Drivers of Entrepreneurship and Small Businesses: Project overview and database design

Robert J. Bennett, Harry Smith, Carry van Lieshout and Gill Newton

rjb7@cam.ac.uk hjs57@cam.ac.uk cv313@cam.ac.uk ghn22@cam.ac.uk

Working Paper 1:
Working paper series from ESRC project ES/M010953:
Drivers of Entrepreneurship and Small Businesses

University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure, Downing Place, Cambridge, CB2 3EN, UK.

May 2017

Comments are welcomed on this paper: contact the authors as above.

Keywords: Entrepreneurship, Employers, Self-employment, Small businesses, Census

JEL Codes: L26, L25, D13, D22

Drivers of Entrepreneurship and Small Businesses: Project overview and database design

Robert J. Bennett, Harry Smith, Carry van Lieshout and Gill Newton

Working Paper 1: ESRC project ES/M010953: Drivers of Entrepreneurship and Small Businesses, University of Cambridge.

1. Introduction.

This paper introduces the ESRC-supported project ES/M010953 'Drivers of Entrepreneurship and Small Businesses'. Its aim is to give a project overview, to introduce the design of the database that has been developed, and provide a guide to the decisions that have been made in its construction. It also acts as introduction to more detailed documentation of the database deposited at UK Data Archive (UKDA)[1] and other research developments and analysis that are discussed in a series of other working papers and publications. These other papers also develop the academic background; the purpose here is to introduce the database and how it can be used and analysed.

It has been long recognised that a significant gap in data on the business population in the nineteenth century exists. This has been a major barrier to empirical research and also to the development of theoretical understanding. Some efforts to fill that gap have been made through case studies, individual business histories, by use of the data on company's registration, Factory Inspectorate information, and a range of the official and other surveys that exist in Parliamentary Papers, and other sources. However, none of these give a fully satisfactory coverage of all business, and it is difficult to align sources with each other to produce a consistent series over time. The aim of this working paper is to show how use of population census information can fill most of the gaps that have previously impeded research. It is recognised from the outset that the population census was not a business census

[1] *User Guide*: target for database deposit is mid-2018.

and the results will be imperfect. The census was administered by the General Register Office (GRO) to count the population, with information on industry and the economy a secondary consideration. However, the census made significant attempts to obtain systematic information on all employers and own account self-employed, and for early years it also attempted to obtain information on the workforces of employers (including farmers), and the acreage of farms. Because there was also a legal obligation to reply and to provide accurate information the census also has the advantage of near-complete coverage and a high level of accuracy of what it collected.

Being a source that describes characteristics of each individual person, the census provides information that is unique in its scale of coverage of the whole population. Furthermore, it can be linked to other sources on individuals at census dates. The scale of the data, containing over 1-2 million individual entrepreneurs per year (and comparative information on the rest of the whole population) offers unique new research opportunities. The ESRC project which is introduced in this working paper seeks to develop this potential through data extraction and data enrichment, leading to a database deposit that should open the way for a wide range of analysis by the current authors, and other researchers. This includes the potential for further data enrichment that can allow the information on 19th and early 20th century entrepreneurs to be developed further.

The population census allows individual employers and the self-employed to be identified, provides scope to estimate their total numbers, and for census years 1851-81 gives the employee numbers of their businesses and the acreages of farms. Information on employers was gathered by the census in all years from 1851, but only for some years was any result published from the enquiry: in 1851 in summary tables nationally for males only and for regions, and then in a variable and rather limited format from 1891 onwards. However, even in the years for which some results were published, the information was very limited and subject to awkward aggregations and exclusions, which are also inconsistent over time. For example, the national aggregate tables distinguishing between employers and workers published in 1891-1911 only give employer, own account and worker numbers for certain sectors and this changed from census to census; thus, farmers were not distinguished as employers and non-employers in 1891 or 1901, but were in the 1911 report.

Despite the lack of formal census publications, the data collected survives and is contained in the original Census Enumerators Books (CEBs). These data have recently become available electronically and can be used for research and analysis in a way not previously possible. However, for employers and entrepreneurs, the CEB data are difficult to extract and process, and the transcription and encoding process that has been used to obtain the electronic data has many gaps and deficiencies. Indeed a key contribution of the ESRC project is to overcome these deficiencies so that others can use the census data on entrepreneurs without having to confront its problems and the sources of bias deriving from non-available, missing or truncated entries, so that users will have no additional need to infill gaps, or embark on complex re-coding of the original records. The user will therefore be able to go directly to a set of data files on individual business people that are cleaned, corrected, re-coded in various ways, and supplemented to infill lost, truncated and missing material. Without this approach users of the census for identifying entrepreneurs are likely to extract very partial and biased samples and draw very misleading conclusions.

The starting point is the electronic census data that have recently been made available through an ESRC project based at the University of Essex which has produced the database deposit of the original CEBs and made available at the UKDA: *The Integrated Census Microdata* (I-CeM).[2] This resource is an extremely valuable starting point which took its constructors many years to create. It is a huge tribute to their energies and perseverance that I-CeM exists at all. Since 2014 it has provided the key resource for anyone wanting to undertake research on the historic British censuses for 1851-1911. However, for a variety of reasons, and despite all the efforts made, I-CeM gives imperfect coverage of the original CEBs, and it is particularly imperfect for the records of employers that the census originally collected. This paper outlines the decisions that have to been made in order to extract these data, infill gaps, and the cleaning and coding necessary to make them more accurate and generally useable. The paper also describes the way in which they have been supplemented in order to yield a more complete database of business proprietors and entrepreneurs for all years 1851-1911. This supplementation covers two aspects:

(i) Supplementation to ensure as complete as possible coverage of what was originally contained in the census but is imperfectly captured in transcriptions used by I-CeM; and

[2] Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481

(ii) Supplementation to include as far as possible what was systematically omitted in the original census itself.

The aim of the ESRC project and this paper is to contribute to five main project outputs:

1. Construction of an aligned, complete, quality-controlled, and consistent database of business proprietors/businesses for 1851-1911.
2. Deposit of this database at UKDA to provide an open and generally accessible new resource for economic and business research.
3. Inclusion in the database, as far as possible, of enrichment from other major sources to ensure that the business population identified is as complete as possible.
4. Interpretation of the main developments in businesses, self-employment and economic development using this new information.
5. Contributions to theories of business and economic development over the long term.

A major outcome is a database that for the first time provides a fully comprehensive source for businesses over the nineteenth and early twentieth centuries. The material seeks to include as fully as possible *employers* (those who employed others), *sole proprietor own account self-employed* (who employed no-one else), and *corporations* (either as the *company entity* itself and/or for its *individual directors*).

It is not claimed that the final result is a database that can include all businesses, but it seeks to come as closely as possible to meeting this target with the information available. Two overlapping elements are represented: individual entrepreneurs; and businesses. The database is available in three parts: first, for *individuals*, numbering 200,000 to 2 million records per year depending on their category; second, for *partnerships and companies,* owned or led by more than one individual; and third, *spatial aggregates* at the geographical level of either parishes, Registration Sub districts, or Counties.

The rest of the paper is divided into four sections. Section 2 discusses the breadth of the business population as a whole and how this was captured by the census. It also introduces the sources needed to supplement where a more complete coverage is needed. Section 3 summarises the census sources used and how they are aligned to ensure a quality-controlled and consistent data base deposited at the UKDA. Section 4 discusses the comparator

measures that can be used to calculate entrepreneurship rates. This paper provides the overview of the data and project; subsequent working papers and publications outline the different methods and stages of analysis required to use and interpret these data.

2. Definitions and targets of analysis, business categories and sources

For the purposes of this working paper and the project 'Drivers of Entrepreneurship and Small Businesses', business proprietors and entrepreneurs are defined as widely as possible, with information and coding added so that different groups and categories can be selected by users as required to answer different research questions. The core of available information is the statement or imputation of employer status in the census. From this and other resources, the base objective is to identify *business proprietors*: those that were, at the date of the census (or a nearby date where other data have to be used) *currently* the responsible individuals bearing the risks of running private business enterprises, however small, and even if they did this alone with no other individuals involved (as self-employed individuals). This definition is drawn very broadly to include all the major types of business proprietors that have been the focus of most previous research. However, there are some categories of entrepreneur that are excluded because no information is recorded or survives, or because they do not fully meet the definition of business proprietor outlined above.

2.1 Target groups

The business proprietors currently active at any point in time and responsible for the decisions and risks of running private business enterprises, however small, generally fall into five categories:

1. Sole proprietor with no employees
2. Sole proprietor with employees
3. Partnership with no employees
4. Partnership with employees
5. Company directors

The first category, of sole proprietors with no employees, is generally referred to in the nineteenth century as those *working on own account*.

