

A Note on the Design and Structure of a Small Firms Relational Database*

Gavin C. Reid

ABSTRACT. The design of a database for analysing field-work data on small firms is reported upon. The 'flat file' relational database methodology is adopted. This facilitates the flexible handling of large volumes of both quantitative (numerical and categorical) and qualitative (textual) data. Manipulation of the database is illustrated using some small firms examples.

1. Introduction

The great economists from Adam Smith, through to Alfred Marshall and to Ronald Coase have engaged in field work. This paper is concerned with how to handle the large volumes of data generated by field work. Points of principle are illustrated by reference to a field work study of small business enterprises. Field work methods generate a great deal of information which can easily put the investigator on "informational overload". Here, I discuss how this can be avoided. First, I look at how this information may be stored in a flexible way and efficiently retrieved (database design). Then I look at how it may be manipulated in a rapid and consistent fashion, within an economically meaningful analytical framework (relational database analysis).

2. The data

Before turning to abstract principles of data handling, let me set the scene with my example. Initially all the data gathered in the field were directly recorded by pen on paper during inter-

views. There were three instruments: an administered questionnaire; a semi-structured interview and a re-interview questionnaire.¹ Seventy-three small firms (or 'sites') were investigated.

The data gathered (on pricing, competitive strategy, survival etc.) were recorded on upward of one hundred and seventy physical schedules, which occupied some twenty large box files. The types of data ranged from 'hard' numerical (e.g., profits) through 'soft' numerical (e.g., estimated market share), to categorical (e.g., ordering of actions taken when demand rises), to qualitative (e.g., views expressed on the enterprise culture). Broadly speaking, they were *numerical* or *textual*. In order to answer even the simplest questions about average values of employment, sales and profits etc. one is involved in much labour, with a fair likelihood of confusion and error arising. Considerably more complicated questions are often contemplated, beyond enquiring into averages, standard deviations etc. One might ask qualitative questions like what is the strategic segment favoured by those firms which in following a focus strategy, emphasise low cost and product differentiation.² Or one might ask quantitative questions like how market share, as a dependent variable, can be explained by independent variables like advertising, price flexibility, competitiveness and degree of product differentiation. For such questions, the performance of data analysis using physical files is out of the question, and, using conventional computer files rapidly becomes unmanageable.

This is true of both qualitative and quantitative data analysis. The time honoured technique of qualitative analysts, which involves laying out on the floor or carpet, pieces of paper with certain written evidence on them, and re-arranging them to display sets and relationships between them, will not work as the number of sites becomes at all

Final version accepted on December 15, 1991

Department of Economics
University of St. Andrews
St. Salvator's College
St. Andrews, Fife, KY16 9AL
Scotland

large, say more than ten. Also, the compact computer data files traditionally used by econometricians in quantitative work very often provide poor service. When one is dealing with hundreds of variables (e.g., over two hundred per firm, for the administered questionnaire), which themselves might be combined and modified or transformed to define many hundreds more, the number of files one might wish to manipulate, and to perform calculations upon using statistical and econometric software, rapidly runs into many thousands. In short, one runs into 'informational overload'.

What one needs in order to use this abundance of data efficiently, accurately and economically, and in a way which facilitates scientific discovery, is a database management system (DBMS). The design and structure of a database to facilitate work on the small firms data must clearly take into account both the sorts of tasks to be performed (e.g., cross-site qualitative analysis of interview probes from the semi-structured interviews; cross-section regression analysis using data generated by the administered questionnaire), and the way in which data were originally gathered (e.g., using the three instruments mentioned above).

With a computerised database in place, containing the data from the field work on small firms in a form which facilitates easy access, issues of the manipulation of the database, using its management system, become significant. A variety of general principles govern the design of databases, of which the most important for the small firms work presented in this paper is that of the "flat file".³ As this term suggests, the principle is rather similar to the time honoured one used by field workers who have few sites to investigate. They examine the properties of their sites by arranging and re-arranging sheets of paper on the floor, seeking novel patterns which can be represented by two-dimensional tables. Introducing this principle to database design implies that qualitative (and specifically textual) as well as quantitative data can go onto each flat file; and that there is no limit of floor space, metaphorically speaking, restricting to very few the number of files that may be arranged in this way. Databases organised along these lines are said to be *relational* (as opposed, for example, to *hierarchical*),⁴ as the flat files impose very little structure on the data. Thus

the investigator is left to explore and discover structures for himself, by manipulating flat files in an organised way (e.g., by an algebra, a logical calculus, or a language).

