CBR DATABASE METHODOLOGY

Companies in the wider Cambridge region

This document describes how the database of over fifty thousand companies operating in the wider Cambridge region was created. The database contains only companies and limited partnerships (LLPs) and therefore excludes sole proprietorships and other forms of unincorporated businesses. The database provides data for turnover and employment of locally based companies, but it must be noted that these figures represent their employment and turnover in Cambridge and in other parts of the country and abroad. Further work is being carried out to enable us to provide an estimate of the local content of their employment. The source of the data is the FAME database produced by Bureau van Dijk which includes over 3 million active UK-based companies and a further 6 million that are dormant or who have died.

The database was first formed in 2015 and included any companies in the Cambridge Ahead area (see below) that were alive in any year from 2010/11 to 2014/15. We have now tracked births and deaths (and location changes) of this population from 2010/11 to 2019/20, but by default the Cambridge Cluster database shows only those alive at the end of the latest time period.

Since 2016 the database was extended to include all fourteen districts included in the Greater Cambridge Greater Peterborough Local Enterprise Partnership (LEP), backdated as before to 2010/11. This is now called the Business Board area (see below). This area has been updated to 2019/20 for the Combined Authority area (see below), but (due to a lack of sponsorship) only to 2017/18 for the other eight (now seven) districts in the Business Board area.

Cambridge Ahead area and the meaning of Cambridge companies

This is defined as a twenty-mile radius from the centre of Cambridge (see the outer circle in **Figure 1**). All postcodes (about 20,000) that fell within this area were selected. Each postcode was assigned to an electoral ward.

Figure 1 Cambridge Ahead area

The **Cambridge based** companies were then selected as companies with either their registered office, or their primary trading address, in one of the Cambridge Ahead postcodes. The location of each company for mapping purposes was taken to be the postcode of its primary trading address. In many cases the primary trading address was not provided. In general, the registered office address was then used as the location postcode. However, in many cases local knowledge and internet searches enabled us to identify a primary trading address despite it not being found on FAME. The whole of a company's operations is displayed at its primary trading address even if it operates at several locations in the Cambridge area. The database has over 26,000 Cambridge **based** companies in the Cambridge Ahead area alive in 2019/20.

Another group of businesses was identified as **Cambridge active** companies. These companies do not have their registered office, or their primary trading address, in one of the Cambridge Ahead postcodes but do operate within the Cambridge Ahead area. Most of these were identified since they gave a non-primary trading address within the Cambridge Ahead area. Others were identified from local company network lists, but were included only if the postcode of their trading location in the Cambridge Ahead area could be identified. In general, such companies are included on the database only if they have 20 or more employees. Whilst the financial data for these companies are collected, they are not reported since their global figures would be misleading for the Cambridge region. The database has over one thousand Cambridge **activ**e companies in the Cambridge Ahead area alive in 2019/20.

**Combined Authority area and the Business Board companies**

The Combined Authority area includes the following Local Authority Districts (see the grey shaded area in **Figure 2**): Cambridge; South Cambridgeshire; East Cambridgeshire; Huntingdonshire; Peterborough; Fenland.

The Business Board area includes the Combined Authority area and the following additional Districts (see the red line in **Figure 2**): King's Lynn and West Norfolk; North Hertfordshire; Rutland; South Holland; South Kesteven; Uttlesford; West Suffolk (formerly Forest Heath and St Edmundsbury). Data for any of these seven districts outside the Combined Authority area are provided only up to 2017-18.

The same distinction is made between those based in these areas and those operating there, but based elsewhere. The Combined Authority has over 26,000 **LEP based** companies alive in 2019/20, of which about 8,000 are outside the Cambridge Ahead area. It also has about eleven hundred **LEP active** companies alive in 2019/20.

If we look at the full Business Board area it has over 50,000 **LEP based** companies alive in 2019/20; and about seventeen hundred **LEP active** companies alive in 2019/20.

