

**SWP 35/90 MANAGING THE CONSTRUCTION OF A
MANUFACTURING ESTABLISHMENT DATA BANK**

**DR PAUL WESTHEAD
Research Officer
Cranfield School of Management
Cranfield Institute of Technology
Cranfield
Bedford MK43 OAL
UK**

(Tel: 0234 - 751122)

Copyright: Westhead 1990



MANAGING THE CONSTRUCTION OF A MANUFACTURING ESTABLISHMENT DATA-BANK

PAUL WESTHEAD

(Research Officer, Cranfield School of Management, Cranfield Institute of
Technology, Cranfield, Bedford, England, MK45 OAL)

INTRODUCTION

Over the last decade there has been an increased interest in the development of an 'enterprise culture' and the formation and development of new independent small establishments within the United Kingdom. Recently policymakers have requested detailed information surrounding spatial variations in new firm formation rates (Frank et al., 1984). The need for more detailed research into the nature and extent of spatial variations in new manufacturing firms has also been noted by researchers (Mason and Harrison, 1985; Johnson, 1986; Mason, 1987); not least because such information is an essential prerequisite in justifying the case for a spatially selective small firms policy (Storey, 1982).

This paper is part of a wider study of new independent manufacturing firm founders established in contrasting environments in Wales between the 1st January 1979 and the 31st December 1985 (Westhead, 1988). Construction of the manufacturing establishment data-bank (EDB) was to provide a lead in to a sampling or at least survey frame for new firms in contrasting labour market environments. In this study the environment was conceived in spatial terms as a labour market area or a Revised (1978) Travel-to-Work Area (TTWA) - in which people live and work and ultimately seek premises for the establishment of their own business. Since the identity of these contrasting TTWAs could not be determined until after a classification of TTWAs into 'ecological incubator' environments (Westhead, 1988), and since the layout of the 'atomic' sources (e.g the new independent manufacturing establishments) rarely fitted a TTWA framework, it was decided to build an EDB for the whole of Wales.

A wide range of manufacturing establishment data including business name, establishment address and postal code, Revised (1978) TTWA location code, telephone number, description of the establishments activities, activities listed as Standard Industrial Category (1968 and 1980) Minimum List Headings, ownership details and the data sources which had recorded the identity of the establishments were obtained during three stages of data collection. The first stage entailed the identification of the total population of manufacturing establishments in each of the 40 TTWAs. The

second stage identified and eliminated branch and subsidiary establishments through a detailed checking of secondary data sources. During the third stage, based on predominantly 1983 secondary data sources, the EDB was up-dated and 'cleaned' through 'direct surveys' of establishments in selected TTWAs. At this stage the total population of surviving new independent manufacturing firms established between the two time periods in 20 out of 40 TTWAs were identified. This list of new firms provided the survey frame for a questionnaire survey of new firm founders conducted in 1986.

RESEARCH OBJECTIVES

This paper has three objectives. First, the methodological advantages and disadvantages of constructing a manufacturing EDB are detailed. Second, an evaluation of the fifteen data sources used in constructing the EDB is presented. Third, the methodological problems associated with identifying the new independent manufacturing firm is explored.

WHAT IS A MANUFACTURING ESTABLISHMENT DATA-BANK?

The inadequacy or non-existence of economic data below a regional level has encouraged a number of researchers to invest considerable amounts of time, labour and computing resources into creating establishment based data sets which can overcome many of the deficiencies of official data. Basically, an EDB is defined as a file of data held on all establishments from a specified population (e.g. particular data sources) in a specified area (Firn, 1976, p.9). A manufacturing establishment data-bank that is constructed with individual establishments rather than at the more aggregated level of the firm provides the 'industrial lowest denominator' (Lloyd and Dicken, 1968, p.311), which defines physically distinct premises and, at least initially, avoids the confusion between single and multi-plant firms.

METHODOLOGICAL ADVANTAGES OF ESTABLISHMENT DATA-BANKS

A major problem involved in the use of virtually all officially collected data is that of the spatial units over which the data are aggregated. These units - planning regions, local authority areas, enumeration districts, employment exchange areas - are invariably arbitrary administrative constructs bearing little, if any, relation to the 'spatial facts of life'. It is universally acknowledged that such units rarely if ever correspond to the specific requirements of a research project. One obvious difficulty is the irritating tendency for the boundaries of areas to change over time rendering

awkward, or even impossible, an accurate study of trends. Less widely appreciated but certainly no less important are the implications of the variations in the size and shape of the different spatial units: in other words, the problems posed by the nature of the aggregation. Yet, although the existing units are widely regarded as unsatisfactory, data based upon them have been subjected to a variety of analytical techniques apparently without regard for the distortional influence of the manner of the aggregation (Lloyd and Dicken, 1968, p.308). Access to data at the establishment level is essential in many situations to enable the total number of new manufacturing firms to be identified. Establishment data collected in an 'atomic' form has been shown to overcome some of the distortional effects of aggregation (both spatial and temporal) inherent in published data (Gudgin, 1978; Mason, 1981; Healey, 1983; Birley, 1984; Aldrich et al., 1989).

Furthermore, the spatial distribution of establishments in a realistic 'atomic' sense gives the flexibility of aggregation into meaningful categories with which to explore and test the numerous hypotheses in new firms research which relate to individual units. The availability of firm and establishment data at two or more points in time would also allow the 'components of change' technique to be adopted. Components of change is an accounting framework which disaggregates net change into its individual parts: openings, closures, migrants and survivors. Perhaps the most useful property of an EDB with or without the time dimension is that an EDB can represent a population of establishments or firms from which a new firm survey can be drawn. Moreover, events at the establishment level can be related to higher order processes in the management of business enterprise, through an extension of the characteristics of the establishments to include the locus of corporate control. This would enable an important distinction to be drawn, for example, between an independent indigenous establishment and a branch plant formation (Mason, 1981, p.2).

An EDB can provide an important input to policy decision-making because it can supply information on industrial structure, the outcome of firm strategies, the effects of state intervention and the impact of individual entrepreneurial initiatives. Furthermore, Keeble (1978, p.322) noted that a particularly fruitful avenue of progress in industrial geography would be to extend EDB coverage to "areas and regions which exhibit a different industrial environment and whose manufacturing performance differs from that of declining conurbations on which attention has so far tended to focus".