These categories coincide with the same categories used in most modern formal definitions: for modern UK population censuses and similar surveys (such as the LFS, GHS and FES) the first four categories are identical (although technical detail may vary), although company directors are more imperfectly included in censuses and other surveys until very recent years. As another comparator, categories 2 and 4 are equivalent to modern tax regulations for employers with responsibility to deduct tax at source (PAYE), whilst categories 1, 3 and 5, and the proprietors in categories 2 and 4, fall under self-assessment for income tax, or (if they are venture capitalists) are remunerated through income from dividends or capital gains and taxed on those sources through self-assessment or corporate taxation. Of course these categories may not be entirely distinct, and a given person may be simultaneously in two or more categories, e.g. if they run differently structured portfolio businesses, or own one business and act as a director in others, or run a business and also operate as a worker in other enterprises.

The main source used in the project 'Drivers of Entrepreneurship and Small Businesses' is the population census. This allows most of the business-owning population to be identified more fully than previous analyses. However, the census was not designed as a business census so that its coverage is imperfect. To fulfil the aim of being as complete a business coverage as possible, this project also uses other sources to supplement the census to ensure that gaps are filled.

Table 1 shows how the census for two main periods within 1851-1911 covers the five business categories listed above. The table also shows how far those individuals with portfolio of involvement in more than one business and those who were company directors were included. It also shows how or if employees and workers were differentiated from business proprietors and entrepreneurs, and compares the census to three alternative sources. The table shows that the census has the advantage of clearly differentiating business proprietors and workers. Although own account self-employed are not differentiated from workers for early years, this deficiency can be largely overcome. The census however, does not properly identify company directors. Various directories allow the census to be

supplemented, and also offer comparison sources against which checks of the census can be made. The most useful for filling the major gap for company directors are directories of directors. These provide the main potential for enrichment of census records to expand to a more complete coverage. Professional and sector directories such as the *Law List*, *Medical Register*, *Directory of Chemical Manufacturers*, etc., provide important further potential for enrichment where specific sectors have to be supplemented. Local directories are also a valuable supplement for case study locations, but until they are available as an electronic and coded database for the whole country they are too costly to use on a large scale.

Type of business		**Censuses 1851-81**	**Censuses 1891-1911**	**Local directories**	**Prof/sector directories**	**Co. & director directories**
1	Sole proprietor no employees	Some masters identifiable; otherwise incl. with employees	√	√ but may incl. partners not listed	Some √ some not	n/a
2	Sole proprietor with employees	√	√			n/a
3	Partners no employees	Some	Some incl. with sole proprietors	√ but bus. names may incl. sole proprietors	Some √ some not	n/a
4	Partners with employees	Some				n/a
5	Company directors	Very few	Very few	√ Co. but not usually their directors	√ Co. but not usually directors	√
6	Portfolio businesses	√	√	Some	Some	√ for most other Cos.
7	Workers/employees	√	√	'Private addresses' but some occupations listed which may confuse	x	x

Table 1. Relation of census and directory sources to business categories; √ = available from that source; n/a not applicable.

As clear from Table 1, as well as some exclusions, the census also changes its format between 1851-81 and 1891-1911. The details of the specific differences and their implications are discussed in Working Papers 2, 3 and 4. Apart from minor changes in question wording and design, the critically important issue that arises is two changes in

methodology in 1891: for recording employers; and for individuals who were self-employed own account. For the whole period most employers were explicitly designated in the CEBs, and for 1851-81 their number of employees was also recorded (and for farmers the acres farmed). Employers can therefore be extracted from the census as a consistent time series (but specific representativeness questions still have to be dealt with as detailed in later studies). However, for 1891-1911 there was no attempt made to gather information on employee numbers or acres farmed. Hence, although the category of employers can be tracked continuously over the whole period, there is a major discontinuity that generally restricts the analysis of firm size to the early period. A second discontinuity occurs in 1891 for the self-employed own account. Prior to 1891 own account self-employed were not explicitly identified, but from 1891 they were. This is a less serious deficiency than for employers because it is possible in many cases to reconstruct own account responses for the earlier period from the occupational descriptors that individuals used in the census. The results of this reconstruction are imperfect, and the accuracy varies between business sectors, but for many individuals and many sectors a reasonably accurate assessment can be made.

<table>
<tr><td rowspan="6">Individual entrepreneurs</td><td>1851-81</td><td>1891</td><td>1901</td><td>1911</td></tr>
<tr><td colspan="4">Employers identified in all years</td></tr>
<tr><td>Firm size: Employee nos. given</td><td colspan="3">Firm size: Employees only in case studies and panel samples</td></tr>
<tr><td>Own account reconstructed</td><td colspan="3">Own account identified explicitly</td></tr>
<tr><td colspan="4">Partnership and partners identified for some businesses</td></tr>
<tr><td colspan="4">Directors identified in directories are matched to census records</td></tr>
<tr><td colspan="4">Most portfolio businesses identified</td></tr>
<tr><td>Non-entrepreneurs</td><td>Workers and own account have to be separated</td><td colspan="3">Workers explicitly identified</td></tr>
</table>

Table 2. Database structure constructed for each census year.

Table 2 summarises the resulting structure of the database that can be assembled for each census year to identify individual business proprietors and entrepreneurs. It also shows how non-entrepreneurs, which we loosely refer to as ‘workers’, can be separated for comparative purposes. The irrecoverable data is that on employee numbers from 1891 onwards (except

through data enrichment from other sources). For the other records a nearly complete or reasonably complete record can be constructed that can support most types of subsequent analysis. The level of coverage and how deficiencies can be managed is discussed in other working papers. For other users of the database it is important to be aware of the variation in coverage of the different entrepreneur categories in each year and to use appropriate strategies that are provided in the database documentation *User Guide*.

Note that partnership businesses have to be handled carefully. They are a large category, but the census record of them is imperfect, especially outside of the broad census category of 'trade and manufactures'. As business proprietors all partners are distinct and can be treated as separate individuals in analysis, with additional codes to identify their partner status. But as businesses they constitute multiple individuals within the same business. To count all such individuals and believe that this is equal to the number of business clearly results in over-estimation. For analysis of businesses, therefore, it is often appropriate to pass over the duplicate partners or directors and create a reference person for the whole business. This would normally be the senior partner, if identified in the census or other record. These are coded in the database as the assumed senior partner or 'lead. From the creation of this reference person, each business is uniquely identified and the database can thus be interrogated at the level of distinct businesses as opposed to individual entrepreneurs. Alternatively all partners or directors can be scrutinised as individuals if this is more appropriate for the analysis. More detailed discussion is given in a pilot analysis for 1881,[3] from which the definitions of partnership coding used in the database are developed. For partners co-resident within the same household, additional effort is made to code these within family business structures.

Workers are not the main subject of this project or database. But they are an important comparator category of those who were *not* business proprietors or entrepreneurs at the time of the census records who can be investigated to assess how far their personal or other characteristics differ from those who were entrepreneurs at the time. Certainly it is important to be able to distinguish clearly those who were workers from those who were own account to ensure that entrepreneurs at a given time are fully identified. Hence the project database contains codes to classify workers as well as self-employed at various levels of certainty as to

[3] Bennett, Robert J. (2016) Interpreting business partnerships in late Victorian Britain, *Economic History Review*, 69, 4, 1199–1227.

their status. Definitions are of course complex, and individuals often moved category across their lifecycle between employee and self-employed status, particularly those at the lowest income levels where so-called 'survival entrepreneurs' or 'necessity entrepreneurs' often differed little in economic terms from workers, or indeed in some cases form the unoccupied or partially occupied.

Portfolio businesses where an individual runs two or more identifiably separate businesses at the same time are another special case, for which the census gives a very good record. About 10% of identified business people have portfolios. In addition, an incomplete sub-sample of entrepreneurs with by-employments can be identified, these are those who had a business and are also identifiable as an employee (e.g. someone who owned a grocery business and was also a post master employed by the post office).

2.2 Decisions on inclusion and exclusion

The broad definition of entrepreneurship adopted aims to include as wide a range of individuals and business proprietors and entrepreneurs as possible. However, given the constraints of historical records, some categories of entrepreneurs that have been the subject of important modern literatures can be only partially included, or are unavoidably excluded. This section discusses the decisions made on the main categories where different definitions have been used in other literature.

(i) ***Nascent entrepreneurs***: Those at the formative stage where they are bringing together resources but have not yet started a recognisable business are an important subject of modern entrepreneurship research. Unfortunately, these people usually cannot be detected from historical data, certainly not *en masse*, since until they have done something definitive about starting their business they usually do not leave a record. Those that do begin a businesses can be investigated for their former occupations and characteristics and are included in the panel tracking in this project; but those who were nascent but never implemented entrepreneurial actions generally cannot be detected in the census data.