Figure 2 Combined Authority and Business Board areas

Contains Ordnance Survey Data. (c) Copyright & Database rights 2017.
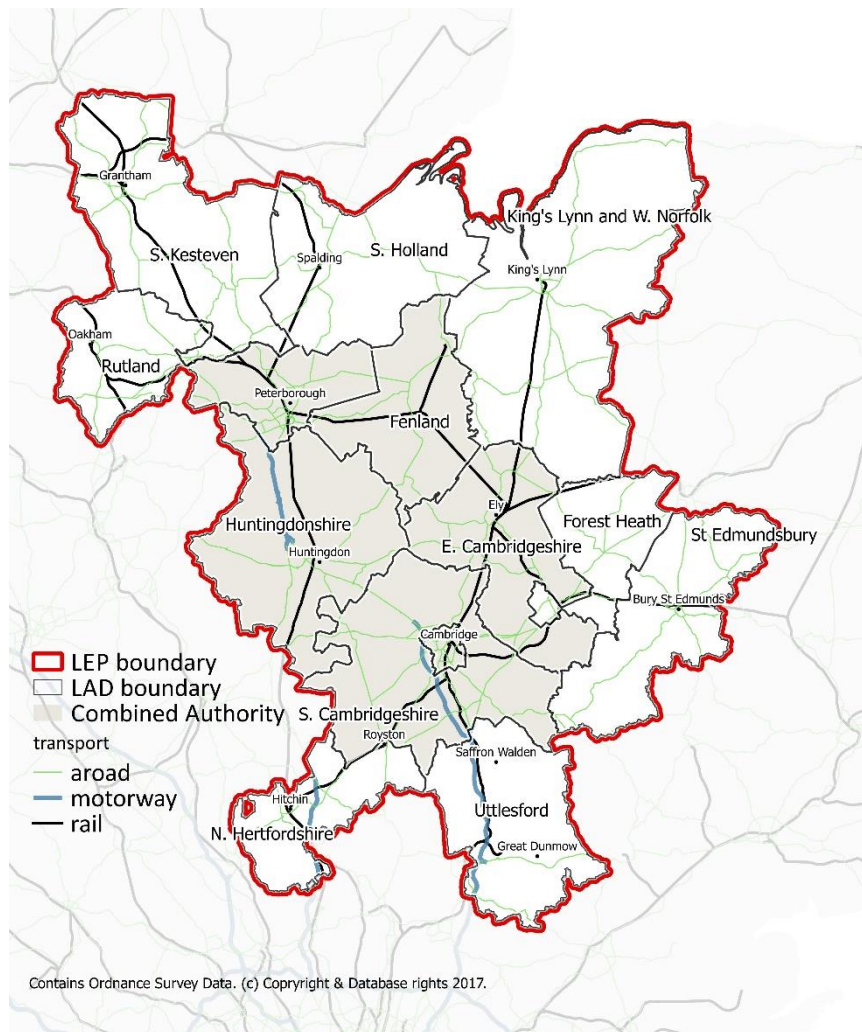
Updating the database and special cases

The updating over the years has introduced some new elements. First, there are those that moved into the area in the last year and those that moved out. The companies that have moved out are removed from the Cambridge Cluster database and their data for previous years is not shown. Companies that have moved into the area are now included, along with their data for any years during the last ten that they were alive. This methodology provides a vibrant picture of the performance of companies that currently exist in the area over the past ten years.

A further complication of a longitudinal version of the Cambridge Cluster database is that some foreign owned businesses in the area have stopped reporting financial data since

2014/15. This means that we could show significant falls in employment and turnover if we were to continue to include them as Cambridge based companies (which they are!). Instead, we decided to move them to Cambridge active companies on the database (for all of the years they were alive) in order to avoid this distortion.