3. Design and structure of the database

It was stated that relational databases may be designed to have what some call a "flat file"⁵ structure, in which no, or as few as possible, relationships are imposed (e.g., by hierarchical organization) but data are simply a set of two-dimensional tables. It is particularly fitting to store data in this way for this study of small firms, because data were gathered by field work methods, and a basic tenet of this technique is that as little structure should be imposed upon the data as possible. This allows for an interplay between data analysis and theorising. In this way, one hopes to develop theories which are well 'grounded' in business reality.⁶

An important aspect of the relational database is that it breaks down the distinction between entities and attributes. The small firm is an organizational form (an entity) which has certain characteristics like sales, assets, and market share (all attributes). A set of attributes may nominally be attached to an entity, but that is not necessarily important for manipulation of the database. In cross-site analytical case studies, for example, one might be interested in the way in which cost leadership and product differentiation combine to form what Porter calls a focus strategy. The relative costs of competitors and their sources, and customers' purchasing criteria, are then instances of the way in which one would want to explore relationships between attributes without any necessary reference to specific entities (here, small firms).

Many varieties of sophisticated software are available for database construction, and the best impose few restrictions on implementation. Indeed, it would fly in the face of database philosophy to impose the form of implementation, for this would not encourage features like responsiveness to new requests and 'evolvability'.⁷ Thus each operational database is unique, even if the software used in constructing it is an industry standard. It is useful to develop two mental pictures of one's unique database. The first picture is of the

contents of the database and the way it is organized; and the second picture is of what can be done with the contents. Briefly, in the case of the small firms database, the contents were the qualitative and quantitative data obtained with the three instruments and the way in which they were handled was governed by a special kind of language.⁸ In this section, I shall confine comment to the contents of the database; to how it was designed and structured. In the next section, I shall look at how the contents of the database are manipulated.

The database was initially created in 1985 and then extended in 1988. It had the general form indicated in Table I.⁹ In 1985 the numerical and categorical data collected for each small firm using the administered questionnaire were stored in one record type (type 1 record). They were a coding of the replies recorded (by pen on paper) during field interviews with owner-managers.

Every small firm had one of these records. For a sub-set of the small firms there were in addition a set of agenda records (type 2 records) which contained the textual information of summary notes constructed by the investigator for each agenda item of a semi-structured interview carried out later in 1985. Each firm was given a unique identifier and an additional dummy firm, number 99, was used as a technical device to hold the agenda records containing the probes of the semi-structured interview.

In 1988 two further record types were added to the database, one containing numerical and categorical data (type 3 record) and the other containing textual data (type 4 records), from the

re-interview questionnaire. This indicates the importance of the property of 'evolvability'. The final form of the database involved seventy three firms each treated as a case, with up to four record types associated with each firm, two record types containing numerical and categorical data and two record types containing textual data.¹⁰

4. Manipulation of the database

I mentioned above the way in which a flat file approach to database construction minimized, by design, the extent to which structure was imposed on the data. The data are simply a set of tables. These tables may be manipulated by an algebra or a logical calculus. On the grounds of simplicity, a language which is as similar to English as possible is to be preferred. The best known of such languages is SQL (Structural Query Language). It is this language that was used for database retrievals and manipulations.

Suppose one is interested in small firms at the very lower end of the size distribution, for which employment is less than or equal to two persons. Such firms are typical of the membership of the National Federation of Self Employed and Small Businesses which has 50,000 U.K. members. Suppose, further, that one wants to know whether these very small firms take on trainees (and if so, how many). Also, within which industrial classification (SIC code) does each of these small firms lie? This sort of information is relevant to an analysis of the training function in an enterprise culture (e.g., the assimilation of the Training

TABLE I
Flat-file structure of small firms database

	Numerical and categorical data	Textual data
Administered questionnaire and semi-structured interview	<i>Type 1 record</i> Administered questionnaire, numerical and categorical data	<i>Type 2 records</i> Semi-structured interview textual data
Re-interview questionnaire	<i>Type 3 record</i> Re-interview questionnaire numerical and categorical data	<i>Type 4 records</i> Re-interview questionnaire textual data

Note: Regarding each firm as a "case", every "case" may have four record types.

Agency into Scottish Enterprise). Using the database language for a retrieval¹¹ reveals that twenty six firms (35%) employed two or less full time workers. The database also provides all the requested SICs and full information on trainees. Briefly, just four of these firms with less than three employees also employed trainees. That is just 5% of the total and only 15% of this sub-sample of very small firms. Two of these small firms were in manufactures and construction, and two were in transport and services. By contrast, 47% of firms employing more than two persons did also employ trainees, the great bulk (82%) of these being in manufactures and construction. Here, the average number of trainees was two. Thus the training function in this sample tends to be fostered mainly in the larger of the small firms in manufactures.

