Business Perceptions of Confidentiality[1]

Nick Greenia
Statistics of Income Division
Internal Revenue Service

J. Bradford Jensen
Center for Economic Studies
U.S. Bureau of the Census

Julia Lane
American University,
The Urban Institute
and
LEHD program
U.S. Bureau of the Census

[1] This study would not have been possible without the extraordinary efforts put into the development of the survey by Diane Willimack and Kristin Stettler, U.S. Census Bureau. The authors would also like to thank Dr. Frederick Knickerbocker for facilitating this work, as well as providing extremely helpful comments, and Faye Schwartz for excellent research assistance. All views expressed represent those of the authors, and not necessarily the institutions they represent.

I. Background/Introduction

The core mandate of statistical agencies is to collect data on businesses and households. In so doing, each agency enters into an implicit pact with its respondents that the data will not only be used well, but will be protected from unauthorized access and use. This promise of confidentiality is not only a legal and ethical mandate, but also an important contributor to optimal data quality and response rates. However, little is known about the substance of this pact – particularly with respect to businesses. This ignorance can have serious potential consequences. If government misconstrues its role or the nature of the pact, and consequently businesses do not trust government to protect their data because they either mistrust or misunderstand the pact, it will be difficult for government not only to maintain high quality and timely response rates but also to frame new ideas on data collection, protection, and access.

This chapter reviews the current state of knowledge of business perceptions and presents the results of a survey of businesses perceptions of confidentiality protections. This is one of the first quantitative analyses of the sensitivity of different types of business data and businesses' assessment of the protection provided by different agencies – as well as their assessment of the quality of statistical work performed by the agencies. It also examines the knowledge of the business community of the financial and criminal penalties associated with breaches of confidentiality and its willingness to permit data to be shared among different federal and private agencies.

This information can be used not only for the maintenance and improvement of current collection systems but also for framing ideas on new data collection - such as data sharing initiatives and data dissemination and access systems. In each of these cases, business concerns are likely to differ from those of households. Businesses, unlike households, have multiple requests from different government agencies for information, so data sharing across agencies may be an attractive way of reducing response burden.

Similarly, businesses may be more aware of the importance of data quality and data access for their own research needs.

The study is particularly timely given the heightened interest in confidentiality and privacy issues. Media attention has increasingly focused on lapses in security in the private sector – for example, the private marketing of personal dossiers compiled from consumers' electronic sales records as well as the dissemination of medical records. The federal statistical community should do all that is possible to convey to the respondent community that they have addressed both real and perceived concerns.

The paper is organized into four core sections. The first of these describes and discusses the role of business confidentiality in statistical data collection. This is followed by a description of the questionnaire design and the sample frame. The next section provides the quantitative results, and we conclude with some preliminary suggestions for extensions of this research that can ultimately be used to affect data reporting, collection, protection, and access decisions.

II. **Business Confidentiality Issues**

"In general we need less government intrusion into business as well as personal life Still sensitive about

"Because I don't understand the ""system"" I am generally distrusting Especially of any governmnet agency"(Survey Respondent)

The lack of quantitative research does not mean that no attention has been paid to the confidentiality[2] of business data provided for federal statistical purposes. In 1992 , the U.S. Office of Management and Budget (OMB) established a working group that not only noted the differences between household and business perceptions of

[2] The protection of confidentiality in this chapter is defined as the restriction of access to information about the individual party/entity once it has been provided—for statistical or administrative purposes--to a second party charged with the collection responsibility. The confidentiality protection responsibility is traditionally viewed as residing with the collecting party, even when the law permits third and fourth parties to access the data. Indeed, the consequences of any breach of confidentiality would almost always be borne by the collecting party in the form of reduced response rates and less precise responses as the cost exacted for such violations

confidentiality[3], but also identified several factors that were likely to affect business trust in the protection afforded their data. These factors are more fully developed in a series of papers by Willimack and colleagues. This paper examines a subset of these factors – namely, the sensitivity of individual items queried; the perceived benefits of the data collection (e.g., survey objectives); the costs of data collection (e.g., survey completion time); and the protection provided respondents.

a) Sensitivity

Knowing what types of data businesses consider to be most sensitive might be used by statistical agencies to accord different levels of protection and permit broader access to, and analysis of, subsets of data. It is well known that different types of household data have different levels of sensitivity – item response levels on income measures vary substantially from those on age and number of children. Although no hard evidence exists, the sensitivity of business data is likely to be different for a number of reasons: the existence of publicly available information; the structure of business entities and the existence of competitors.

Businesses, unlike households, are often routinely required to provide information for administrative or regulatory purposes, some of which quickly becomes publicly available (see Willimack et al, 1999). For example, publicly traded companies are required annually to provide to the U.S. Securities and Exchange Commission extensive financial information, much of which also ends up on commercially available datasets such as Compustat. Further, all employers sponsoring an employee benefit plan (retirement, health, etc.) are required to file an annual Form 5500 series information return so that the Internal Revenue Service (IRS), Department of Labor (DOL) , Pension Benefits Guaranty Corporation (PBGC), and Social Security Administration (SSA) can administer their respective provisions of the 1974 Employee Retirement Income Security Act. Private corporations, such as ABI/INFORM, use information in the yellow pages to create business lists with name, address, employment, payroll, and industry information

[3] These are similar to the differences between collecting household and business data generally (see, eg, Box and Chiannapa, 1995.

on the universe of businesses. In sum, virtually all of the data provided for these various purposes are publicly available, raising the question of **whether** confidential datasets containing subsets of the provided information really need to—or even can--protect all of their information equally. It may well be that the overriding issue to businesses in such cases is to avoid further reporting burden, rather than with how to obtain maximal "confidentiality" protections for data already in the public domain.

Other items, not in the public domain, may not be sensitive simply because of the structure of business entities. For example, the taxpayer identification number, which is quite sensitive for individuals (the SSN or Social Security Number), may not be sensitive for firms, since the EIN or Employer Identification Number (assigned by the IRS) often appears on publicly available datasets and also may be changed several times over the business' lifetime. This difference makes businesses more elusive to track and monitor over time than individuals and hence does not enable instant access to complete lifetime data.

The inherently competitive nature of business is also something to be considered in analyzing the sensitivity of data. Many businesses may consider some information, such as name and mailing address, less sensitive than individuals and households, but other items necessary for profitable strategic planning, such as sales at the establishment level or trade secrets, very sensitive indeed.

