusters are Specialised if at least 30% of the organisations in the cluster belong to one broad SIC code sector.
- Co-located – Clusters identified through the spatial clustering algorithm are considered as Co-located. Other clusters are Co-located if the average radius of the cluster is at most 30km, equivalent to the radius of Greater London (see Figure 2.1).
- Internally Collaborative – if there are at least 5 pairs of organisations within the cluster that co-applied for IUK funding. The cluster is "Marginally Collaborative" if there is at least 1 pair. Collaboration communities identified through the community detection algorithm are also considered as Internally Collaborative. All clusters that satisfy conditions for Internally Collaboration also satisfy the RD&I-active criterion by definition, but not vice versa.

See Table 2.6 for a summary of datasets and methodologies on classifying clusters using these criteria. Discussion on the selection of thresholds and sensitivity analysis are in Annex B.

Figure 2.1 Map of London demonstrating the cluster radius satisfying the Co-location criteria



[2] Innovation Score is developed by The Data City. See the following section for definitions.

Table 2.2 Number of firms by broad SIC code sectors

Broad SIC code Sector	No. Firms
Accommodation	22,957
Arts and Recreation	156,533
Bioscience	11,806
Business Support Services	254,585
Chemicals and Materials	21,905
Construction	357,019
Electricity	7,982
Electronic Devices	27,158
Engineering	129,192
Finance	131,255
Food and Agriculture	163,600
Healthcare	71,778
Higher and Further Education	36,780
Hospitality	192,549
Legal & Accounting	73,325
Logistics & Freight	122,674
Machinery	39,233
Management and Social Science	195,918
Media and Publishing	97,517
Metal Products	36,000
Mining and Extraction	5,140
Primary and Secondary Education	51,459
Public Administration	24,375
Real Estate	147,664
Retail	304,517
Social work and Care	78,808
Software and IT	140,888
Textiles Products	25,127
Transport Equipment	97,065
Transport Services	16,171
Water & Waste	13,877
Wood Products	46,468
Grand Total	**3,101,325**

Table 2.3 Number of firms by RTIC sectors

RTIC Sector	No. Firms
AdTech	872
Advanced Manufacturing	10,096
Advanced Materials	2,403
Agency Market	13,284
AgriTech	1,556
Artificial Intelligence	2,911
Autonomy and Robotics	1,012
CleanTech	4,829
Computer Hardware	1,513
Cryptocurrency Economy	1,555
Cyber	7,879
Data Infrastructure	14,286
Data Intermediaries	974
Design and Modelling Technologies	1,978
Digital Creative Industries	13,265
E-Commerce	2,615
EdTech	1,486
Electronics Manufacturing	7,009
Energy Generation	5,124
Energy Management	1,014
Energy Storage	3,786
FinTech	8,242
Food Tech	2,292
Gaming	2,436
Geospatial Economy	4,618
Immersive Technologies	2,132
In-Orbit Space Manufacturing	1,677
Internet of Things	978
Life Sciences	25,180
MedTech	1,653
Modular Construction	1,014
Net Zero	22,070
Omics	1,329
Pharma	2,358
Photonics	1,064
Quantum Economy	557
Research & Consulting - Physics & Engineering	14,248
Sensors	3,243
Software as a Service (SaaS)	4,323
Software Development	8,111
Space Economy	2,732
Space Energy	1,881
Streaming Economy	791
Supply Chain Logistics	4,112
Telecommunications	2,680
Wearables and Quantified Self	442
Grand Total	**219,610**

2.4 The Spatial Approach

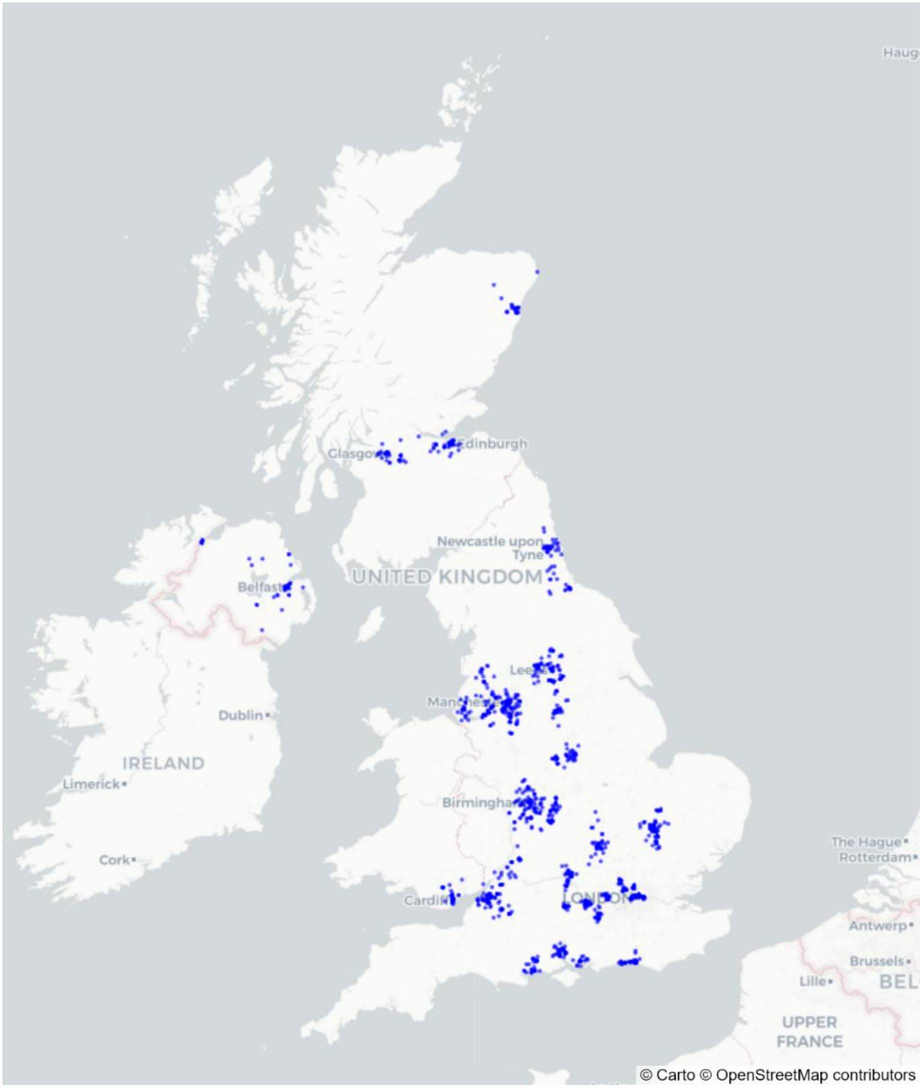
First, we identify spatial clusters for the IDBR and RTIC datasets:

1. Organisations are partitioned into 32 and 46 broad sectors of related economic activities based on SIC codes and RTIC emerging sectors, respectively. See Table 2.2 and Table 2.3 for the number of firms in each sector.
2. For each sector, we create a spatial map using the geospatial coordinates of the IDBR and RTIC business sites. All sites are allocated an even share of the business's total estimated turnover/employee count, weighted by employee count on each site if the data is available.
3. Lastly, we identify spatial clusters using the HDBSCAN clustering algorithm[3] on each sector's spatial map.

HDBSCAN clustering algorithm is a density-based algorithm – it assumes clusters for densely located firms and classified the sparsely located as 'noise'. The HDBSCAN algorithm has a few key advantages:

- Compared to other clustering methods like K-means and Spectral clustering, HDBSCAN does not require that every organisation be assigned to a cluster and hence does not partition the data. Therefore, organizations are only classified as a cluster if and only if they are truly densely co-located.
- HDBSCAN identifies clusters with varying density – clusters in sparsely populated areas are identified along with those in major cities. Whereas algorithms like DBSCAN are either going to fail to identify them, split them up, or lump some of them together depending on the selection of density threshold.

Figure 2.2 RTIC Artificial Intelligence spatial clusters identified using HDBSCAN



Source: RTIC dataset from The Data City

[3] How HDBSCAN Works — hdbscan 0.8.1 documentation

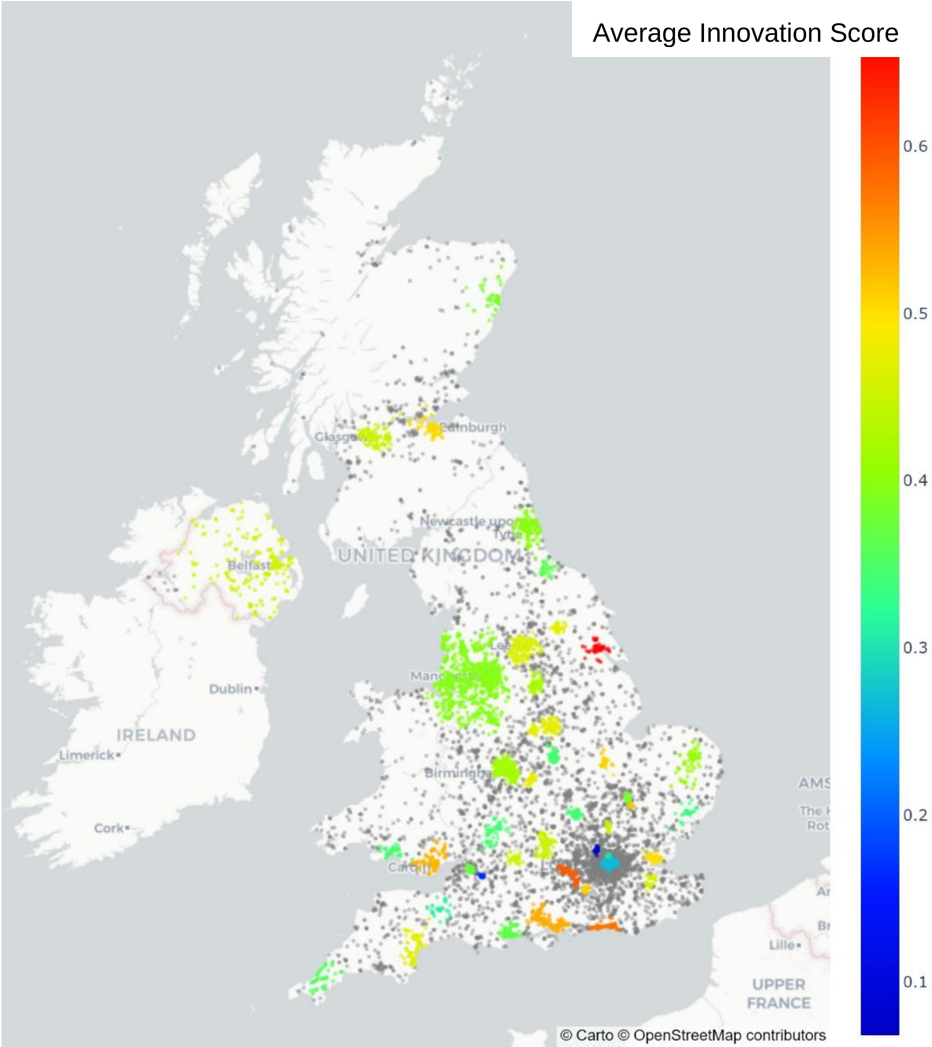
- HDBSCAN only has one parameter which determines whether organisations are 'falling out of a cluster' or splitting up to form two new clusters. The parameter lets us select approximately the minimum number of organisations in a cluster.

One limitation of HDBSCAN is that geographically separated areas will often be identified as a single cluster, making it more difficult to distinguish potential clusters. In the case of the UK, this shows up most prominently in Northern Ireland, which, for example, is identified to be a single large cluster in the IDBR Bioscience sector because it is separated from Great Britain by the Irish Sea. While the Northern Ireland cluster is an artifact of the methodology, the large Northwest England cluster which spans several cities is not: this is simply the true pattern of the actual geospatial grouping of firms in this sector.

For the IDBR dataset, the HDBSCAN parameters are based on the number of organisations in each sector – a higher parameter value is selected for a denser sector (with more firms), resulting in clusters that are spatially denser. The parameter ranges from 23 for the least populous sector (Mining and Extraction sector) to 199 for the most populous (Construction sector) in IDBR dataset. For RTIC sectors, the parameters are selected based on expert knowledge. Selecting a low parameter value means fewer number of firms are required for the collective to be classified as a cluster. For a sector with many firms, a low parameter value would result in many small clusters. In contrast, for a sector with less firms, a high parameter value would result in one large cluster spanning the country. See Annex B for more technical details and sensitivity analysis on the parameter selection.

5,565 spatial clusters were identified using HDBSCAN: 3,965 from IDBR and 1,600 from RTIC. Because of the nature of their activity in emerging sectors, we assume that all RTIC clusters are RD&I-active. However, this is not necessarily true of spatial clusters identified from

Figure 2.3 IDBR Bioscience spatial clusters, coloured by Innovation Score

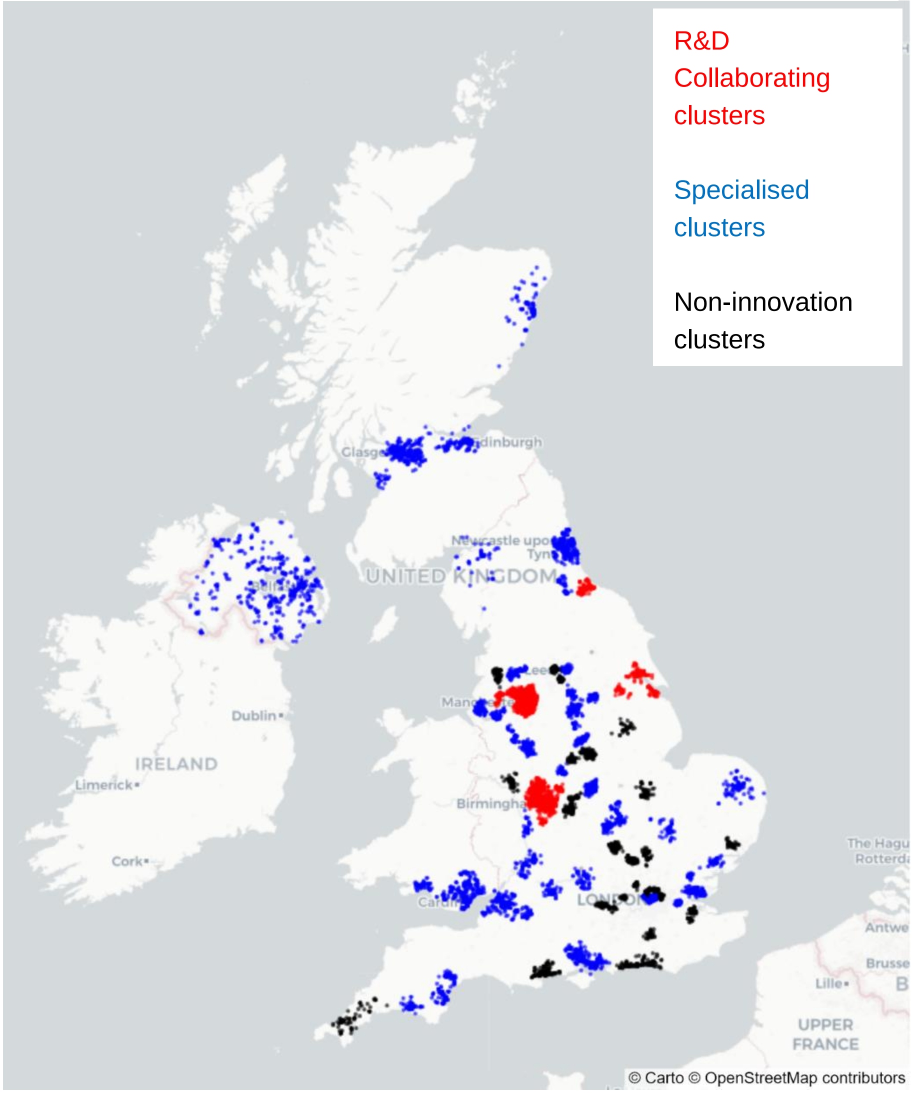


Sources: IDBR dataset analysed by Cambridge Econometrics, using Innovation Score from The Data City

IDBR data. To filter our co-located IDBR clusters down using RD&I-active criteria, we appended to organisations their Innovation scores[4] – A confidence ranking in the evidence of innovation from a TDC's data model trained on 800 firms with known R&D intensities. To augment this, we also treat organisations as RD&I-active if they appear in the IUK application dataset. If a cluster has less than 50 organisations with an Innovation score or less than 20 belonging in the IUK dataset, it is treated as non-RD&I-active. **After filtering, it leaves 3,229 innovation spatial clusters (1,629 IDBR and 1,600 RTIC).** See Figure 2.3 for a spatial map on the spatial clusters identified in the IDBR Bioscience sector. Each coloured dot on the map represents a firm's operational location. Firms are coloured by their cluster's Average Innovation Score, whereas non-clustered firms are in grey.

To check whether these clusters are also Internally Collaborative on R&D projects, we counted the number of organisation pairs collaborating on public R&D grants within each cluster. This is done by matching firms from IDBR and RTIC datasets to IUK grant collaboration data using their Company Reference Number (CRN). As an indicator for collaboration, the number of collaborating pairs was selected over the total number of collaborations because the latter is biased by repeated collaborations among the few. Note that this approach is currently reliant on a relatively sparse dataset and provides evidence of only one type of collaboration. Furthermore, it was not possible to identify which specific sites of a multi-site firm were actively involved in collaboration: this may have introduced false positives into the results. Additional datasets on other types of collaboration such as patents or site-by-site collaboration would be an improvement on this approach.