Other retrievals are possible. For example, one could look at the distribution of the small firms across industrial classifications. SICs of 60 or over relate to transport and services. Suppose one were interested in the number of small firms in transport and services that employ less than three people. There were twelve of these (46% of the sub-sample for which employment ≤ 2). Further, one could focus on more qualitative issues. Referring now to the re-interview questionnaire (type 4 records), one might ask whether the owner managers of the same firms had sensed the emergence of an enterprise culture in Scotland since 1985.¹² To be even more specific, what was the attitude to enterprise of such a small firm in, say, wholesale distribution (SIC = 61) compared to a small firm in, say, recreational services (SIC = 97)? Record type 1 of the database informs us that two firms were both in the relevant size range and operating in wholesale distribution in 1985. Record type 4 of the database informs us that one of these owner managers thought that the government had 'led the way' and that more people would 'have a go on their own'. He observed that older small firms had better subsequent survival rates than younger small firms. From record type 3 of the database we find that the other small firm in wholesale distribution (whose firm distributed cane and basket ware) had ceased trading by 1988 and its owner manager could not be traced for comment. There were also two firms in the sub-sample operating in recreational services in 1985. From record type 4 we find that one owner

manager, a tour operator, thought the successful setting up of many small firms selling tours in recent years was indicative of an emerging enterprise culture. The other small firm in recreational services, supplying kit-form acrobatic aircraft, had ceased trading by 1988 and the owner manager could not be traced for comment.¹³

Pursuing qualitative evidence further, one could extract text from the type 2 records which related to the probes of the semi-structured interview.¹⁴ Consider, as an example, Porter's category of *rivalry*. For the firm in wholesale distribution that survived, a database retrieval indicates that the owner manager gave a detailed account of the nature of rivalry. His small firm was a wholesale wine business. The owner manager had felt the increased pressure of competition mounting daily. A particular competitive pressure was felt from brewers who had diversified into wine sales because wine had become fashionable and beer sales were on the wane. It was felt that the profit margins of small firms in wine distribution were being squeezed by these new big players in the market. Brewers who had diversified into wine were thought to have had high strategic stakes and to be here to stay. For the firm in recreational services that survived, the holiday tour operator, a similarly full account of rivalry was given. The fact that so many small firms operated in this line of business was explained by low fixed costs (low entry barriers) and ample opportunities for specialization (niche exploitation). It was observed that there was as much specialization amongst tour operators as there were kinds of holidays. Markets were so highly fragmented that small firms could always find niches that did not overlap with other firms' markets. This clustering of products in holiday tours can be explained by the information externality analysed by Pepall (1990). Small firms can observe information about demand by looking at the market for similar products produced by rival firms. At low cost they can market new goods in those parts of characteristics space which are 'spanned' by the qualities of existing products (over characteristics like transport, meals, accommodation and entertainment).

Both the examples on rivalry are of firms that survived to 1988. One could go on to enquire why these two firms survived but their partner firms in the same SIC classifications did not. The failed

firm in wholesale distribution did not complete the semi-structured interview, but the failed firm in recreational services did. The owner manager in the latter case ran a small firm which provided "kit" aeroplanes for aerobatics.¹⁵ Looking at the agenda item on rivalry, one notes the owner-manager expressing doubts about his product and the market. It depended on the existence of a special type of customer who had "flying in his blood" and was "very self-indulgent". Investment was over a long period of time and during that time there was no return; only four kits could be constructed a year; and government regulations on components and construction were very strict. Clearly survival for this firm was not going to be easy. One can probe further turning from qualitative analysis to quantitative, using two good predictors of survival, the gearing ratio and cash flow. Survival predictions are based on a binary probit analysis of data from the administered and re-interview questionnaires, as described in Reid (1991). Higher gearing lowers the probability of survival, as does the existence of cash flow problems. Retrieving from record type 1 of the database, one discovers that of the two firms in recreational services both had experienced cash flow difficulties, but the one which failed (the kit aircraft firm) was much higher geared (250% compared to 75%) and arguably, therefore, more exposed to risk. Looking at the two firms in wholesale distribution, the one that survived (the wine distributor) had not experienced cash flow problems, whereas the one that failed (the cane and basket ware distributor) had. Further the one that survived had a zero gearing ratio (i.e., no debt at all) whereas the one that failed had a 400% gearing ratio.

5. Conclusion

This note has been largely concerned with what economists would regard as research design: with the systematic organization of the data obtained from the field-work instruments in a relational database. Diverse examples have been given which, because of the flexible database design, were all based on retrievals during the same session at a computer terminal. However, they were no more than illustrative. The important part of economic analysis is to move on from mere

illustration to more serious and systematic analysis. What this paper has aimed to display is a method which makes such analysis both more feasible, and in principle, more innovative. Feasibility is facilitated by speed and ease of handling of data without sacrifice of accuracy or consistency. Innovation is facilitated by the flat file structure, which avoids viewing data in a rigid way, and breaks down the traditional distinction between qualitative and quantitative data.

