
Challenges and Opportunities in Administrative Records

Gordon E. Priest, Statistics Canada

In January 1995, the Conference of European Statisticians (CES) held a work session on Registers and Administrative Records. Participants represented 24 countries, as well as Eurostat and the International Labour Office. Over 30 papers were presented, which provided both status reports on activities of statistical agencies and discussions of issues related to the derivation of data from administrative sources.¹

As a result of the meeting, Eurostat and the Secretariat of the CES agreed to publish a summary report of the status of the work of statistical agencies with respect to key issues, such as frequency of updating, centralized versus non-centralized sources, statistical uses, type of identification used, legal constraints, quality problems, and use of identification numbers and coverage.

The participants concluded that demands to improve the timeliness of data, to reduce costs, and to reduce respondent burden were all stimuli that would precipitate increased activity in the derivation of statistical information from administrative records. They noted that legal constraints, as well as public concerns about privacy and confidentiality, were a hindrance to more effective exploitation of such sources. They also noted, however, that public perceptions and fears were often ill-founded, given existing legislation related to confidentiality and statistical applications.

The group further concluded that work needed to be done to promote the development of standards in administrative records. Integration has been difficult because of the lack of standards between sources and the resultant conceptual and definitional disharmonies.

Confidentiality was cited as another area that needed to be addressed in terms of providing safeguards and assurances to the public with regard to the legitimacy of the statistical mining of administrative records.² While other issues, such as quality and coverage, were also discussed, it was apparent that the matter of standards

on the one hand and of confidentiality on the other is critical to the fuller exploitation of such sources.

■ Challenges

There are compelling reasons to address the matter of the exploitation (or mining) of administrative records, not the least of which is the matter of both past and anticipated cuts to the funding of all government programs, including statistical services. Statistical agencies must simply become more efficient and more effective in the exploitation of their data sources. Not only must agencies improve their data management practices through improved documentation, improved harmonization, and improved integration of existing sources, but they must also seek lower cost collection alternatives.

The current mood of both business and the general population suggests that we need less government interference. Business must become more competitive and cost-conscious to survive. Households face stress and a time crunch as they juggle family responsibilities with work activities. Increasingly, not only are both spouses in the labor force, but they may also hold more than one part-time job. As a result, statistical agencies must be ever more mindful of reducing respondent burden on firms, households, and individuals and seeking alternative sources, such as those offered through administrative records.

Ironically, forces precipitating debt reduction, downsizing, and cost effectiveness in government also impact upon the demand for data. While budgets to statistical agencies are likely to continue to erode, it may be expected that demands for their services will continue to escalate. An increasingly global and competitive economy, economic and social restructuring, and scarcer resources all conspire to elevate the demand for timely and accurate statistical information because the implications of making uninformed decisions now carry much higher risks. In effect, statistical agencies must

strive to meet rising demands with diminishing budgets.

Furthermore, the client base of statistical agencies has both broadened and become more sophisticated. The introduction of the personal computer, the networking of those computers, and, indeed, the Internet itself have all contributed to the significantly enhanced numeracy of the population at large. Once the domain of government, big business, and academia, statistical use and analysis have spread to a burgeoning consultancy of hired guns ready to assist clients, particularly those in small businesses and not-for-profit organizations, in reaching informed decisions. The use of statistical information has even extended to private citizens who may be merely trivia junkies or curiosity seekers but are more likely to be legitimate users of information. On a related note, to the degree that a statistical agency has the mandate to ensure that the democracy within which it operates has an informed electorate, there is a growing need to provide the media with accurate, objective, and easily understood information, which illuminates the issues of the day.

■ Opportunities

While the above-noted factors present a challenge for action, other factors are what we might refer to as enabling and, therefore, offer opportunities. Significant developments in information technology have brought us not only the networked personal computer but also new, powerful, and user-friendly software that facilitates the mining and sharing of data bases. Much has recently been written about reengineering and data warehousing. The linkage and searching of extensive and diverse data bases are easily attainable. High-speed communications also permit both the browsing of off-site bases and the movement of data between them. This technology is ideally suited to the integration and exploitation of administrative sources. At the same time, we have made methodological advances, such as the extension of sampling theory from censuses and surveys to administrative records.