Figure 2.4 IDBR Chemicals and Materials clusters, coloured by cluster type



Sources: IDBR dataset, supplemented with IUK dataset and Innovation Score, analysed by Cambridge Econometrics

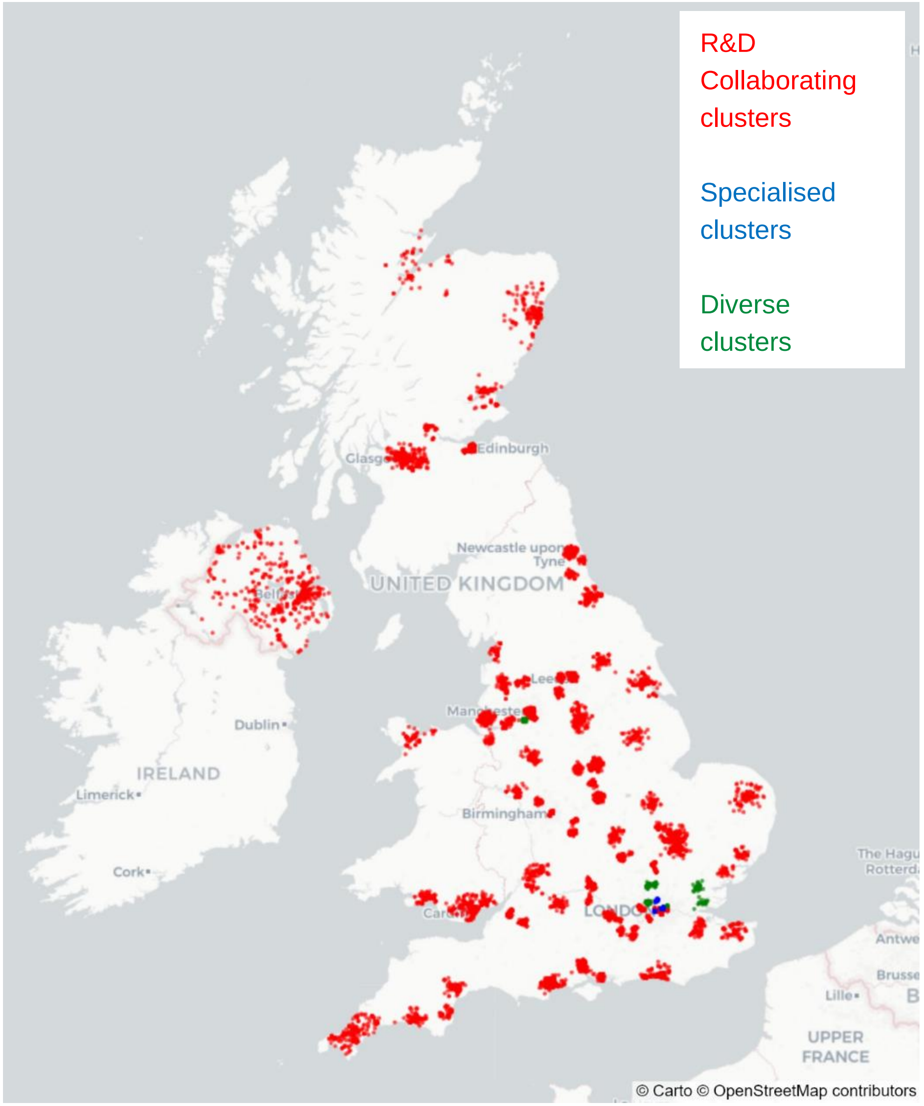
[4] Introducing our company innovation measure - The Data City

These caveats aside, we found that **251 IDBR and 85 RTIC clusters met the criteria for Internal Collaboration. In total, there are 336 R&D Collaborating clusters and 2,893 Specialised clusters.** See Figure 2.4 for the clusters in IDBR “Chemicals and Materials” sector.

To identify Diverse clusters, a similar Spatial approach is performed on the set of organisations in IUK and MAKG datasets, but this time *without* pre-categorising the datasets into industrial sectors. The HDBSCAN parameter was selected based on the number of organisations in the IUK and MAKG datasets (See Annex A for technical details). The clusters were then tested for Internal Collaboration. **94 IUK and 14 MAKG spatial clusters were identified, of which 79 IUK and 14 MAKG clusters met the criteria for Internal Collaboration, adding 93 R&D Collaborating clusters.** See Figure 2.5 for the IUK spatial clusters.

Using the CRN provided in the IUK (and MAKG) dataset, we mapped the associated SIC codes for each organisation and classified them into the 32 broad SIC code categories as discussed above. If the majority of firms in a cluster belong to the same industrial sectors, the cluster is Specialised.[5] **8 IUK clusters are Specialised but not Internally Collaborative. Thus, there are 8 IUK Specialised clusters and 7 IUK Diverse clusters** (Figure 2.5). IUK spatial clusters are presented in the accompanied interactive mapping tool while MAKG spatial clusters are excluded from the tool due to their focus on academic institutions.

Figure 2.5 IUK spatial clusters, coloured by cluster type



Source: IUK dataset analysed by Cambridge Econometrics

[5] Academic institutes are classified as the “Higher and Further Education” sector which is excluded when identifying the largest sector in a cluster to prevent skewing the sector speciality of a cluster.

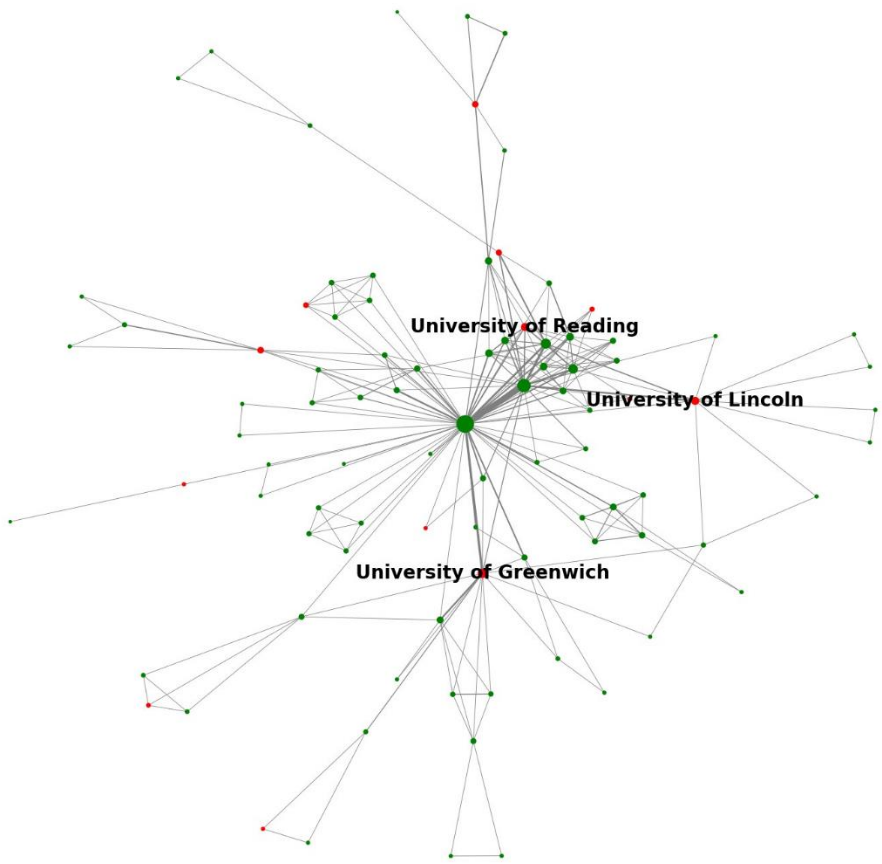
2.5 The Network Approach

In parallel with the spatial clustering approach, we also performed network analysis on the IUK project funding application data (and MAKG publication data). Collaboration networks between co-applicants (and co-authors) were created after identifying university departments and removing academic only collaborations. The resulting collaboration network based on the IUK dataset contains about 48 thousand organisations with 88 thousand collaboration pairs. Every pair included at least one firm with the other partners either being firms, universities or other R&D organisations. The most frequent collaboration pair of organisations collaborated 28 times between 2016 and 2023.

We analysed the network of organisations that collaborated with at least two other organisations at least twice, by removing one-off collaborations and one-off collaborators. This method removes a lot of the noise out of the collaboration data and make it easier to see the most important and frequent connections. The resulting collaboration network based on the IUK dataset has 222 groups of partnerships (of at least size 3). See Annex A for the descriptive statistics of the network.

On the filtered collaboration network, we identified collaboration communities using the Louvain algorithm[6] (Blondel et al. 2008) for community detection. This approach partitions the network in order to maximise "modularity" – high modularity means having dense connections between nodes within communities but sparse connections across communities. The algorithm starts with assigning each node in a network to a unique community. Then, each node is removed from its community and placed into another neighbouring community in order to maximise modularity. Repeat this step for all nodes until modularity

Figure 2.6 Collaboration network of the second largest Louvain community



Source: IUK dataset analysed by Cambridge Econometrics

Notes: Link thickness represents collaboration frequency. University and firm nodes are coloured in red and green. Top 3 university collaborators are named.

[6] Louvain — scikit-network 0.30.0 documentation

does not increase any further. Afterwards, each community is regarded as a single node. Then, the entire procedure is repeated until no further improvement of modularity is achieved.

The Louvain algorithm provides a fair compromise between low computational complexity and high accuracy in estimating modularity maximum. Other algorithms such as the Edge Betweenness (EB) algorithm (Girvan and Newman 2002) were computational too expensive to run on the large collaboration network.

99 IUK collaboration communities (with at least 10 members) were identified using the Louvain algorithm. Figure 2.6 shows the collaboration network of the second largest community where nodes represent collaborators and links collaborations. It presents a hub-spoke structure where a single firm facilitates collaboration across many other firms and academic institutions. We repeat the analysis on the MAKG dataset to identify **10 MAKG collaboration communities**. By definition, they are RD&I-active and inwardly Collaborative. There are only a few MAKG collaboration communities because the dataset consists of mostly academic publications between a small number of UK academic and medical institutions. Despite only having a total of 1,472 organisations, there are more than 16 thousand pairs of collaboration within the MAKG clusters, collaborating 6 million times.

96 IUK and 10 MAKG communities are Internally Collaborative, none of which are Co-located. This resulted in 106 Dispersed communities and no additional R&D Collaborating clusters. These Dispersed communities cannot be represented on the interactive map.

To better understand the research conducted within a cluster, we use Natural Language Processing techniques to study the project texts. We first collecting the titles, descriptions and abstracts of the projects where the cluster members collaborated in. Then we removed the stop words and lemmatised the words.[7] Lastly, we identified keywords with the highest frequency of occurrence.

Using the project keywords, we were then able to classify the IUK collaboration communities into 11 approximate research areas. Table 2.4 provides the project keywords and the associated Research area of the top 10 largest communities in size based on the IUK dataset.

Table 2.4 Project keywords of the top 10 IUK collaboration communities in size

No.	word1	word2	word3	word4	word5	Research Area
1	high	material	energy	application	manufacturing	Manufacturing
2	crop	farm	food	grower	high	Agrifood
3	cell	patient	drug	high	disease	Medical
4	drone	service	operation	air	vehicle	Automotive
5	energy	offshore	wind	turbine	power	Energy
6	energy	network	power	carbon	low	Energy
7	energy	low	building	high	material	Materials
8	patient	clinical	treatment	health	disease	Medical
9	farm	farmer	food	animal	emission	Agrifood
10	quantum	application	high	laser	key	Electronics

Source: Project texts from the IUK dataset analysed by Cambridge Econometrics

[7] Stop words are commonly occurring words that can be safely ignored without sacrificing the meaning of the sentence, such as ‘the’, ‘and’, ‘it’. Lemmatisation is the process of grouping together the different forms of a word so they can be analysed as a single item in context of the sentence.

The largest research area is “Automotive” with 871 organisations and 25 clusters, of which 11 are Dispersed communities (see Table 2.5).

Table 2.5 Number of organisations and communities by the 11 IUK research area

Research area	No. of Organisations	No. of Clusters	No. of Dispersed communities
Automotive	871	25	11
Manufacturing	530	17	6
Agrifood	443	12	5
Materials	313	8	1
Medical	251	8	4
Electronics	272	8	4
Infrastructure	119	7	3
Digital	131	6	4
Healthcare	72	4	3
Energy	142	3	2
Others	27	1	0
Total	**3,171**	**99**	**43**

Collaborations generally occur between academic institutions spread across the UK, leading to clusters that are not co-located. Collaboration communities have an average diameter of at least 150km, approximately as wide as England. Figure 2.7 shows the collaboration network of a collaboration community in the “Automotive” research area overlayed on a geospatial map. The figure shows all the organisations (including universities) that are collaborating in the community and that they are not co-located. From the keywords analysis, we can determine that the community collaborates on electronic vehicle battery research.

Figure 2.7 Dispersed Automotive collaboration community

Source: IUK dataset analysed by Cambridge Econometrics

Notes: Link thickness represent collaboration frequency. Universities are labelled.

2.6 Summary of Results

After all the methods were used, 3,443 innovation clusters were identified in the UK (of 10 or more firms). There are 429 R&D Collaborating clusters, 2,901 Specialised clusters, 7 Diverse clusters, and 106 Dispersed communities. Table 2.6 presents a summary of the datasets, clustering methodologies, filtering criteria, clusters and cluster types identified. Discussion on the selection of thresholds and sensitivity analysis are in Annex B.

One caveat is that clusters that we have failed to identify as "Collaborative" may still collaborate with others. The absence of evidence in our dataset does not suggest evidence of absence. The IUK research funding application dataset only provides a subset of collaborations in the UK economy. Other examples include collaborations on filing for patents and applying for funding in UK Research and Innovation or less formal collaborations through face-to-face interactions or online communities. In a future iteration, we would like to introduce these datasets to build a more comprehensive collaboration network and identify more Collaborative clusters.

Table 2.6 Summary table of clustering methodology and results

	IDBR Spatial	RTIC Spatial	IUK & MAKG Spatial	IUK & MAKG Network
Data points	3 million business sites	5 million business sites	111 thousand projects & 239 million publications	111 thousand projects & 239 million publications
Pre-clustering	Classified by broad SIC sectors, thus Specialised	Classified by RTIC sectors, thus Specialised	RD&I-active by definition	RD&I-active by definition
Clustering Method	HDBSCAN algorithm, thus Co-located	HDBSCAN algorithm, thus Co-located	HDBSCAN algorithm, thus Co-located	Louvain algorithm, thus Internally Collaborative
Criteria 1	Check RD&I-active using Innovation indicators	RD&I-active by definition	Check Specialised using share of SIC sector	Check Specialised using share of SIC sector
Criteria 2	Check Internally Collaborative using IUK collaboration network	Check Internally Collaborative using IUK collaboration network	Check Internally Collaborative using IUK collaboration network	Check Co-located using cluster radius
Total number of clusters	3,965	1,600	94 IUK & 14 MAKG	99 IUK & 10 MAKG
- R&D Collaborating clusters	251 R&D Collaborating clusters	85 R&D Collaborating clusters	79 IUK & 14 MAKG R&D Collaborating clusters	0 R&D Collaborating clusters
- Other Innovation clusters/communities	1,378 Specialised clusters	1,515 Specialised clusters	8 Specialised & 7 Diverse IUK clusters	96 IUK & 10 MAKG Dispersed communities

Source: Analysis by Cambridge Econometrics and The Data City
Notes: The accompanied interactive mapping tool represents the IDBR, RTIC and IUK spatial clusters.

3 Analysis of clusters

3.1 Introduction

In this section, we present some high-level analysis of the various types of clusters identified within the UK. We analyse their:

- Composition – organisation and sectoral make up,
- Knowledge relationships – collaboration and overlaps,
- Geography – regional analysis of types of clusters,
- Patterns of Investment.

Looking at clusters' sectoral make up, we find that:

- Cluster sizes are not strictly comparable and growth targets should be tailored to industrial characteristics.
- The low correlation between number of clusters per sector, average employment, and turnover should challenge the one-size-fits-all perspectives regarding cluster development. It also highlights the sectors in which more can be accomplished with fewer centres of activity.
- Clusters that collaborate on IUK projects are correlated with higher average turnovers across a wide variety of sectors.

When investigating cross-cluster knowledge relationships, we find that:

- There are strong collaborations and firm overlaps between clusters of the same sector. There is also a significant degree of interaction between clusters across different related sectors.
- Strong collaboration networks span across long distances of at least 150km between research intensive universities, research organisations, and firms.
- Outside of these research-intensive poles, places tend to cooperate more frequently with partners in their own broad geographic region.

Examining the geographical patterns of clusters, we find that:

- The type and location of clusters we observe largely depend on the spatial and resource requirements of the sectors within those clusters: resource-extracting sectors such as mining or offshore power generation mostly occur in limited geographies and cover wider territories; knowledge intensive activities (both services and non-services) tend to require access to large and diverse labour pools and therefore concentrate in denser urban environments and occupy a smaller spatial footprint.
- Industrial co-location generally occurs because sectors have spatial and resource needs that are satisfied by similar locations. Industries that need more space to operate are more likely to co-locate in suburban locations where office parks and manufacturing sites are more prevalent.