Another dimension that might differ between household respondents and business respondent is the time sensitivity may well be different. While individuals may feel that their personal information such as medical records or earnings histories remain sensitive throughout their life, the nature of business respondents might mean that after some period of time the information is no longer sensitive. If particular types of data provide a competitive edge in a rapidly changing business world, then data are more than one or five years old may need much less protection than current data. Other types of data could provide information in legal cases regardless of the time frame, and may need

more protection. The survey questionnaire examines business sensitivities to both the types of information collected and the time period the data should be protected.

b)Perceived Benefits to Respondents

"Have no clue what the government collects and what they do with it- this information should be for private industry to use, collect, buy and sell- not a role for federal monies"(Survey Respondent)

"I would be willing to provide information to any agency if they could show how it would benefit the public or my company I am a Health Care provider; it costs me money to provide extra reports, but if they can use what we already have, it works!" (Survey Respondent)

One concern raised by the OMB working group, and supported by the research of Willimack et al (forthcoming), is that the direct benefits to businesses of the data that are collected from them by government statistical agencies may not always be readily apparent to them. In fact, the major producers of statistical data on businesses, such as the Census Bureau, the Bureau of Labor Statistics, and the Bureau of Economic Analysis have as their primary mandate to produce data on the economy for government policymakers (e.g., the Congress and the Administration), not for businesses. Although Census notes that economic census data can be used by businesses to, among other things, study their industry, gauge competition, calculate market share and study business markets, these data are not intended to be a substitute for data used in market research and competitive analyses carried out by private sector firms. A brief perusal of private company websites such as Dunn and Bradstreet, American Business Information, and the Donnelly Information Files provides convincing evidence that businesses can access and analyze detailed and quite current firm level information on competitors[4] as an alternative to aggregate, federal government data, albeit at considerable expense.. The survey explores business perceptions about the usefulness of government data products and the correlation between business perceptions about usefulness and confidentiality.

c) Respondent Burden and Data Sharing

The burden imposed on businesses by the federal government's requiring them to fill out surveys and censuses is a clear concern of both OMB and Congress (per the

[4] See, for example, http://www.mscnet.com/prodserv/nationaldatabases/index.htm

Paperwork Reduction Act of 1995), yet the full cost of such an imposition is neither easy to measure nor well known. Examples merely suggest the magnitude of the burden. To illustrate, the Census Bureau's Economics Directorate has estimated that the cost to business of filling out surveys in non economic census years is about 2 million hours; economic census years add an extra 5 million hour burden.

The burden associated with filling out surveys is evident from business response rates. Mail response rates of 1% to 2% can mean a highly successful mailing for some credit card offers. Market research surveys are usually much higher, but 10% to 15% response rates are common[5]. Our own anecdotal experience supports a general reluctance to provide any information . We had originally intended to explore the topic of confidentiality with the Chief Information Officers of the 200 largest US corporations, but the initial set of 15 phone calls revealed not one company willing to participate in a voluntary survey. In fact, almost all of these calls were unable to make it past the initial company screener. Federal statistical agencies, which typically are able to invoke mandatory compliance rules, experience much higher response rates, but there is evidence that American businesses are particularly unhappy about providing the same or similar information to different agencies (Nichols et al., 1999).

One means of reducing respondent burden is to share more data between administrative and statistical agencies. While data sharing does, to some degree, compromise the confidentiality of respondent data, businesses might be willing to trade less protection for lower respondent burden. The survey explores business perceptions on this issue.

d) Knowledge of protection provided respondents

"Have no clue what the government collects and what they do with it- this information should be for private industry to use, collect, buy and sell- not a role for federal monies"(Survey Respondent)

Although federal statistical agencies make much of the legal protections that are afforded respondents by statute and practice, it is not known whether businesses know about or value these protections. This lack of knowledge may be due to the decentralized

[5] See, for example, DSS research, http://www.dssresearch.com/

nature of the US statistical system, since data collection instruments from different agencies carry both different privacy and confidentiality protection statements or pledges[6]. Alternatively, it may reflect trust in the federal statistical system, which has a long history of data protection and no history of breaches of confidentiality.

This situation raises several questions. Do business respondents understand what is meant by confidentiality protection? Do they realize and understand the variation that exists across different collection agencies? An additional implicit question is how and even whether the federal statistical system should address—or even acknowledge--the authorized secondary disclosures that occur among the collection agencies[7]. Should the agencies spell out what data are re-disclosed, to whom and for what purpose? Should they re-affirm that these secondary disclosures are not to the public, and only for statistical, not regulatory purposes? How much "comprehensive" information should be provided to respondents to bolster their informed consent status, without adversely affecting response rates at the same time? Should the same data collected from different sources (administrative or survey) be treated differently from a confidentiality standpoint?

Suggesting answers to any of the foregoing questions is complicated by the absence of data about perceptions of the problem within the respondent community itself—a reasonable point of departure in any discussion of the subject. Confounding the issue are findings such as those by Singer et al. (1997) suggesting that the volume of information supplied in survey questions can itself adversely influence response rates. Obviously, such a result would seem to be in the ultimate interest of neither respondents nor policy makers, so some optimal combination of information and brevity (the right amount of the right sort of information) would appear to be the objective.

---

[6] See appendix A for the privacy and confidentiality pledges published by the federal statistical agencies, Census and BLS, as well as the federal income tax agency, IRS. Penalties for confidentiality violations (also known as unauthorized disclosures) are provided for the same three agencies in Appendix B.

[7] For example, universe extracts of records with limited item content are provided annually by IRS to Census primarily to reduce both respondent burden and collection costs, but with the secondary usage of providing an additional input source against which to check and verify some survey and even census information. While such redisclosures are statutorily authorized (USC Title 26 section 6103 j1A, Title 13,

III. Questionnaire Design and Data Collection

Although this brief discussion raises many questions, it is clear that there is no quantitative base on which to answer them. We developed a survey questionnaire[8] to begin to answer a subset of these questions, while recognizing that this represents only a first step towards the development of a much broader research agenda. We administered the survey through the mail using a commercially available business database (Dunn and Bradstreet) to designate the respondent population.

a*) Questionnaire design*

The questionnaire itself was designed to inform two discrete components of the confidentiality knowledge base:

1) What kinds of data/information do businesses consider sensitive – and for how long are they perceived to be sensitive?

2) What are businesses' perceptions of collection agencies' ability to collect and protect data? In particular, do businesses believe there are differences in the quality of data collection and protection across both private and public agencies, and are they aware of (and do they have confidence in the efficacy of) penalties for disclosure violations?

These are complex concepts – particularly given that most businesses may not think much about confidentiality issues (Nichols et al., 1999; Willamack et al., 2000). Because information is in itself known to influence perceptions, we decided to define

---

etc.) such arrangements are often not mentioned at all or else only briefly in the actual collection instruments used by these agencies.

[8] A copy of the survey instrument is provided in the appendix.

confidentiality only minimally in this first pass at data collection[9]. This strategy (minimally defining "confidential") also parallels research on household perceptions of confidentiality in order to identify the dimensions of confidentiality as understood by respondents.

The first set of questions deals directly with the question of whether businesses believe that different types of data have different levels of sensitivity – despite the fact that statistical agencies treat all these data with virtually the same level of protection. For example, we ask whether the business considers its primary identifiers--name, address, and phone number—to be sensitive. We then ask its views on the sensitivity of amount items such as employment, payroll, sales, profits, and tax liability. We expect, *a priori*, the former items (name, etc.) to be not as sensitive, since they are typically available in the phone book and are even advertised to promote the firm or its activities, but the latter set of items to be considered more sensitive. We also ask whether similar data are more sensitive at the establishment or company level.

Building on this foundation, we then probe whether businesses feel that there is a time dimension to the sensitivity of their data (again – statistical agencies typically treat business data as sensitive regardless of the age of the data). That is, are some items

[9] Nevertheless, the questionnaire does use the item responses themselves to begin to establish differences in perceived definitions. For example, item 7 seeks answers to core questions which frame business belief systems regarding confidentiality, including whether statistical collection agencies—release identifiable data to anyone, keep collected business data confidential, release any collected data outside government, and share collected data with other agencies.

considered sensitive longer than other items? We differentiate again between types of data and age of the data (1, 5, 10 and 30 years).