PROBLEMS OF ESTABLISHMENT DATA-BANKS

Although establishment data possess superior attributes to data that is aggregated by geographical area, sector or size, there are a number of problems which can be extremely difficult to resolve if such data are to provide more than a highly sophisticated and costly inventory of industrial stocks. Indeed it is the critical need to explore the dynamic nature of industrial events and processes over space and time that warrants the development and maintenance of an EDB. The technical and methodological problems of such an undertaking are considerable, but it is argued that they are not insurmountable (Thompson, 1983, p.67-68).

Resources

Setting up an EDB is expensive in terms of time and cost. The required resources are significant when the attributes assigned to the EDB are not simply confined to being extracted from a common source. The data collection phase alone can be a long and tedious clerical exercise particularly when collection demands the collating of materials from a number of sites. The initial stage must invariably be followed by the lengthy process of data 'cleaning' because much of the data is derived from sources not designed for use in an EDB (see Firn, 1976, p.12; Hubbard et al., 1979; Thompson, 1983, p.70). This will entail checking for inconsistencies, illogical data (i.e. 'ineligible classifications'), identifying gaps in the data and eliminating duplications. Secondary matching of the data with previously collected records can present its own unique and highly labour intensive problems. Further supplementary information with which to augment the raw data may be necessary in order that the context for the research may be defined in relation to each establishment. Commonly, this includes a sectoral classification, spatial referencing and characteristics of organisational control, each of which consumes vast quantities of time, since they require a number of sources of information. A final phase in the construction of an EDB is the procedure for dealing with incomplete or dubious data, where the source has not provided the necessary information. It is therefore necessary to adopt a structural convention for enhancing this type of data, even if it is a temporary feature during the search for the correct information. The expense of constructing EDB's has led to some questions about their cost-effectiveness (e.g. Swales, 1976). However, the rapid increase in recent years in the number of EDB's set up outside government, suggests that the consensus of opinion is that the potential insights into the patterns and processes of industrial location, as well as new firm formation, are worth the time and effort involved (Healey, 1983, p.15).

Access and the Constraints of Confidentiality

Although government agencies are empowered to collect information from industrial establishments, this does not become publicly available until it has been aggregated to such a level that no individual plant or firm can be identified. The level of aggregation is administratively determined, often related to the notions of commuting patterns, the inadequacy of which, even for this limited purpose has been laid bare (Coombes and Openshaw, 1982). When access is granted to bona fide researchers, then the same restrictions under the Statistics of Trade Act, 1947 (this act states that information collected by an official Government department on an individual business cannot be disclosed to a third party due to confidentiality constraints), still apply to the published material. Moreover, as Sprott (1977, p.13) stated, "In accordance with Section 9 of the 1947 Act, care must be taken to ensure that any subsequent report, summary or communication compiled from this information is so arranged as to present any particulars relating to any individual person or undertaking except with the previous consent in writing of that person or the person carrying on that undertaking, as the case may be".

Coverage and the Deficit Caused by Exemption

No microdata source claims to include every manufacturing establishment in an area or industry. Invariably some very small establishments are missed out or do not respond to the request for information. Coverage between various sources can therefore vary widely. Sometimes small establishments are omitted deliberately. For example, the Business Statistics Office (BSO) Classified List of Manufacturing Businesses (PO 1007) excludes all establishments employing less than twenty people. In some commercial directories establishments have been omitted because some companies refuse to pay the entry fee. Surprisingly, some of the non-official directories can give a coverage as comprehensive as the official ones. For example, in a comparison of four data sources, Foley (1983, p.103) found for the electrical engineering industry in West Yorkshire - Market Facts File (MFF) and Industrial Market Location (IML) directories gave almost as good a coverage as the Factory Inspectorate (FI) data and better for his purposes than the Annual Census of Employment (ACE), though the differences were not large. But for a number of research reasons the deficiencies in coverage pose considerable problems. For example, in order to monitor the process of new firm formation difficulties arise through a policy of exemption of establishments employing a small number of workers, especially when the average employment in indigenous enterprise within the first five years of its life cycle is likely to be less than ten (Lloyd and Dicken, 1982).

Comparability, Time Specificity and the Problems of Updating Cycles

There are difficulties in combining data from different sources. Moreover, there are problems with comparing information between data sources. A major reason for the difficulties in comparability is that the definitions, classifications and procedures used vary between sources. A related difficulty is the different time periods to which the data relate. The time dimension may pose several complications in constructing a data-bank. Unless the data collection is originally by census, it is not strictly correct to assign information collectively to a specific point in time. Where the updating of records is conducted on a rolling programme, the time span of each cycle can be extended to several years, but even this does not imply that each establishment is treated equally in this respect. In fact, establishments could be treated differently on the basis of characteristics which are only tangentially related to employment, but which may have significant effects on the reliability of data from them. If the data collection is by census, the intercensal period may be excessively long and produce its own methodological problems, restricting the ensuing analysis to a comparison of cross-sectional slices. The delays at source in collecting and reporting new information may restrict investigation to an historical perspective.

The Problems of Flows

A key part of an analysis of establishments by the components of change technique is the ability to distinguish between those establishments which are new, have closed down, have remained locationally static and those which have migrated. The detail required to ascertain whether an establishment has changed its operational name rather than being replaced by a new enterprise is often not available, or only on a patchy basis, from the original data source. Similarly, in order to determine whether an establishment has migrated from its original location, as opposed to 'dying', and conversely, to distinguish between immigrants and genuine new start ups requires substantial supplementary industrial data.

Economic Indicators and their Limitations

Most sources of data for EDB's provide only a range of economic indicators for each industrial establishment, beyond a few non-economic 'factor' input specifications. Generally, this is restricted to the magnitude of employment, with a possible subdivision by sex and occupational type (Swales, 1976, p.3). Whilst this may be of primary interest for research, it provides little scope of an understanding of the changing strategy and likely prospects of a firm, beyond the recent manifestations in total workforce.