Comparing investment funding received by clusters, we find that:

- The unevenness of funding distribution across sectors reflects the different propensities and levels of competence across sectors in seeking innovation funding.
- Emerging sectors like Fintech received more private than public funding per firm, whereas the opposite is true for traditional sectors like Machinery.

3.2 Cluster composition

Applying the methodology described above yielded a total of 3,398 innovation clusters. However, there are considerable variations in the size, composition, and potential of clusters across sectors and places. These variations hold interesting insights for policy and contextual considerations for future iterations of this dataset.

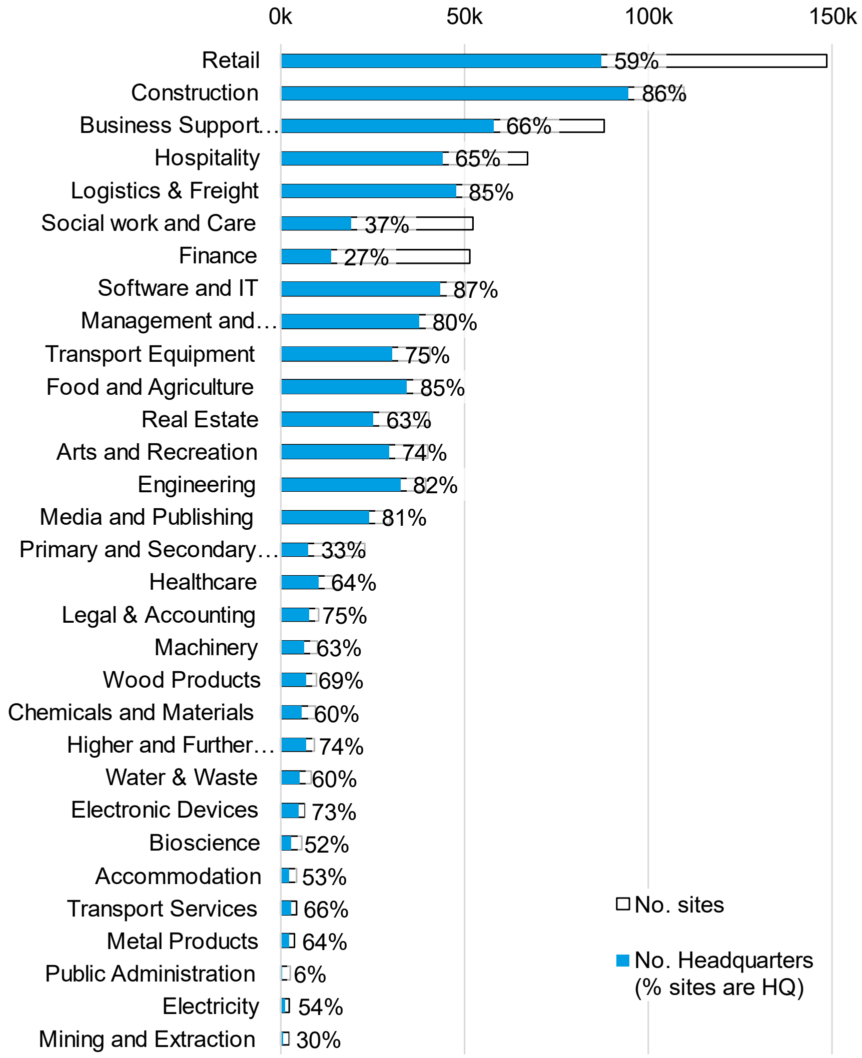
We found that the sectors with the largest number of clusters tend to be those that have many office sites. Figure 3.1 depicts the number of IDBR headquarter and non-headquarter sites while Figure 3.2 and Figure 3.3 depict the total number and types of clusters identified for each SIC sector and RTIC sector, respectively. For example:

- Sectors like retail, construction, logistics, hospitality, finance, and social work tend to have more, and smaller, clusters.
- Emerging sectors tend to be concentrated in fewer specialist centres (e.g., space related sectors, net zero and clean energy, and emerging Fintech sectors).
- Geography dependent sectors (e.g., energy generation and storage, utilities and waste, agriculture/agritech, and mining) tend to also have relatively fewer clusters because their activities are definitionally limited to a smaller number of places that have the appropriate resource mixes.

These variations demonstrate that having more clusters is not necessarily a mark of "success" and highlight how many different factors, including the methodological decisions in the identification process, contribute to the total number of clusters in any given industry.

Clusters also vary significantly in terms of number of participant organisations identified in our dataset. IDBR spatial clusters range from 25 organisations (Mining and Extraction) to 18,000 (Food and Agriculture), while RTIC spatial clusters ranges from 10 (Wearables

Figure 3.1 Number of sites and headquarters by IDBR sectors



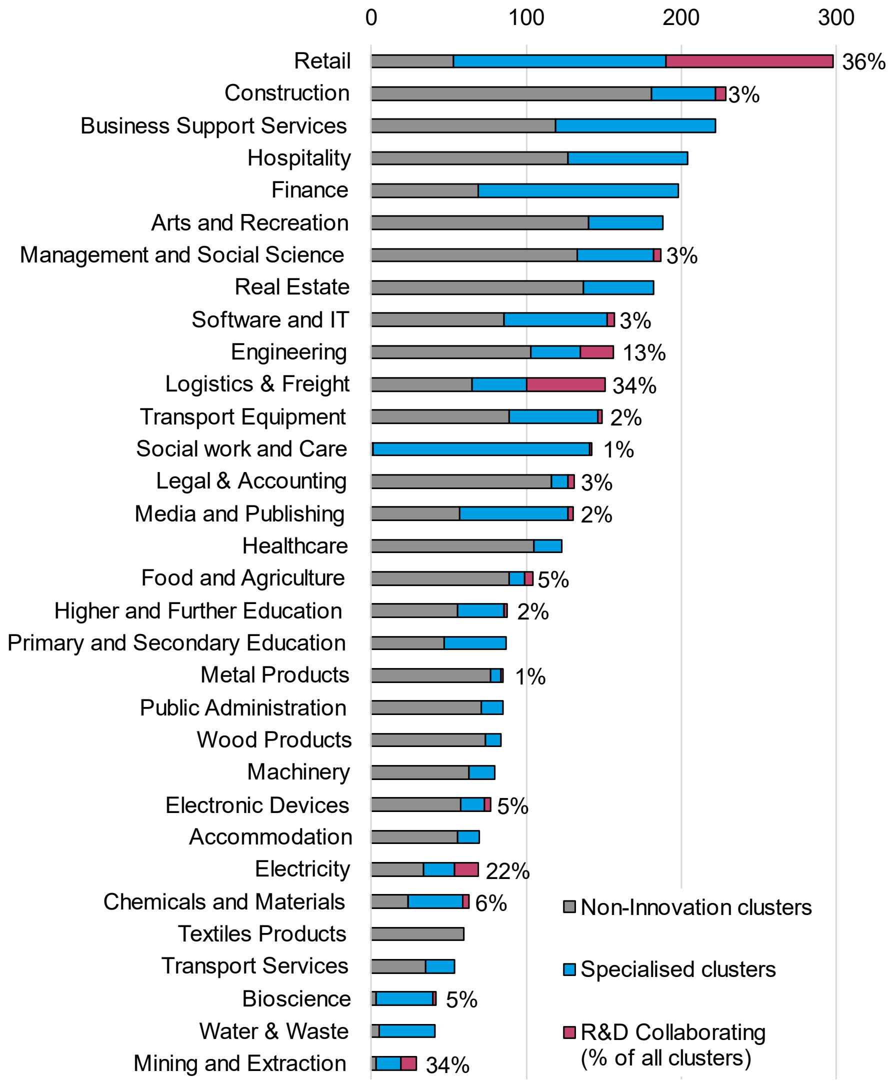
Source: IDBR dataset

and Quantified Self) to 7,000 (Digital Creative Industries), while IUK spatial clusters range from 10 to 99 participants. The number of participants is closely tied to factors such as the overall size of the industry (by number of business sites), and how loosely and tightly defined the industrial classification definition is. For instance:

- Sectors that are both larger overall and present in most cities – like construction, real estate, healthcare, and hospitality – tend to be made up of more sites.
- Sectors that have high numbers of smaller producers and occupy larger geographies, like agritech, also appear near the top of this list.
- Spatially constrained sectors, such as mining and energy generation have the smallest average numbers of sites.

This signals that cluster sizes are not strictly comparable and suggests that growth targets should be tailored to industrial characteristics.

Figure 3.2 Number of clusters by IDBR sectors



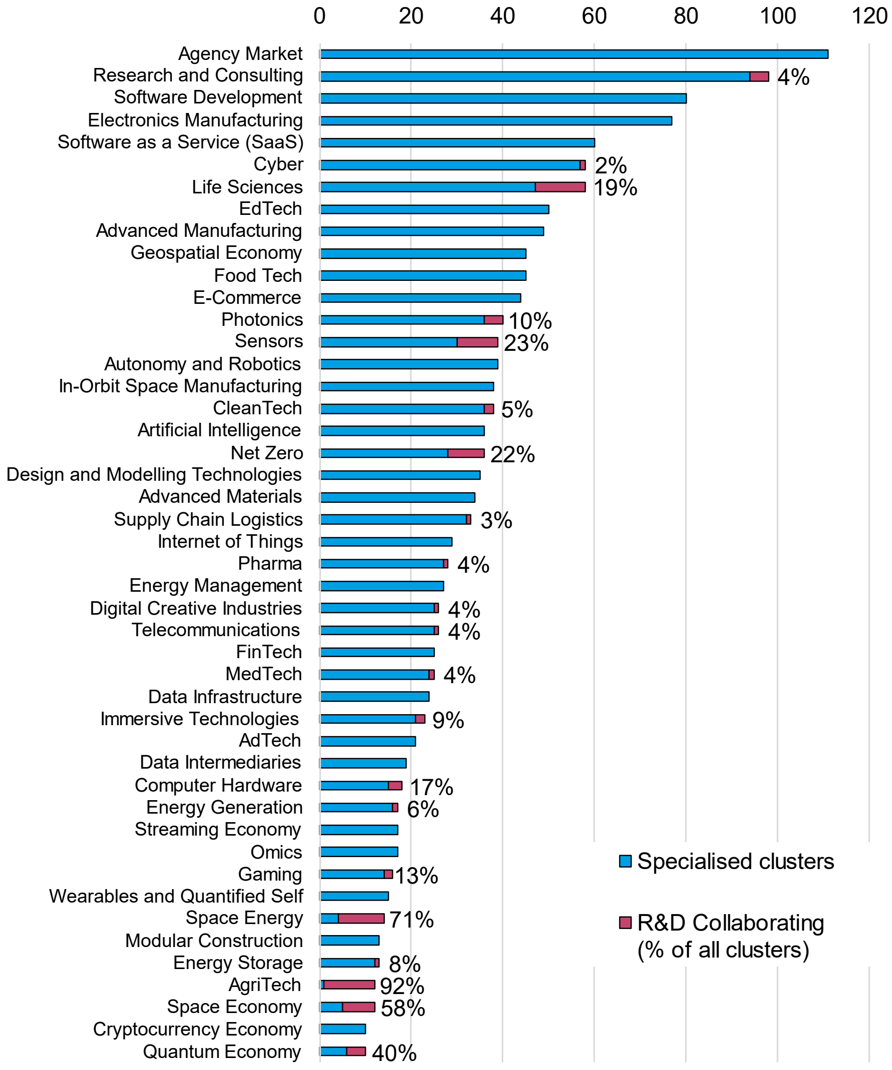
Source: IDBR dataset, analysed by Cambridge Econometrics

The proportion of clusters that satisfy the criteria for R&D Collaborating also differ across sectors in our sample (See Figure 3.3 for RTIC sectors). More research is necessary to fully unpack the logic behind these patterns, but we venture some explanations here.

- Space economy and quantum economy sectors in the RTIC dataset are still relatively small, with fewer total clusters, and tend to be highly concentrated at and around R&D centres – such as Harwell. These research centres are active in seeking and securing RD&I funding with local partners and so appear to exhibit higher than average collaboration.
- Agritech in the RTIC dataset, while a much larger sector, encompasses a wide variety of verticals across its value chain and often centres on highly active research organisations.
- Logistics and freight is a large sector in the IDBR dataset that relies on networks to function. As such, collaboration between entities may be relatively more common.

Many of the UK's most innovative sectors have not produced R&D Collaborating clusters at very high rates. One possible reason is limitations of the primary source of collaboration data. Collaborations that do not show up in the IUK dataset do not imply it is not happening. These results should be taken as indicative and might underestimate collaborations in several sectors.

Figure 3.3 Number of cluster by RTIC sector



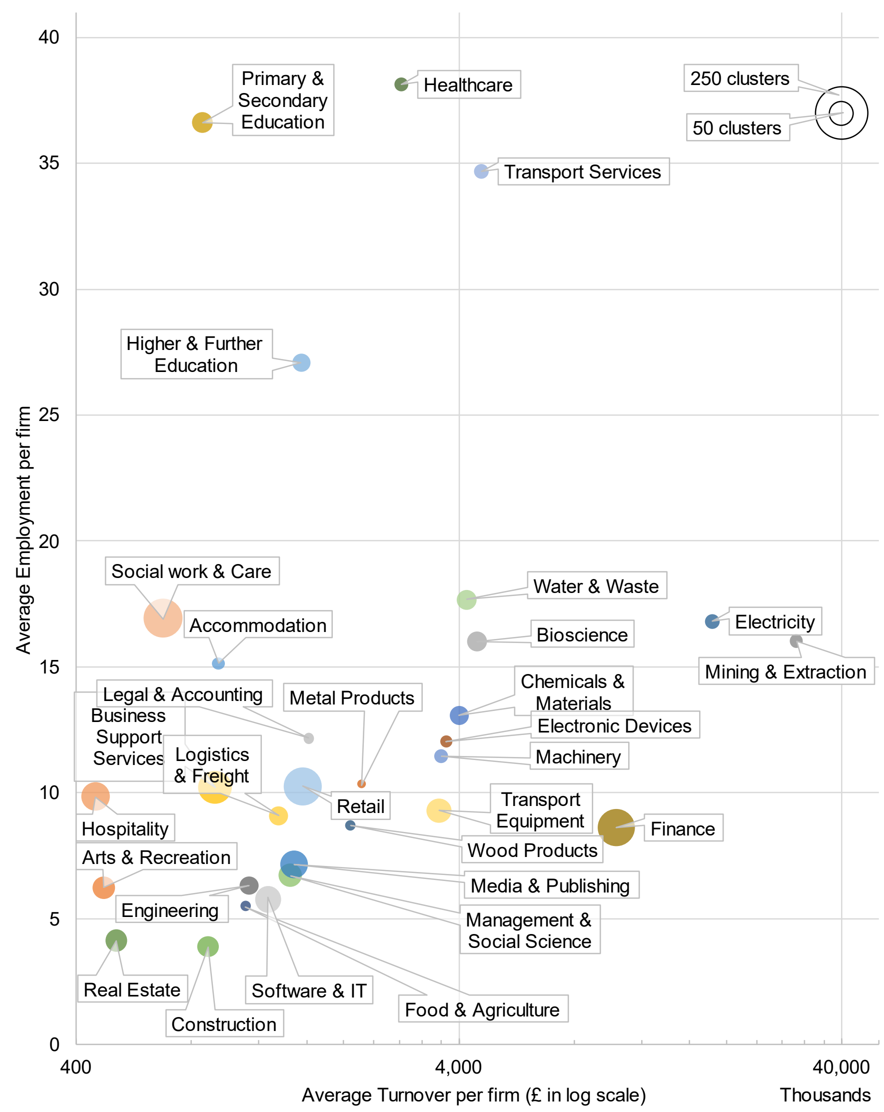
Source: RTIC dataset, analysed by Cambridge Econometrics and The Data City

Looking at the annual performance of firms, some of the sectors with the highest numbers of clusters in the IDBR dataset tend to have lower employment and lower turnover per cluster (see Figure 3.4) – that is to say we see lots of clusters of smaller firms, rather than a smaller number of clusters of larger firms, for example:

- The large number of clusters in lower-productivity sectors like social work and care, hospitality, and business support services has not translated into significant economic impact in terms of turnover or employment.
- In contrast, high-productivity sectors like mining and electricity, with an average employment and a smaller number of clusters, generate high levels of turnover.
- More specialised sectors like finance, biosciences, chemicals, and machinery, produce moderate-high turnovers with comparatively fewer clusters (although finance is a notable exception with many).

The low correlation between number of clusters per sector, average employment per sector, and turnover per sector should challenge the one-size-fits-all perspectives regarding cluster development. It also highlights the sectors in which more can be accomplished with fewer centres of activity. For some sectors, it might be appropriate and desirable for every town to have its own cluster; for others, you might only expect one cluster per region, and policy efforts to fight this tendency may be unwise.