The next set of questions asks whether businesses are more or less concerned about the data collecting agent/recipient – whether they distinguish among federal regulatory or statistical agencies, not-for-profit researchers, for profit researchers, other businesses or the general public.

We then attempt to capture how businesses feel about the performance of federal statistical agencies, and later correlate this with their other responses. In particular, we ask whether the federal government is better than the private sector at collecting information, providing information, and protecting information. This is followed by a more detailed set of questions to find out how businesses perceive the protection provided by federal statistical agencies – whether their data are kept confidential, or whether their data are disclosed to or shared with other agencies. We also try to test our idea of absolute respondent cynicism regarding government trust by asking whether the respondent believes that any federal agency, including the IRS, can access data provided by businesses any time it wants. The analytical section will examine the interactions across the responses to these different items.

An important - hitherto unexplored – issue is whether businesses are aware of the legal and financial penalties imposed by federal agencies on employees who divulge confidential information without authorization. We address not only this contingency but

also whether these penalties should be increased, or more stringently enforced, and, if either, whether the business would be more inclined to provide confidential data as a result.

The last set of questions generally investigates whether the respondent might view responsible data sharing as a solution to some of their concerns about burden and invites respondents to identify any other concerns they may have about the collection or use of their data by federal statistical agencies.

b) *Data Collection and Response Rates*

The sample frame for the survey was obtained from Dunn and Bradstreet's commercial database, which represents 11,300,000 businesses in the U.S.[10]. This database has several advantages. It provided us with information on the name and title of up to four levels of management (e.g. the Owner, Chief Executive Officer, Chief Information Officer and Chief Financial Officer), information on industry, employment size and single/multi unit status, as well as being quite current (updated on a daily basis). Dunn and Bradstreet implements a number of checks to improve the quality of the data. In particular, they either make an on-site visit or a telephone investigation to each business in their database at least once a year . The average age of a record in D&B's U.S. business database is 7.5 months. The most obvious disadvantages are that the sample frame is not necessarily representative of all businesses in the U.S., and the quality of this sample frame relative to the major federal sampling frames in the U.S. (the Business

[10] This total is derived from the Dunn and Bradstreet report and differs from the size of the Census Bureau's Business Register, most likely due to differences in definition of a "business".

Register of the Bureau of Labor Statistics and the Business Register at the U.S. Census Bureau) is unknown.

From this sample frame, we selected a stratified random sample of 5000 cases; 1,250 in each of four strata defined by the number of employees – 0-49 employees; 50-249, 250 –499 and 500 plus. Of the 5,000 total businesses, 2,530 were multi-unit businesses with headquarters locations[11], and 2,470 were single unit companies with only one business. While the geographic coverage was national, and the sample was representative of Dunn and Bradstreet's database, we do not attempt to weight the sample to assign any kind of representativeness to the sample and simply view our results as preliminary evidence.

Two types of pilots were conducted. The first was a mailout of 25 questionnaires to a random subset of the survey sample. The second was a set of cognitive interviews conducted by survey methodologists from the U.S. Census Bureau on a subset of 8 out of 25 respondents who were interviewed by telephone about their understanding of the questions. These results were used to clarify and reformat some of the survey questions.

We employed the Standard Total Design Method (TDM) methodology (Dillman, 1978 and sent out the first wave of questionnaires by first class mail on November 28, 2000. A stamped, first class envelope was provided to the respondents. A follow-up reminder postcard was sent to non-respondents on December 8, and a second mailing of the questionnaire and cover letter was transmitted on December 28. .

Of the 5,000 questionnaires, 213 were returned as undeliverable, so that the response total of 509 questionnaires resulted in a response rate of just over 10%[12]. While

[11] We did not ask Dunn and Bradstreet to strip out subsidiaries from the file
[12] We did not separate out subsidiaries or non-headquarters responses

this response rate is quite reasonable for a private sector survey of businesses, the response rates for mandatory, government surveys often exceed 80%. In hindsight, we erred in not excluding subsidiaries or the non-headquarters units of multi-unit enterprises from our sample: the response rate for this group (929 survey units) was only 6%; that of other survey units was almost 12%.

We provide a more detailed analysis of the response rates by different types of business in Table 1. Response rates decline as business size increase - possibly because the survey did not reach the appropriate person, because large businesses are already burdened by numerous surveys or because small businesses are more interested in voicing their confidentiality issues than are large ones. The effect of firm size is quite substantial: even controlling for industry characteristics, the likelihood of response from firms with employment of 250-499 drops four percentage points; that of firms in the largest size class drops five percentage points.

There are some substantive differences in response rates across industries as well – businesses in the service sector were far more likely to respond than those in most of the other sectors – notably manufacturing. Again, even controlling for size of firm, the likelihood of firms in the manufacturing sector responding to the survey was more than four percentage points lower than in the service sector, and the same held true for firms in transportation, communications and public utilities, and retail trade.

Table 1: Response rates and sample counts by Industry and Employment Size class					
Major Industry	1-49	50-249	250-499	500+	Total
Agriculture, Mining and Contract Construction	10.10% 199	12.80% 109	7.60% 66	3.80% 52	9.60% 426
Manufacturing	8.20% 85	9.90% 232	8.00% 289	7.00% 355	8.10% 961
Transportation, Communication and Public Utilities	7.32% 41	16.95% 59	4.84% 62	13.40% 100	11.20% 262
Wholesale Trade	13.50% 74	13.70% 95	3.60% 84	11.50% 52	10.50% 305
Retail Trade	10.90% 221	6.50% 215	9.50% 137	5.80% 104	8.40% 677
Finance, Insurance and Real Estate	19.20% 73	9.20% 109	8.70% 126	8.10% 123	10.40% 431
Services	17.20% 459	14.70% 382	11.90% 444	8.90% 440	13.20% 1,725
Total	13.63% 1,152	11.66% 1,201	9.19% 1,208	8.26% 1,226	10.64% 4,787

Clearly, one of the early lessons learned from this relatively low response rate is that it is very difficult to conduct business surveys privately, even with government support, on such a sensitive topic as confidentiality issues.

IV. Survey Results

i) Sensitivity of Data Items

Many of our *a priori* notions were upheld, but with some surprises, with regard to types of data considered sensitive [13]. Entity information—such as name, address, etc.—was not considered very or extremely sensitive, except when the data pertained to the company's employees. This is not surprising, given that it is obvious that the information existed in the public doman (through Dunn and Bradstreet). Interestingly, of all the data

[13] There is little variation across industry classifications, so we do not continue to report these results. They are available from the authors upon request, however.

types queried in the survey, the employee identity data were considered by respondents to be the most sensitive, and the larger the company, the more sensitive these data were considered. One reason sensitivity increased with company size might be that such companies have legal departments more aware of the serious problems which can result from disclosing such information—such as workplace or domestic violence from estranged spouses.