Accuracy

Even when the coverage is good the sources may include inaccuracies. One of the weakest areas lies in the classification of establishments into industrial groups. Some of the errors occur because only limited business activity information is available from some employers and some detailed descriptions are themselves difficult to classify. For example, establishments on the manufacturing-service interface could be incorrectly included or excluded (e.g. those described as undertaking repair, bottling, bakeries attached to small shops - classified as non-manufacturing activities). In many cases, notably involving small jobbing engineering establishments and multi-plant establishments, there may be no 'correct' Standard Industrial Category Minimum List Heading (MLH). Moreover, analysis of change over time can be markedly affected by these inaccuracies in industrial classifications. A similar problem may occur from errors in the location coding of establishments. Further problems can occur because of the inclusion of out-of-date information which sometimes happens because of the inclusion of unchecked data from one edition of a directory to the next.

Evaluation and Explanation

Results from establishment based research often seem to have been a good first step, giving a good description of certain phenomena, without being able to explain or evaluate these phenomena. In fact, Swales (1976, p.6) claimed that, "although databanks have provided a fine degree of disaggregate description, they may not lead to an explanation and evaluation of the phenomena they describe". Lloyd and Mason (1976, p.55) have also noted that "Realistically, the data should be seen as a starting point and not a complete record. It permits the academic researcher to begin rather than to finish his search for the detail of industrial processes at the micro level".

Summary

The extent to which these limitations of establishment based research are significant clearly varies from one data-base to another. The problems of resources and access to data affect the viability of the enterprise and are influenced by the ability to obtain research funding and the quality of the contacts made with those who provide the data. The effects of the limitations of the coverage and the accuracy of the sources used can be reduced, principally by using as many different sources as possible, and by vigorous checking of the data used. Of course, the extent to which

the strategy can be followed depends upon access to data and resources. It also raises again the comparability limitation in the form of the difficulties of combining information from different sources (Foley, 1983, p.104; North et al., 1983, p.121). Furthermore these data sets represent second best in many cases, yet they may possess advantages that, while not removing their inadequacies, a researcher can certainly reduce to manageable proportions especially in the light of the objectives of the study.

THE PROCESS OF CONSTRUCTING AN EDB FOR WALES

Official Data Sources

This researcher initially tried to gain access to establishment data held by the following Government departments: the Annual Census of Employment, Watford; the Factory Inspectorate of the Health and Safety Executive (HSE), Cardiff; the Industry Department of the Welsh Office, Cardiff; and the Welsh Development Agency (WDA), Pontypridd. Unfortunately, due to the confidentiality restrictions of the 1947 Statistics of Trade Act and resource constraints access was not granted. However, one official data source - the Classified List of Manufacturing Businesses (PO 1007) compiled from the UK Register of Manufacturing Businesses under the administration of the Business Statistics Office (BSO) for the Government was available. The BSO provided establishment information regarding name, address, and MLH (1968) classifications. Thompson (1983, p.73) has however identified two major drawbacks when using this data source. Firstly, the BSO, from 1973 onwards, only covers establishments which employ twenty or more workers, thus excluding most new establishments. Secondly, the classified list is voluntary, thus providing a basis to the list which cannot be estimated, except at the gross level. The BSO list for the UK contains some 27,000 manufacturing businesses, which is estimated to cover approximately two thirds of all manufacturing employment. BSO also claim that the shortfall is mainly accounted for by refusals for inclusion in the list and less by the small enterprises which were exempt. Moreover, no estimate has been made of the spatial bias in the classified list. Even with these weaknesses this data source provided a starting point for the construction of an EDB for Wales.

Supplementary Unofficial Data Sources

Fourteen other data sources were also identified and used in the construction of the EDB for the 40 Revised (1978) TTWAs in Wales. The majority of sources used were not derived from Government departments but from supplementary unofficial data

sources such as telephone directories, local newspapers, County Councils, Chambers of Commerce, development agencies and industrial directories compiled by private agencies. At this point it must be stated that there were no hitherto unknown sources with magically high yields of information (Forster, 1981, p.43). A number of sources had a relatively low yield of information but there was still a need for considerable effort in checking the data (Table 1). As can be seen, the density of information per source varied greatly. As a result it was a very labour intensive exercise which, although essentially clerical, still required a high degree of initiative and a considerable amount of commitment.

Construction of a Pilot List of Manufacturing Establishments

A pilot list of manufacturing establishments was constructed for the Aberystwyth and Lampeter TTWAs in order to ascertain the most productive sources as well as the level of overlap between data sources in identifying establishments. Eleven different data sources were identified and utilised in the two selected pilot study areas for the construction of the EDB. Tables 2(i) and 2(ii) indicate that two of the sources were official, Factory Inspectorate (FI) [code 7] and Business Statistics Office (BSO) [code 10] data, whilst the remaining nine were supplementary unofficial data sources. The FI data was supplied in order to identify establishments in the Ceredigion District waste survey (Moyes and Heal, 1985) for the Aberystwyth and Lampeter TTWAs. For each establishment identified in the Aberystwyth and Lampeter TTWAs the single data source or the multiple data sources that had recorded the existence of that establishment were noted. On the basis of this information it was possible to state how many establishments were identified by a single data source as well the number identified by more than one data source alone. Table 2(i), for example, indicates that one establishment in the Aberystwyth TTWA was solely recorded in the Welsh Chamber of Commerce Directory (WCC) [code 2], whilst another three establishments were recorded in the Industrial Directory for Dyfed (IDD) [code 5] alone. In contrast, two establishments were each recorded in each of the following data sources: the Industrial Directory for Wales (IDW) [code 1], the WCC [code 2], Yellow Pages [code 3], the Mid Wales Industrial Directory (MWD) [code 4], the IDD [code 5], the Cambrian News / Cambrian Trader newspapers (CN/T) [code 6] and the rateable value rolls (RV) [code 9].