Figure 3.4 IDBR cluster: Turnover vs. Employment vs. Num. of cluster by sector

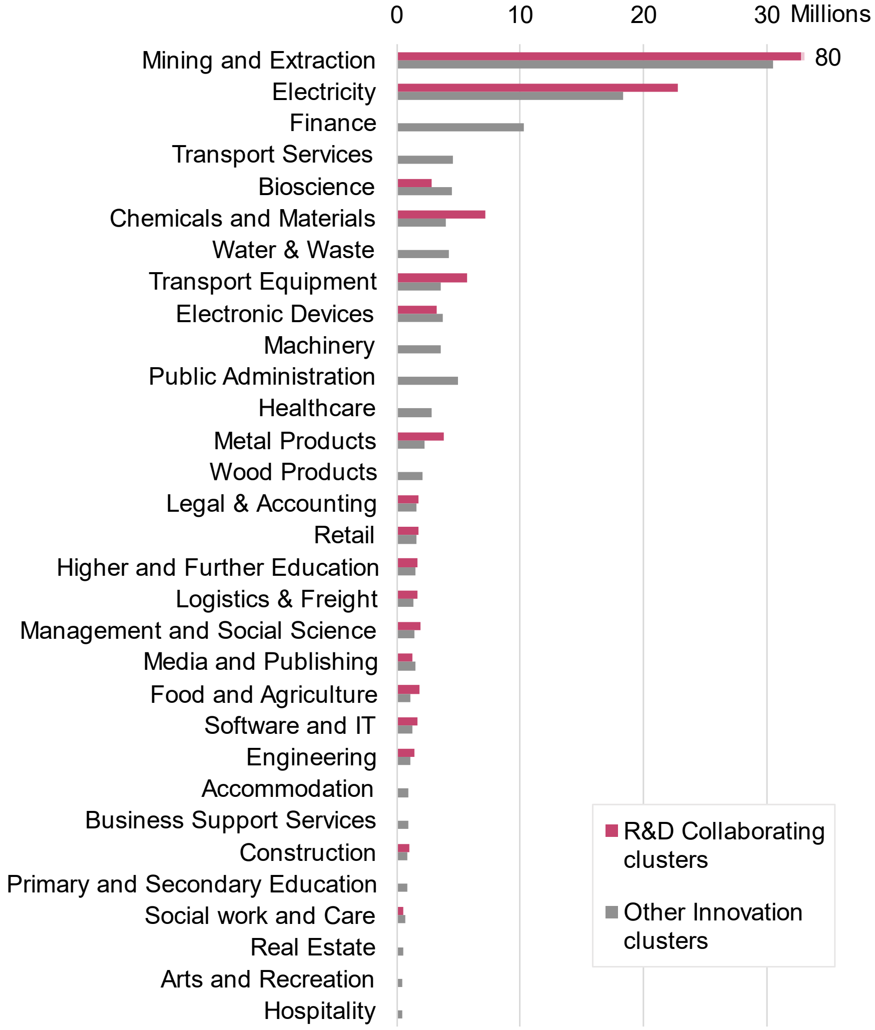


Source: IDBR dataset, analysed by Cambridge Econometrics

One question worth probing is whether being part of an R&D Collaborating cluster makes a difference. Figure 3.5 shows demonstrates that R&D Collaborating clusters tend to have higher average turnovers than their Specialised clusters counterparts. This effect is particularly pronounced in chemicals and materials, metal products, and electricity. In other sectors, like biosciences and electronic devices, Specialised clusters appear to have an edge. This effect may be an artifact of our collaborative dataset, which may not be effectively reflecting relationships funded by councils other than IUK (e.g., the BBSRC or MRC). If not, these cases are interesting puzzles that merit further exploration. Overall, however, **these findings suggest a correlation between R&D Collaborating clustering and higher average turnovers across a wide variety of sectors.**

Another source of variation is the degree to which clusters in our dataset have attracted public R&D funding and Venture Capital funding. We explore this further in Section 3.5.

Figure 3.5 Turnover per firm averaged across clusters by IDBR sectors (£)



Source: IDBR dataset, analysed by Cambridge Econometrics

3.3 Cluster knowledge relationships

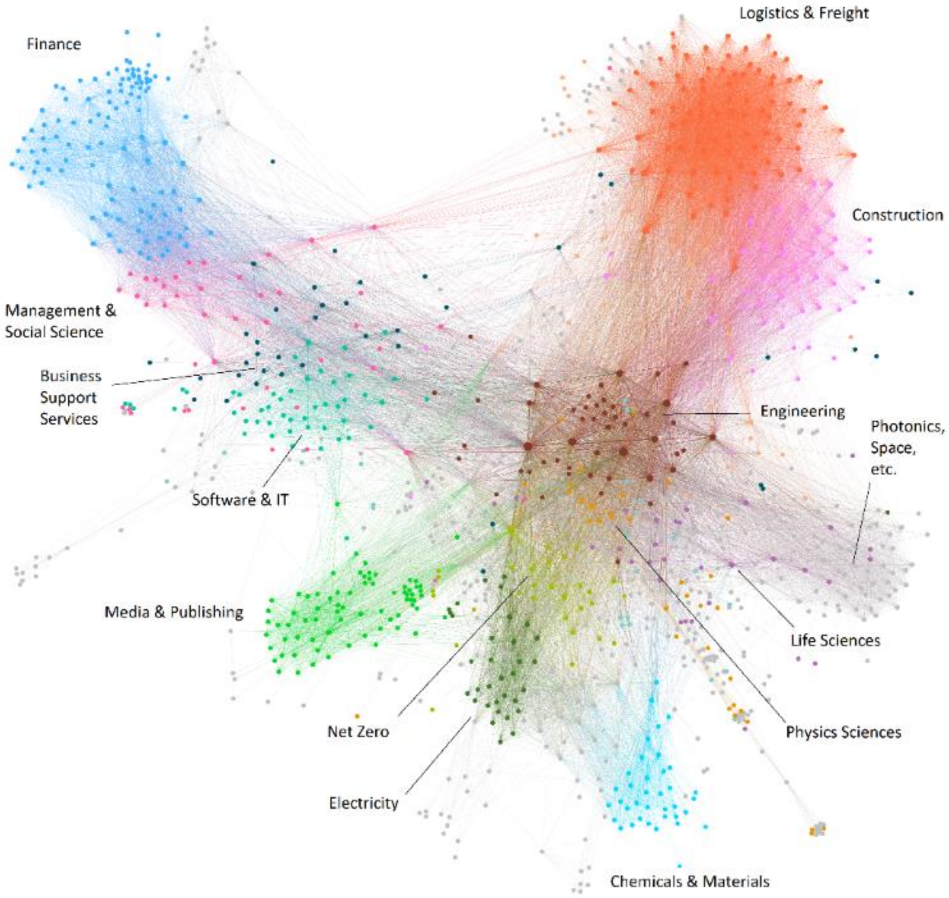
R&D Collaborating cluster, as noted above, appears to be correlated with higher average turnovers across a variety of sectors. Less is known, however, about collaboration between clusters in different sectors. This section explores patterns of collaborative relationships between clusters, while the following section (3.4) investigates spatial correlation of clusters and potential implications for theory and practice.

Inter-cluster collaboration is important because it opens up the potential for new combinations of knowledge, which can multiply potential for innovation. While clustering tends to focus on specialisms, the principle of related variety holds – regions[8] will tend to develop along innovation pathways that combine and recombine elements of different existing industrial specialisation (Whittle & Kogler 2020; Boschma, Balland, & Kogler, 2014; Neffke, Henning, & Boschma, 2011). From this perspective, new specialisms are constructed from the building blocks of previously developed capabilities and this recombination of knowledge drives innovation and industrial evolution.

The emergence of recombinant knowledge and the process technological branching through relatedness relies on spatial concentration of economic activities that share some similarities, for example in markets, supply chains, knowledge bases, or skillsets. This underpins the smart specialisation focus on local specialisms and competitive advantage and in developing innovation strategy.

However, research has also noted that external knowledge flows can introduce complementary knowledge, fill gaps, and mitigate the risk of lock-in (Trippl et al. 2015; Balland & Boschma 2018). More

Figure 3.6 Network of IDBR, RTIC cluster collaboration, coloured by sectors



Sources: IDBR, RTIC and IUK dataset analysed by Cambridge Econometrics
Notes: Cluster collaboration is the sum of the number of cross cluster collaboration and the number of overlapping members.

[8] Note that while the term region is used here, dynamics of relatedness and diversity can be explored at a variety of scales.

geographically distant collaboration has been found to increase the likelihood of radical innovation as more distant knowledge pools are likely to be dissimilar and complementary (Kesidou et al. 2023; Hsieh et al. 2018; Bertrand & Mol 2013) as different places have evolved different types of specialisms, even within the same industry.

Exploring inter-cluster relationships provides a test of expectation around knowledge flows as well as insights about which pathways are currently highly active. We defined inter-cluster collaboration through participation in collaborative research projects (IUK database) and overlaps in membership between clusters. Figure 3.6 visualises a meta network where nodes represent clusters, edges represent cluster collaborations and node colours represent sectors.

Unsurprisingly, we find strong collaborations between clusters in the same sector, as evidenced by clear groupings of each colour. There is also clearly a significant degree of interaction between clusters in different sectors. For instance, engineering clusters collaborate with a wide variety of clusters such as life sciences, physical sciences, energy, and construction, suggesting that engineering provides important bridging technologies and services. The relationships between these clusters stand out because they are among the largest sectors, both in the number of clusters and the number of organisations.

Collaborations between these clusters, however, also exhibit a degree of relatedness. Running a further Louvain community detection process over the meta-map of sector-sector collaboration revealed that sectoral collaboration patterns fell into six approximate collaboration groups, which we labelled “Digital”, “Foundational”, “Engineering”, “Life Sciences” “Hardware”, and “Services” (see Table 3.1).

Table 3.1 Six Louvain communities of sector-sector collaborations

Digital	Foundational	Engineering	Life Sciences	Hardware	Services
Media and Publishing	Chemicals & Materials	Construction	Bioscience	Advanced Materials	Accommodation
Agency Market	Electricity	Electronic Devices	Life Sciences	AgriTech	Arts and Recreation
Cyber	Mining and Extraction	Engineering	Omics	Computer Hardware	Business Support Services
Digital Creative Industries	Water & Waste	Logistics & Freight	Pharma	Electronics Manufacturing	Finance
EdTech	Wood Products	Machinery		MedTech	Food and Agriculture
FinTech	Advanced Manufacturing	Metal Products		Photonics	Healthcare
Gaming	CleanTech	Transport Equipment		Quantum Economy	Higher and Further Education
Immersive Technologies	Energy Generation	Transport Services		Sensors	Hospitality
Software as a Service (SaaS)	Energy Management	Autonomy and Robotics		Space Economy	Legal & Accounting
Software Development	Energy Storage	E-Commerce		Space Energy	Management and Social Science
Streaming Economy	Food Tech	Geospatial Economy			Primary and Secondary Education
Telecommunications	Net Zero	In-Orbit Space Manufacturing			Public Administration
Artificial Intelligence		Internet of Things			Real Estate
Data Infrastructure		Modular Construction			Social work and Care
Software and IT		Physical Sciences & Engineering			Design & Modelling Technologies
		Supply Chain Logistics			

There are, of course, many more interesting inter-cluster intersections and interdependencies, each of which can provide clues to new directions of technological evolution. These are difficult to discern in a broad analysis of meta-data but can provide interesting insights about emerging and novel industrial configurations. Diving more deeply into this relational data may also reveal which clusters are more successful at forging and leveraging external knowledge partnerships and offer clues as to best practices, potential pitfalls, and advance research on related and unrelated variety.

Figure 3.7 shows the strongest collaborations of the Net Zero RTIC cluster. As might be expected, the most frequent collaborations are with clusters in Clean Tech, Electricity, Logistics & Freight, Mining & Extraction, Engineering, Energy Generation, and Water & Waste.

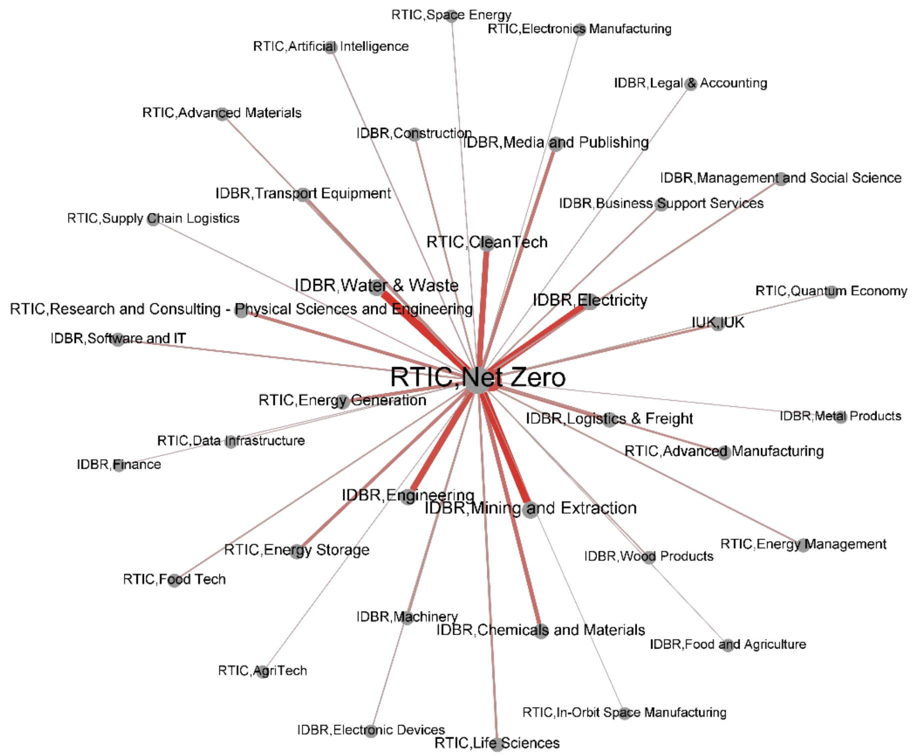
Spatial relationships

The inter-cluster meta network does not show spatial relationships. By grouping spatial clusters based on their location, we can study how knowledge is flowing between places.

Figure 3.8 shows a matrix in which the counties with more numerous collaborations (and participant overlaps) are depicted in darker shades. The matrix is ranked by the top 20 counties based on the number of collaboration they have with Greater London. As expected, Greater London emerges as the core knowledge "hub" in collaborative networks.

A perhaps more notable insight, however, is that *internal* collaboration is strongest in some of the most innovative places anchored by numerous and/or strong research universities and private R&D operations – e.g., in Cambridge, Oxford, Edinburgh, Manchester, and Matrix of counties by counties with cells shaded by the number of

Figure 3.7 Sectoral collaborations of the RTIC Net Zero sector



Sources: IDBR, RTIC and IUK dataset analysed by Cambridge Econometrics

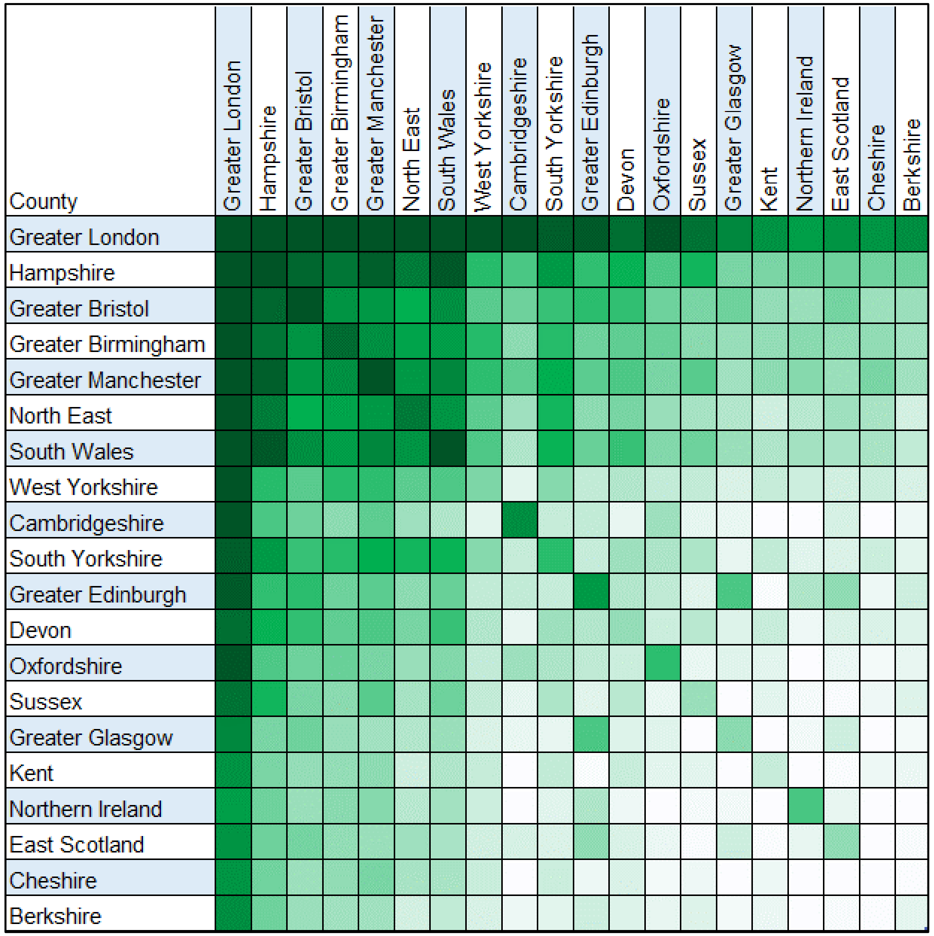
Notes: The number of cross sector collaboration is calculated by summing the cross-cluster collaborations by sectors.

cross-county collaborations. Highest collaborations are between Greater London and Hampshire. South Wales has high internal collaboration. Glasgow and Edinburgh has high cross-county collaboration.Hampshire[9]. These are also the places that collaborate relatively more frequently with each other. This suggests that **there are strong networks, sometimes across quite long distances, between research intensive universities, research organisations, and the firms around them.**

While less pronounced, it appears that, outside of these research-intensive poles, places tend to cooperate more frequently with partners in their own broad geographic region. For instance, relatively strong relationships between Edinburgh and Glasgow suggest a regional effect. Places in the North are slightly more likely to have partnerships with other northern places – e.g., outside of London, places like South Yorkshire are most likely to connect with partners in North East England and Manchester. However, this effect is not even which raises questions as to why certain places are more likely to pair than others.