Table 2: Proportion indicating that these data items are "Very or Extremely Sensitive" (Question 1)

Employment Size Class	1-49	50-249	250-499	500+	Total
Name, address and phone number of your business	9.20%	4.70%	1.00%	2.10%	4.60%
Type of industry/business operated	6.90%	3.20%	2.00%	5.10%	4.40%
Name, home address and phone number of employees	78.50%	86.40%	89.00%	87.80%	85.00%
Number of establishments	13.10%	7.10%	3.00%	4.10%	7.30%
Number of managers and executives	15.60%	14.30%	13.00%	19.40%	15.50%
Number of non-managerial employees	23.40%	14.30%	12.00%	14.30%	16.40%
Your company's total payroll	69.00%	66.70%	55.00%	53.60%	61.90%
Your company's total sales	59.20%	52.40%	40.00%	35.10%	47.90%
Your company's total operating costs	66.20%	67.50%	57.00%	52.00%	61.50%
Your company's total profits	73.80%	77.00%	67.00%	58.80%	70.00%
Your company's total tax liability	70.00%	77.80%	77.00%	64.30%	72.5%

Employment size data were also not considered sensitive, even when distinguished by categories of management and non-management. This result is not surprising, given the public availability of this information – indeed, Dunn and Bradstreet—from whom we obtained the sample frame data—provided employment data on the records we obtained for the survey sample.

Not too surprisingly, financial data, such as company payroll, operating costs, profits and tax liability were considered quite sensitive, with tax data lagging only employee entity information in perceived sensitivity. For multi-unit respondents, the types of data surveyed were generally considered slightly more sensitive at the company level than at the establishment level. Both payroll and profits data, which we already

knew were considered very sensitive by most businesses, were considered much more sensitive at the company level—almost by a factor of two--which, at least for publicly traded companies, which is surprising given that a lot of company level information is reported to other, publicly available, sources.

The variation in response across firms' employment size classes is worth noting because of the quite distinct views exhibited. Broadly speaking, smaller firms seemed to believe that financial information is much more sensitive than did larger firms, perhaps because their data are less likely to be publicly available, or because they are less likely to have been sampled in a government survey.

ii) Time Sensitivity of Data

Not all data need to be current in order to be useful. While some data users, such as policymakers, analysts and researchers, need access to entity level data in order to better understand the working of the economy or society, these data can be historical. However, some important data have no time limitation at all on data release – not even for business name, address, and employment size e.g. Federal tax data which constitute a significant portion of the Census Bureau's business register. In addition, the Census Bureau maintains the confidentiality protections on business data it collects permanently.

We examined businesses' perceptions of the time sensitivity of their data in question 4, and report the results in Table 3. Two general impressions stand out. First, the sensitivity of data is clearly affected by its age. Those businesses that responded had a wide variety of views spanning the spectrum of available time periods - and the views were clearly different depending on the data items concerned. Second, there is a great

deal of either indifference to or misinformation about the issue. Unlike the earlier questions, there was a substantial proportion of responding businesses that skipped this question – particularly when it came to profits data, where almost one in three either had no opinion or did not respond. However, since this phenomenon applied particularly to the company level questions, it is possible that single unit businesses felt they had already answered the question.

Turning to the details, the least sensitive data item was again the number of employees at both the establishment and the company level. Well over half of the respondents, and almost 80% of those with an opinion, believed that this item was no longer sensitive after 5 years. As we would expect, the most sensitive data element in terms of timing was profits – only 8% thought establishment level profits were not sensitive after one year; with 11% having the same view of company level profits. Almost one third of respondents, and over half of those having an opinion, thought that company and establishment level profits were sensitive for more than 10 years.

This time qualification of data sensitivity by respondents could be a useful avenue for future research, particularly if certain companies would permit expansion of researcher access to some of their financial data after a suitable time period.

Table 3: Time Period after which data are no longer sensitive (Question 4)

	After 1 year	After 5 years	After 10 years	After 30 years	Don't know, not applicable, no response
Number of employees at each establishment	27.03%	28.76%	10.42%	4.05%	29.73%
Establishment-level payroll	13.83%	39.46%	19.50%	12.47%	14.74%
Establishment-level sales	16.36%	37.05%	19.09%	11.59%	15.91%
Establishment-level profits	8.30%	27.80%	20.27%	14.86%	28.76%
Number of employees at the company	36.10%	23.17%	7.53%	5.79%	27.41%
Company-level payroll	12.77%	30.06%	17.88%	12.57%	26.72%
Company-level sales	18.66%	26.52%	16.11%	10.61%	28.09%
Company-level profits	11.20%	24.17%	21.22%	15.72%	27.70%

iii) Trust and Respect for Federal Agencies

"I have very little faith that the government can accurately gather such data and less that it can maintain it

"*As a highly regulated industry, we provide REAMS of data regularly to a variety of federal agencies- both operational and statistical We have never had a problem "(Survey Respondent)*

An important component of confidentiality is the trust that respondents have in the agency that collects the data. Since the U.S. statistical system is somewhat fragmented in nature, we asked respondents about their degree of trust in the statistical system in general, and specific types of agencies in particular. We also asked respondents how they felt about providing data to other entities, by way of a benchmark for comparison. Once again, the data were not silent, providing useful information about not only the level but also the differences in trust across agencies and the differences by size of firm.

In general, respondents seemed least concerned about providing the types of data to the core of federal statistical agencies—such as the Census Bureau and the Bureau of Labor Statistics. Respondents seemed most concerned about providing these data types to other businesses (presumably, their competitors) followed closely by commercial or for-profit researchers which they perhaps viewed as funded by their competitors or at least possibly supporting competitors' interests (especially if their competitors provided no data at all).

Table 4: Proportion of respondents who were very or extremely concerned about providing business data to the following entities (by employment size) (Question 5)					
Type of recipient	1-49	50-249	250-499	500+	Total
Federal Regulatory Authority (EPA, SEC, FTC)	48.40%	43.20%	25.00%	32.00%	38.20%
Federal Statistical Agency (BLS, Census, BEA)	34.10%	33.10%	20.00%	27.80%	29.30%
Not for profit researchers (universities, think-tanks, research organisations)	42.20%	44.80%	41.00%	37.50%	41.60%
For-profit researchers	60.90%	62.40%	68.70%	56.70%	62.10%
Other businesses	73.60%	76.00%	77.80%	67.70%	73.90%
The general public	70.50%	73.60%	66.00%	53.60%	66.70%

The level of trust is quite interesting. Even though the federal statistical agencies score the highest in relative trust, almost one third of respondents were very or extremely concerned about providing data to core statistical agencies such as the Census Bureau. At the same time, however, almost the same proportion (37%) reported feeling not at all concerned about this issue. The analysis by size class reveals that, while the concern is widespread, by and large small businesses seemed more concerned than large businesses.

While the responses to question 5 indicated that it matters who collects the data, the responses to question 6 were intended to help understand why. Instead, we get apparently conflicting answers. After having just reported that the core statistical agencies are most trusted as data collectors, respondents seem to change their minds by stating that the federal government is worse than the private sector at everything: collecting data, protecting the data's confidentiality, and converting the data into useful information. One interpretation of such apparently contradictory information is that the term "federal government" has attained demonized status with many companies in the private sector, which understand such a term as virtually synomynmous with anti-business. Future surveys might well substitute "federal statistical agencies" (such as

Census Bureau and the Bureau of Labor Statistics) for the term "federal government" to see if the same results hold.