Finally, we are delighted to note that some major international businesses like AstraZeneca PLC, Aveva and Arm have their international headquarters in Cambridge; and as such they are classified as Cambridge based. However, to include their global figures would swamp the total employment and turnover of Cambridge based companies. Therefore, we have managed to obtain data for their local employment over the past ten years and have included this on the database.

Period covered and data collected

The 2020 draw gathered data for ten years from 2010/11 to 2019/20. The year 2019/20 includes data for companies for accounting year ends between 6th April 2019 and 5th April 2020. It was decided to select any company in our areas that had been operating in any of these ten years and these are included on the database. However, by default Cambridge Cluster Insights ([Cambridge Cluster Insights - Cambridge Ahead)](Cambridge Cluster Insights - Cambridge Ahead) shows only active companies, not those that have died, or moved away.

Since we have to wait for companies to audit and approve their accounts and then file them with Companies House, there is an inevitable lag in the data portrayed on the Cambridge Cluster. The annual update is carried out in the autumn and becomes available at the start of the following year.

The data gathered for each company included: Company name; Registered number; Registered accounts type; Legal form; Current Market Capitalisation; Trade Description; its Sector according to the Standard Industrial Classification (SIC) 2007; Peer Group; IPO date; current Immediate Shareholder information; current Domestic and Global ultimate owner; Registered office address and postcode; Primary trading address & employment when provided; and Other trading address in Cambridge area and employment when provided.

The financial information included: Accounting date; total assets; employment; and turnover for each of the ten years if available. It should be noted that employment represents the average number of employees (including part-time) over the accounting year. All changes of

accounting dates were identified and adjustments were made where necessary to annualise the accounting information.

Some companies will move their trading activity between holding companies and subsidiaries. Where appropriate (and rather than showing an artificial death and birth) the past data for the former company is attached to the new company to give a continuous flow of annual data.

Treatment of parents and subsidiaries

This is probably the most difficult aspect of building the database and requires judgements to be made. First, Cambridge based companies with no parent and no subsidiaries on the database can be retained in the dataset. Second, for the remaining companies, all parents and their subsidiaries are gathered together. Third, if the parent is on the database and includes the information about its subsidiaries in its consolidated accounts, then the subsidiaries are dropped and only the parent is retained on the database. Fourth, for those companies with a parent that is a holding company and does not report the consolidated accounts for itself and its subsidiaries, the parent company is removed from the database and the subsidiaries are retained. Fifth, the same procedure is carried out for Cambridge active companies. This means that it is possible to have a parent as Cambridge active and its subsidiaries as Cambridge based since this will not involve double-counting of financial data.

There are also cases where a company ceases to trade because it transfers its activities to a new company registration. This would show up on the database as a death and a birth whilst no real change has occurred in its economic activity. We have attempted to identify such cases and to show the company as continuing. Another problem is the acquisition of a Cambridge based company by a company from outside the area. In some cases the Cambridge based company may continue to trade and report as before, but in other cases it may be wound up despite its activity in Cambridge being unchanged. We continue to work on this latter problem which will be helped if this database is maintained over time so that these acquisitions can be tracked.

Marshall of Cambridge is a special case and is treated as such. Part of the business, Marshall Motor Holdings PLC has a stock market listing and so this is separated from the rest of the business (largely Marshall Aerospace). We treat the former as non-KI and the latter as KI.

Sector reclassifications

Companies are classified into their principal activity. The SICs for the companies are self-reported. Several hundred companies had not reported their SIC. In these cases their activity descriptions and their names (and occasionally internet searches) allowed us to assign SICs to them.

Another group of companies were assigned to generic activities such as 'head offices' or 'holding companies'. These were reassigned into SICs that represented their economic activities.

We decided to have 'Life Science and Healthcare' and 'Information Technology and Telecoms' as two of our sectors owing to their importance in the Cambridge region. This involved identifying these businesses across a wide variety of SICs through the use of word searches within their business name and stated activities. We were helped in achieving this (and in other decisions) by an expert panel of local business representatives and economic consultants.