Future research might focus on exploring the spatial dimension of knowledge networks between clusters and places by sector. Learning which places are connecting most effectively and why across different sectors could deepen understanding about the value of encouraging the development of external networks as a complement to specialisation strategies and identify potentially complementary pairings that might not be being effectively exploited. Embracing approaches that take these broader system dynamics into account could enhance the toolkit of policymakers engaged in place-based policy.

Figure 3.8 Top 20 broad regions in cluster collaboration with Greater London



Sources: IDBR, RTIC and IUK dataset analysed by Cambridge Econometrics
Notes: Colour represents the number of cross-county collaborations – White (<500) to Dark green (2,000+), with median 1100 and max. 10,000

[9] Hampshire is the 5th largest county by population in the UK and has 4 major universities that participate in many collaborations. It is also home to IBM with its R&D laboratories at Hursley and UK headquarter at Cosham, both in Southampton, Hampshire. IBM alone participated in approximately 100 collaborations, ranking 80th among organisations across the UK.

3.4 Cluster geography

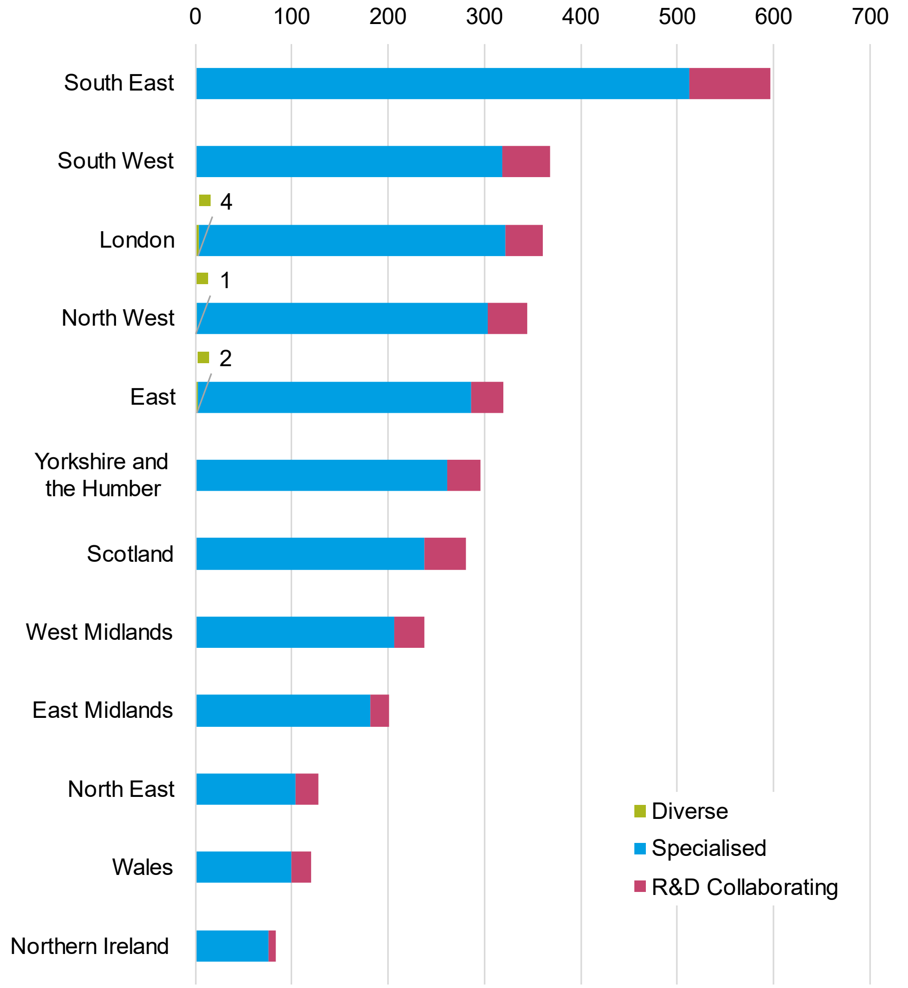
Understanding patterns in the geography of clusters can unlock valuable insights to more effectively target investments and tailor supportive policies. The conventional wisdom is that clusters are highly concentrated, but that is not always the case. While some degree of spatial concentration is important to fulfil our criterion of spatial co-location, the HDBSCAN methodology enabled us to identify clusters of variable densities to ensure that our findings did not exclude clusters that might occur in more rural areas.

Figure 3.9 shows the number of Innovation clusters by International Territorial Level (ITL1) statistical region. The share of different types of Innovation clusters are approximately the same across regions. Roughly 12% of Innovation clusters are R&D Collaborating clusters and the rest are Specialised clusters. Diverse clusters, identified through the Spatial approach using IUK dataset, are mostly located in London and East of London.

A few caveats when interpreting analysis on cluster geography:

- In a cluster, business sites can locate in more than one region. Extra care should be taken when performing regional analysis.
- The cluster regions are identified based on the average centre of their sites. Given that London is small, large clusters that include London firms may be located in surrounding regions (such as East and South East). This underestimates the number of clusters in London.
- Firm-level data such as turnover and employment are assigned equally across sites registered when headquarter information is not available. This underestimates the importance of clusters with high concentrations of headquarters, such as London.

Figure 3.9 Number of Innovation cluster by ITL1 regions



Sources: IDBR, RTIC and IUK dataset analysed by Cambridge Econometrics

Notes: Dispersed communities are not grouped in one region and thus excluded. Regions of clusters are identified based on the average geographical centre of their sites. Small regions like London are likely underestimated.

Specific Regional Strengths

The data also allows us to identify specific regional strengths. There are a variety of ways to do this, and we invite readers to look through the accompanying database. Our preferred method is to look where within the country the top 3 clusters by size (number of firms) in each sector (RTIC or SIC) are located. Although London dominates this analysis, this is likely still an underestimate, as the HDBSCAN algorithm subdivides the densest areas into multiple smaller clusters. For example, rather than having a single finance cluster spanning London, we see multiple smaller finance clusters in the City, in Westminster, and in Canary Wharf.

With those caveats in mind, we mention below the revealed innovation cluster strengths in each region in turn:

- **London** hosts 54 top-3 clusters by size, with the largest cluster in 43 sectors. This is by far the most of any ITL1 region.

- The **South East** hosts 10 top-3 clusters, including the largest clusters in Space Energy and Geospatial Economy.

- The **South West** hosts 11 top-3 clusters.

- The **East of England** hosts 5 top-3 clusters.

- The **East Midlands** hosts 3 top-3 clusters, and the largest cluster in is the Textiles sector.

- The **West Midlands** hosts 36 top-3 clusters, with the largest clusters in Construction, Arts and Recreation, Transport Equipment, Wood Products, Machinery, Electronic Devices, Life Sciences, Advanced Manufacturing, and Advanced Materials.

- The **North West** hosts 39 top-3 clusters, with the largest clusters in Legal & Accounting, Higher & Further Education, Bioscience, Food Tech, and Omics.

- **Yorkshire and the Humber** hosts 18 top-3 clusters, including the largest clusters in Metal Products and Clean Tech.

- The **North East** hosts 14 top-3 clusters, including the largest clusters in Engineering, Healthcare and Electricity.

- **Scotland** hosts 20 top-3 clusters, including the largest clusters in Management & Social Science, Real Estate, and Software & IT.

- **Wales** hosts 3 top-3 clusters.

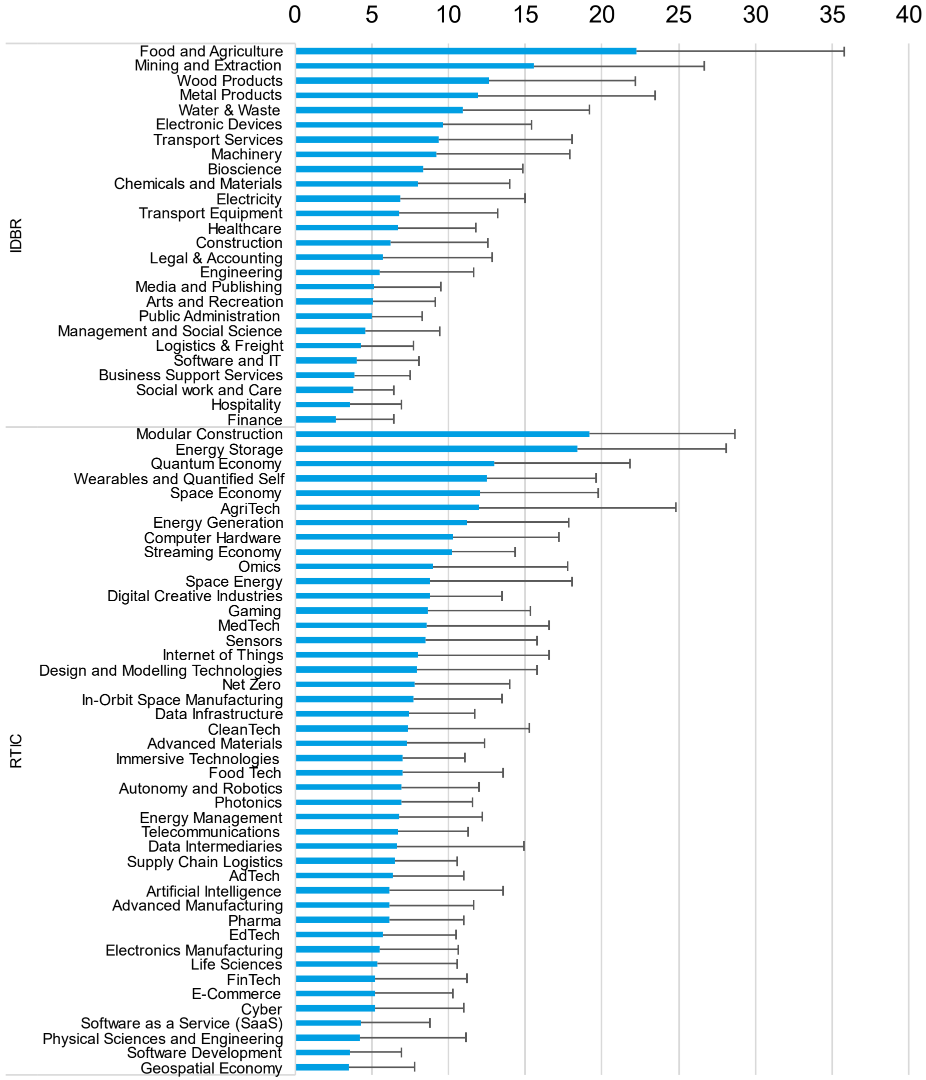
- **Northern Ireland** hosts 8 top-3 clusters, with the largest clusters in Food and Agriculture and Finance.

Outside of London, the impact of large cities able to host dense clusters containing large numbers of firms – particularly the UK's twin second cities of Birmingham and Manchester – is clearly visible. Other ways of judging clusters' strengths would likely reveal different spatial patterns.

The type and location of clusters we observe largely depend on the spatial and resource requirements of the sectors within those clusters. In general, resource-extracting sectors such as mining or offshore power generation mostly occur in limited geographies and cover wider territories; knowledge intensive activities tend to require access to large and diverse labour pools and therefore concentrate in denser urban environments and occupy a smaller spatial footprint.

The average distance of a firm from the centre of its spatial cluster (hereby called radius) ranges between 0.1 and 60km. In denser areas, clusters are generally smaller as a result of the HDBSCAN method. For instance, clusters' radii in Greater London average around 1.5km, compared to 5km in Manchester and 26km in Warwickshire.[10] Figure 3.10 depicts the average radius by sectors. It shows that extractive and heavy manufacturing sectors that require larger industrial campuses tend to have larger clusters than urban-centred knowledge-based sectors.

Figure 3.10 Average cluster radius by sectors (km)



Sources: IDBR and RTIC dataset analysed by Cambridge Econometrics

Notes: Whisker lines represent the standard deviation of the radius.

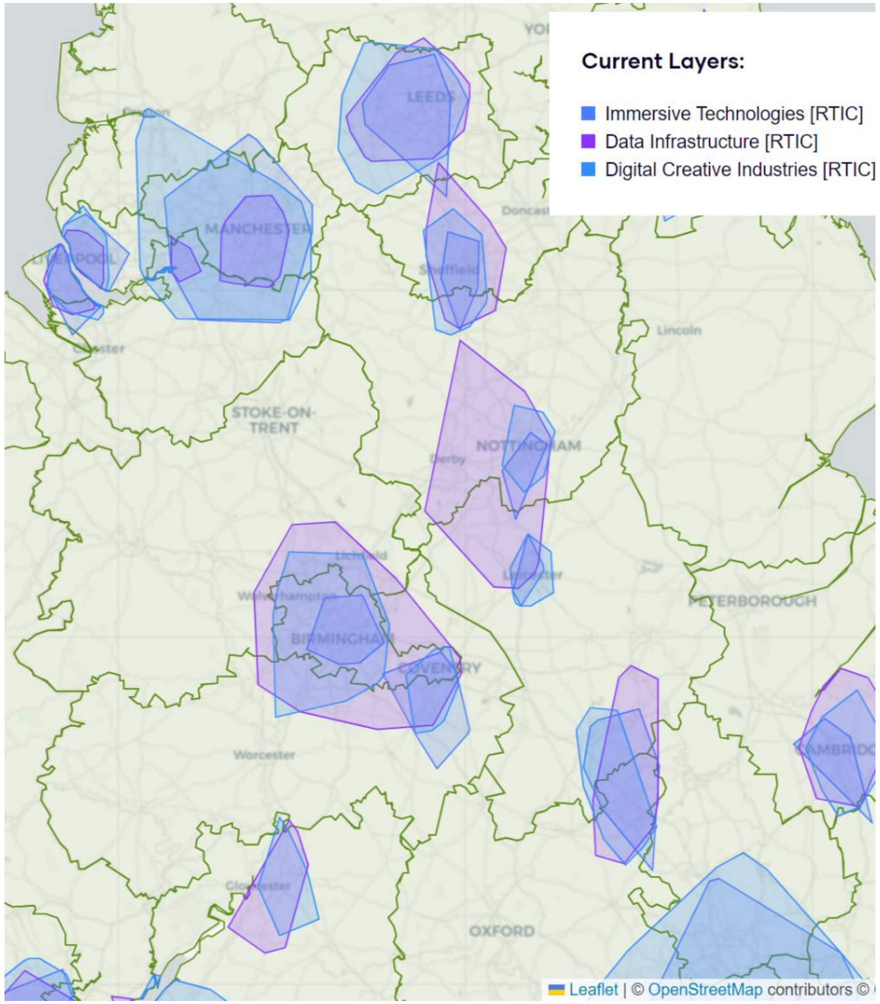
[10] Areas that are segregated are naturally identified as a cluster. For example, Northern Ireland is often a cluster due to it being separated from Scotland by the Irish sea.

These variations in spatial footprint are important for applying lessons of agglomeration in the development of place-based innovation policy. Research on agglomeration economies and the spatial dynamics that trigger and sustain virtuous cycles of innovation has been criticised for its lack of precision in space and scale. Frequently, researchers and policymakers default to convenient administrative or statistical geographies (Nelles et al. 2022). However, in practice, clusters often spill over these boundaries and occupy spaces in which policy levers are divided between different political authorities. See Figure 3.11 for an example where a “Digital Creative Industries” cluster spans across Merseyside, Lancashire, Greater Manchester and Cheshire. Understanding variation in spatial expanse by industry can help to challenge or, where appropriate, justify research and policy based on dynamics within specific administrative units.

The geographies of which sectors tend to locate close to or overlap each other provides important insights for place-based innovation policy. These proximities and overlaps are indicative of which local specialisms are most likely to experience synergies from co-location and forge new technological paths through knowledge recombination.

Generally speaking, industrial co-location occurs because sectors have spatial and resource needs that are satisfied by similar locations. Industries that need more space to operate are more likely to co-locate in suburban locations where office parks and manufacturing sites are more prevalent.