It can be inferred from the information presented in Tables 2(i) and 2(ii) that no single data source identified all the establishments in the two TTWAs. Evidence presented in Tables 2(i) and 2(ii) suggests that Yellow Pages (YP) and local and county industrial and trade directories [codes 4, 5, 12, 13, 14 & 15] were the most prolific identifiers of establishments. Rateable value rolls were quite a good source

but due to the administrative problem of gaining access to similar rateable value rolls from other districts it was not used again. Similarly, local newspapers and trade newspapers identified a number of establishments, but the rate of return from time spent scanning these was relatively low. Kompass Industrial Directory (KOS) [code 11] and the BSO's Classified List of Manufacturing Businesses (PO 1007) were of limited use in identifying a large number of establishments but PO 1007 was used throughout Wales in order to identify larger establishments. Results of the pilot studies suggested that a 'single data source can only approximate reality' and it is essential to use more than one data source in order to identify manufacturing establishments when constructing an EDB.

Evidence presented in Table 3 indicated that Yellow Pages (YP), the Welsh Chamber of Commerce Directory (WCC), the Industrial Market Location Directory (IML) [code 8] and the Industrial Directory of Wales (IDW) were the most significant identifiers of establishments in the two selected labour markets. Again, it can be inferred from Table 3 that local data sources are of considerable relevance in the construction of an EDB. Moreover, the comprehensive deletions records of the FI allowed the EDB data to be further 'cleaned' resulting in the removal of recorded establishment closures from the EDB card index.

Characteristics of the EDB

The constructed EDB contained 7,653 manufacturing establishments and as expected the majority of establishments were located in traditional industrial TTWAs South and North-East Wales but a variety of rural TTWAs did have a considerable number of establishments recorded (Table 4). This compared with 5,354 manufacturing establishments identified by the Factory Inspectorate (FI) in 1985 (FI, personal correspondence) and 2,141 establishments (with 11 or more employees) and 1,341 establishments (with 20 or more employees) identified by the Business Statistics Office (BSO) (BSO, Business Monitor PA 1002, 1979 and 1983) or 1979 and 1983, respectively.

IDENTIFYING NEW MANUFACTURING FIRMS IN SELECTED TTWAs

The New Independent Manufacturing Firm

Identification of new manufacturing firms has always been a difficult problem for research workers (Mason, 1983). The start-up process has been described as a series of stages involving the developing of skills, the recognising of opportunities and the assembling of resources (Stanworth and Curran, 1976; Birley, 1982). Some of these

stages, which have been described by Cooper (1981) as the 'incubation period', are passive and can take place over many years. Indeed, even the active period after the final decision to start has been made can take an extended period of time (Birley, 1984). Birley (1984) has listed eight possible points of 'newness' with regard to the time at which the new firm can be said to exist. For example, from the owners decision to start to when the founder takes on the first full-time employee. She noted that choosing any of the eight points involved a somewhat arbitrary decision about when a firm actually comes into existence, and each poses problems for researchers seeking a representative sample of new firms. For the purposes of this study, the definition of the new firm adopted was that the wholly new manufacturing firm had been established independently and had no "obvious parent in any existing business organisation" (Allen, 1961). The new firm had to be established between the 1st January 1979 and the 31st December 1985. Moreover, the start-up date of the new firm was taken as the date of commencement of production on a full-time basis. Also, each new firm had to employ one (i.e. the founder) or more workers.

Identifying the New Firms

The data-bank assembled for the whole of Wales was used as a basis for field survey in selected TTWAs which had demonstrably different environmental conditions and characteristics (Westhead, 1988) and twenty out of the forty TTWAs were selected for a survey of new manufacturing firm founders. From the total number of establishments in each of the selected TTWAs externally-controlled branch and multi-plant establishments in the EDB based on secondary data sources were identified by careful checking through Who Owns Whom, Industrial Market Location Directories and Companies House Register of Limited Companies. The remaining establishments were personally contacted in order to ascertain their precise status in terms of ownership, date of start up on a full-time basis and industrial activity. This stage is detailed below.

Through detailed fieldwork (e.g. 'walking the streets') in the selected TTWAs in 1986 it was possible to up-date the EDB which had been constructed on predominantly 1983 data sources (a direct enumeration approach to identifying new businesses has also been undertaken by Aldrich et al., 1989 in Durham County, North Carolina in the United States). This direct survey enabled 240 surviving new independent manufacturing firms established between 1979 and 1985 to be identified and recorded (Table 5). Establishments omitted from the data sources referred to may have been left out for a variety of reasons including: they were too small to be included; they were too new to be included; some because of the nature of their

business, will have chosen to be ex-directory; those relying on word of mouth or subcontract work may have considered it unnecessary; and some will merely have used their private number. A further 195 'other' manufacturing establishments were identified during the 1986 survey which had been established prior to the 1st January 1979 or after the 31st December 1985. Moreover, the importance of conducting a direct survey in selected environments is indicated by the evidence presented in Table 6 which shows that 240 out of the 335 surviving new firms (71.6%) in the study period were identified by the direct survey approach.

The resulting EDB based on 1983 data sources and the direct survey conducted in 1986 enabled the total population of 335 surviving new independent manufacturing firms in the twenty selected TTWAs to be identified and recorded. The EDB constructed for selected TTWAs in 1986 was a comprehensive data-base of not only all new independent establishments but branch and subsidiary establishments. Again, branch and multi-plant establishments were identified by careful checking through Who Owns Whom, Companies House Register of Companies and direct personal contact with individuals. Moreover, it must be noted that establishments were withdrawn from the EDB if they could no longer be identified either due to 'closure' or 'transfer' out of the TTWA under analysis.

The New Firm Survey

Between January and June 1986 personal interviews were conducted with new firm founders in the selected TTWAs. The unarranged 'knock-on door' approach achieved a high response rate (80.3%: 269 out of 335 new independent firms contacted). In two TTWAs (Machynlleth and Monmouth) no surviving new independent firms were identified in 1986. Spatial variations in response rates do exist but they are not so large as to invalidate the use of the survey results for analysis. For example, a response rate over 85% was recorded in Llangollen, Llanrwst, Rhyl, Tywyn, Lampeter, Holyhead and Milford Haven, whilst in Aberystwyth, Llandrindod Wells and Welshpool response rates were as low as 57% (Table 7). Also, no attempt was made to control for industry type in the 'grab' field survey but as Table 8 indicates the surveyed firms were not seriously over-represented in any direction. Response rates over 88% were recorded in Metal Manufacture (SIC 6), Instrument Engineering (SIC 8), Timber, Furniture, etc (SIC 17) and Electrical Engineering (SIC 9), whilst in Other Manufacturing Industries (SIC 19) and Vehicles (SIC 11) rates between only 65% and 75% were recorded. On the basis of the results presented in Tables 7 and 8 it is not unreasonable to suggest that a representative survey of surviving new independent firms in selected TTWAs was conducted in 1986.