The sectors assigned to the largest companies in the area, and to those with the largest changes in employment over time, are also reviewed every year through 'eyeballing' sessions with economic development officers and other local experts.

Classification of sectors as knowledge intensive (KI)

The classification of sectors is based on the London Analysis, Identifying Science and Technology Businesses in Official Statistics that can be found at:

www.ons.gov.uk/ons/rel/regional-trends/london-analysis/identifying-science-and-technology-businesses-in-official-statistics/index.html

We rely primarily on the science and technology indicator provided in this publication and their analysis of how each SIC07 code is classified according to the OECD and Eurostat "High-tech statistics" publication available here:

https://ec.europa.eu/eurostat/documents/3217494/5651313/KS-57-03-104-2-EN.PDF/923b5c6d-d32b-47d1-88f6-5b8b7ecd5047?version=1.0

We classify KI sectors as Information technology and telecoms; Life science and healthcare; High-tech manufacturing; and Knowledge intensive services.

The OECD definition classifies knowledge intensive services (KIS) into four groups: KIS financial services; KIS market services; other KIS; and KIS other high-tech services. As explained below, in general we take only the last of these to be KI in our definition.

In general, the KI sectors are: High technology and medium-high technology manufacturing and KIS high-tech services. Low and medium-low technology manufacturing, KIS market services, KIS financial services and Other KIS are generally classified as non-KI.

There are a few exceptions to these general rules and these are discussed below.

SIC 42220

This concerns construction projects for telecoms and electricity and we have taken these seven companies to be KI.

SICs 53100 – 53202

These are postal courier services and despite them being classified as KIS high-tech services by the OECD, we do not regard them as KI.

SICs 59111 – 59120

These sectors involve TV and video production and despite being classified by the OECD as Other KIS, we have classified them as KI.

SICs 59200 – 60200

These sectors involve TV and radio broadcasting and despite being classified by the OECD as Other KIS, we have classified them as KI.

SICs 69109 & 77400

These sectors cover intellectual property protection and despite being classified as KIS market services, we have classified them as KI.

SICs 71121 – 71200 and 74901

Despite these being classified as KIS market services, we have classified them as KI. In support of this decision, their science and technology category is given as 'Other scientific/technological services'. However, to be consistent we should also have classified quantity surveyors as KI, but we did not.

SIC 75000

This sector covers veterinary activities. After some debate it was decided to not classify as KI despite the importance of the racing industry in Newmarket.

SICs 86101 – 86900

These sectors involve human health activities and whilst some of the companies within these sectors may be classifiable as KI, we decided to classify these sectors as a whole as non-KI.

Of course, it should be noted that we have classified sectors, not companies. There will be KI companies operating in non-KI sectors and vice versa.

Estimation of global employment and turnover figures

The vast majority of our companies are SMEs and take the full exemption in filing their accounts. This means that not much more than balance sheet figures were presented up until 2016/17; but since that year approximately 75% of the companies provide employment data (and the remaining quarter are largely one-person businesses). This means that we need to adopt a variety of approaches to the estimation of employment and turnover. First we tackle employment, using accounting information for the approximately 39,000 companies that provide it. Second, when this is missing we frequently have the number of employees at their primary trading address provided by the company (over 8000 cases). The third approach begins by estimating employment from total assets for the remainder. To do this we collect employment, turnover and total assets for all companies on FAME. From these we calculate the median employment to total assets ratio and the median turnover to employment ratio for each four-digit SIC averaged over the past three years (amalgamating adjacent SICs when the number of observations is too small). We then use the appropriate employment to assets median to estimate employment for each company that has not reported it. Occasionally, companies reported turnover figures but not employment and in these cases employment was also estimated on the basis of turnover.