Thus, the environment of today presents many challenges and opportunities, which suggest that the time is right to more actively pursue the matter of mining ad-

ministrative records. Tapscott and Caston have written in their book, *Paradigm Shift: The New Promise of Information Technology*,³ that “the isolated technology applications of an earlier time are no longer adequate. Companies are discovering that they have to establish enterprise capabilities that will create new opportunities for sharing and reusing information and information technology at all levels. More and more organizations are becoming aware that the technical and structural barriers that have previously prevented or hindered internal communication and the sharing of resources must be dismantled. There is a growing need for direct links between sources of information and the people who use it...” Those who have attended any of the conferences and workshops on information technology and data warehousing over the past few years will be aware of many success stories where corporations have effectively integrated and mined diverse data bases in support of statistical analysis.

Not that such success has come easily: even within a given corporation, links between different computing platforms had to be forged or rebuilt. Sacred cows had to be neutered and stovepipe thinking had to be overcome. In addition, two important conditions were met. The first was that a corporate will developed to integrate and mine the collective data holdings. The second was that jurisdictional corporate control could be exerted over all the source data bases.

While the statistical agency may have the technology, the methodology, the incentives, and the corporate will to exploit, reuse, and share administrative data, it lacks one important ingredient: that is, corporate control over the data bases. This is a very familiar problem for decentralized agencies that have had to acquire much of their data from statistical activities in other government departments. There remains, however, a common concern for all agencies, both decentralized and centralized, in the amount of potentially valuable administrative data that reside in other jurisdictions, often at different levels of government or perhaps even in the private sector. It is in this context that the two issues raised by the CES work session come to the fore, that is, the lack of standards and the concerns about confidentiality.

■ Resolutions

Gordon Brackstone in his 1987 paper in *Survey Methodology*⁴ has well articulated the objectives for improving administrative records data mining. He has also set out proposed and desired changes and mechanisms for meeting those objectives. He notes the difficulty of working with non-standardized inputs, the problem of discontinuity in time series or disharmonies between sources, and the issue of data gaps. All of these problems could be reduced if the statistician had an opportunity to influence the design of the administrative system. He goes on to note a number of mechanisms that could be implemented to achieve such an objective. Most reflect sound common (and business) sense. Also discussed is the issue of privacy and confidentiality and related public perceptions. He further proposes public communication strategies, which might make the public more amenable to the manipulation of administrative records for statistical purposes.

One might add two other strategies to those expressed in the paper. The first is the matter of responsible data management on the part of statistical agencies. Statistical agencies can scarcely become champions of the exploitation of data from administrative sources when they have so often failed to effectively exploit their own data sources. These agencies are usually organized on the lines of collection methodologies (census, post-censal survey, household sample survey, business survey, etc.), with each area developing its own methodologies, systems, concepts, and dissemination programs. As a result, the agency more resembles a consortium of independent producers rather than a corporation. This leads to conceptual disharmonies, lack of uniform standards, biased single-source outputs, lack of thematic integration based on issues, populations, or geographies, and the absence of any corporate knowledge of the totality and comprehensiveness of the data holdings. In effect, the very problems in administrative records that are lamented by statisticians exist in data sources they, themselves, have collected. Their case, therefore, lacks credibility. The one hope here is that, belatedly, we are seeing signs that the agencies are beginning to recognize that they must give priority to putting their own houses in order. A demonstrated com-

mitment to effective management of their own data resources, therefore, must be the first strategy.

The second strategy, predicated upon progress on the first, is to address the problem that any given enterprise can only dictate standards within its own boundaries. How can a statistical agency hope to effectively exploit non-standardized administrative records that are held across many private and public sector enterprises? Based upon the discussions held at the CES meeting, it was evident that some agencies had little legislated support in mining administrative records, and of those that did, few had any authority to influence their contents. Three questions arise. Is legislation needed to provide statistical agencies with the authority to not only exploit administrative records but also to negotiate standards and influence content? Are there sufficient economic incentives for record holders to collaborate between themselves (as well as with the statistical agency) in the development of standards? Or is it possible that the information technology industry will develop generic software packages (with built-in standards) that will be so cost-attractive to the record holders that many of the problems will be overcome?

Is it possible to develop enabling legislation that very clearly not only extends the statistical activity to the mining of administrative records but also provides authority to the statistical agency to determine data fields and standards within those records? That would include records held by all levels of government and designated private sectors or non-governmental organizations.