Figure 3.11 Overlapping clusters from the Immersive tech, Data infrastructure, and Digital creative industries RTIC sectors



Sources: Cluster database plotted on the interactive map developed by Cambridge Econometrics and The Data City

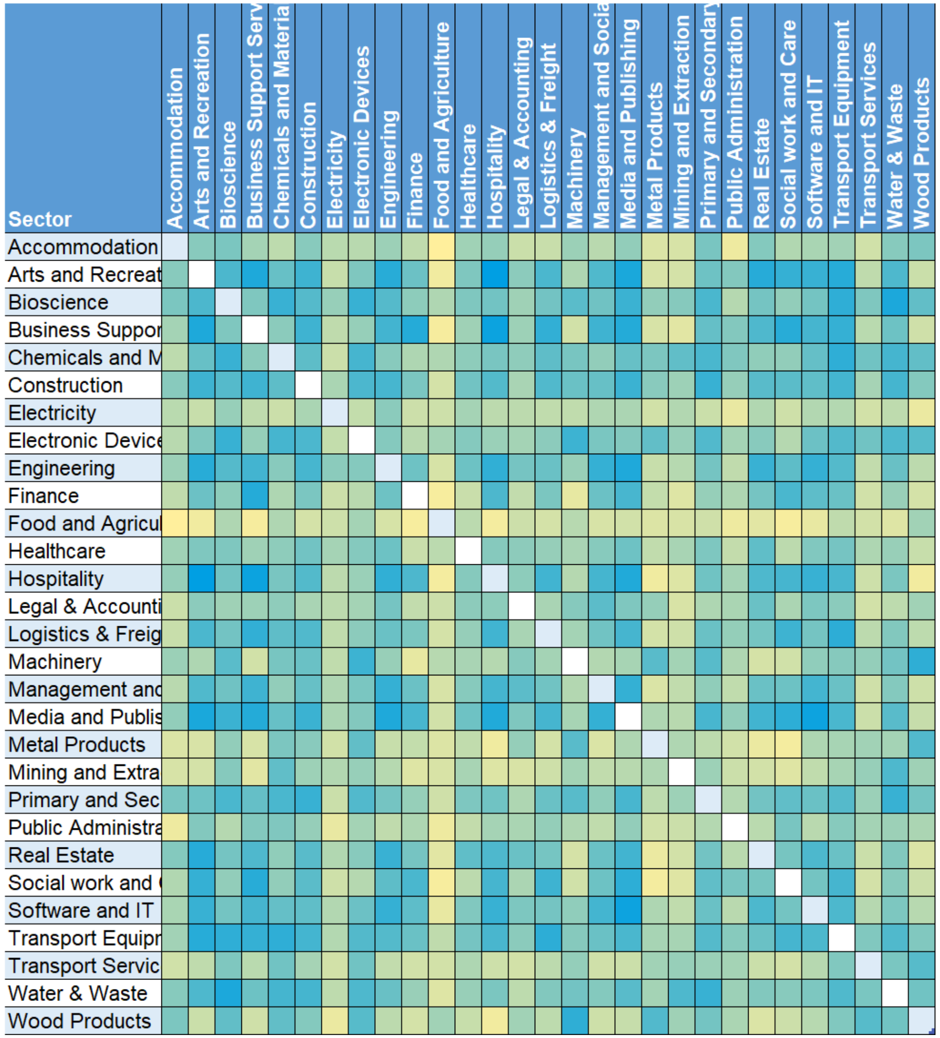
Notes: The polygons represent spatial clusters layered on UK ITL1 regions (green).

For instance, Figure 3.12 shows that:

- Food and agriculture clusters tend not to overlap much with other sectors. These clusters generally require a lot of space and are located in agricultural areas that don't host a wide range of other sectors.
- Electricity, mining and extraction are also quite place-specific and are naturally distanced from other sectors that concentrate around population centres.
- We found strong overlaps between media and publishing, software and IT, hospitality, arts and recreation, and business support services – all of which are typically found in urban areas and centres.
- Similarly, digital creative sectors, immersive technologies, and data infrastructure tend to frequently co-locate. The strength of overlap is stronger than with some other sectors. This is because Digital Creative sectors – quite a broad category – tend to have larger spatial footprints, so multiple more focused immersive technology and data infrastructure clusters can occur within them in larger metropolitan areas (See Figure 3.11).
- Adtech, streaming, data infrastructure, and telecommunications also frequently overlap. However, only a few places have all four at once (Newcastle, Leeds, London, and Belfast).
- Agritech and space economy also co-locate, which could equally be driven by space considerations or potentially technology/skills overlaps (e.g., in the use of satellites).

It is tempting to infer that the correlation between proximity and industrial similarity must mean that there are strong interactions between co-located sectors. While true in some of these cases, it cannot be definitively proved by this data alone. The co-location of

Figure 3.12 Share of cluster geographical overlap by IDBR sectors



Sources: Cluster database developed by Cambridge Econometrics and The Data City

Notes: Cluster overlap is calculated as the share of every pair of clusters between two sectors that strongly overlap each other – Max. 60% (Blue) and Min. 5% (Yellow) sectors like electricity and mining suggests again that space and resource proximity requirements may be just as powerful explanatory factors. What this data does tell us is which sectors, in which places, might have developed or have the potential to develop inter-industry synergies. This can help anchor place-based interventions that link existing capabilities and suggest potential targets for interdisciplinary funding strategies.

As an example of this, some combinations of sectors that we found were likely to both co-locate *and* collaborate include obvious supply chain partner sectors, for example:

- Engineering, Construction, and Machinery
- Omics and Pharmaceuticals
- Finance and Legal & Accounting
- Metal Products and Transport Equipment

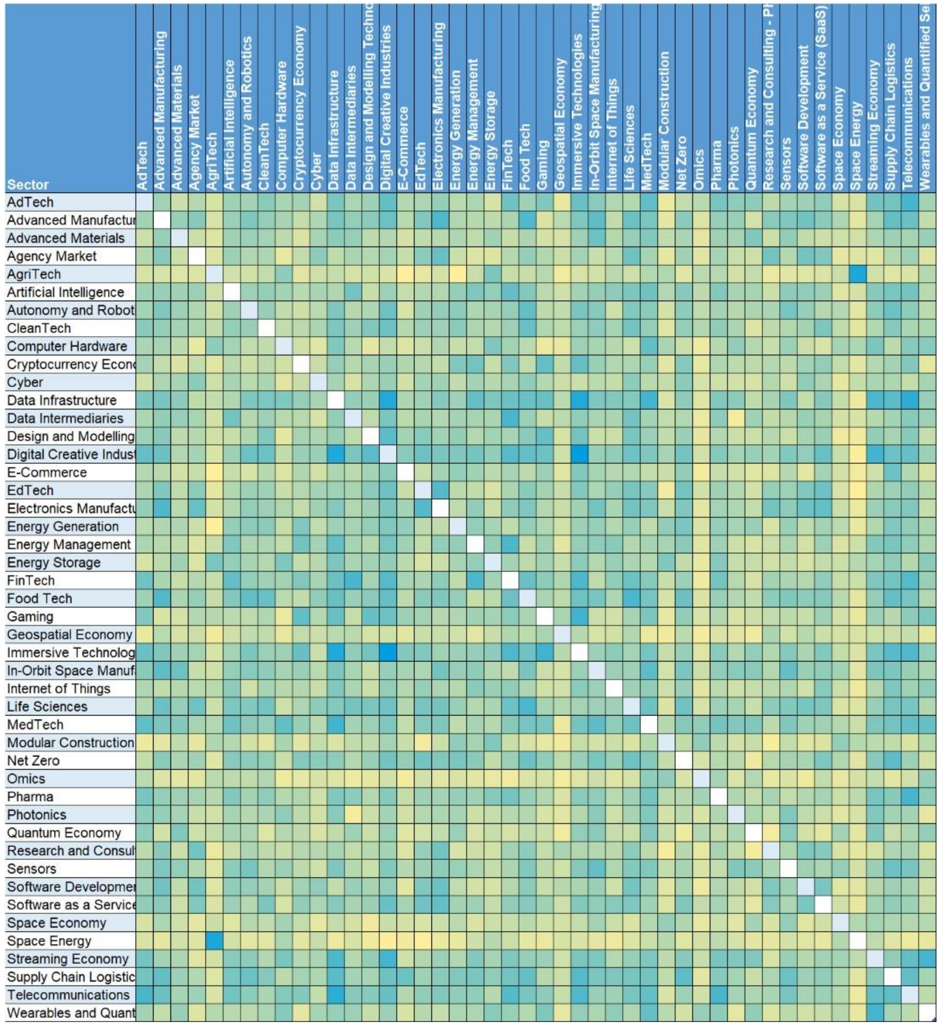
While sectors that tended to co-locate but *not* collaborate include otherwise unrelated sectors with similar locational needs:

- Cyber and Transport Services (both based in city centres for different reasons)
- Energy Generation and Logistics & Freight (both have positive correlation with coastal areas)

Lastly, sectors that tended to collaborate but *not* co-locate include sectors with supply chain relationships but very different locational preferences, e.g.,:

- Transport Services and Transport Equipment
- CleanTech and Energy Generation

Figure 3.13 Share of cluster geographical overlap by RTIC sectors



Sources: Cluster database developed by Cambridge Econometrics and The Data City

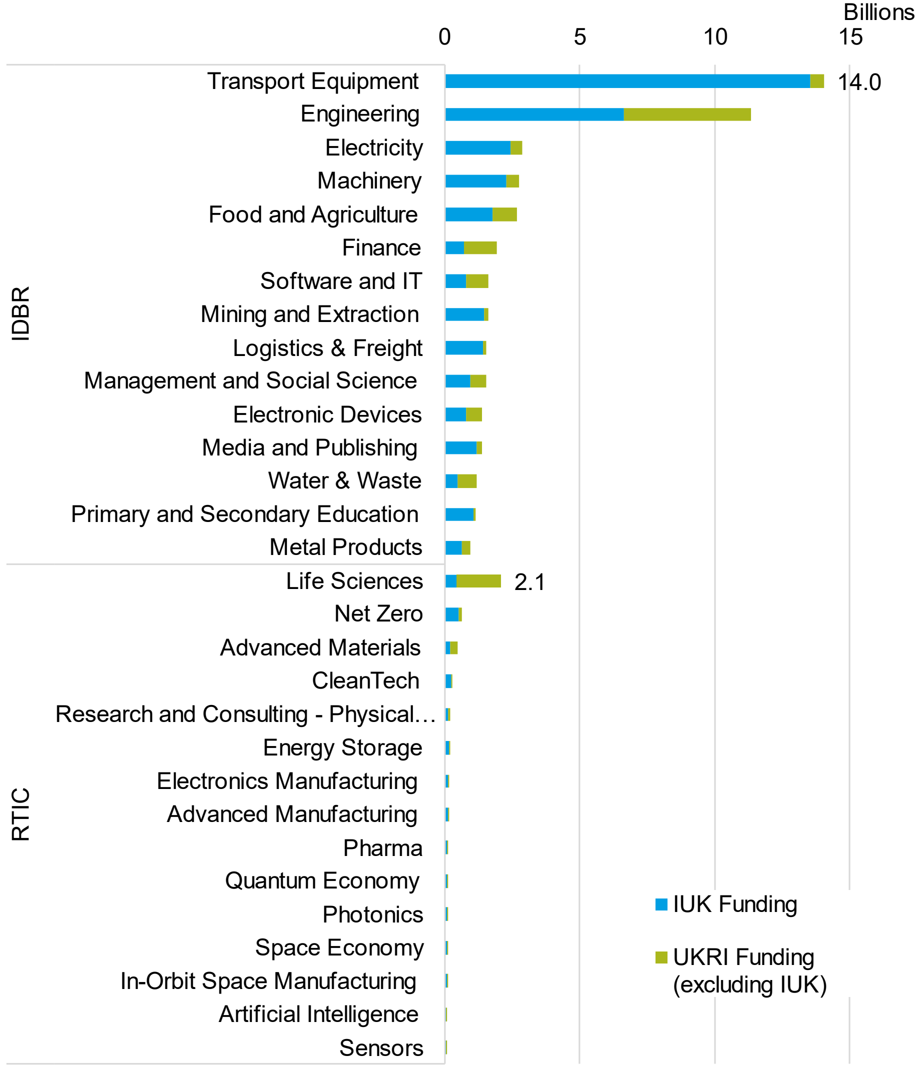
Notes: Cluster overlap is calculated as the share of every pair of clusters between two sectors that strongly overlap each other – Max. 60% (Blue) and Min. 5% (Yellow)

3.5 Patterns of investment

Figure 3.14 depicts the sectors that have received the most total IUK and other UKRI funding (Sep 2016 – Jan 2023). This is a proxy for public funding to innovative firms. Data on all public funding received by firms is not available at the granular level.

Transport equipment (which includes automotive and aerospace), engineering, electricity, machinery, food and agriculture, and life sciences have amassed the most impressive totals and IUK funding exceeds funding from other UKRI sources by a significant margin (except for life sciences). The total funding received is not comparable between IDBR sectors and RTIC sectors due to different degrees of data loss when assigning funding to organisations by name and registered location. See Annex A for details on funding allocation to the clusters. In addition, firms in RTIC emerging sectors may have less access to resources and experience in applying for UKRI funding.

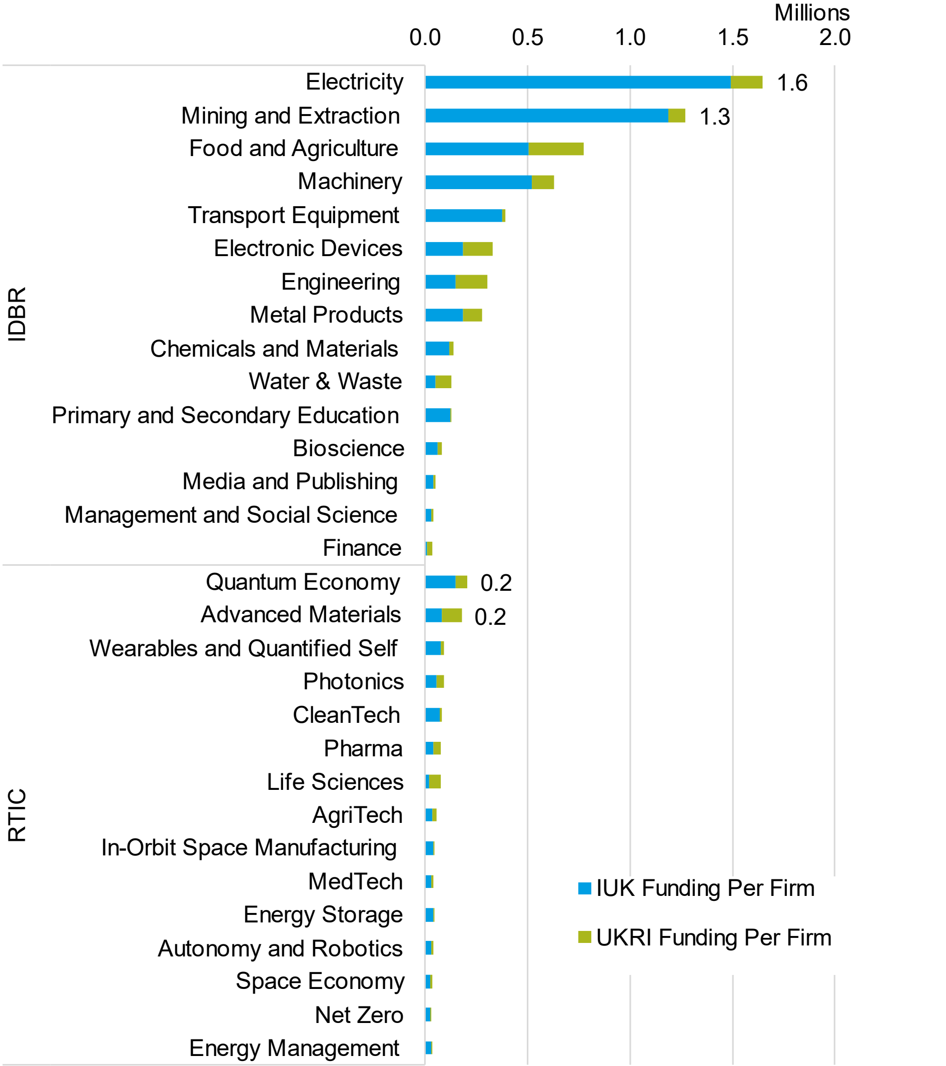
Figure 3.14 Top 15 IDBR and RTIC sectors by public funding received (£)



Sources: Cluster database developed by Cambridge Econometrics and The Data City, public funding from UKRI and IUK database.

Controlling for industry size (Figure 3.15) shows similar patterns in IDBR clusters – Primary and Secondary industries such as Food and Agriculture, Mining and Extraction, and Electricity and Machinery received the most public funding per firm. However, smaller and emerging RTIC clusters – in sectors like quantum and photonics – are capturing proportionally more funding per firm. The unevenness of funding distribution reflects the different propensities and levels of competence across sectors in seeking innovation funding but is also likely somewhat skewed by thematically specific funding calls that have targeted certain sectors more than others.