(ii) ***Financial entrepreneurs:*** Those whose contribution was to bring together financial assets that others used to develop businesses (such as brokers, bankers, investment agents) are often

excluded from explicit recognition in the census. These may be recognisable from other historical records, and they are included in the analysis where they are recognisable as company directors who were legally the risk-bearers in that business, but in general this form of entrepreneurship is 'hidden' within activities covered by an individual's general occupational descriptor. Also within this group are many that were not the prime decision maker in terms of business strategy. For example, individuals such as asset managers, insurance fund managers, etc. that we would not wish to include as entrepreneurs because they were not the final decision takers within the businesses covered by the investments. However, substantial numbers of these asset managers were developers, co-preneurs, and venture capitalists who were joint investors and risk takers in specific projects. These are often recorded in historical records and were included in the database where there are recognisable from titles such as 'contractor', 'developer', 'broker' or through their status as company directors. They were most common in the nineteenth century in construction and infrastructure development and foreign-based enterprises.

(iii) ***Land and asset owners***: Asset owners who deploy their assets to develop businesses are an important class of proprietor that are sometimes difficult to identify in the census; e.g. mine/quarry owners, shipowners, barge owner, machine owners, and others with physical or land assets as their main business basis. These should be included with all other business proprietors but often the census provides little employee information, and many of these owners were resident (and hence enumerated by the census) in locations far from their businesses. This means it is important to attempt to match owner and business together as far as possible, given that the number or extent of the assets owned is rarely stated in the census. A special challenge is for those only referred to as landowners and house proprietors, for which the census again usually gives no employee information. These are ambiguous cases. Many land owners, even very large ones, are not proprietors of businesses but occupy the space for personal use. While they were likely to employ individuals to maintain the household (domestic servants), such employment relations are usually of a different character to the employer-worker relationship in private businesses, though their role as employers of estate workers qualifies them as business proprietors where these worker unambiguously suggest they are contributing to the estate as a 'business'. House proprietors may or may not be actually letting their property. The census classification and terminology is unhelpful in both cases because of a lack of sufficient recorded information. For database purposes these

categories are coded as separate categories that can be analysed along with other entrepreneurs, or excluded, depending on the purpose of the study.

(iv) ***By-employment with minor entrepreneurship:*** Business proprietors who have entrepreneurial activity that is a small part of a range of activity with their 'main' occupation are difficult to identify as they are usually recorded as something else and are therefore not recognisable. They are instead normally identified e.g. as waged or salaried employees (even though they may have some self-employed trade 'on the side'); retired, unemployed or unoccupied where their entrepreneurship is only a minor or part-time activity; and those whose income is mainly derived from investment or other assets which may be in part used to support a level of entrepreneurship, but where their main activity is managing their investments and living on their assets (such as passive investors, rentiers, and those who were not direct venture capitalists). This distinction becomes important in the historical record for some categories that were classed in the census as 'living on own means'. Some of this group have been identified via data enrichment through their directorships, even if their given occupation in the census did not hint at entrepreneurial activities. An important group is also sometimes identified as own account, and sometimes even as employers, but where this was not their main activity. One of the most numerous of these categories is vicars and other church representatives of any denomination. Whilst often earning a small supplementary income from fees for ceremonies and other functions, or raising support from congregations, their 'franchise' was almost always though appointment to a post by church authorities, from which they could be dismissed for poor service; hence they are treated here as employees.

(v) ***Managerial entrepreneurs***: These are of employee status and are excluded. The contribution of managers to a business can be substantial, and they may often be the dominant strategic decision makers, but as they were not the ultimate proprietors they were not generally included as entrepreneurs in this project and database; the only categories where this is nuanced is for farm bailiffs,[4] and some special categories of those running activities where they appear to be the main risk-takers (see below).

[4] We include farm bailiffs where they give other farm information because they are frequently the sole census respondent who lists the farm's acreage and workforce. In this way they act as 'surrogate' farm proprietors, but they managerial status is coded and their personal characteristics are analysed separately from other farmers where required.

(vi) ***Outworkers and agents***: Outworkers were an important part of the population throughout the 19th century. They took risks and were prime decision-takers in their activities to the extent of being theoretically able to take in work from a range of clients. However, they were usually entirely dependent on the extent to which work was put out to them by entrepreneurs and intermediary managers and had little or no ability to set contract rates or manage volumes. They are generally excluded from the entrepreneurs database. Indeed both the census, and the individuals themselves in their census responses, generally responded as workers rather than as own account self-employed. This covers a large range of activities, of which the most numerous were often women in industries such as glove, hosiery, clothing, basket and hat making. For other categories of outworker there is sometimes more difficulty separating them from contractors. This also arises in modern situations, e.g. in the so-called 'gig economy'. Generally 'contractors' and 'consultants' are included in the database as entrepreneurs where they appear to be distinguished from outworkers by the level of independence they have, and the range of clients they work for, as far as it is indicated in the occupation descriptor. Another difficult group to classify is agents of companies such as life assurance agents, bank agents and managers, and agents of specialised businesses such as fertiliser sales, seeds sales, building contractors, coal suppliers, etc. These were sometimes fully employed as agents in that occupation; although many were agents as well as having another 'main' occupation (for example many insurance company agents were solicitors). In all cases where identifiable this type of agent was coded to their main occupation, or if only an agent they were excluded as essentially an employee and under the control of the business or company for which they were agent. This approach also excludes travelling salesmen. The aim is to identify those contractors that offered a varied, high-value added and tailored service, rather than being held to account by a single client for a contract at a set price by the piece to closely scrutinised and managed specifications. It is acknowledged that in this field definitions and delimitations are difficult for the past, as indeed they remain today.

(vii) ***Employers of domestic servants:*** One of the key identifiers used in database construction is the title and concept of being an 'employer'. A very large number of employers in the nineteenth century were employers of domestic staff, but not in any other way identifiable as employers: they were themselves labourers or employees of other businesses. We take this type of employer as very different from entrepreneurship as defined here, because there was no marketed product being sold as a result of the employment of the servant. Of course, servants freed resources of heads and other family members to do other

things. But employing servants was essentially a choice affecting redistribution of income and resources within the household (including the cases of servants living out) and not risk-taking or a direct productive activity in itself. Similarly, in the many cases where employers or own account individuals also employed domestic servants, the servants are not seen as adding to the overall number of the business's employees, nor changing the definition of an own account proprietor to an employer, because in each case the servants were not direct contributors to the business itself. In all these cases, therefore, the employers of domestic staff were not included in those identified as business proprietors or entrepreneurs, and their employees not counted within business employees. There are of course difficult cases, where domestic staff, including family members, may be both domestic servants and also assisting in the business (as with many farms and shops), or where a business employee was mis-labelled as a servant. However, the census process was supposed to differentiate individuals through occupational classification based on 'main' occupation, so census terminology is used as the discriminator in all cases, though this may be inaccurate at the margins. But an important part of analysis is examination of the subset of servant employees and/or family members co-resident with own account individuals and employers. These are examined in detail through separate analyses of real (recorded) and potential entrepreneurship relationships within households and family businesses. This seeks to identify those who were 'assistants', 'shopmen' or had some similar relationship making them business employees rather than purely as domestic staff. Apprentices and journeypersons are not considered domestic servants and are therefore identified in relevant parts of the database and analysis under those explicit titles. Comparison is also made of total business counts based on including or excluding domestic servants; this seeks to facilitate comparison with some modern business data sets that include employers of domestic staff as 'employers' (see 'comparators' section below).

(viii) ***Other types of entrepreneurship:*** The primary focus of this analysis is on private sector enterprises. Hence, although making valuable contributions to their organisations, other forms of entrepreneurship which often form an important part of modern entrepreneurship research are generally excluded: such as social entrepreneurs involved with welfare and/or not-for-profit organisations, administrative entrepreneurship, and public sector entrepreneurship (but distinctions with the public sector are discussed further below).

(ix) ***Foreign-based business:*** The concern of most of this project is with domestic industries and entrepreneurs. However, foreign-based businesses were also an important part of UK entrepreneurship in the 19th century and it is important to identify those involved so that they can be either explicitly included or excluded depending on the focus of the analysis. Foreign-based entrepreneurs are only rarely explicitly identified in the census; and if identified there is usually no information on where their business operated. Generally it has to be assumed, and it is safe to assume, that entrepreneurs identified in the census were operating locally and within the British domestic economy. Since in the 19th century over 99% of businesses and entrepreneurs were proprietors of very small concerns (as also true today), their activities were typically very circumscribed geographically. Where someone calls themselves a 'baker employing one lad' it safe to assume that this was local to where the census recorded them; and it is highly unlikely that they were a baker operating abroad. Similarly, in the census if a foreign-based individual was recorded who was only a short-term visitor to the UK they were supposed to be recorded as a 'visitor' or might be a hotel 'boarder'. They can thus be identified and included or excluded in subsequent analysis as relevant. However, in a few industries there were small numbers of employers and self-employed in the census whose operations were abroad and this was not fully recorded. One large category is foreign-based ships that happened to be in a UK port on census night. These have to be specifically distinguished from other ships and appropriately coded. A second category was entrepreneurs who were absentee owners of businesses who ran foreign-based operations from the UK. This group includes the directors of many large Stock Exchange-listed companies and was important in foreign mining, railway, shipping, plantation agriculture, and land development industries. Absentee foreign owners were sometimes explicitly identified in the census, but for the most part are identified and included in the database through enrichment of identified company directors.