Table 5: Proportion of respondent who somewhat or strongly agreed with the following statements (Question 6)

	1-49	50-249	250-499	500+	Total
The federal government is better than the private sector at collecting business data for statistical purposes	32.60%	23.80%	38.00%	33.30%	31.50%
The federal government is better than the private sector at providing useful information	38.80%	26.20%	31.30%	29.20%	31.60%
The federal government is better than the private sector at protecting business data from being released to those without authority to have it	35.90%	39.20%	42.00%	24.00%	35.60%

We expected there to be a link between business trust in protection and their concern about providing data. In order to verify this notion, we correlated the responses to question 5 – particularly the response to federal statistical agencies – with the responses to question 6[14]. Those businesses who seemed least concerned about providing data were in all cases significantly more likely to agree to the third statement in question 6 – that the federal government was better at protecting business data. This was independent of size of business and of industry classification.

iv. Respondent Burden and Data Sharing Concerns

The fragmented nature of the U.S. statistical system has generated a fundamental respondent burden problem in that multiple statistical agencies (e.g. Bureau of Labor Statistics and Bureau of the Census) request similar or identical data from businesses but these data are often not permitted to be shared. In questions 7 and 10, we directly asked businesses whether they knew about the extent and legality of datasharing across agencies, and the responses provided two very interesting results. First, respondents

seemed evenly split on whether they believe that federal statistical agencies keep the data confidential (the first and second rows of table 6). Second, they believe that datasharing already occurs (the third row in table 6), and in general, there is no problem with the core of federal statistical agencies sharing data among themselves. The clear implication is not only that respondents did not think such data sharing is a problem, they think it is currently occurring and probably authorized! Obviously, considerable caution has to attend an interpretation such as that rendered above, due to the small response rate for this survey.

Table 6: Proportion of businesses somewhat or strongly agreeing with the following statements (Question 7)

	1-49	50-249	250-499	500+	Total
I believe that federal statistical agencies keep data provided by businesses confidential	48.10%	69.80%	69.70%	60.80%	61.60%
I believe that federal statistical agencies do not release information by which a company or its data can be identified	41.10%	66.40%	65.70%	57.70%	57.10%
I believe that federal statistical agencies do not share data provided by businesses with other government agencies	21.10%	21.60%	27.60%	21.60%	22.80%
I believe that federal statistical agencies do not release data provided by business to people outside the government	28.30%	41.70%	43.30%	38.50%	37.60%
I believe that any federal agency, such as the Internal Revenue Service, Small Business Administration and Federal Trade Commission can access business data my company has provided to other federal agencies whenever it wants	71.10%	68.50%	67.30%	73.20%	70.00%

What is not in doubt is that most respondents were cynical in their views of the IRS - indeed any federal agency - when it comes to whether they can and do access any government data whenever they wish. This view of the world is particularly discouraging in view of the fact that federal statistical agencies are extremely careful to protect their data from even perceptions of IRS access, and because IRS is circumspect not only about accessing any other agency's data (and only through authorized channels and if they

[14] These detailed correlations are not reported in the text, but are available from the authors on request.

exist) but also about its own employees accessing tax data when they have no need to know[15]. Thus, it is probably an understatement to say that the education and public relations tasks before IRS, and to a lesser extent federal statistical agencies, which may be suggested by such survey responses are considerable to the point of being daunting.

v) Knowledge of Penalties

"*Training, discipline and top quality people management skills go farther than threats of fines and prison Hiring and retaining personnel with high morals is better than jails The Government is generally weak in effective supervision and developing objectives and holding personnel accountable for goals and accomplishment Many large bureaucracys are weak in this area" Survey respondent*

"I firmly believe all information should be highly confidential and sensitive at all times Large penalties should continue to be enforced" Survey respondent

One way of addressing concerns about disclosure is to provide steep fines and penalties for breaches of confidentiality. Indeed, Title 13 of the U.S. code, which governs the U.S. Census Bureau, provides for fines of up to $250,000 and prison time of up to 5 years if such a violation occurs. While federal administrators believe that this is both an important deterrent, and provides an important reassurance to the public, in Question 8 we directly ask businesses whether they know that such penalties exist. To our surprise, only 25% of respondents knew they existed; 55% did not know, and the balance was unsure.

[15] Annual mandatory briefings at IRS drill into employees that unauthorized browsing or access of tax information is both reason for dismissal and even criminal prosecution.

Table 7: Proportion of respondents somewhat or strongly agreeing with the following statements (Question 9)					
	1-49	50-249	250-499	500+	Total
Federal penalties for the release of business data by federal statistical agencies or employees to those without authority to have it are adequately enforced	28.10%	23.80%	20.20%	16.50%	22.70%
Higher penalties would make my company more willing to provide business data to federal statistical agencies	51.90%	50.80%	43.40%	43.30%	47.90%
A higher likelihood of conviction would make my company more willing to provide business data to federal statistical agencies	51.90%	58.70%	47.50%	53.60%	53.20%

Given the existence of penalties, however, very few –fewer than one in four - respondents believed that the penalties were enforced. Interestingly, small firms generally seemed more trusting in the efficacy of penalties than large firms. Although the link between enforcement and response rates has often been touted, our preliminary results suggest that this link is complex. While penalties and convictions would only reassure about half of the respondents, we established that almost 80% of respondents don't believe current penalties are enforced. Since the second question deals with higher penalities (not necessarily their enforcement) and the third question deals with conviction, which is part of enforcement, it is possible that people who are skeptical about current enforecement would continue to be skeptical about it for even more stringent penalties.

V. Summary

"I feel comfortable with how things currently work" (Survey Respondent)

We began by noting that statistical agencies have an implicit pact with their respondents. Both the quantitative results and the written comments that we have selected for inclusion in this analysis suggest that the current state of the pact is an uneasy one, in many, but not all, respects. While it would be ideal to have all respondents write in comments like the one that began this section, this comment is but one in a sea of disgruntlement. In other words, agencies have some work on their hands to convey the extent to which data are already protected. It is not clear what work needs to be done, however. Both the survey response rate as well as some inconsistency in the responses received demonstrate that while we now know more, we still know very little about

business perceptions of confidentiality. A full scale survey mounted by a federal statistical agency, with both its imprimatur and the advantage of its extensive sampling frame, is probably necessary to fill the knowledge gap. . Existing surveys and censuses would also make reasonable vehicles for these purposes, often to their own immediate benefit.

We also noted that agencies had a mission to disseminate data. This mission might be met with even more success than currently, without compromising confidentiality, if businesses were asked whether their less sensitive data might be released.to academics, policymakers and other researchers. Some business data items, such as company name, address, and employment size might be candidates for this, as might other data items old enough to be non-sensitive. By broadening access to more researchers and analysts, the entire statistical knowledge base could benefit from exposure to a richer institutional skill set. Additional benefits would include, of course, the reduction of both respondent burden and data collection costs.

Respondents also indicated that they distinguish between collection entities in the statistical data process—they do not regard the federal statistical community, commercial survey takers, non-profit researchers as homogenous. Thus, statistical agencies might want to consider the possibility of different degrees of confidentiality protection for different types of analysts. Moreover, based upon our survey responses, it seems that much work needs to occur in the federal statistical community to inform the respondent community of not only what is currently done, but its relationship to what is authorized by law, and what criminal penalties apply , as well as how they are enforced. Related efforts tailored to the needs of the user community are also probably advisable, but we will leave that task to other researchers/projects.