Figure 3.15 Top 15 IDBR and RTIC sectors by public funding received per firm (£)



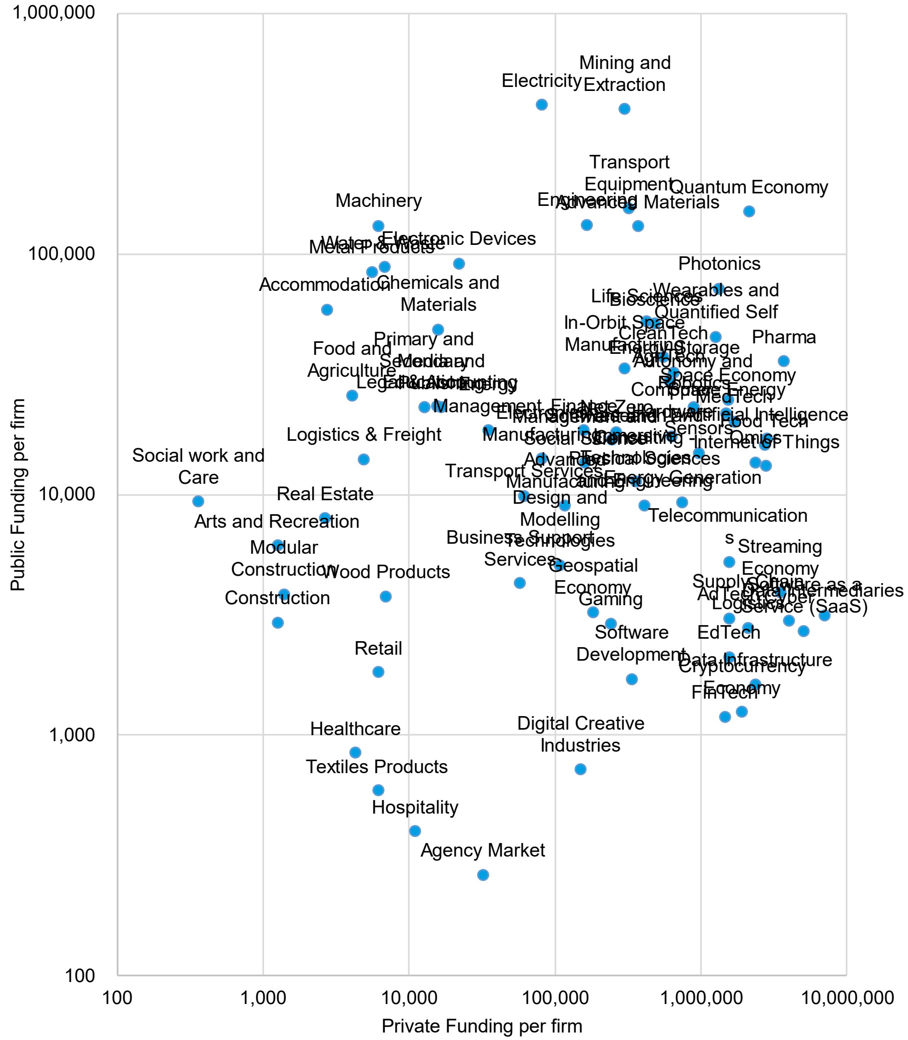
Sources: Cluster database developed by Cambridge Econometrics and The Data City, public funding from UKRI and IUK database.

We use venture capital (VC) investment per firm (Aug 2019 – Aug 2023) sourced from Dealroom to estimate private funding. This includes private funding from different rounds ranging from, for example, Seed to Unicorn. Figure 3.17 depicts that RTIC digital sectors such as Data Infrastructure, Cyber and Software as a Service received the highest total VC funding. Controlling for industry size, Figure 3.18 shows that Pharma, Cryptocurrency, and Food Tech instead received the most funding per firm. VC funding received by IDBR sectors are not compared due to data loss in matching Dealroom data to IDBR dataset.

Figure 3.16 compares public and private funding per firm across sectors. Overall, emerging sectors like Fintech received more private than public funding per firm, whereas the opposite is true for traditional sectors like Machinery. Physical capital heavy sectors, such as Mining and Extraction, Engineering, Materials sectors, receive a lot of both UKRI and Venture Capital funding. Sectors with greater levels of positive social externality, for example social work and care, arts and recreation, or water and waste, receive relatively higher levels of public compared to private funding. Whereas sectors with less obvious positive social externalities, and in which consumer surplus capture is easier, such as E-commerce and Advertising (Agency Market), receive relatively more private sector funding.

R&D tax credit data is not available at this geographical level but would be a valuable addition in the future as it covers all growth stages of R&D firms.

Figure 3.16 Public vs Private Funding per firm by sectors (£ in log scale)



Sources: Cluster database developed by Cambridge Econometrics and The Data City, public funding from UKRI and IUK database and private VC funding from Dealroom.

Figure 3.17 RTIC sectors by private funding received (£)

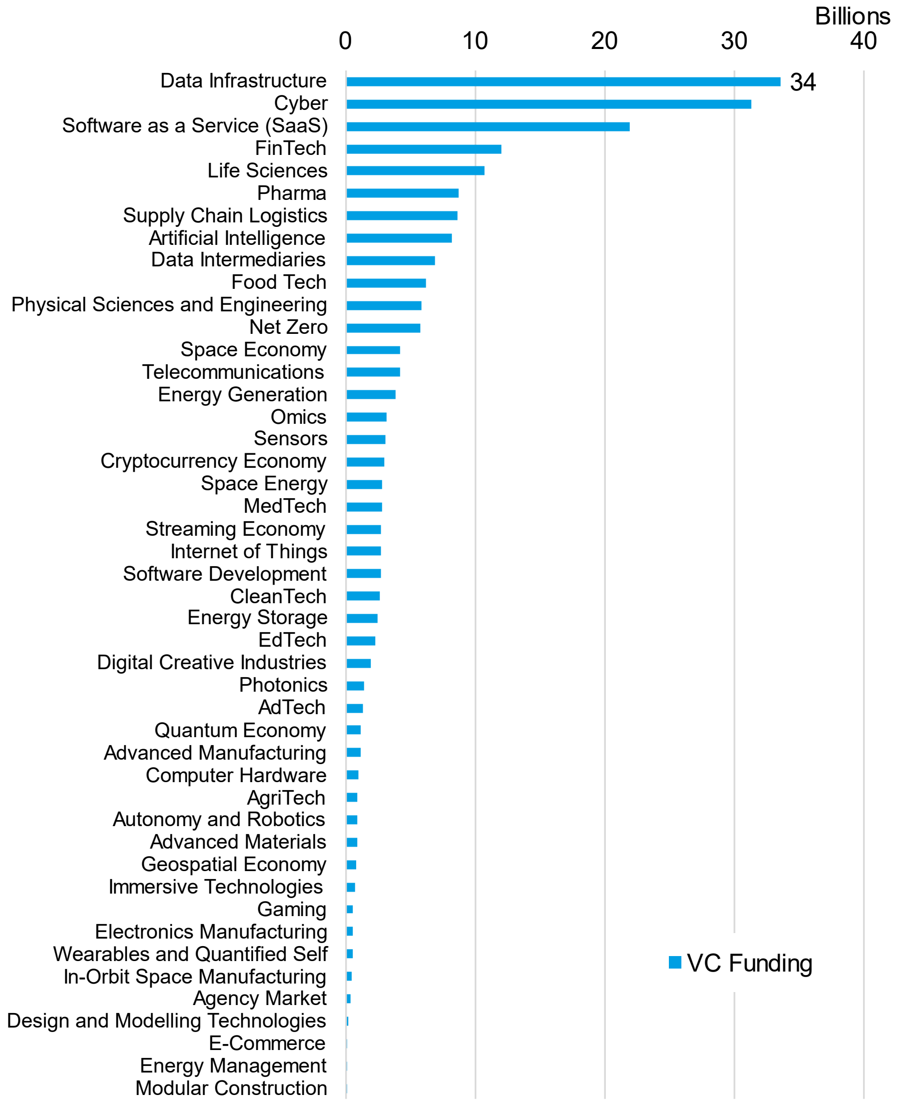
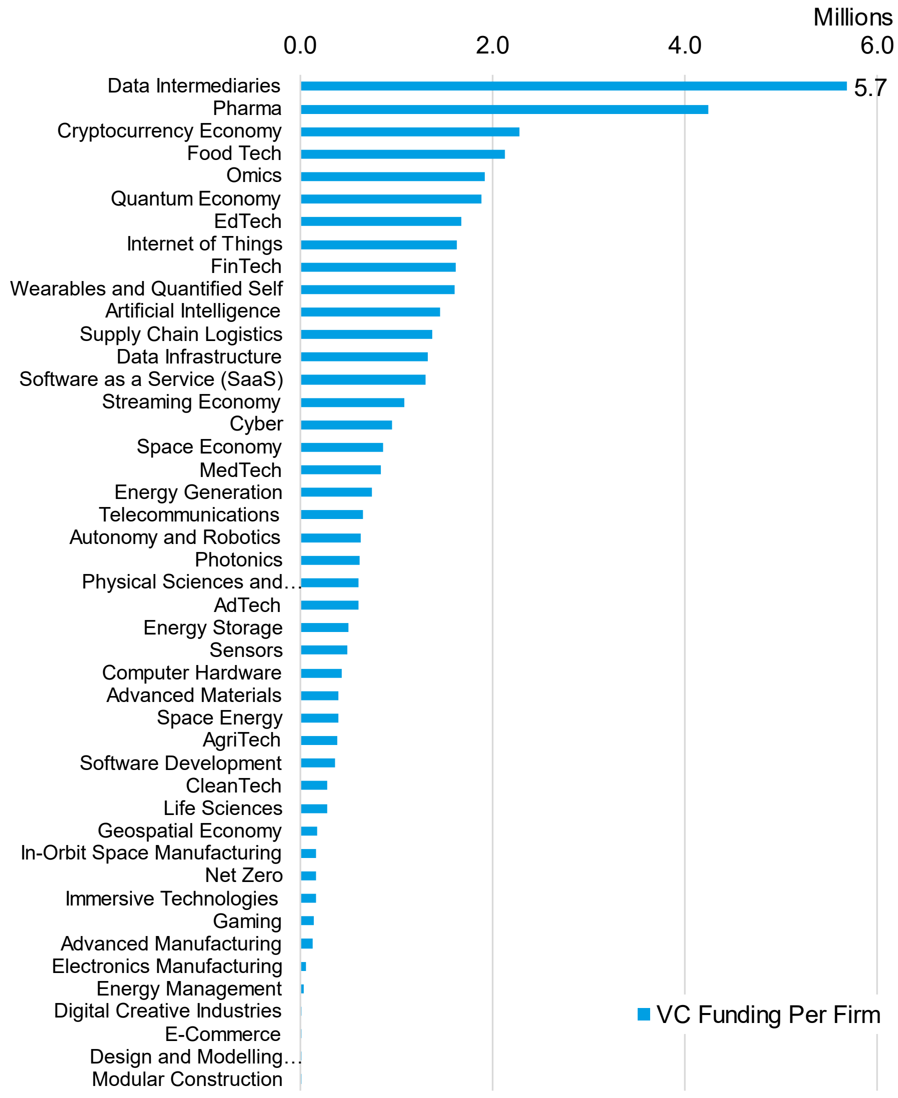


Figure 3.18 RTIC sectors by private funding received per firm (£)



Sources: Cluster database developed by Cambridge Econometrics and The Data City, private VC funding from Dealroom.

4 Conclusion – How to use this new evidence and future research

The analysis reflections in this report are just a small representation of the types of questions, challenges, and insights that this new and innovative data set enables. In addition to providing a more comprehensive spatial map of clustering activities – this data set is a rich resource for policymakers and researchers interested in exploring and enhancing the spatial dynamics of innovation. The accompanied interactive mapping tool visualises clusters identified with the dataset and allows users further exploration. However, in promoting its potential, it is also important to acknowledge this is the start and not the end for understanding RD&I clusters in the UK.

This data driven analysis should be supplemented with local intelligence and qualitative research to explore the enablers and barriers to maximise the benefits and growth of RD&I clusters. We cover most of these in the methodological discussions above, but it is important to consider them in perspective to emphasise what this research can effectively enable and suggest ways in which it might evolve.

Data sets like this are only as good as the data that underpins them. The data and the maps that result from them, should be regarded as the first step in a process that could result in deeper and more refined data and analysis. Our maps were produced using a combination of data sets – each of which come with limitations. Our methodology, while sound, required making decisions and compromises that others' might have approached differently. Subsequent iterations can build on this foundation, include additional sources of data as they become available, and go further to integrate them. More detailed and comprehensive data on RD&I activity would improve our estimates of innovativeness. Additional data on input-output and other relational data would enhance our ability to map networks of collaboration. There are also opportunities to increase the scope of insights about employment and skills.

The tool that we have produced provides a good indication of where cores of clustered activities are located and estimates across a variety of indicators. It is particularly useful for identifying industry hotspots and places with growth potential. It will help places understand their strengths – some of which might come as a surprise – and be a useful tool for developing innovation strategies that will increase and leverage potential synergies. That said, some may not recognise the geographies depicted or agree with industrial classifications. In some cases, our empirical analysis will challenge preconceptions in productive ways. In others, cluster geographies may differ from expectations due limitations in data or methodological decisions. For example, IDBR and RTIC clusters in similar sectors – such as chemicals and pharmaceuticals – often overlap. In some cases, this could be interpreted as different manifestations in the data of the same cluster. Relatedly, in some places, clusters that appear to be separate sectors may, in from some perspectives, be related. For instance, one investigation already identified that the co-location of transport equipment, computer hardware, and cyber indicated a single cluster centred on the defence and security sectors. Similarly, multiple clusters that are spatially separated on our maps may be considered by some to be different poles in a single larger cluster.

In recent analysis carried out for West Yorkshire Combined Authority for example, we combined analysis of the RTIC and IDBR clusters found here to centre within that geography with more traditional SIC-code based location quotient analysis and a literature review of previously identified priority clusters and sectors. This gave a more holistic overview as to the way in which firms across a range of traditional sectors were combining to form collaborative innovation clusters in knowledge areas that crossed the underlying sectors, for example we found how firms from the finance and software sectors were coming together to jointly form an emerging fintech cluster within the region. Equally importantly, we identified opportunity areas that were not currently being exploited, but perhaps could do so with greater policy support. Rather than using this data in isolation, we recommend using it to add additional depth and insight to existing qualitative and quantitative methods.

We encourage users to consider this new mapping tool and underlying data as building blocks that, with the benefit of contextual knowledge, can be arranged to tell stories about each place's unique offerings. These clusters are flags – invitations to investors and policy makers to explore opportunities and learn how local dynamics can catalyse and multiply growth.

References

Balland, P. A., Boschma, R., Crespo, J., & Rigby, D. L. (2018). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional studies*.

Bertrand, O., & Mol, M. J. (2013). The antecedents and innovation effects of domestic and offshore R&D outsourcing: The contingent impact of cognitive distance and absorptive capacity. *Strategic Management Journal*, *34*(6), 751-760.Blondel et al. 2008

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.

Boschma, R., Balland, P. A., & Kogler, D. F. (2015). Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010. *Industrial and corporate change*, *24*(1), 223-250.

Centre for Cities. (2014). Industrial revolutions: capturing the growth potential.

Centre for Cities. (2023). Innovation hotspots: Clustering the New Economy.

Cottineau, C., & Arcaute, E. (2020). The nested structure of urban business clusters. *Applied Network Science*, *5*(1), 1-20.Delgado and Porter 2021

Delgado, M., & Porter, M. E. (2021). Clusters and the great recession. *Available at SSRN 3819293*.

Department for Levelling Up, Housing and Communities. (2022). Levelling Up the United Kingdom. *Department for Levelling Up, Housing and Communities Policy Paper*.

Department for Science, Innovation and Technology. (2020). UK Research and Development Roadmap. *Department for Science, Innovation and Technology Policy Paper*.

Department for Science, Innovation and Technology. (2021). UK Innovation Strategy: leading the future by creating it. *Department for Science, Innovation and Technology Policy Paper*.

Department for Science, Innovation and Technology. (2023). UK Science and Technology Framework. *Department for Science, Innovation and Technology Policy Paper*.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, *99*(12), 7821-7826.

Goyal, S. (2023). *Networks: An economics approach*. MIT Press.

Hsieh, W. L., Ganotakis, P., Kafouros, M., & Wang, C. (2018). Foreign and domestic collaboration, product innovation novelty, and firm growth. *Journal of Product Innovation Management*, *35*(4), 652-672.

Kesidou, E., Love, J. H., Ozusaglam, S., & Wong, C. Y. (2023). Changing the geographic scope of collaboration: Implications for product innovation novelty and commercialization. *Journal of Product Innovation Management*, *40*(6), 859-881.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, *2*(11), 205.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic geography*, *87*(3), 237-265.

Nelles, J., Vallance, P., Vorley, T. & Wallace, P. (2022). *Understanding Cluster Growth Potential*. Retrieved from: https://innovationcaucus.co.uk/app/uploads/2022/06/ClusterReport_Final.pdf

Nelles, J., Walsh, K., Papazoglou, M., Nyanzu, E., Abdul Rahmen, S., & Vorley, T. (2023). *Understanding Cluster Growth Potential* - Place Based Innovation in the UK: Synthesis Report.

Sunley, P., Harris, J. L., Pike, A., Harris, R., Martin, R., & Evenhuis, E. (2022). Industrial policies, strategy and the UK's Levelling Up agenda. *Local Economy*, *37*(5), 403-418.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific reports*, *9*(1), 5233.

Treasury, H. M. S. (2021). Build Back Better: our plan for growth. *HM Treasury Policy Paper*.

Trippl, M., Grillitsch, M., & Isaksen, A. (2018). Exogenous sources of regional industrial change: Attraction and absorption of non-local knowledge for new path development. *Progress in human geography*, *42*(5), 687-705.

Tuckerman, L., Nelles, J., Walsh, K., & Vorley, T. (2023). Sustainable innovation policy: Examining the discourse of UK innovation policy. *Environmental Science & Policy*, *145*, 286-297.