2.2 Distinctions in practice

Some of the distinctions necessary to define the current responsible individuals and risk-bearers in private business enterprises are difficult to draw, especially from historical data, but the definitions chosen should give a fairly clear differentiation between those concerned with business entrepreneurship, those who were entrepreneurs in other fields (i.e. they were

not private sector businesses), or those not involved as the major risk-takers or decision takers such as workers, managers and passive investors.

One area of potential difficulty is overlap between definitions of entrepreneurship in the private and public sectors. The definition used here focuses on private business enterprises. Hence activities in the public sector that are purely service provision based on general revenues and taxation, with no business or trading activity, are excluded. This excludes activities such as public sector state schools, state hospitals, state welfare organisations, and public roads. However, the distinctions at various points in the historical past are not always clear-cut, and definitions of the same activity, sometimes undertaken by the same people, may change over time. Thus schools run by private tutors, as private academies or commercial schools, should be identified as private sector trading activities, and were indeed an important field of activity especially for early female entrepreneurship. Similarly, many individuals in the past, especially professionals, musicians and artists worked for public as well as private organisations just as they do today (such as doctors in private as well as state hospitals, or musicians in private or municipally and state-supported orchestras). In many cases these individuals were self-employed whatever the organisation they worked for, since they were peripatetic between several different 'employers'; they acted as contactors to many independent organisations across different sectors. This also applies to many engineers and consultants. All such cases must be included in an entrepreneur database. Others had both a salary and also operated privately (as for many doctors). This is a dimension that overlaps with outworkers. Where these distinctions can be identified, individuals are assigned to their 'main' occupation (which is the same definition as used in the census), with other activity coded as by-employment. In general the defining characteristic was, as far as possible from the records available, to include business proprietors and self-employed that were independent contractors within the database, even if this included contractors for public sector contracts, provided that they were independent professionals who were not of employee status.

2.3 Municipal and public 'enterprises'

Another important definitional issue that must be considered is public sector activities that were truly risk-taking through trading, and supported by charges or fees for services that aimed to make them self-financing or make a surplus. These can be treated as businesses in

many cases; as with municipal gas works, water works, tramways, covered markets, road tolls, harbours and docks, and a range of other municipal enterprises. However, they also differ from purely private enterprise because they usually had a wider foundation of finance to fall back on, provided by local taxation (the Rates on property) or other public revenues and assets; they could often borrow more readily and at lower interest rates because of the security they could offer from their guaranteed tax and other revenue streams; and they usually operated as de facto monopolies in their territory. For the purposes of our analysis these public enterprises are included where possible for the cases where they were entities that were separate from the general activities of state or municipal provider's other service activities financed by general taxation and other revenues. Thus, for example, the analysis of corporations includes municipal gas companies or water companies where possible. But this different type of business is, as far as possible, treated as a separate additional sector. Those who were mainly responsible for running these municipal businesses were, however, less clearly entrepreneurs. As legal entities municipal enterprises had boards of directors or fell under the jurisdiction of pubic boards and councils. Given that the businesses were municipal, these individuals were not personal risk-takers in the same sense that business proprietors or directors were in the private sector; indeed they often occupied positions for only a few years, ex officio from another office, and may have had no expertise in the sector at all but were merely appointed to provide oversight. Although some may have been truly 'entrepreneurial', often municipal enterprises were run by their managers, with only the broadest of overall strategy falling on the municipal governing entity and its directors, if they chose to exercise it; but nevertheless the risks did not fall directly upon the managers and directors. For this reason, in those cases where an individual can be identified as the manager of these entities, they are treated in the same way as other managers and excluded for the entrepreneurs database. But directors, where identifiable, are included but coded separately so that subsequent analysis can include or elude as desired. The two different elements of the database (entrepreneurs and businesses) must be borne in mind throughout the use of the data, and the interpretations also kept distinct.

2.4 Summary

The entrepreneurs database provides a large scale coverage of businesses and their proprietors. The census as a primary source allows the main categories of own account self-

employed and employers to be identified, but corporate information on directors and their business has to be added mainly from elsewhere. There are important discontinuities in data definitions used in the census over time before and after 1891, but most can be overcome. Important decisions have to be made in the data extraction from the census on how some categories are classified to identify portfolio businesses, financial entrepreneurs, contractors and developers, or those trading with assets; and to distinguish domestic from foreign-based enterprise, public from private, and to exclude managers, and those outworkers with insufficient autonomy to be classified as self-employed.

3. The Census database.

A primary output from this research is an open and generally accessible *entrepreneurs database* of aligned, quality-controlled, and consistent data on businesses extracted from the censuses records 1851-1911. The electronic versions of the census used derive from several sources, of which the most important in terms of number of individuals included is I-CeM. This is discussed first before engaging with how it was supplemented.

3.1 The Integrated Census Microdata (I-CeM).

I-CeM is a database constructed by a research team at the University of Essex using ESRC funding and other support. The data they used derives from the original CEBs for 1851-1901, and from the original household returns for 1911. The original records are held at The National Archives (TNA). They were originally transcribed by FindMyPast (FMP) (part of BrightSolid) using a variety of transcriptions conducted by FMP itself, by Genealogical Society of Utah for 1881, Federation of Family History Societies, local Family History Societies, and other sources. The Essex team turned the FMP genealogical database into a generic database structure, adding coding for occupations, household structure, locations and other features. The data were deposited at the UKDA in 2014.[5] A *User Guide* summarises the structure and coding of the database.[6] An update of the *Guide* was made in 2015 warning of

[5] Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481.
[6] https://www.essex.ac.uk/history/research/icem/documents/icem_guide.pdf

some of the deficiencies in using the data. Some improvements to the original data deposited were made in 2017.[7] Further improvements to the I-CeM data have been made in-house at the University of Cambridge, Cambridge Group for the History of Population and Social Structure (Campop) as part of this ESRC project and by the research team on another ESRC project concerned with analysis of fertility using the I-CeM data. Important guidance and assistance for these improvements has come from the members of the original I-CeM teams, notably from Kevin Schürer, the original PI for the I-CeM project. The in-house Campop version has gone through several enhancements.[8]

For the Drivers of Entrepreneurship and Small Businesses project the enhanced versions of I-CeM developed at Essex and Campop have been used as a main source. However, for the 1881 census year, the original electronic version constructed in another ESRC project by Schürer and others has been employed as an alternative to the modified version contained in I-CeM.[9] This was transcribed by Genealogical Society of Utah and Federation of Family History Societies and is referred to here as the Essex-GSU database.

These electronic versions of the census allow the occupational descriptor that was filled in by enumerators (or for 1911 filled in by householders) to be accessed. However, the occupational descriptors as transcribed by FMP or GSU and entered into I-CeM are just as they were entered by census enumerators or householders (sometimes with additional clerical marks transcribed within strings in a disjointed and misleading manner). The transcriptions are mixed alphanumeric strings such as 'Alderman, JP, cotton manufacturer employing 200 men, 100 women and 3 boys, captain of volunteers'. The strings include the employer and employee information required, and the occupation(s) or industry(ies) of the employer, but in a complex format, mixed with other descriptors and 'occupations' and 'status' categories that are not necessarily relevant, and can be presented in any order, plus any clerical marks also in any order. To be useable for analysis relevant records containing such information have to be first accurately detected, and then precisely split into different fields for the industry information on the business of the employer, its employees by gender and age (men, women, boys, girls), and any other information either separately coded or cleaned and removed. Employers have to be definitively recognised through the presence of keywords representing

[7] https://www.essex.ac.uk/history/research/icem/documents/icem-guide-version-2-2015.pdf
[8] Schürer, K., Higgs, E., Reid, A.M., Garrett, E.M. (2016) *Integrated Census Microdata V.2 (I-CeM.2).*
[9] Schürer, Kevin and Woollard, Matthew (University of Essex) (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)* [computer file] UKDA, SN-4177.

employer and employee types (allowing for spelling and gender variants, abbreviations or synonyms), and different ordering of employee types, numbers and the terminology to describe workers (such as 'hands', 'labourers', 'persons', 'assistants', 'apprentices', 'servants', 'clerks', etc.). For farmers the additional census information available on the number of acres occupied also has to be recognised, separated from employee information and parsed into the correct field. The scale of the analysis, which requires the entire population of 17-40 million records for each census year to be scrutinised, means that automated data extraction is essential. A pilot of the methodology to do this was developed by Bennett and Newton for 1881.[10] Working papers 3 and 4 describe the process by which algorithms were further developed, refined and deployed for the two census formats 1851-81 and 1891-1911, and how the results were checked and supplemented by hand checks against the images of the original CEBs.