In sum, we hope that this chapter serves as a starting point for the development of a continuing dialogue among respondents, data collectors, researchers/analysts and policy/decision makers. Perception studies could be used to monitor response climate, routinized so that surveys/censuses and other data gathering constructs (including administrative data uses and time series linkages) are not undertaken without the underpinnings of an understanding of how these affect firm perceptions. The results from such studies could be used to demonstrate good faith efforts on part of stat community.

They could also facilitate new collection systems, particularly the collection and storage of, and access to, administrative data. Finally, such surveys could be used to evaluate the value and believability of accountability standards of statistical agencies.

References

Biemer, P.B. and Fecso, R.S. (1995), "Evaluating and Controlling Measurement Error in Business Surveys." Chapter 15 in B. Cox, et al. (eds.), Business Survey Methods. New York: Wiley

Christianson, A. and Tortora, R. (1995), "Issues in Surveying Businesses: An International Survey." Chapter 14 in B. Cox, et al. (eds.), Business Survey Methods. New York: Wiley.

Cox, B.G. and Chinnappa, B. N. (1995). "Unique Features of Business Surveys." Ch. 1 in B. Cox et al. *Business Survey Methods*. New York, Wiley

Dillman, D. A. 1978. *Mail and telephone surveys: The total design method*. New York: John Wiley & Sons

Edwards, W.S. and Cantor, D. (1991), "Toward a Response Model in Establishment Surveys." Chapter 12 in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S. (eds.), Measurement Errors in Surveys. New York: Wiley.

Mackie, C and N. Bradburns (eds) 2000 Improving Access to and Confidentiality of Research Data: Report of a Workshop National Research Council

Nichols, E., Willimack, D. K. and Sudman, S. 1999 "Balancing Confidentiality and Burden Concerns in Censuses and Surveys of Large Businesses". Paper presented to the Washington Statistical Society, U.S. Bureau of the Census, Washington, D.C., September.

Singer, Eleanor, S. Presser and J. Vanhoewyk 1999 "Public Attitudes Toward Data Sharing By Federal Agencies", in Record Linkage Techniques - 1997 Proceedings of an International Workshop and Exposition National Academy Press

Wallman, K and J. Coffey 1999 "Sharing Statistical Information for Statistical Purposes" in Record Linkage Techniques - 1997 Proceedings of an International Workshop and Exposition National Academy Press

Willimack, Diane K., Elizabeth Nichols and Seymour Sudman (forthcoming) "Understanding Unit and Item Nonresponse In Business Surveys," in Survey Nonresponse, Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little (editors), New York: John Wiley and Sons.

Willimack, Diane K, Seymour Sudman, Elizabeth Nichols and Thomas Mesenbourg 1999 "Cognitive Research on Large Company Reporting Practices: Preliminary Findings and implication for Data Collectors and Users" mimeo, the Census Bureau

Appendix

1. Census Bureau Confidentiality Pledges

Item 1: STANDARD CONFIDENTIALITY PLEDGE

The following confidentiality pledges are representative of the pledges used in business/employer-based censuses and surveys.

Mandatory - Census: Current for 2002 Economic Census

YOUR RESPONSE IS REQUIRED BY LAW. Title 13, United States Code, requires businesses and other organizations that receive this questionnaire to answer the questions and return the report to the U.S. Census Bureau. By the same law, YOUR CENSUS REPORT IS CONFIDENTIAL. It may be seen only by U.S. Census Bureau employees and may be used only for statistical purposes. Further, copies retained in respondents' files are immune from legal process.

Mandatory - Survey: SA-42A Annual Trade Survey 1999

NOTICE -- Response to this inquiry is required by law (title 13, U.S. Code). By the same law, your report to the Census Bureau is confidential. It may be seen only by sworn Census employees and may be used only for statistical purposes. The law also provides that copies retained in your files are immune from legal process.

Voluntary - Survey: Advance Monthly Retail Trade Report

NOTICE - Your report to the Census Bureau is confidential by law (title 13, U.S. Code). It may be seen only by sworn Census employees and may be used only for statistical purposes from which no firm or establishment may be identified. The law also provides that copies retained in your files are immune from legal process.

Voluntary - Survey: Business and Professional Classification Report

NOTICE -- Your report to the Census Bureau is confidential by law (Title 13, United States Code). It may be seen only by sworn Census employees and may be used only for statistical purposes from which no firm or establishment may be identified. The law also provides that copies retained in your files are immune from legal process.

Item 2: Expiration Period Title 44, United States Code (U.S.C.), which includes the National Archives Act (the "Archives Act"), governs the ultimate disposition of personal and business information. Under the Archives Act, the National Archives and Records Administration (NARA) determines what Government records have permanent value. The NARA is obliged to preserve records that have such value. It has the authority to take legal possession of such records after a specified period of time. A 1952 interagency agreement between the Census Bureau and the NARA established the 72-year rule, which covers population census and survey records. In 1978, Congress recognized this

agreement through an amendment to Title 44, which incorporates the 1952 Agreement by reference, thus providing for the 72-year protection unless the two agencies mutually agree on some other arrangement.

Specifically, the retention periods for Title 13 permanent records are reflected in Title 44 U.S.C., Section 2108(a) and (b).

Title 44, Section 2108(b) states

"With regard to the census and survey records of the Bureau of the Census containing data identifying individuals enumerated in population censuses, any release pursuant to this section of such identifying information contained in such records shall be made by the Archivist pursuant to the specifications and agreements set forth in the exchange of correspondence on or about the date of October 10, 1952, between the Director of the Bureau of the Census and the Archivist of the United States, together with all amendments thereto, now or hereafter entered into between the Director of the Bureau of the Census and the Archivist of the United States. Such amendments, if any, shall be published in the Register."

Section 2108(a) of Title 44 permits the Archivist to withhold other Government records from examination or use for thirty years if they are subject to statutory restrictions. Data collected by the Census Bureau in the various economic censuses and surveys are also classified as confidential by Title 13.

Willimack, Diane K, Seymour Sudman, Elizabeth Nichols and Thomas Mesenbourg 1999 “Cognitive Research on Large Company Reporting Practices: Preliminary Findings and implication for Data Collectors and Users” mimeo, the Census Bureau

2. Survey Instrument

Form CQM-46 OMB No. 0670-0760

Businesses' Perceptions of Confidentiality

The Urban Institute

A Nonpartisan Economic and Social Policy Research Organization

1. Some types of business data may be considered more sensitive than other types of data. The sensitivity of data may be related to its strategic, legal or security importance to the business or to whether it is released to those without authority to have it.

   For your company, please indicate the sensitivity for each of the following types of data. *Circle one response for each line. DK = Don't Know. NA = Not Applicable.*

	Not at all Sensitive	Somewhat Sensitive	Very Sensitive	Extremely Sensitive	
a. Name, address and phone number of your company	1	2	3	4	DK/NA
b. Type of industry/business operated	1	2	3	4	DK/NA
c. Name, home address and phone number of employees	1	2	3	4	DK/NA
d. Number of establishments (locations/ stores/plants) operated by your company	1	2	3	4	DK/NA
e. Number of managers and executives	1	2	3	4	DK/NA
f. Number of non-managerial employees	1	2	3	4	DK/NA
g. Your company's total payroll	1	2	3	4	DK/NA
h. Your company's total sales	1	2	3	4	DK/NA
i. Your company's total operating costs	1	2	3	4	DK/NA
j. Your company's total profits	1	2	3	4	DK/NA
k. Your company's total tax liability	1	2	3	4	DK/NA

2. Does your company have more than one location/store/plant? *Mark one.*

   ☐ Yes, multiple locations/stores/plants - continue with Item 3.