We now have employment best estimates that are robust and these can be used to estimate turnover figures by using the turnover to employment ratios. This was done for all companies with turnover not provided and calculated for all ten years (with missing values where the employment figure is missing).

Andy Cosh, Senior Research Associate

Giorgio Caselli, Research Fellow

Centre for Business Research, University of Cambridge

January 2021

Annex 1   Non-corporate Research Institutions – Methodology

The significance of research-based non-corporate organisations in the Cambridge region led to the decision to include them on the Cambridge Cluster database. This note concerns how the data was gathered.

1. Institutions were approached by email and/or telephone for their employment data. Only universities and some hospitals provided sufficient detail to enable us to separate knowledge intensive from non-knowledge intensive employment. Figures for all other institutions are total employment.

2. **University of Cambridge data**

a) **University staff** headcount data are provided by the University. The figures differ from those in the University's Databooks by virtue of inclusion here of zero-hours contracts.

The figures shown on the Cambridge Cluster database are for Academic plus Academic-Related plus Research staff employed in all Schools and in Council/General Board/Unified Administrative Service institutions.

Institutions *excluded* from the University headcount are:

>   "External" employees of the ADC
>   Cambridge Enterprise [NB included in Corporates]
>   Gates Cambridge Trust
>   Malaysian Commonwealth Studies Trust
>   Judge Executive Education
>   Needham Research Institute
>   Strangeways Research Lab [= PHG Foundation in Research Institutes list]
>   Cambridge Commonwealth, European and International Trust
>   The Dental Practice
>   The Sports & Social Club
>   Great St Mary's
>   Cambridge University Press
>   Cambridge Assessment
>   MRC Mitochondrial Biology Unit
>   MRC Cognition and Brain Sciences Unit
>   MRC Biostatistics Unit

Most of these excluded units appear separately on the Cambridge Cluster. Figures shown in the non-corporate section for **Cambridge University Press** and **Cambridge Assessment** are for Cambridge-based staff employed by "The Chancellor, Masters, and Scholars of the

University of Cambridge"; Cambridge-based staff employed at corporate subsidiaries are excluded but are shown on the Cambridge Cluster as corporate entities.

b) **Cambridge University research students**: the source for our numbers here is the Student Numbers dashboard (https://www.prao.admin.cam.ac.uk/data-analysis-planning/student-numbers/snapshot-courseschooldepartment), published by the University. We include full time research students (PGR) in all faculties but we exclude part time students and those writing up, who are likely to be present only occasionally in Cambridge.

The PGR classification was set in 2015 using the then existing HEFCE definitions. It includes the following degrees:

> Ph.D
> Ed.D
> Eng.D
> M.D.
> M.Litt
> M.Sc
> M.Phil excluding a list of 30 programmes (see list below)
> M.Res
> M.St
> M.Ed
> Students registered as Probationary Ph.D
> Diploma in International Law
> Diploma in Legal Studies
> Certificate of Postgraduate Study

The list of M.Phils classified as Teaching (PGT) rather than Research degrees is as follows:

> Advanced Chemical Engineering
> Advanced Computer Science
> Archaeology
> Architecture and Urban Design
> Assyriology
> Bioscience Enterprise
> Computational Biology
> Conservation Leadership
> Development Studies
> Economic Research
> Economics
> Egyptology
> Energy Technologies

Engineering for Sustainable Development
Environmental Policy
Epidemiology
Finance
Finance and Economics
Industrial Systems, Manufacturing, and Management
Innovation, Strategy, and Organization
Management
Management Science and Operations
Micro- and Nanotechnology Enterprise
Nuclear Energy
Planning, Growth, and Regeneration
Public Health
Public Policy
Real Estate Finance
Scientific Computing
Technology Policy

Starting from 2016/17, a number of MPhil courses were reclassified from research to taught. Student numbers consistent with the 2015 PGR classification are provided to us by the Planning and Resource Allocation Office at the University of Cambridge, so as to avoid any impact on student numbers that is purely due to classification changes.