Legislation of this dimension may be viewed as heavy-handed, but it may well be the measure of last resort if any progress is to be made (for all the good reasons noted above). While many statistical agencies can probably point to some measure of success in coaxing collaboration on the exploitation of one file or another, such successes have been limited and usually won only after an inordinate investment in time, hand-holding, and resources. Furthermore, these successes have generally been isolated, and little progress has been made on the issue of standardization, given the diversity of sources.

Legislated changes do not come easily, can often be

painfully slow, and always present the risk of opening Pandora's box. Nevertheless, given the compelling reasons for expanding the derivation of data from administrative records, is it time to start laying the groundwork and developing proposals for the legislators? Alternatively, inasmuch as corporations are beginning to reap the benefits of integrating, standardizing, and mining their own data bases, is it possible that they will be more amenable to collaborating with statistical agencies? Will enterprises in the public and private sectors see benefits in sharing in the use and re-use of information, particularly if it lowers respondent burden or reduces the overhead of the statistical system? Will enterprises see economic benefits to themselves in partnering in administrative records exploitation? Conversely, is it possible that any initiative that might be taken through legislation or collaboration and partnering might simply be overtaken by developments in the information technology industry? We now have many examples of where the industry has established *de facto* standards through the development of generic products, which become almost universal in application (be it operating systems, word processing software, or spreadsheet software). If the industry sees economic opportunities in the development of management information systems (a common source of administrative data), transaction systems, or registry systems, is it possible that the industry, in its drive for a cost-effective product, might develop its own standards? If the products have mass appeal and mass applicability, is it possible that the statistical agency gains without having to push for legislation or plea for collaboration?

Assuming, however, that the problem of standards, or the lack thereof, is resolved, there remains the issue of privacy. Legislation under which statistical agencies work varies significantly from country to country. Where restrictive or highly restrictive privacy legislation exists, little progress can be made. At a minimum, legislation must provide the authority to both access and link records from various sources. At the same time, such legislation must provide some assurances that such work is undertaken only for statistical purposes and that society benefits as a result. In this activity, it is important that not only must the statistical agency be squeaky clean, but, more importantly, it must be seen to be so.

■ Summary

In summary, there are presently a number of factors conspiring to force statistical agencies to consider more fully the exploitation of administrative records for statistical purposes. Agencies face shrinking budgets, rising demands for their services, a need to reduce respondent burden, and the need to better service an increasingly sophisticated and demanding clientele. At the same time, developments in information technology and developments in statistical methodology present opportunities in records management, transfer, and manipulation that did not exist even a year or two ago. Now is an opportune time to consider new directions.

A prerequisite to new directions, however, is the need for agencies to put their houses in order with regard to data management. Agencies do not now generally manage and exploit their own data resources well. Significant improvements will have to be made before agencies gain the credibility and the confidence that they can effectively manage the data holdings of others.

At a minimum to administrative records mining, overly restrictive privacy legislation needs to be modified, with access authority provided to statistical agencies. With regard to the issue of the standardization of record keeping, three possible scenarios emerge. One is the introduction of legislation that provides the agency with the authority to influence the content and format of records. Another is the possible collaboration of record holders in developing standards (ideally in conjunction with the agencies) because of perceived benefits to themselves. A third is the possibility of the information technology industry developing generic products (with built-in standards) that become universal in application, thereby introducing the standards needed for more effective data mining.

■ References

- Brackstone, G. (1995), "Issues in the Use of Administrative Records for Statistical Purposes," *Survey Methodology*, Vol. 13, No. 1.
- Conference of European Statisticians (1995), *Report*

of the January 1995 work session on Registers and Administrative Records for Social and Demographic Statistics, CES/1995/R.10.

Tapscott, D. and Caston, A. (1994), *Paradigm Shift: the New Promise of Information Technology*, New York: McGraw-Hill.

¹ *Report of the January 1995 work session on Registers and Administrative Records for Social and Demographic Statistics*, Statistical Commission and Economic Commission for Europe, Conference of European Statisticians, CES/1995/R.10, January 30, 1995.

² Data mining has generally been described as the process of extracting useful, previously unknown or unused information from datasets, which had been created for non-statistical (usually administrative) purposes.

³ Tapscott, Don and Caston, Art, *Paradigm Shift: The New Promise of Information Technology*, New York, McGraw-Hill, 1994.

⁴ Brackstone, G. J., "Issues in the Use of Administrative Records for Statistical Purposes," *Survey Methodology*, June 1987, Vol. 13, No. 1, pp. 29-33.