   ☐ No, only one location/store/plant - GO TO Item 4 on next page.

3. For each of the following types of data, please indicate which level of data is more sensitive: establishment/location-level data or company-level data. *Circle one response for each line. If your company has only 1 location/store/plant, please go to Item 4 on next page.*

	Establishment/Location-Level Data	Company-Level Data	
a. Number of employees	1	2	DK/NA
b. Payroll	1	2	DK/NA
c. Sales	1	2	DK/NA
d. Profits	1	2	DK/NA

4. For each of the following types of data, please indicate the time period after which they are no longer sensitive to the operation of your business. *Circle one response for each line.*

		After 1 year	After 5 years	After 10 years	After 30 years	
a.	Number of employees at each establishment	1	2	3	4	DK/NA
b.	Establishment-level payroll	1	2	3	4	DK/NA
c.	Establishment-level sales	1	2	3	4	DK/NA
d.	Establishment-level profits	1	2	3	4	DK/NA
e.	Number of employees at the company	1	2	3	4	DK/NA
f.	Company-level payroll	1	2	3	4	DK/NA
g.	Company-level sales	1	2	3	4	DK/NA
h.	Company-level profits	1	2	3	4	DK/NA

5. In your opinion, how concerned is your company about providing business data (e.g., number of employees, payroll, sales, profits, etc.) to each of the following? *Circle one response for each line.*

	Not at all Concerned	Somewhat Concerned	Very Concerned	Extremely Concerned	
a. Federal regulatory authorities (e.g., Environmental Protection Agency, Securities and Exchange Commission and Federal Trade Commission)	1	2	3	4	DK/NA
b. Federal statistical agencies (e.g., Census Bureau, Bureau of Labor Statistics and Bureau of Economic Analysis)	1	2	3	4	DK/NA
c. Not-for-profit researchers (e.g., universities, think tanks and research organizations)	1	2	3	4	DK/NA
d. For-profit researchers (e.g., market researchers and consulting firms)	1	2	3	4	DK/NA
e. Other businesses	1	2	3	4	DK/NA
f. The general public	1	2	3	4	DK/NA

6. Business data are collected for statistical purposes by both the federal government and the private sector (e.g., universities, think tanks, research organizations, market researchers, consulting firms, etc.). For each statement below, please indicate if you strongly agree, somewhat agree, somewhat disagree or strongly disagree. *Circle one response for each line.*

	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	
a. The federal government is better than the private sector at collecting business data for statistical purposes.	1	2	3	4	DK/NA
b. The federal government is better than the private sector at providing useful information to government policy makers.	1	2	3	4	DK/NA
c. The federal government is better than the private sector at protecting business data from being released to those without authority to have it.	1	2	3	4	DK/NA

7. For each statement below, please indicate if you strongly agree, somewhat agree, somewhat disagree or strongly disagree. *Circle one for each line.*

	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	
a. I believe that federal statistical agencies (e.g., Census Bureau, Bureau of Labor Statistics and Bureau of Economic Analysis) keep data provided by businesses confidential.	1	2	3	4	DK/NA
b. I believe that federal statistical agencies do not release information by which a company or its data can be identified.	1	2	3	4	DK/NA
c. I believe that federal statistical agencies do not share data provided by businesses with other government agencies.	1	2	3	4	DK/NA
d. I believe that federal statistical agencies do not release data provided by businesses to people outside the government.	1	2	3	4	DK/NA
e. I believe that any federal agency, such as the Internal Revenue Service, Small Business Administration and Federal Trade Commission, can access business data my company has provided to other federal agencies whenever it wants.	1	2	3	4	DK/NA

8. Currently, federal regulations do provide penalties, such as fines and prison time, for the release of business data by federal statistical agencies or employees to those without authority to have it.

   Were you aware that these penalties existed? *Mark one.*

   ☐ Yes

   ☐ No

   ☐ DK

9. For each statement below, please indicate if you strongly agree, somewhat agree, somewhat disagree or strongly disagree. *Circle one for each line.*

	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	
a. Federal penalties for the release of business data by federal statistical agencies or employees to those without authority to have it are adequately enforced.	1	2	3	4	DK/NA
b. Higher penalties would make my company more willing to provide business data to federal statistical agencies.	1	2	3	4	DK/NA
c. A higher likelihood of conviction would make my company more willing to provide business data to federal statistical agencies.	1	2	3	4	DK/NA

8. Currently, federal regulations do provide penalties, such as fines and prison time, for the release of business data by federal statistical agencies or employees to those without authority to have it.

   Were you aware that these penalties existed? *Mark one.*

   ☐ Yes

   ☐ No

   ☐ DK

9. For each statement below, please indicate if you strongly agree, somewhat agree, somewhat disagree or strongly disagree. *Circle one for each line.*

	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree	
a. Federal penalties for the release of business data by federal statistical agencies or employees to those without authority to have it are adequately enforced.	1	2	3	4	DK/NA
b. Higher penalties would make my company more willing to provide business data to federal statistical agencies.	1	2	3	4	DK/NA
c. A higher likelihood of conviction would make my company more willing to provide business data to federal statistical agencies.	1	2	3	4	DK/NA

10. Currently, except as authorized by law, federal statistical agencies, such as the Census Bureau, Bureau of Labor Statistics and Bureau of Economic Analysis, cannot share the business data they collect with any other government agencies. If data sharing were legal, among which federal statistical agencies would your company be willing to allow its data to be shared, for statistical purposes only? *Please circle one response option for each line.*

a.	Census Bureau	Yes	No	DK/NA
b.	Bureau of Labor Statistics	Yes	No	DK/NA
c.	Bureau of Economic Analysis	Yes	No	DK/NA
d.	Statistics of Income Division of Internal Revenue Service	Yes	No	DK/NA
e.	Other federal agencies			
	Please specify: _____	Yes	No	DK/NA
	_____	Yes	No	DK/NA
	_____	Yes	No	DK/NA

11. Please use this space to provide any additional comments you may have about the sensitivity of company data, confidentiality pledges, federal statistical agency use and protection of business data, or related topics.

3 : Quality of Dunn and Bradstreet Database (from Dunn and Bradstreet)

Dun & Bradstreet's Worldbase database is a global marketing database that contains over 53 million business records, including linkage, in 200 countries. Our U.S. marketing database includes 11,300,000 business establishments in the 50 United States plus the territories of Puerto Rico and the Virgin Islands.

Frequency of master database update

Under D&B's Full File Maintenance Strategy, all businesses in the U.S. marketing database are investigated either via a site visit or a telephone investigation at least once a year with many records being touched much more frequently -- as often as several times a month -- through various triggered maintenance programs. The average age of a record in D&B's U.S. business database is 7.5 months.

Procedure for database update

To maintain the dynamic nature of marketing information, D&B utilizes a fully integrated approach towards data maintenance:

D&B's Full File Maintenance Strategy is a planned and integrated full file data maintenance process where all businesses in the U.S. marketing database are investigated either via a site visit or a telephone investigation at least once a year. D&B conducts these scheduled updates. Professionals located in our 50+ field offices around the country or by multiple touch calling from one of our four Telecenter locations.