UK Research and Innovation. (2022). UKRI strategy 2022 to 2027: transforming tomorrow together. *UK Research and Innovation Strategy Paper*.

Whittle, A., & Kogler, D. F. (2020). *Understanding regional branching knowledge diversification via inventor collaboration networks* (No. 2006). Utrecht University, Department of Human Geography and Spatial Planning, Group Economic Geography.

Annex A – Data Analysis

Cleaning IDBR dataset: Combine enterprises and local units

The IDBR dataset contains in 3 files:

- ents.csv – enterprises by enterprise reference number (entref)
- lus.csv – local units reference number (luref) by entref
- crns.csv – company reference number (CRN) by entref

The first two files contain details of the sites of the enterprises. The following table shows the contents they include respectively.

File	Firm Name	Postcode	Employment	Turnover	SIC codes
ents.csv	✓	✓	✓	✓	1+
lus.csv	✓	✓	✓	✗	Only 1

Enterprises with only one local unit but no head office (entref in lus.csv but not in ents.csv) have no turnover data. The single local unit is treated as a head office.

Enterprise with a head office and local units (entref in both ents.csv and lus.csv) will have their data combined and adjusted.
If the enterprise has:

1. A single local unit with 0 employment,
    a. with the same postcode as the head office, disregard the local unit.
    b. with different postcodes as the head office, treat the head office and local unit as two different offices, and the latter as having 0 turnover.
2. A single local unit with some positive employment,
    a. with the same employment and postcode as the head office, disregard the local unit.
    b. with different employment or postcodes as the head office, treat the head office and local unit as two different offices.
3. Multiple local units,
    a. if the total employment of the local units sums to that of the head office and at least one local unit share the same postcode as the head office, disregard the head office and treat the local unit that share a postcode as the new head office.
    b. if the total employment of the local units does not sum to that of the head office or they don't share postcodes, treat the head office and local units as different offices.

Given that turnover is missing in the local units file (lus.csv), the share of employment across the local units of the enterprise is used to split the enterprise's total turnover and impute the missing values.
Where the head office and local units are treated as different offices,

1. Subtract the total employment of local units from the enterprise employment to get employment of the head office, potentially leaving the head office with 0 employment.
2. Split out the enterprise turnover by the employment shares of all the offices.

Where one of the local units is treated as the head office,

1. Split out the enterprise turnover by the employment shares
2. Set one of the local units that share a postcode with the enterprise as head office.

Creating collaboration networks using IUK & MAKG dataset

When creating collaboration networks using the IUK and MAKG datasets, one challenge we faced was that universities were often listed as a single entity, despite the nature of their collaborations being departmentally specific. Left unaddressed, this would introduce false positive connections between university departments working in very different sectors. To avoid this, we used a data identifier "Innovation Area" to help us split universities out into approximate departments. The analysis steps were:

1. Identify any missing "Innovation Area" for the IUK funding applications using a Random Forest classifier trained on the project texts, such as titles, abstracts, and summaries.
2. Create pseudo university departments by combining universities and Innovation Areas for IUK dataset (or Research field for MAKG dataset) associated with the application.
3. Convert the application (publication) data into a collaboration network graph where:
   - Nodes represent organisations,
   - Lines represent collaborations, and
   - Line weights represent the frequency of collaboration between organisations.
4. Remove academic collaborations (university-to-university) to focus on private sector collaborations (firm-to-university and firm-to-firm).

The resulting IUK collaboration network contains about 48 thousand firms and university departments with 88 thousand collaboration pairs. The most frequent collaboration pair of organisations collaborated 28 times between 2016 and 2023. There are 2514 network components – connected subgroups in the IUK collaboration network. See Table 0.1 for the detailed breakdown of the number of components and their sizes after filtering the network by collaboration frequency.

Table 0.1 Number of components and their size in IUK collaboration network

	Number of components	Largest Component	2nd Largest Component	3rd Largest Component
No Filter	2,514	19,167	16	15
Collaborated at least twice	978 (at least size 2)	4,507	27	25
Collaborated at least twice with 2 others	222 (at least size 3)	3,236	18	16

Table 0.2 Number of organisations and collaborations in MAKG collaboration network after filtering

No. of Organisations	No. of Collaborating Pairs	No. of Collaborations
1,472	59,892	5,426,345

Business data on RTIC database

Reported business counts, turnover and employee data for the firms in the RTIC clusters are based on Companies House. This data is limited in that only approximately 100,000 of the 5.3 million registered firms submit their full financials to Companies House, and only 75% of them submit employee data. Estimates are made based upon this limited data source. RTIC firm location data is restricted to (1) registered addresses; and (2) operating addresses, which The Data City have identified as listed on the firm's website.

Other supplementary datasets and mappings

Other supplementary mappings are used to convert fields in the datasets, such as:

- Postcode to latitude and longitude
- Company name and Postcodes to Company Reference Number (CRN) from Company House
- CRN to SIC code from Company House
- CE's mapping of SIC2007 codes to 32 broad SIC sectors

Finding CRNs and postcodes by firm names

For data preparation, the goal is to identify the Company Reference Number (CRN) for each business in all the available datasets.

1. Using the Company House mapping from Company name to CRN, locate all business with exact matching names.
2. For the unmatched, perform a fuzzy match on the names from the set of firms that share the first 3-digit postcodes. Approve the match if the match quality is high[11].
3. For the rest of the unmatched firms, search for the firm name on the Company House API. The searched match is approved by the same criterion as step 2.

[11] Rapidfuzz package provides token_ratio and token_set_ratio scorers. Two names are matched if and only if they rank at least 90% in token_set_ratio and 80% in token_ratio.

Converting postcodes to x-y coordinates

Convert postcodes into longitude and latitude using pre-existing mappings from ONS[12]. To fill in the remaining postcodes, all postcodes are firstly grouped by the first 3-digits of the postcode (outcode), then sorted in alphabetical order. Within the same outcode group, the missing values are filled in using the nearest available postcode's longitude and latitude. This ensures organisation locations within the same outcode area have similar longitude and latitudes.

The longitudes and latitudes are then converted into x-y coordinates accounting for the earth's curvature using the haversine formula. Lastly, organisations are plotted onto a geographical map using the coordinates where spatial clustering can be performed.

Extracting UKRI and IUK funding for organisations

We distinguish projects by the source of their funding: IUK and UKRI (excluding IUK funding). If the funding allocation of a project to each collaborator is not available, the total funding is shared equally between its collaborators. The funding each organisation received is summed together, identified by its Company Reference Number (CRN) or the IUK Applicant ID.

When appending funding per organisation to the clusters, we split the funding equally by the number of sites of that organisation. For example, if enterprise X has 5 sites and received £10,000 in funding, each site is assumed to have received £2,000 in funding. This avoids double counting funding received, especially if more than one site appears in the same cluster.

[12] UK Postcodes with Latitude and Longitude – Downloaded in 2023-04-15

Some funding assigned to academic institutions are not identified because they often have no CRN. This undercounted funding received by IDBR and RTIC clusters.

Extracting Dealroom VC funding for organisations

The total investment that has been received in the past 4 years (Aug 2019 – Aug 2023) by firms with a registered address in the cluster. The total investment for each firm is allocated to the registered address, and thus the cluster containing the registered address, in absence of information on how the investment was deployed. This includes funding from many different rounds, ranging from, for example, Seed to Unicorn.

As with UKRI funding, if the funding allocation of a project to each collaborator is not available, the total funding is shared equally between its collaborators.

Creating meta collaboration network of clusters

There are collaborations between organisations in different clusters based on the IUK collaboration network. In addition, clusters can overlap, meaning one organisation identifier can appear in two clusters.

Using the identified clusters, we create a meta-collaboration network of clusters (called a quotient graph) where:

- Nodes represent clusters formed by groups of organisations,
- Edges represent collaborations and overlaps between clusters,
- Edge weights represent the number of cross-cluster collaboration and overlap.

Similarly, the meta-collaboration network of clusters by sector/county sums up the edges between each pairs of clusters between two sectors/counties. The final network represents the collaboration and overlaps between sectors/counties, revealing which sectors/counties are closely linked.

Radius of cluster from centroid

We estimated the geographical radius of a cluster from its centroid to study the size of a cluster. For a given cluster, we first calculated its centroid – the arithmetic mean of all the latitude-longitude of organisations within the cluster. The centroids were also used to approximately locate the counties where the clusters reside. We then estimated the average distance of all the organisations from the cluster centroid using the Haversine formula. This average radius method is computationally more efficient than other methods such as measuring the mean geodesic distance of all organisations.

Share of geographical overlap between two sectors

To understand geographical overlap of clusters between sectors, we count the number of overlapping clusters between two sectors and scale it by the sectoral totals.

We define two clusters as “weakly” overlapping if the separation between their centroids is less than the sum of their radii, and “strongly” overlapping if their separation is less than or equal to both radii. See Figure 0.1 and Figure 0.2 for a graphical representation of the definition. For this report, we focus on the “strong” overlaps.

Next, we count the number of overlapping clusters between the pair of sectors, scaling it by the geometric mean of the number of clusters in the pair of sectors. The average overlapping share is 30%.

Figure 0.1 “Weakly” overlapping clusters – separation less than sum of radii

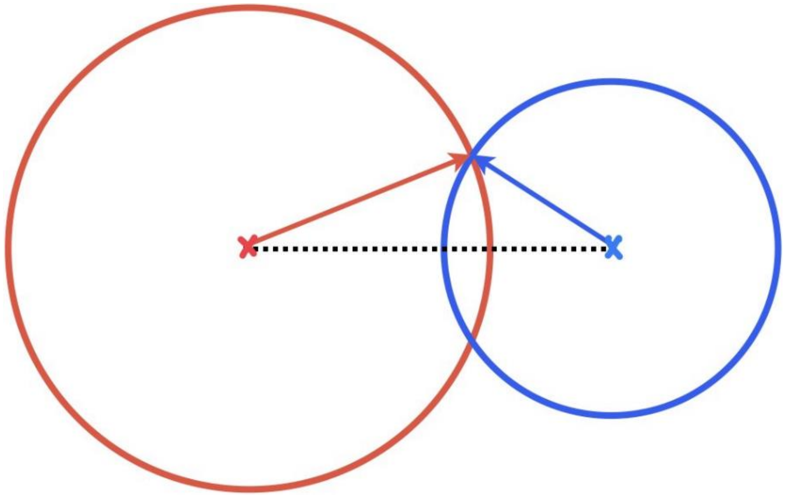
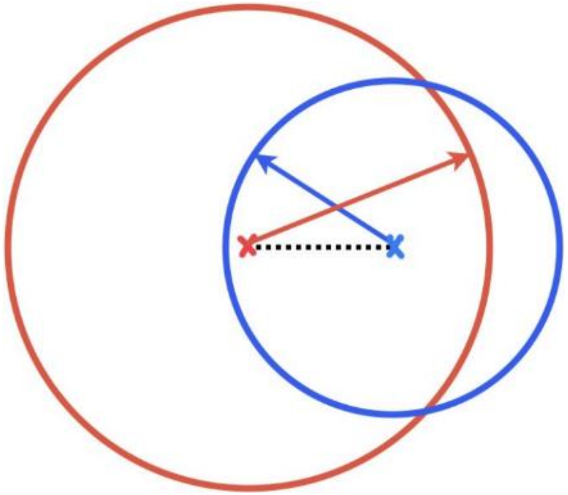


Figure 0.2 “Strongly” overlapping clusters – separation less than both radii



Annex B – Sensitivity Analysis

Selecting criteria thresholds: RD&I-active

Because of the nature of their activity in emerging sectors, RTIC clusters assumed to be RD&I-active. Similarly, IUK and MAKG clusters are also RD&I-active as they participate in innovation and research.

A cluster in the IDBR dataset is RD&I-active if:

1. at least 50 organisations in the cluster have a non-zero Innovation Score, or
2. at least 20% of the organisations in the cluster have a non-zero Innovation Score, or
3. at least 20 organisations applied for IUK grants, or
4. at least 10% of the organisations applied for IUK grants.

These four conditions are additive: the RD&I-active criterion is satisfied if at least one condition is satisfied. Conditions 2 and 4 correspond to conditions 1 and 3, but for clusters with less than 200 firms, looking at the proportion of the organisations in the cluster that satisfy them.

1642 out of 3965 IDBR clusters are RD&I-active by the four conditions. Condition 3 is the most critical in which removing it would reduce innovation clusters to 1323. Therefore, we test the sensitivity of the thresholds on condition 3.

Table 0.1 Number of innovation (RD&I-active) clusters, varying by threshold

		Thresholds		
Dataset	**Approach**	**10**	**20**	**30**
IDBR	**Spatial**	2194	1642	1441
RTIC	**Spatial**	1600	1600	1600
IUK	**Network**	99	99	99
IUK	**Spatial**	94	94	94
MAKG	**Network**	10	10	10
MAKG	**Spatial**	14	14	14
Total		4011	3459	3258

Selecting criteria thresholds: Specialised

Clusters from the IDBR and RTIC datasets pre-classified by broad sectors (SIC and RTIC) are considered as Specialised.

A IUK and MAKG cluster is specialised if at least 30% of the organisations in the cluster belong to one broad Industrial Classification category (SIC). Increasing the threshold reduces the number of IUK and MAKG clusters that satisfy the Specialised criteria. But they only account for a small fraction of the total number of clusters identified.

Table 0.2 Number of Specialised clusters, varying by threshold

		Thresholds			
Dataset	**Approach**	**20%**	**30%**	**40%**	**50%**
IDBR	**Spatial**	3965	3965	3965	3965
RTIC	**Spatial**	1600	1600	1600	1600
IUK	**Network**	86	45	15	7
IUK	**Spatial**	64	17	8	0
MAKG	**Network**	9	6	3	2
MAKG	**Spatial**	14	11	9	8
Total		5738	5644	5600	5582

Selecting criteria thresholds: Co-located

Clusters identified through the Spatial approach are Co-located by definition.

A cluster identified through the Network approach is Co-located if the average radius of the cluster is at most 30km, equivalent to the radius of Greater London (from Heathrow to Westminster). Varying the threshold has no impact on the number of Co-located clusters because IUK and MAKG clusters identified through the Network approach are highly spatially dispersed.

Table 0.3 Number of Co-located clusters, varying by threshold

		Thresholds		
Dataset	**Approach**	**20**	**30**	**40**
IDBR	Spatial	3965	3965	3965
RTIC	Spatial	1600	1600	1600
IUK	Network	1	1	1
IUK	Spatial	94	94	94
MAKG	Network	0	0	0
MAKG	Spatial	14	14	14
Total		5674	5674	5674

Selecting criteria thresholds: Internally Collaborative

Clusters identified through the Network approach are considered as Internally Collaborative.

A cluster identified through the Spatial approach is Internally Collaborative if there are at least 5 pairs of organisations within the cluster that applied for IUK funding together. The cluster is "Marginally Collaborative" if there is at least 1 pair. Increasing the threshold greatly reduces the number of Collaborative clusters because there only a few collaborating pairs of organizations in a cluster, which is due to data loss when matching IUK collaboration network to IDBR and RTIC datasets. Table 0.4 shows that the sensitivity to thresholds is limited to the IDBR and RTIC spatial clusters.

In the future, we can reduce the sensitivity to thresholds by supplementing collaboration indicators with datasets on patents applications and other forms of collaboration. This also increases the number of clusters satisfying the Collaborative criteria.

Table 0.4 Number of Collaborative clusters, varying by threshold

		Thresholds		
Dataset	**Approach**	**2**	**5**	**8**
IDBR	Spatial	546	251	115
RTIC	Spatial	147	85	23
IUK	Network	99	99	99
IUK	Spatial	92	79	71
MAKG	Network	10	10	10
MAKG	Spatial	14	14	14
Total		908	538	332

Selecting the HDBSCAN parameter

Selecting a low HDBSCAN parameter value means fewer number of firms are required for the collective to be classified as a cluster. For a sector with many firms, a low parameter value would result in many small clusters. In contrast, for a sector with less firms, a high parameter value would result in one large cluster spanning the country.

For the Spatial approach using the IDBR and IUK datasets, the HDBSCAN parameter was selected by a function of the number of organisations in each sector for IDBR and in the dataset for IUK. The function ensures that for the density of firms within clusters grow with the number of firms in the sector, but at a diminishing rate.

$$parameter = \sqrt{number\_of\_organsiations} \div 3$$

For example, the Chemicals and Materials IDBR sector has 22 thousand firms. The parameter value is 49 according to the function which identifies clusters between 50-1000 firms with spatial footprint about 10km wide (See Table 0.5). Furthermore, it identifies a reasonable number of clusters around Greater London. Increasing the parameter merges multiple clusters into one large cluster with about 4 thousand firms. Whereas reducing the parameter breaks up clusters, particularly in Greater London, into smaller clusters. Therefore, the parameter value should vary according to the number of organisations within the sector.

For RTIC sectors, the parameters are selected based on expert knowledge from The Data City.

Table 0.5 Number of clusters and size of clusters in Chemicals and Materials sector, by HDBSCAN parameter

HDBSCAN Parameter	Num. clusters	Num. clusters in Greater London	Largest cluster	Smallest cluster
37	85	15	1059	37
49	63	5	1087	49
74	38	1	3855	76

Page intentionally left blank