CONCLUSION

The above discussion has detailed the advantages and disadvantages associated with the construction of an EDB. Empirical evidence presented in this study indicates that wide variety of data sources have to be consulted if a researcher wishes to identify the population of establishments in a TTWA. A major lesson to be learnt from the above exercise is that a population of new firms cannot be identified by using secondary data sources alone because of the problems of omission of various data sources as well as the volatility in the pattern of the 'birth' and 'death' of establishments. A variety of official as well as supplementary unofficial data sources should be utilised. Moreover, a highly productive way of identifying new independent manufacturing firms was through detailed fieldwork (e.g. 'walking the streets') in the selected TTWAs. This field research was made easier to organise and conduct because of the previously conducted long hours of detailed clerical desk research.

The above evidence indicates that a key issue in doctoral research is choosing an appropriate method for collecting data on the problem of interest. A number of different organisational stages in the management of data gathering have been illustrated. Once the subject of investigation has been defined and justified the next step was to collect data in order to test presented research questions or hypotheses. During the research design stage of the doctoral project decisions have to be made on what information to generate, the methods of data collection, the coverage of the data, and the way in which data are to be analysed. This paper has primarily focussed on the problems and issues surrounding the methods of data collection.

In order to obtain information on new firm founders the sampling process followed a series of stages. First, the population of interest [manufacturing firms] and individuals [new firm founders] were defined. Second, the sampling frame was developed which located the individuals within the population (e.g. new independent manufacturing firm founders in contrasting labour market areas). This stage was associated with the construction of an EDB which listed the location of manufacturing firms in Wales. Third, the sampling unit, the selection of which depends on the nature of the study was designed. In this study the selected sampling unit was the firm founders in the new firm located in 'defined' contrasting 'ecological' incubator environments. Fourth, the next stage in the sampling process was to choose the method by which the sampling units [firm founders in new firms] were to be selected. A non-probability rather than a probability sampling method was chosen because it was an aim of the study to survey a sample which represented the characteristics of the total population. Fifth, the size of the non-probability sample was determined. For the purpose of this study it was decided on subjective

grounds to interview a minimum of 40 new firms founders in each of the 5 defined contrasting 'ecological' incubator environments. In four out of the five 'ecological' incubator environments this objective was achieved but in one environment this target was not met because of the sparsity of new firms in this 'ecological environment' (in this environment 9 out of a possible 10 new firm founders were interviewed). Sixth, it was decided to administer the questionnaires to new firm founders by direct 'unarranged' personal interview. In conclusion, the combination of desk and field research organised on a logical and consistent basis enabled a representative survey of new firm founders to be conducted in selected environments.

Table 1 A Comparison of the Data Collected by Official and Unofficial Data Sources in Wales

Information content	Data sources														
	IDW	WCC	YP	MWD	IDD	CN/T	FI	IML	RV	PO1007	KOS	CLID	MGID	CID	WGID
Company name	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Address	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Postcode	X	X	X	X	X		X	X	X	X	X	X	X	X	X
Telephone number	X	X	X	X	X	X		X			X	X	X	X	X
Title and name of location manager								X							
Description of company activities	X	X	X	X	X	X		X			X	X	X	X	X
Activities listed as SIC (1968) Orders								X							
Activities listed as SIC (1968) MLHs							X			X					
Total employment size								X			X	X	X		
Male / female employee breakdown															
Office / operative employee breakdown								X							
Ownership details of the establishment								X							
Maps to show the location of an establishment								X							

Notes: X= Indicates occurrence.

IDW Industrial Directory for Wales, 1983.

WCC Welsh Chamber of Commerce Directory, 1983.

YP Yellow Pages Telephone Directory, 1983.

MWD Mid Wales Industrial Directory (No Date).

IDD Industrial Directory for Dyfed, 1983 (IDD).

CN/T Cambrian News / Cambrian Trader (Various issues).

FI Factory Inspectorate Establishment Data, 1983.

IML Industrial Market Location Directory (Various issues).

RV Rateable Value Lists, 1984.

PO1007 Business Statistics Office, Business Monitor PO 1007 List of Manufacturing Establishments, 1979.

KOS Kompass Directory, 1983.

CLID Clwyd Industrial Directory, 1983.

MGID Mid Glamorgan Industrial Directory, 1983.

CID Cwmbran Industrial Directory, 1983.

WGID West Glamorgan Industrial Directory, 1983.

Table 2 The Coincidence of Sources in Identifying Manufacturing Establishments in the Aberystwyth and Lampeter TTWAs

(i). Aberystwyth

[2]= 1	[3]= 5	[4]= 4
[5]= 3	[7]= 5	[8]= 1
[9]= 1	[1,2,3,4,7]= 1	[1,2,3,4,5,6,7,9]=3
[1,2,3,4,5,6,9]= 2	[1,2,3,4,5,7]= 1	[1,2,3,4,6,7,10]= 1
[1,2,3,4,6,8,9]= 2	[1,3,4,5,6]= 1	[1,4]= 1
[1,4,5]= 2	[2,3,4]= 1	[2,3,4,5,6,7,8,9,10]= 1
[2,3,4,5,7]= 1	[2,3,5,6]= 1	[2,4,5]=1
[2,4,5,7]= 1	[3,4]= 4	[3,4,5,6]= 2
[3,4,5,6,7,8]= 1	[3,4,5,6,7,9]= 1	[3,4,5,6,7,9,10]= 1
[3,4,5,7,9]= 1	[3,5]= 1	[3,4,5,8,9]= 1
[3,6,9]= 1	[3,4,6,7,9]= 1	[3,6]= 1
[3,4,5,9]= 1	[3,7]= 1	[4,5]= 2
[4,6]= 1	[4,6,7,9]= 2	[4,9]= 1
[5,6]= 1	[5,7]=1	[7,10]= 1
[7,9]= 1		