The D&B Telecenters leverage the latest in predictive dialing technology that links the D&B investigative associates into a fully self-contained scripted calling environment. The Telecenter calling environment contains important validations and cross-checks designed to validate the information the Communicator is gathering while they are still on the phone with the information customer. These efficiencies provide D&B with a capacity to call over 1 million records per month in our Telecenters.

Any updates or additional information gathered during our conversations is automatically fed into daily database update systems. Before any information is updated, those records are subjected to over 2,000 edits and validations including duplicate screening. These validations are in place to ensure our information customers receive only the highest quality information possible.

In addition to the Full File Maintenance Strategy, maintenance investigations can be triggered by many other internal and external information sources such as:

- Private Third Party Data Files
- Public Record Data
- Customer Files
- Newsworthy Services
- Internal Telephone / Mail Activities
- National Change of Address
- Address Standardization Processes

Additional information on these and other triggered programs is as follows:

⇨ Newsworthy Services Group - The group is responsible for monitoring and collecting information from news sources as well as conducting the investigations of these changes before entering them into the database. Changes can consist of business closings, business name change, address change, mergers/acquisitions, etc.

⇨ National Change of Address - Based on information supplied by postal service files, a D&B record will either be excluded from the marketing file after it has been telephone verified that the company is out of business, or an address change will be made. Unlike many other companies, D&B also takes the NCOA "soft-matches" or records the Postal Service is unsure of and telephone investigates to determine if changes have occurred.

⇨ Address Standardization Updates - Address standardization software is used on a daily input as well as during monthly full database sweeps to ensure our addresses are standardized and to apply bar-code and Zip+4 information.
⇨ NewsHound - This is a program where D&B associates report new businesses or changes to existing businesses via fax or phone to a central update location. These changes are then investigated before they are entered into the database.
⇨ Top 1,000 Review - The top 1,000 companies in the information base by sales and employees are manually and machine reviewed for inconsistencies. Updates are applied as needed.
⇨ Linkage Updates - All "family trees" are updated on a 12 month cycle using a "top-down" confirmation of the headquarters locations. Additionally, all branch locations are investigated in the same 12 month cycle using a "bottom-up" confirmation of the marketing information specific to that branch location and the linkage to its headquarters is checked.
⇨ On-Demand Programs - In addition to all of the scheduled maintenance programs, D&B recognizes some changes which occur to businesses that need to be updated in the information base right away. For example, business moves, mergers, bankruptcies, or CEO changes which D&B learns of will trigger an immediate investigation by a D&B analyst.

Formal process

Prior to adding records to the database, investigations are conducted at either one of our 50+ field offices or at one of our four Telecenters where our Communicators work using state of the art technology that includes a scripted calling environment. These investigations serve to confirm information as well as gather additional information such as contact names, number of employees, year started, linkage data, tradestyles, additional address information, home based status, just to name a few. Investigations are conducted during normal business hours as well as evening and Saturday calling.

Each record is then subjected to over 2,000 edits and validations including screening for duplicate information against the entire database prior to inclusion in the marketing database. A team of associates is then responsible for resolving all error conditions via automated process flows and/or reinvestigation. The new business listings are then entered into our database daily using fully automated systems.

Source of data elements

Dun & Bradstreet leverages a variety of sources for new record identification, including:

⇨ Directly from the business principal during investigation.
⇨ Response to inquiries from D&B credit customers.
⇨ Customer Files – D&B partners with many of its' customers to identify new businesses and add them to the marketing file after investigating and enhancing the information.
⇨ Personalized Investigations – These new businesses enter our database as a result of customer requests for specific investigations.
⇨ Public Record Sources (Local, State and Federal) - Including New Business Registrations, Corporate Charter Details and Public Bulk Source Files.
⇨ D&B Bounty Program – This program utilizes our associates in over 50+ field offices located throughout the country by motivating them to proactively identify new business start-ups in their area.
⇨ Private Third Party sources – Many of these files are acquired via niche specific compilers of information.
⇨ Business Directories

Please see Attachment A for source information on the specific data elements D&B is recommending for your company.

Quality of data

Marketers require the most-up-to-date information in targeting their markets and reaching their most likely prospects, in order for their marketing campaign to be cost-effective. D&B has numerous quality checks in place to help ensure this. Some of the programs include:

- ⇨ Quality Review Telecenter Output files
- ⇨ A system of over 2,000 Edits and Validations, which occur between our master database (AOS), and the marketing database at the data element level. These edits and validations are cross-relational: For example, they compare the listed city, state and zip code to ensure valid address information and they compare sales, number of employees and SIC to identify businesses with abnormally high figures for a given industry.
  - ⇨ Address Standardization is performed by using a software program that we receive from Group One called Code 1 Plus. This software program is run on a monthly basis to validate and correct 5-digit zip codes, as well as assign Zip+4 extensions. Address validation is also performed along with city correction and delivery point assignments for bar-coding.
  - ⇨ A variety of data cosmetic procedures are implemented to synchronize upper/lower casing, invert business names into standard format (appear as John Smith Industries in the marketing database), and CEO name parsing for genderization and letter personalization.
  - ⇨ Monthly audits are performed since most of the edits and validations occur on an individual record-by-record basis, a monthly audit and standard report are created. Frequency counts, distribution analyses, and profiles are produced and compared to the previous month's results. Unusual occurrences or abnormally high or low frequencies are identified for investigation and correction.
  - ⇨ Accuracy Measurements of the marketing database are performed on a quarterly basis through audits. These are performed by using statistically valid sampling techniques to extract marketing database records, which are then mailed and telephoned, to achieve accuracy levels of plus or minus 1%. These results are then further analyzed to identify segments of the file that may need additional attention.
- ⇨ Customer Feedback, we are eager to receive feedback from our customers on ways to improve the quality and content of the marketing database. This feedback is often incorporated into change activity targeting the improvement of procedures, processes and overall data integrity.
- ⇨ Area Code changes in the marketing database begin with a monthly receipt of data from BellCore, the company that manages the North American phone numbering system. Approximately 30 days before the split is to occur, all of the exchanges moving to the new area code are pulled from the BellCore file. These exchanges are put on a tape, which is processed at the National Business Information Center (NBIC), effectively changing all of the records in our database (AOS). This is performed on the same weekend as the split actually occurs. The marketing database is then fed these changes from AOS and they are included in the subsequent cut of the marketing database. The timing of these splits is the determining factor for when the new area codes will appear in the marketing database. For most splits, the next update of the marketing database will contain the new updated area codes. In addition, at the end of the permissive dialing period for a particular split, another sweep is performed by NBIC to change any records that do not reflect the new area code and validation tables are updated in the data entry system to check the new area code and exchange combination.
  - ⇨ Each month the United States Post Office makes changes to existing ZIP Codes, creates new ZIP Codes or discontinues existing ZIP Codes. The marketing database is swept each month to update the ZIP Codes on our records with the USPS changes.
  - ⇨ D&B utilizes a product from Group One Software called Code One Plus to standardize street addresses and update ZIP Code changes within the database. Group One receives the postal changes from the USPS each month. Group One then formats the new information to be compatible with the Code One Plus software package and sends us monthly updates. Before the marketing database is cut each month, the software is used to update the ZIP Codes.