c) **Cambridge Colleges**: the Cambridge Cluster shows full-time equivalent (FTE) College Teaching Officers (CTOs) plus stipendiary Research Fellows for all 31 colleges. CTO data until 2015/16 are from Table 44 of the (unpublished) Cambridge Colleges accounts collected by the Bursars' Committee, while data from 2016/17 onwards are provided to us by the Planning and Resource Allocation Office at the University of Cambridge; however, the census date for this table varies from year to year and it is unclear for example whether a 1 October census shown in Table 44 of the 2015 accounts refers to 2015 or 2016. The Bursars' Committee survey of Research Fellows was last conducted in 2010-11, when 180 such Fellows were identified (excluding 6 Title B Fellows at Trinity). We assume that this number has held steady over time across all Colleges.

Non-academic staff (FTE) totals are taken directly from the (unpublished) Cambridge Colleges accounts and/or from their annual reports.

3. **Cambridge University Hospitals (CUH) NHS Foundation Trust**: data are provided to us by Addenbrooke's. We exclude from the headcount shown on the Cambridge Cluster database all staff we assess to be in non-knowledge intensive occupations. Figures from

2017/18 onwards include staff who were employed by The Pathology Partnership before it was dissolved at the end of April 2017.

4. **Cambridgeshire and Peterborough NHS Foundation Trust (CPFT)**:

Peterborough units are excluded, and previous years (which covered only the Fulbourn site) have been amended to include Huntingdon and Fenland units. Note that in April 2015 CPFT absorbed parts of Cambridgeshire Community Services.

5. **EMBL-European Bioinformatics Institute (EBI); and the Wellcome Genome Campus**:

We are grateful to these organisations for the data provided and the clarifications of their relationship to the companies at that site on the Cambridge Cluster.

Figures for the Wellcome Sanger Institute include staff working on the Connecting Science programme.

6. **MRC**: the data for all five MRC units around Cambridge are provided by the central unit until 2015/16, while data from 2016/17 onwards is sourced as follows:

- MRC Laboratory of Molecular Biology and MRC Elsie Widdowson Laboratory: MRC central unit.
- MRC Mitochondrial Biology Unit, MRC Cognition and Brain Sciences Unit, and MRC Biostatistics Unit: University of Cambridge.

The MRC Elsie Widdowson Laboratory is included on the Cambridge Cluster database until 2017/18 as it closed in December 2018.

Figures for the three MRC units that became part of the University of Cambridge in 2016/17 are reported separately on the Cambridge Cluster and are excluded from the University headcount.

Organisations included:

Anglia Ruskin University - Cambridge Campus
Anglia Ruskin University - Cambridge Campus research students
British Antartic Survey
Cambridge Arctic Shelf Programme (CASP)
Cambridge Assessment (UCLES)
Cambridge Colleges
Cambridge Commonwealth, European and International Trust
Cambridge University
Cambridge University Hospitals (Addenbrooke's)
Cambridge University Press
Cambridge University research students
Cambridgeshire and Peterborough NHS Foundation Trust
Cancer Research UK Therapeutic Discovery Laboratories (CRUK-TDL)
European Bioinformatics Institute (EMBL-EBI)
Gates Cambridge Trust
Hinchingbrooke Hospital
Hitachi Cambridge Laboratory
MRC Biostatistics Unit
MRC Cognition and Brain Sciences Unit
MRC Elsie Widdowson Laboratory
MRC Laboratory of Molecular Biology
MRC Mitochondrial Biology Unit
Needham Research Institute
PHG Foundation (Foundation for Genomics & Population Health)
Public Health England
Public Library of Science (UK European Editorial Office)
Royal Papworth Hospital
Royal Society of Chemistry
The Pathology Partnership (until 2016/17)
UNEP-WCMC (UN Environment Programme World Conservation Monitoring Centre)
Wellcome Sanger Institute