The results from the extraction method allow, as far as possible, identification of all economically active employers and their businesses, their employees, and in the case of farmers the acres farmed. This should satisfy a central objective of the Drivers of Entrepreneurship project to identify as complete as possible a database of employers. However, during the initial phases of data extraction from I-CeM it was discovered that there were significant truncations and omissions of the employer occupational strings that had to be filled, some entirely missing data, and numerous miscoding by sector and business status that had to be corrected in order to meet the project's objective of constructing a consistent database that is as complete as possible for all businesses.

3.2 Supplementation of I-CeM to fill gaps and overcome deficiencies.

Preliminary investigations of the I-CeM data evidenced that there were four main shortcomings that had to be overcome in order to satisfy the target of an aligned, complete, and consistent database for what the original census itself recorded:

1. For 1851 there are a large number of missing records, and/or a complex and variable truncation of some lines of the occupational descriptors in I-CeM. This appears to have

[10] Bennett, R. J. and Newton G. (2015) Employers and the 1881 population census of England and Wales, *Local Population Studies*, 94, 29-49.

arisen because FMP for the 1851 census either did not fully transcribe, or used some transcripts from family history societies that were incomplete. In some cases groups of individuals are missing altogether. In other cases the occupation given in the CEB was not fully transcribed, with the information on employment status excluded; for example, 'Cotton manufacturer employing 10 men' was transcribed as just 'Cotton manufacturer'. This may be understandable for genealogy purposes, but is fatal for allowing identification of employers or separating employers from self-employed or workers with the same descriptors, and omits all the business size information required. Analysis of potential gaps demonstrated that this was non-random, and geographically specific, reflecting the batch inputting by FMP and family history societies. It was calculated that truncation affected most or all of 14 counties (including the populous and economically crucial counties of Lancashire, Cheshire and London). In addition truncation affected all or part of a more dispersed set of 210 Registration Sub districts (RSDs) in 87 Registration District (RD) areas. The total missing records of possible missing employers numbered about 75,000, of which it was fund 58,000 were missing. Their identification and infill is discussed more fully in WP 3.

2. For 1861 a different truncation problem arises because, although in this case all the original occupation descriptors appear to have been transcribed, entries over 49 or 50 characters have been lost at some stage in the database management by FMP. Checks with FMP confirmed that the rest of the field is no longer available. Because of the complexity and hence length of many employer descriptors, over 90% of the truncated entries are targets for extraction as employers for the required entrepreneurs database. The records affected are identifiable from the termination of lines at 49-50 characters. Of these about 35,000 were identified as probable employers. Their identification and infill is discussed more fully in WP 3.

3. For each census year there are census records that are now no longer extant, having been lost or destroyed in the preservation at TNA and its predecessor bodies. These records are irrecoverable. The locations of these lost records are not documented in I-CeM. They are partially listed by FMP. For most years the lost records are very small in number and do not matter for most analyses. However, for 1861 the lost information is substantial, and locationally specific to important areas of the country: about 3-5% of the households are missing, mainly in London and the Home Counties. For analytical purposes it is

critically important to know as far as possible the locations and any other available information on these lost records. Because of the importance of this issue, a substantial effort has been made to identify the records affected and analyse their geographical distribution. This is discussed more fully in WP 3 and other publications.

4. In addition there were numerous data assembly shortcomings from the FMP data, and coding errors of occupational strings. There are five main categories of error:

(i) For 1851 I-CeM version 1 contains about 250,000 duplicated records where FMP entered the same people two or more times in different versions of keying. These records have been deleted in the new versions. The same issue affects some other years to a lesser extent.

(ii) A very significant problem for accurate identification of employers and other entrepreneurs is that a non-trivial proportion of individuals are ascribed to the wrong occupational code and hence to the wrong industry. This arises across all the census years for all categories of worker, employer and own account.

(iii) An additional specific occupational coding error affects employers because in a significant number of cases an employee occupational category is wrongly assigned to the employer and vice versa. For example, a cotton manufacturer employing carriers may be assigned to the carrier category; conversely the domestic servant of a baker may be coded as a baker. Hundreds of thousands of records have had to be corrected for miscoding errors (i) – (iii) for each of the census years.

(iv) Additionally, for 1891, 1901 and 1911, where the census has specific identifiers for employer, own account and worker, there can be significant numbers of these miscoded between categories for reasons deriving from both the original census enumeration process and the encoding method used in I-CeM for this information.

(v) Some of the additional enrichment coding added by Essex to I-CeM had a significant degree or error and misplacement of records in I-CeM version 1. This affected chiefly the household classification, conjugal family units, place of birth, and geographical location attributed to the household. I-CeM version 2 corrects most of these defects and has been used as the starting point for analysis.[11] However, users need to be aware that some errors persist and care needs to be exercised in all interpretations. Also

[11] Schürer, K., Higgs, E., Reid, A.M., Garrett, E.M. (2016) *Integrated Census Microdata V.2 (I-CeM.2)*.

the geographical attribution of households in the new versions of I-CeM is accurate mainly at the RSD level and contains some significant mis-allocations for a few parishes within some RSDs.

All these errors, as far as possible, have been corrected in the new versions of I-CeM or in the entrepreneurs database deposit allowing research users to be less concerned about miscodes. However, resources available in most cases allowed checks for this sort of problem to be run only down to a specific level where the percentage error was very small; e.g. for occupations down to the level of distinct strings within each prospective employer occupational category that accounted for at least 25 individuals. The remainder generally constitutes well below 1% of any occupational category so will not be of concern in large scale statistical analysis. But the user needs to be aware that lurking within the data there will be some remaining mis-codes of occupation. A high level of care should be exercised when seeking to undertake analysis with uncommon or small occupational categories, in small localities, or using cross tabulations that result in very small numbers of individuals in any data cell.

It should be noted that for users of I-CeM who do not require full data on employers and business proprietors, the truncations and missing material in 1851 and 1861 may not be so consequential. The existing occupational code, although subject to error, may be sufficient. However, it is advisable for researchers to check on completeness of occupational coverage in case study areas, for specific groups, or for specific occupations and sectors that might be among those that are truncated, missing or miscoded. This includes many common occupations. It should also include checks on all those individuals that have long or complex occupational information, among which are some of the more prominent people of the time. These individuals may hold a range of offices and 'ranks' that were usually listed in the census records before the occupation was noted, creating special challenges to the principal that first-listed occupation was most significant. For 1851 a map of the main areas where truncated records occur is given in WP 3, which provides a guide to what might be expected for that year.

Because of the truncations and omissions in I-CeM, a substantial effort was mounted to assess the availability of alternative sources to infill the missing records. This involved checking all family history society publications at the Library of the Society of Genealogists, which covers a large sample of transcripts from the census for the whole UK. In addition all

family history societies for the 14 counties that were identified as fully or partly missing in 1851 were contacted directly. Although a few full transcripts are available from these sources, it was possible to find only one RD and a few parishes that covered the required missing areas. This reflects the gaps in the extant documents that FMP itself used for data entry. For 1861 there were no sources found that allowed the truncations to be readily infilled. As a result two strategies were followed to supplement I-CeM for 1851 and 1861, as discussed more fully in WP 3.

First, for 1851 another genealogical supplier of fully transcribed information was used (S&N Genealogy). Second, for 1861 the truncated portions of the lines in I-CeM are irrecoverable. The only way to complete them was to refer to the original CEB images. This was undertaken for all the approximately 35,000 truncated lines, achieving a final full and complete record in each case.

For 1871 an additional difficulty arises because there are almost no usable employer records that were transcribed or exist in FMP. A Version 3 of I-CeM is aimed at providing cover of this census year, but it excludes all but a few occupational descriptors, and all marital status, and birthplaces, because these data were not transcribed by FMP or if transcribed were subsequently lost. The lack of occupational descriptors, which also contain the information on employers and employees, means that it is of no utility to identify business proprietors. As a result the 1871 data were sourced from an alternative supplier, S&N. For this year the extraction algorithm described below was applied directly to the data by S&N and then subsequently parsed and cleaned. However, there are various challenges resulting from differences in the structure of the S&N and I-CeM databases and transcriptions. This required an additional significant research input for which additional support was provided by the Isaac Newton Trust. The full documentation of the 1871 extraction is discussed in another working paper. For the extraction of employers, however, it is believed that the final results from the methods of extraction, cleaning and parsing achieved similar levels of coverage to that for the other census years.