Number of establishments= 67

(ii). Lampeter

[2]= 7	[3]= 28	[4]= 2
[5]= 11	[6]= 6	[7]= 3
[8]= 13	[9]= 2	[1,3,5,11]= 1
[1,2,3,4,6,8]= 1	[1,3,5,11]= 1	[1,2,3,4,6,8]= 1
[1,2,3,4,8]= 1	[1,2,3,4,5,6,7,8]= 1	[1,2,3,5,8,11]= 1
[1,3,4,5,8]= 1	[1,2,7,10]= 1	[1,2,3,4,5,7,10]= 1
[1,2,3,4,6,7,8,9,10,11]=1	[1,2,3,4,6,7,9]= 2	[1,3,4,9]= 1
[1,3,5,7,8]= 1	[1,3,4,5,7,8,9]= 1	[1,3,5,6]= 1
[1,2,4,5,8]= 1	[1,2,3,4,5,7,8,10]= 1	[1,2,3,4,5,6,7,8,9]= 1
[2,5]= 1	[2,8]= 2	[2,3,4,5]= 5
[2,3,4,5,7,8,9]= 3	[2,3,4,5,7,9]= 1	[2,4,5]= 2
[2,3,4,5,7]= 2	[2,3,5,8]= 1	[2,6,8]= 1
[2,4,5,7]= 1	[3,5]= 10	[3,4]= 2
[3,9]= 2	[3,5,6]= 2	[3,4,5]= 1
[3,6]= 3	[3,8]= 1	[3,5,7]= 1
[3,4,5,7,9]= 2	[3,7]= 2	[3,5,8]= 2
[3,5,6,8]= 1	[3,5,6,7,9]= 1	[3,6,7,8]= 1
[3,7,9]= 1	[4,5,6]= 1	[4,5]= 2
[4,6]= 1	[5,6]= 1	[5,6,7]= 1
[8,10]= 1		

Number of establishments= 138

- Notes:
- [1] Industrial Directory for Wales, 1983.
 - [2] Welsh Chamber of Commerce Directory, 1983.
 - [3] Yellow Pages Telephone Directory, 1983.
 - [4] Mid Wales Industrial Directory (No Date).
 - [5] Industrial Directory for Dyfed, 1983.
 - [6] Cambrian News / Cambrian Trader (Various Issues).
 - [7] Factory Inspectorate Establishment Data, 1983.
 - [8] Industrial Market Location Directory (Various Issues).
 - [9] Rateable Value Lists, 1984.
 - [10] Business Statistics Office, Business Monitor PO 1007 List of Manufacturing Establishments, 1979.
 - [11] Kompass Directory, 1983.
 - [12] Clwyd Industrial Directory, 1983).
 - [13] Mid Glamorgan Industrial Directory, 1983.
 - [14] Cwmbran Industrial Directory, 1983.
 - [15] West Glamorgan Industrial Directory, 1983.

Table 3 The Number of Establishments Identified by Individual Data Sources

Data source	TTWAs				Total Wales	
	Aberystwyth		Lampeter			
	No	%	No	%	No	%
IDW	14	6.5	10	9.6	2,082	14.0
WCC	17	7.9	38	10.4	2,233	15.0
YP	36	16.7	88	24.2	4,747	31.8
MWD	40	18.5	37	10.2	517	3.5
IDD	31	14.4	64	17.6	461	3.1
CN/T	23	10.6	26	7.1	57	0.4
R	26	12.0	28	7.7	73	0.5
IML	5	2.3	39	10.7	2,176	14.6
RV	20	0.0	18	4.9	50	0.3
PO1007	4	1.9	5	1.4	741	5.0
KOS	0	0.0	3	0.8	16	0.1
CLID	0	0.0	0	0.0	544	3.6
MGID	0	0.0	0	0.0	733	4.9
CID	0	0.0	0	0.0	146	1.0
WGID	0	0.0	0	0.0	336	2.3
Total	216	100.1	364	99.9	14,912	100.1

Notes:

IDW	Industrial Directory for Wales, 1983.
WCC	Welsh Chamber of Commerce Directory, 1983.
YP	Yellow Pages Telephone Directory, 1983.
MWD	Mid Wales Industrial Directory (No Date).
IDD	Industrial Directory for Dyfed, 1983.
CN/T	Cambrian News / Cambrian Trader (Various Issues).
R	Factory Inspectorate Establishment Data, 1983.
IML	Industrial Market Location Directory (Various Issues).
RV	Rateable Value Lists, 1984.
PO1007	Business Statistics Office, Business Monitor PO 1007 List of Manufacturing Establishments, 1979.
KOS	Kompass Directory, 1983.
CLID	Chwyd Industrial Directory, 1983).
MGID	Mid Glamorgan Industrial Directory, 1983.
CID	Cwmbran Industrial Directory, 1983.
WGID	West Glamorgan Industrial Directory, 1983.