Notes

* Many persons have contributed to the larger project on which this paper is based. On the field work side, I should mention Lowell Jacobsen, Anne-Theresa Lawrie and Jacqueline Campbell. On the computing side, I should mention Sandra Rice and Julian Read. Christopher Corrie provided useful research assistance during my tenure of a Leverhulme Fellowship, when this work was accomplished. I remain responsible for such errors and deficiencies as this paper may contain.

¹ Further details of the instrumentation and method are in Reid (1987a, b) and Reid and Jacobsen (1988).

² Here I simply illustrate, using aspects of the framework of Porter (1985) which are employed in the semi-structured interviews.

³ For further detail on flat files and their use in relational databases, see the contribution by Lorie (1987) in Paradaens (1987, Ch. 2).

⁴ The traditional focus in database management has been on data structures which are hierarchical, network or relational in form. See Everest (1986, Ch. 4) for a detailed statement of this taxonomy, and criticisms thereof.

⁵ In Everest (1987, p. 149) the following definition is given: "A single flat file consists of a set of attributes which describes entities from a single entity class. Each attribute in a flat file structure contains only one value instance for each entity instance". To illustrate, the entity class in my case is the small business enterprise, and an entity is a specific small firm. The set of attributes which describes entities include sales, date of financial inception, profits, and the debt/equity ratio. There is only one value of sales, profit etc. for each small firm. A flat file could therefore be a cross-classification of profits and sales.

⁶ The classical analysis of theory which is well 'grounded' in reality, using field work methods, is Glaser and Strauss (1967).

⁷ See Everest (1986, pp. 43-45).

⁸ In this case, the SIR database software was used, which is manipulated by the language SQL (Structural Query Language). See Frank (1984, Ch. 13), Paradaens ed. (1987, Ch. 2) for the general features of SQL, and for extensive treatment of its implementations, the SIR/Light Manual (1988, Ch. 4), which is a compact manual of the proprietary database management system SIR/DBMS.

⁹ Database design and implementation were provided by

Julian Read and Sandra Rice, and data entries were undertaken by Lowell Jacobsen and Jacqueline Campbell.

¹⁰ Once fully installed the database contained 29,200 data items for the administered questionnaire; 2,635 data items for the semi-structured interview; and 3,519 data items for the re-interview questionnaire making 35,354 data items in all. Allowing a mean value of 4 characters per data item the total disk space used was 141.416 K bytes.

¹¹ The basic building block of the SQL language is the 'SELECT' command. For the example given of a retrieval from the type 1 record, the relevant command would be: SELECT firm sic employ trainno FROM 1 WHERE employ LE 2. Other retrievals mentioned later in the text were carried out in a similar way.

¹² In view of the textual nature of the data the DBMS facility of the SIR database software was used.

¹³ Owner managers of some firms that had not survived *did* recognise the emergence of an enterprise culture. Thus 78% of the sample as a whole recognised it, as against 75% of small firms surviving.

¹⁴ Again, given the textual nature of the agenda and probes data, retrieval was achieved through the DBMS facility of the SIR database software. 'Probes' are a device for guiding the respondent to providing more detail on a specific topic in a semi-structured interview.

¹⁵ Owner managers nominated the relevant SIC code themselves in the administered questionnaire, selecting from a full listing of SIC descriptors. This firm is probably mis-categorised, though I have retained the illustration for its intrinsic interest.

References

- Everest, G. C., 1986, *Database Management: Objectives, System Functions, and Administration*, New York: McGraw-Hill.
- Frank, L., 1984, *Database Theory and Practice*, Reading, MA: Addison-Wesley.
- Glaser, B. G. and A. L. Strauss, 1967, *The Discovery of Grounded Theory: Strategies for Qualitative Research*, New York: Aldine.
- Paradaens, J. (ed.), 1987, *Databases*, London: Academic Press.
- Pepall, Lyn, 1990, 'Market Demand and Product Clustering', *Economic Journal* **100**, 195–205.
- Porter, Michael, 1985, *Competitive Advantage*, New York: Free Press.
- Reid, Gavin C., 1987a, *Theories of Industrial Organization*, Oxford: Basil Blackwell.
- Reid, Gavin C., 1987b, 'Applying Field Research Techniques to the Business Enterprise', *International Journal of Social Economics* **14**, 3–25.
- Reid, Gavin C., 1991, 'Staying in Business', *International Journal of Industrial Organization* **9**(4), 545–556.
- Reid, Gavin C. and Lowell R. Jacobsen, 1988, *The Small Entrepreneurial Firm*, Aberdeen: Aberdeen University Press.
- SIR/Light Manual*, 1987, Deerfield, IL, USA: ISI.

Copyright of Small Business Economics is the property of Kluwer Academic Publishing / Business and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.