3.3 Final census data base.

As a result of the gaps and truncations in I-CeM and their infill from various sources as discussed above, the database referred to in this and other project Working Papers is an

amalgamation of sources. For those sources solely derived from the census we refer to the supplemented database as *I-CeM/S&N*. In addition, other data are added to the entrepreneurs database through various enrichment efforts. The collective result is an entrepreneurs database deposited at the UKDA and made freely available to other researchers that should be as complete and consistent as possible within the constraints of the data that are available. The main sources can be summarised as follows:

Census data 1851-61; 1891-1911: Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481, Version 2 of I-CeM includes a range of valuable additional re-coding and corrections by colleagues at Campop; see Schürer, K., Higgs, E., Reid, A.M., Garrett, E.M. (2016) *Integrated Census Microdata V.2 (I-CeM.2).*

Census data 1851: Supplementation of gaps in I-CeM for up to approximately 58,000 employer records extracted specifically for this ESRC project, derived from the manuscript census enumerators' books, transcribed and supplied by S&N Genealogy Supplies (TheGenealogist.co.uk), and integrated with assistance from Gavin Robinson.

Census data 1861: Supplementation of gaps in I-CeM for approximately 35,000 employer records by original transcriptions from CEBs specifically for this project, supported by ESRC, performed by Mark Latham, Gavin Robinson, and Tiffany Shumaker.

Census data 1871: Supply of all employer records by S&N Genealogy Supplies (TheGenealogist.co.uk), coded as part of the ESRC project with additional support from the Isaac Newton Trust.

Census data 1881: Schürer, Kevin and Woollard, Matthew (University of Essex) (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)* [computer file] UKDA, SN-4177, transcribed by Genealogical Society of Utah and Federation of Family History Societies.

Data enrichment: Census data is enriched (especially for companies) from a variety of sources, mostly by original transcription and data entry, as described in the relevant Working Papers.

A flowchart of the decisions involved in data construction, data infill and enrichment is shown in Figure 1. This illustrates how the original population of businesses is recorded in the census CEBs or has to be derived from other sources; the stages at which data from

additional sources is introduced to infill and enrich the census transcriptions; how the final database coding of categories of businesses has to be managed for the different census periods 1851-1881 and 1891-1911 because of the changed census instructions; and how this yields estimates of the final business population. The estimates at the three levels of individual entrepreneurs, businesses, and various geographical aggregates constitute the structure of the database deposit.

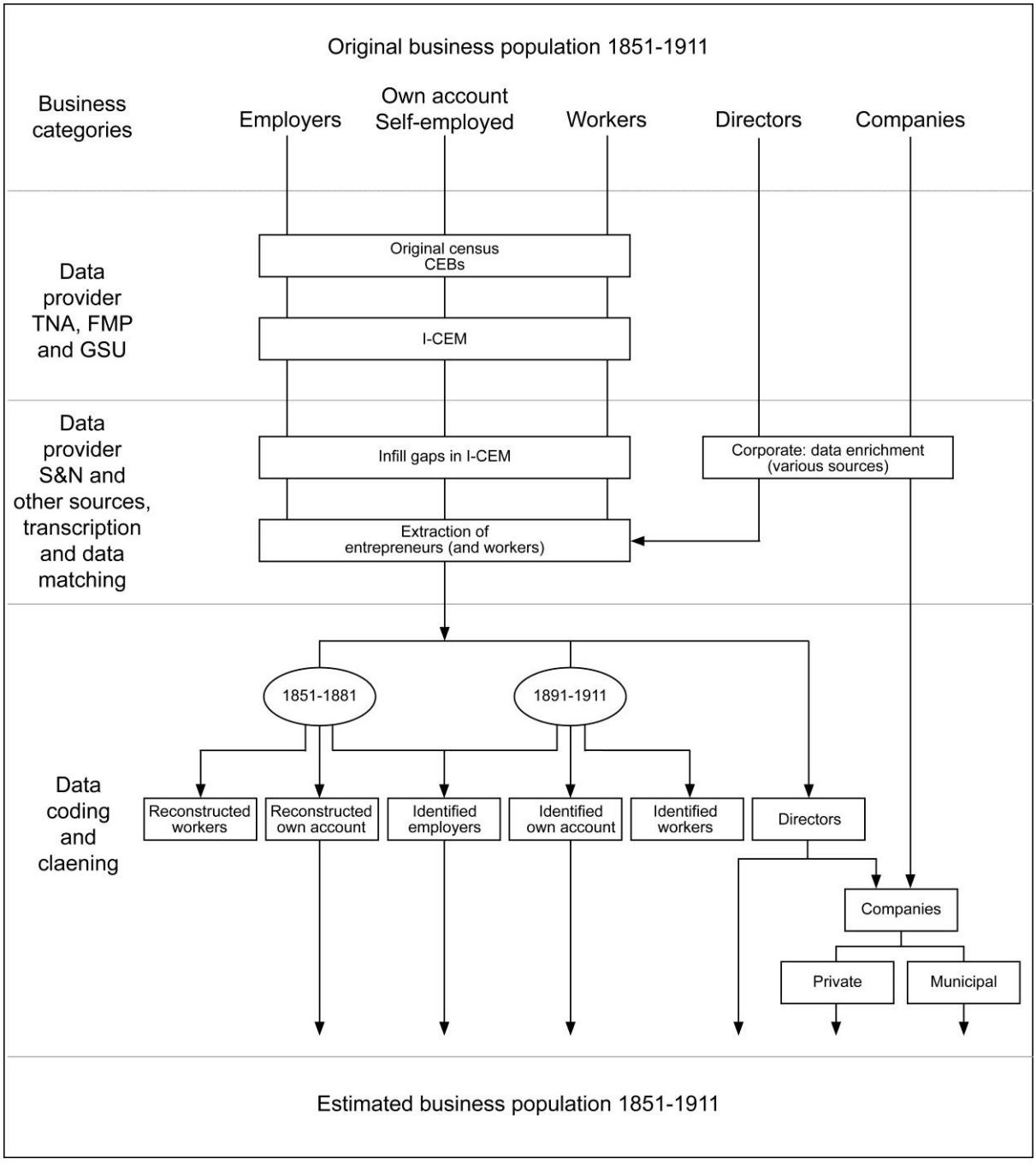

Figure 1. Flowchart of stages in database construction.

4. Comparators and calculation of entrepreneurship rates

To gain a reliable indication of the significance of employer status and self-employment over time it is essential to have a consistent comparative basis to provide entrepreneurship rates and other comparisons that are measured against the same base. Four main comparisons are used: (1) the total population; (2) the economically active; (3) the population over school-leaving age; and (4) the economically active (as (2 or 3)) but excluding domestic servants. Each must be differentiated by gender. The different measures allow comparisons against different bases that permit significant changes to be controlled for in the demographic makeup of the population, gender participation in different forms of employment status, and progressive increases in the school leaving age.

It is also important to refer calculations to the same geographical locations (which for the census are place of residence, unless a visitor), and the correct base dates.

4.1 Total population.

The total population is the simplest comparator that allows the proportion of small firms to be scaled, but gives only a gross entrepreneurship rate. Because the census was primarily focused on counting population numbers, the population counts should be the most accurate of all the data available, but are reported for place of residence (see below). Because total population is often the most widely available source and is broadly most comparable across different nations and cultures, it is often used in international comparisons (e.g. by World Bank, OECD, EU and others). However, it must be borne in mind that there are substantial differences over time and space in the age structure of populations.

4.2 Economically active.

The economically active is a measure of the total of all employed, self-employed and employers. Using it as a comparator allows tracking of the relative proportion of the labour force formed by the self-employed and employers, and also allows some indication of employees within different types of firms. It is the most commonly used base for international comparisons for entrepreneurship rate calculations.

To estimate activity rates that allow calculations of employers per sector, or entrepreneurship rates for businesses per head, we need measures of the economically active population by sector. Whilst the census was good at counting total population, its approach to defining occupations and those that were occupied varied over time, especially in the nineteenth century. Comparisons of the economically active were first attempted in a systematic way by Booth for the 1851-81 censuses.[12] Hakim summarises many of the limitations of making comparisons over time. For employers the constraints centre on (i) the recording of women's work in general, (ii) the variable recording of housewives and other domestic contributions, (iii) how the unoccupied were defined, which has an important interaction but does not entirely overlap with the recording of women's' activity, (iv) correctly identifying the retired, (v) what starting age of work should be recognised, and (vi) whether those employing domestic servants should be measured as businesses (as in many modern statistical sources), which was important when so many households were employers of domestic staff.[13]

There was considerable instability in the early censuses about how the economically active were measured. This was influenced by both the way in which the unemployed, students and retired were tabulated, and also by the treatment of the domestic categories of married women. As Booth noted: 'the domestic class in one Census includes the large part of the population, and in the next is reduced by more than half; 350,000 persons in England alone (consisting of the wives and other relatives of farmers, etc.) are taken from the agricultural class of one Census and placed in the unoccupied of another; the partially occupied wives are in no two successive Censuses classed alike – and generally there is a want of fixity of principle and method'.[14]

The censuses' measure of economically active generally includes those 'intending to get work'; i.e. unemployed and involuntarily inactive.[15] These were referred to as 'occupied' in earlier censuses. In 1851 the tables included the unemployed and retired and cannot be separately identified unless specifically noted in CEBs, which is often possible.[16] They