Table 4 Recorded Overlap In Establishment Entries from Data Sources, 1983/84

TTWAs	Single data source		More than one data source		Total number of establishments	
	No	%	No	%	No	%
1 Aberdare	51	45.5	61	54.5	112	1.5
2 Aberystwyth	20	29.9	47	70.1	67	0.9
3 Bargoed	91	57.6	67	42.4	158	2.1
4 Barmouth	29	50.9	28	49.1	57	0.7
5 Blaenau Ffestiniog	15	44.1	19	55.9	34	0.4
6 Brecon	23	37.1	39	62.9	62	0.8
7 Caernarfon	74	67.9	35	32.1	109	1.4
8 Cardiff	830	60.5	543	39.5	1,373	17.9
9 Cardigan	43	55.8	34	44.2	77	1.0
10 Carmarthen	72	67.3	35	32.7	107	1.4
11 Denbigh	58	61.7	36	38.3	94	1.2
12 Ebbw Vale	92	58.6	65	41.4	157	2.1
13 Fishguard	24	75.0	8	25.0	32	0.4
14 Holyhead	92	66.7	46	33.3	138	1.8
15 Lampeter	72	48.6	76	51.4	148	1.9
16 Llandeilo	26	52.0	24	28.0	50	0.7
17 Llandrindod Wells	82	58.6	58	41.4	140	1.8
18 Llandudno	140	65.4	74	34.6	214	2.8
19 Llanelli	152	50.8	147	49.2	299	3.9
20 Llangollen	20	34.5	38	65.5	58	0.8
21 Llanrwst	17	70.8	7	29.2	24	0.3
22 Machynlleth	14	46.7	16	53.3	30	0.4
23 Merthyr Tydfil	75	61.0	48	39.0	123	1.6
24 Milford Haven	84	57.5	62	42.5	146	1.9
25 Monmouth	28	71.8	11	28.2	39	0.5
26 Neath	57	49.1	59	50.9	116	1.5
27 Newport	296	58.5	210	41.5	506	6.6
28 Newtown	71	48.6	75	51.4	146	1.9
29 Pembroke Dock	40	66.7	20	33.3	60	0.8
30 Pontypool	206	52.8	184	47.2	390	5.1
31 Pontypridd	234	57.6	172	42.4	406	5.3
32 Port Talbot	213	51.1	204	48.9	417	5.4
33 Pwllheli	61	62.9	36	37.1	97	1.3
34 Rhyl	82	70.1	35	29.9	117	1.5
35 Shotton	247	58.9	172	41.1	419	5.5
36 Swansea	309	52.6	279	47.4	588	7.7
37 Tenby	15	51.7	14	48.3	29	0.4
38 Tywyn	23	63.9	13	36.1	36	0.5
39 Welshpool	51	51.0	49	49.0	100	1.3
40 Wrexham	187	49.5	191	50.5	378	4.9
Total	4,316	56.4	3,337	43.6	7,653	99.9

Table 5 Characteristics of the Up-Dated EDB for Selected TTWAs in 1986

Surveyed TTWAs	Number of new independent establishments, 1979-1985 identified by direct survey in 1986		Number of 'other' establishments identified by direct survey in 1986		Total number of establishments identified by direct survey in 1986		Total number of establishments in EDB in 1986
	No	%	No	%	No	%	No
1 Aberystwyth	0	0.0	0	0.0	0	0.0	61
2 Blaenau Ffestiniog	4	20.0	0	0.0	4	20.0	20
3 Caernarfon	17	17.3	8	8.2	25	25.5	98
4 Denbigh	9	14.1	6	9.4	15	23.4	64
5 Holyhead	9	8.3	5	4.6	14	13.0	108
6 Lampeter	23	21.1	2	1.8	25	22.9	109
7 Llandrindod Wells	3	4.1	4	5.5	7	9.6	73
8 Llangollen	4	9.1	0	0.0	4	9.1	44
9 Llanrwst	3	14.3	1	4.8	4	19.0	21
10 Machynlleth	0	0.0	4	22.2	4	22.2	18
11 Milford Haven	6	5.6	8	7.5	14	13.1	107
12 Monmouth	0	0.0	3	16.7	3	16.7	18
13 Newtown	15	14.0	10	9.3	25	23.4	107
14 Pembroke Dock	7	20.0	3	8.6	10	28.6	35
15 Pontypridd	37	10.8	40	11.6	77	22.4	344
16 Rhyl	19	22.6	16	19.0	35	41.7	84
17 Shotton	46	15.6	46	15.6	92	31.2	295
18 Tywyn	5	17.9	2	7.1	7	25.0	28
19 Welshpool	1	1.4	5	7.0	6	8.5	71
20 Wrexham	32	11.2	32	11.2	64	22.4	286
Total	240	12.1	195	9.8	435	21.8	1,991

Table 6 **Surviving New Independent Manufacturing Firms, 1979-1985 Identified by Direct Survey in 1986**

Surveyed TTWAs	Number of new independent establishments, 1979-1985 identified by direct survey in 1986	Total number of new independent establishments, 1979-1985 in 1986	Number of new independent establishments, 1979-1985 as a percentage of all new independent establishments, 1979-1985 in 1986
1 Aberystwyth	0	7	0.0
2 Blaenau Ffestiniog	4	6	66.7
3 Caernarfon	17	20	85.0
4 Denbigh	9	12	75.0
5 Holyhead	9	15	60.0
6 Lampeter	23	31	74.2
7 Llandrindod Wells	3	5	60.0
8 Llangollen	4	7	57.1
9 Llanrwst	3	4	75.0
10 Machynlleth	0	0	0.0
11 Milford Haven	6	7	85.7
12 Monmouth	0	0	0.0
13 Newtown	15	20	75.0
14 Pembroke Dock	7	8	87.5
15 Pontypridd	37	50	74.0
16 Rhyl	19	21	90.5
17 Shotton	46	58	79.3
18 Tywyn	5	9	55.6
19 Welshpool	1	5	20.0
20 Wrexham	32	50	64.0
Total	240	335	71.6

Table 7 Spatial Variations in Response Rates for Surviving New Manufacturing Firms

Surveyed TTWAs	Total number of new firms, 1979- 1985	Surveyed new firms, 1979-1985	Percentage new firm survey response rate in 1986
1 Aberystwyth	7	4	57.1
2 Blaenau Ffestiniog	6	5	83.3
3 Caernarfon	20	17	85.0
4 Denbigh	12	9	75.0
5 Holyhead	15	13	86.7
6 Lampeter	31	27	87.1
7 Llandrindod Wells	5	3	60.0
8 Llangollen	7	7	100.0
9 Llanrwst	4	4	100.0
10 Machynlleth	0	0	0.0
11 Milford Haven	7	6	85.7
12 Monmouth	0	0	0.0
13 Newtown	20	16	80.0
14 Pembroke Dock	8	5	62.5
15 Pontypridd	50	40	80.0
16 Rhyl	21	19	90.5
17 Shotton	58	44	75.9
18 Tywyn	9	8	88.9
19 Welshpool	5	3	60.0
20 Wrexham	50	39	78.8
Total	335	269	80.3

Table 8 Industrial Variations In Response Rates for Surviving New Manufacturing Firms