[12] Booth, Charles (1886) Occupations of the People of the United Kingdom, 1801-81, *Journal of the Royal Statistical Society*, 49, 2, 314-444.
[13] Hakim, Catherine (1980) Census Reports as Documentary Evidence: The Census Commentaries 1801-1951, *Sociological Review*, 28, 3, 551-80.
[14] Booth, Occupations of the People, 318.
[15] Census 1961, Occupations volume, p. ix.
[16] Census 1851, General Report, p. lxxviii-lxxxi

constituted about 2% across all occupations in 1871.[17] For farmers this form of reporting continued over 1851-71, with unemployed and inactive farmers only separately identified in 1881, although the inactive were estimated to be a very minor component, largely because of the family support that developed on most farms. However, published tables are given for 1891-1931 for the unoccupied, who can be removed to produce tables for the economically active that measure only the employed, self-employed and employees, and these can be made reasonably comparable with later years. For farmers the General Register Office (GRO) also undertook detailed estimates that were comparable over time, which can be used for base line purposes. For analysis here, as far as possible CEB extractions can control for the changes in the definitions of the economically active. This will not be entirely successful, but it allows a comparable measure of the economically occupied to be extracted using essentially the same methods as adopted by GRO. For most purpose the main element of this calculation is to use the working population aged between 15 and 65, but the upper age cut-off is not always used in comparisons because of the tendency for many people to continue to work into old age in the 19th century.

4.3 Population of working age.

The progressive increases in the school leaving age, or in early censuses the population accounted for in 'occupations' and not discounted as children, developed as follows, with the tabulations of employers (where present) and occupations in the census generally following the same age limits:

1851-1911	-	10 and over
1921	-	12 and over
1931	-	14 and over
1951-71	-	15 and over
1981-2011	-	16 and over

As far as possible it is desirable to standardise ages used to avoid differently sized population cohorts being used as comparators for different periods. For 1851-1911 comparisons, the use of 10 and over would be consistent with the censuses of the period; however, 15 and over is a better comparator for consistency with later censuses and as a comparator of those genuinely

[17] Census 1881, Occupations Report, p. 36.

employed full time. For employers/own account, 15 and over is also more realistic for calculating entrepreneurship rates. Indeed most 'employers' or 'own account' under 15 that are identifiable in the CEBs could not be of significant employer status except in a very restricted sense. Other Working Papers examine the choice of age for comparisons in more detail.

4.4 Inclusive or exclusive of domestic servants?

The census instructions to 'employers' explicitly *excluded* employment of domestic staff. This is welcome, because it is consistent with the definition of entrepreneurship followed in the database construction, as outlined above. Hence an employer for census purposes should include only those employing staff for business purposes, not for household duties. For example, the 1891 instructions stated that employers 'refer only to employment in trades and industries, and not to the employment of domestic servants'. In 1911 the instructions stated that employers were 'those employing persons other than domestic servants', and for employees that 'No entry needed for Domestic Servants in private employment'. However, there were some difficulties in practice, since domestic staff of a boarding house business were usually counted as employees. There are also inconsistencies in early censuses as to whether domestic staff who were also assistants employed on farms or in shops were included or not, especially for family members; this particularly affects women. The definition of farm servants was especially problematic and ambiguous for the early censuses, with some inconsistencies over time. These special cases have to be controlled for in the data coding and extraction. However, for most cases, exclusion of domestic staff led to exclusion of an employer where their only employment was of servants. This is similar to definitions used in modern censuses; e.g. for 1971 the instructions stated 'Domestic servants and family workers do not count as employees when determining whether or not a self-employed person has employees'.

But these census definitions are quite different from modern non-census statistics on businesses. For example the modern Small Firm Statistics include all private sector employers, whatever their sector; even if employing domestic staff was their sole employer activity, they count as a business. No attempt is made to exclude domestics because the purpose of modern business statistics is to estimate the whole of the UK private sector, the

only difference being that if a care worker, cleaner or nanny worked via an agency the employee would be counted as an agency employee rather than by the family (this is the same as for outworkers). The comparisons between the historical and modern definitions is discussed more fully in a subsequent Working Paper.

It is important to be aware of these different definitions in different comparison datasets. Thus for census comparisons over time employers solely of domestic staff should be excluded from employer categories. But for comparison against modern business statistics it may be a more valid calculation to include all employers of domestic staff as employers. Because of the massive reduction in the scale of employment of domestic staff between the 19th century and modern data, the inclusion in employers of those solely employing domestic staff has a large impact.

4.5 Census place of residence

Employer locations in the census are all based on the place of residence of the employer (or where they made their return, which could be in a second home or as a visitor elsewhere). This would *not* be the same as the business location unless the employer actually lived in the business premises or nearby, which could well have been the case, e.g. for many shopkeepers, millers, or farmers. For almost all small businesses and the self-employed own account this distinction is unimportant because they will generally be recorded in the same parish or nearby. However, for larger businesses there will be a level of disjuncture between the home address of its owner and from where the business operates. This was most significant in the larger towns and cities, especially London, and for some sectors; e.g. mining, which had many absentee owners. It suggests that care must be taken in interpreting location information, particularly for the larger businesses.

For directors it is important to use the residential address for data enrichment purposes as this links the individual to their census record (and hence their demographic and personal characteristics); but this then has to be linked to their company which may have a different geography. For incorporated businesses the registered address was often in London (or Edinburgh for Scottish businesses, or Truro for Stannary companies), and many proprietors were also located in London. However, as discussed above, companies are generally

excluded from the census and only enter analysis by their inclusion through data enrichment from sources other than the census, from which their area of operations and proprietor places of residence (where known) can be coded separately.

For both small and large businesses efforts are made in the entrepreneurs database to give the alternative location where both place of operation and proprietor place of residence can be determined and this is believed to be significant. This is at best partial and provides only a sample. However, since much of the database is provided at Registration Sub district (RSD) level, which were quite large (between 2000 and 2200 persons in England and Wales depending on the census year) and often represented functional economic units, for many small firms this means that the residence and place of business were in the same geographical unit in any case.

For businesses with multiple locations and branches, the census made no attempt to assess different operations until 1931, though multiple shops were separately identified as such from 1901. Over 1851-1911, therefore, there should in theory be no duplication of reporting from multiple business locations; businesses were intended to be identified at the proprietor's place of residence as a single business with no double counting of employers. Plants or branches are not expected be identifiable in the census except for some cases of multiple shops from 1901.

4.6 Boarders, lodgers and visitors

The original and primary purpose of the census was a count of the whole population who were in a place on the Sunday night of the census count, 'who slept or abode in this house on the night of …'; and from 1861 it was to include at their residence 'those who may be travelling or out at work during the night, and who return home on Monday'. On census night many people could be absent from their normal place of residence, had no normal residence, or might be absent from the country altogether. Those absent on census night from one household were not to be entered, but should have been recorded at the house or lodging where they were, unless out of the country. Entrepreneurs could occur in any of the possible alternative categories of boarders, lodgers and visitors.

Visitors were specifically listed as a category from 1851 as part of the column on the 'relation to head of family'. Unfortunately there was no request to state any possible additional family (or business) relationship of visitors to the head. This was not explicit resulting in some CEBs containing extra relationship information to the head as well as recording "visitor", but this is relatively rare. From 1921 it became an explicit instruction to exclude any additional relationship information. Visitors were generally individuals, but those so identified could include visitors' families travelling with them, where again their relationships with each other might be included in the CEB, but was not required. In 1911 'absent' was included for those in military establishments located abroad not present on census night (but was not used in the normal household schedules). Business proprietors who were visitors often gave their business details, including their employee numbers for 1851-81. They have to be included in the entrepreneurs database when seeking a count of all businesses, but since they are definitionally in a different location to that of their home and probably of their business, they have to be treated and coded separately.

Boarders and lodgers were not specifically itemised as a possible category until 1861, where they were included in the 'relation to head of family' column. The column headed the request as 'boarder'; however, the examples in the instructions listed a 'lodger' and no boarders. The general instructions referred to the whole or portions of the house being 'sub-let to different families or lodgers'. The format of the instructions continued until 1901, and then with only a small textual change for 1911 when an instruction to the keeper, manager or other person in charge of hotels, boarding houses, clubs and similar establishments was required to fill up the schedule for all inmates. Generally a distinction that researchers and GRO have tried to draw is between a boarder who takes meals with the family and a lodger who is self-catering. This implies different household structures, where a boarder rents just a room and otherwise shares with the rest of the household, whilst a lodger rents space including a kitchen and bathroom (or has separate access to it) which could include lodging in almost all of the whole housing unit. However, internal housing relationships were often fluid in Victorian times and it is not clear that enumerators were able to consistently distinguish between boarders and lodgers. Boarders and lodgers could include their own families if present ('boarders wife' etc.), which were identified in I-CeM as separate family units. Commercial travellers, who can be quite numerous among boarders, are not generally classified as business proprietors or entrepreneurs in this study as they were usually employed by one or several companies, even if this might be on a tenuous or purely a commission basis.

4.7 Institutions

Where the building enumerated was an institution this was treated differently by the census. Where small it was generally enumerated within the CEBs of the district as an identified unit with its name (in the same way as a separate household), but its occupiers were usually listed as 'inmates' or by their specific titles such as 'patient' in hospital or 'prisoner' in a prison. Such institutions included many hotels and lodging houses which had many business proprietors as visitors, lodgers and boarders. In 1901, for example, there were 49,000 boarders and lodgers and 18,000 visitors who were business proprietors. These 'quasi-institutions' have been usually but not always coded in I-CeM. Large institutions were given separate forms which were included in the CEBs as separate enumeration books with the title of the institution, which should be linked to the geographical enumeration district to which they relate. Some of these records have become displaced and have been difficult to code in I-CeM leading to some uncertainties.

I-CeM gives a range of codes for institutional inmates where they are explicitly identified, but this is not entirely successful so that not all institutions can be drawn out from institutional codes applied to individuals, resulting in the need for care in interpreting some records. In one case study area (Devizes), for example, the local gaol is not separately identified as an institution in I-CeM, but two business proprietors are correctly identified in I-CeM as 'prisoner'. The two, who are a butcher, and a vegetable dealer, appear not to be local, illustrating how many institutional inmates have to be handled in a similar way to boarders and lodgers.

In addition, the business proprietors of institutions will be embedded in different census schedules and have to be identified and extracted separately, as discussed in subsequent Working Papers.

4.8. Census dates

The census was undertaken as referring to where people were resident on a Sunday night, with the papers given out on the day or the day before, and collected in on the following day. These dates fall at the end of March or beginning of April:

1851 30 March
1861 7 April
1871 2 April
1881 3 April
1891 5 April
1901 31 March
1911 2 April

When enriching or comparing with other data it is important to choose the source material as close as possible to these census dates, as far as the sources are actually available. This means that comparisons with directories that were compiled on an annual basis are most likely to be accurate for the previous year, but if there was a substantial lag in production, then the best comparative year will be the same as the census year. Membership directories of organisations (such as for medics and lawyers) are probably normally best taken at census year dates. Trade directories are probably best taken at the census year or the year following as many were not updated quickly. The actual comparison dates used for each comparator or data enrichment source is given in the documentation attached to the entrepreneurs database *User Guide*.

4.9 Census years and the economic cycle

An important economic literature focuses on changes in entrepreneurship and small firm numbers across the economic cycle, generally suggesting that it is pro-cyclical. As far as this is true, over the long term there will be lower levels of entrepreneurship in major depressions such as those in 1814-27 and 1870-85; and increases in boom periods such as 1843-57, 1898-1911; with slow-downs in entrepreneurship and business numbers early in down-turns, and accelerations early in up-turns. This also interrelates with rates of incorporation. However, of necessity, some entrepreneurs engaged in self-employment in desperate circumstances, meaning that there may be changes in entrepreneurship in a counter-cyclical fashion, as an alternative to unemployment. These possible influences are part of the econometric study within the project. But for understanding any cross section, and for checking comparative numbers between sources (especially if not for exactly the same years), care must be

exercised in interpretation to recognise the economic environment of the time; though of course it must be borne in mind that all regions, businesses and sectors may not share the same economic experience at the same time. Two definitions are used below: one derived from the broad economic cycle over the census period classified as to whether unemployment levels over the previous two years were rising or falling rapidly, or were relatively stable, based on ONS historical series; and a second derived from the classification of short-term changes in prices and financial indicators provided by Bordo, Dueker and Wheelock (2003). These two measures yield similar insights for the different census years. The census periods in the database can be interpreted against these changes as follows:

Census year	*Broad economic trend*	*Bordo et al. classification of prior year*
1851	'Normal'	'Normal'
1861	'Normal'	Normal'
1871	Recovery from depression	'Distressed'
1881	Depressed activity beginning to increase	'Expansion'
1891	'Normal'	'Normal'
1901	'Normal'	'Normal'
1911	'Moderate growth'	'Expansion'

6. Conclusion.

This paper examines how the population of businesses in Britain can be estimated by use of the census, and where the census needs to be supplemented by other sources. The paper has outlined how the census allows employers, sole proprietor own account self-employed, and workers to be identified. It also outlines how partnerships and corporations (either as the business entity and/or for its individual directors) can be reconstructed or added to the census records to create a more complete database of the business population. The paper also outlines how the electronic records of the census derived from the original Census Enumerators Books (CEBs) contained in I-CeM can be used, how these records can be supplemented to overcome deficiencies, and how other data enrichment can be developed.

As noted at the outset, the population census was not a business census, with the result that the way in which the information was gathered constrains the business information that can be obtained and its level of completeness. However, as clear in this paper, the census when combined with other resources has great potential to fill many of the statistical gaps that have previously impeded research. As a source of systematic information on most employers and own account self-employed, and for early years also the size of the workforces of employers and the acreage of farms, it provides a much larger scale source than previously available. Being a source at the level of each individual for the whole population, the census also provides information that can be linked with other sources on individuals. Other Working Papers examine how the actual extraction process operated to yield the final database, evaluate the accuracy of the estimates of the business population that can be obtained, describe the data enrichments performed, and various stages of subsequent analysis.

A key objective of the database construction is to open opportunities for wider research by making available the first database of the population of individual businesses in Britain for 1851-1911 through deposit at UKDA. A process of comment and interchange is encouraged. Other researchers who would like to develop links with this project and database are encouraged to make contact.

Acknowledgments:

This research has been supported by the ESRC under project grant ES/M010953: **Drivers of Entrepreneurship and Small Businesses.** Piloting of the research for 1881 draws from Leverhulme Trust grant RG66385: **The long-term evolution of Small and Medium-Sized Enterprises (SMEs).** Additional support for the coding of the 1871 census data derived was supported as part of the ESRC project with additional support for data coding and cleaning from the Isaac Newton Trust research grant 17.07(d): **Business Employers in 1871**.

A special acknowledgement of thanks is made to Kevin Schürer for advice and all his help in developing improved versions of I-CeM, and to Alice Reid, Eilidh Garrett, Joe Day, Hanna Jaadla, Xuesheng You, Leigh Shaw-Taylor and other members of the Campop I-CeM group who, with the authors, have collectively worked on the new versions of I-CeM.

References.

Bennett, Robert J. (2016) Interpreting business partnerships in late Victorian Britain, *Economic History Review*, 69, 4, 1199–1227.

Bennett, Robert J. and Newton Gill (2015) Employers in the 1881 population census of England and Wales, *Local Population Studies*, 94, 29-49.

Booth, Charles (1886) Occupations of the People of the United Kingdom, 1801-81, *Journal of the Royal Statistical Society*, 49, 2, 314-444.

Bordo, M. D., Dueker, M. J. and Wheelock, D. C. (2003), 'Aggregate price shocks and financial stability: the United Kingdom 1796-1999', *Explorations in Economic History*, 40, 143-69.

Hakim, Catherine (1980) Census Reports as Documentary Evidence: The Census Commentaries 1801-1951, *Sociological Review*, 28, 3, 551-80.

Higgs, Edward and Schürer, Kevin (University of Essex) (2014) *The Integrated Census Microdata (I-CeM)* UKDA, SN-7481

Schürer, K., Higgs, E., Reid, A.M., Garrett, E.M. (2016) *Integrated Census Microdata V.2 (I-CeM.2)* [data collection].

Schürer, Kevin and Woollard, Matthew (2000) *1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version)*, UKDA, University of Essex, SN-4177.

Other Working Papers:

Working paper series: ESRC project ES/M010953: *'Drivers of Entrepreneurship and Small Business',* University of Cambridge, Department of Geography and Cambridge Group for the History of Population and Social Structure.

WP 1: Bennett, Robert J., Smith Harry J. van Lieshout, Carry, and Newton, Gill (2017) *Drivers of Entrepreneurship and Small Businesses: Project overview and database design.*

WP 2: Bennett, Robert J., Smith Harry J. and van Lieshout, Carry (2017) *Employers and the self-employed in the censuses 1851-1911: The census as a source for identifying entrepreneurs, business numbers and size distribution.*

WP 3: van Lieshout, Carry, Bennett, Robert J., Smith, Harry J. and Newton, Gill (2017) *Identifying businesses and entrepreneurs in the Censuses 1851-1881*.

WP 4: Smith, Harry J., Bennett, Robert J., and van Lieshout, Carry (2017) *Extracting entrepreneurs from the Censuses, 1891-1911.*

WP 5: Bennett, Robert J., Smith Harry J. van Lieshout, Carry, and Newton, Gill (2017) *Business sectors, occupations and aggregations of census data 1851-1911.*

Full list of all current Working Papers available at:
http://www.geog.cam.ac.uk/research/projects/historyofentrepreneurship/