Standard industrial category (1968)	Total number of new firms, 1979-1985	Surveyed new firms, 1979-1985	Percentage new firm survey response rate in 1986
3 Food, drink and tobacco	13	11	84.6
4 Coal and petroleum products	0	0	0.0
5 Chemicals and allied industries	11	9	81.8
6 Metal manufacture	1	1	100.0
7 Mechanical engineering	27	23	85.2
8 Instrument engineering	5	5	100.0
9 Electrical engineering	17	15	88.2
10 Shipbuilding and marine engineering	5	4	80.0
11 Vehicles	12	9	75.0
12 Metal goods n.e.s.	36	29	80.6
13 Textiles	17	13	76.5
14 Leather, leather goods and fur	1	1	100.0
15 Clothing and footwear	12	10	83.3
16 Bricks, pottery, glass, cement, etc	32	28	87.5
17 Timber, furniture, etc	84	66	78.6
18 Paper, printing and publishing	18	16	88.9
19 Other manufacturing industries	44	29	65.9
Total	335	269	80.3

References

- Aldrich, H., Kalleberg, A., Masden, P., and Cassell, J. (1989), "In pursuit of evidence: sampling procedures for locating new businesses". *Journal of Business Venturing*, 4, pp.367-386.
- Allen, G. C. (1961), *"The structure of industry in Britain"*. London: Longman.
- Birley, S. (1982), *"New enterprises: a start-up case book"*. London: Croom Helm.
- Birley, S. (1984), *"Finding the new firm"*. Boston: Proceedings of the 44th Academy of Management Conference.
- Business Statistics Office (1979 & 1983), *"Report on the Census of Production. Business Monitor PA 1002"*. London: HMSO.
- Coombes, M. G., and Openshaw, S. (1982), "The use and definition of travel-to-work areas in Great Britain: some comments". *Regional Studies*, 16, pp.141-149.
- Cooper, A. C. (1981), "Strategic management: new ventures and small business". *Long Range Planning*, 14, pp. 39-45.
- Firn, J. R. (1976), "Economic microdata analysis and urban-regional change: the experience of GURIE". In Swales, K. (ed.) *Establishment based research: conference proceedings*. University of Glasgow: Urban and Regional Studies Discussion Paper No. 22, pp.8-30.
- Foley, P. (1983), "A comparison of four industrial data sources". In Healey, M. J. (ed.) *Urban and regional industrial research: the changing UK data base*. Norwich: Geobooks, pp.91-104.
- Forster, J. V. (1981), "How to identify and locate new enterprise". In Thomas, K. (ed.) *Employment planning and the establishment of new enterprises*. City of Birmingham Polytechnic: Department of Planning and Landscape, Working Paper No. 11, pp.40-46.
- Frank, C. E. J., Miall, R. H. C., and Rees, R. D. (1984), "Issues in small firms research of relevance to policy making". *Regional Studies*, 18, pp.257-266.
- Gudgin, G. (1978), *"Industrial location processes and regional employment growth"*. Farnborough: Saxon House.
- Healey, M. J. (1983), (ed.) *"Urban and regional industrial research: the changing UK data base"*. Norwich: Geobooks.
- Hubbard, R. K. B., Lambert, E. L., and Skelton, R. J (1979), *"Constructing an employment data-bank: the building research establishment experience for Merseyside"*. University of Glasgow: Centre for Urban and Regional Research, Discussion Paper No.31.
- Johnson, P. S. (1986), *"New firms: an economic perspective"*. London: Allen & Unwin.
- Keeble, D. E., (1978), "Industrial Geography". *Progress in Human Geography*, 2, pp.318-323.
- Lloyd, P. E., and Dicken, P. (1968), "The data bank in regional studies of industry". *Town Planning Review*, 38, pp.304-316.
- Lloyd, P. E., and Dicken, P. (1982), *"Industrial change: local manufacturing firms in Manchester and Merseyside"*. London: Department of the Environment Inner Cities Directorate, Inner Cities Research Programme No. 6.
- Lloyd, P. E., and Mason, C. M. (1976), "Establishment-based data for the study on intra-urban and sub-regional industrial change: the Manchester Industrial data bank". In Swales, J. K. (ed.) *Establishment based research: conference proceedings*. University of Glasgow: Urban and Regional Studies Discussion Paper No. 22, pp.31-58.
- Mason, C. M. (1981), *"The creation of an establishment based databank for South Hampshire"*. University of Southampton: Department of Geography, Final Report, SSRC Grant 6796.

- Mason, C. M. (1983), "Some definitional problems in new firms research". *Area*, 15, pp.53-60.
- Mason, C. M. (1987), "The small firm sector". In Lever, W. F. (ed.) *Industrial change in the United Kingdom*. London: Longman, pp.123-148.
- Mason, C. M., and Harrison, R. T. (1985), "The geography of small firms in the UK: towards a research agenda". *Progress in Human Geography*, 19, pp.1-37.
- Moyes, A., and Heal, D. W. (1985), "*A survey of waste arising in Ceredigion*". Aberystwyth: Department of Geography.
- North, D., Leigh, R., and Gough, J. (1983), "Monitoring industrial change at the local level: some comments on methods and data sources". In Healey, M. J. (ed.) *Urban and regional industrial research: the changing UK data base*. Norwich: Geobooks, pp.111-129.
- Sprott, T. F. (1977), "Annual census of employment data". *BURSIA Newsletter*, 24, pp.13.
- Stanworth, J. and Curran, J. (1976), "Growth and the small firm". *Journal of Management Studies*, 13.
- Storey, D. J. (1982), "*Entrepreneurship and the new firm*". London: Croom Helm.
- Swales, J. C. (1976), (ed.) "*Establishment based research: conference proceedings*". University of Glasgow: Centre for Urban and Regional Research, Discussion Paper No. 22.
- Thompson, A. (1983), "The prospects for establishment - level databanks". In Healey, M. J. (ed.) *Urban and regional industrial research: the changing UK data base*. Norwich: Geobooks, pp. 65-89.
- Westhead, P. (1988), "*New manufacturing firm formation in the context of the economy of Wales*". University of Wales: Unpublished Ph. D. Dissertation.