

Introduction

Jacqueline Mayda and Tiinia Luige

*Operations Research and Development Division, Statistics Canada, 6th floor Jean Talon Building,
Tunney's Pasture, Ottawa, Ontario, K1A 0T6*

Electronic Data Reporting (EDR) is no longer just a vision for the future of data collection but a present reality. Government statistical agencies around the world are facing increasing pressure to offer key services on-line in order to lower respondent burden, while continuing to provide high-quality statistical products. The idea of collecting data electronically interests management at all levels because of its potential: obtaining quality information with reduced respondent burden, and at a lower cost. It is obvious, however, that although the potential is there, many problems remain in this rapidly evolving field, and the benefits to the organization and the respondent are not as clear as one might envisage. These concerns are not limited to one country, nor type of statistical organization, and thus the need arises to have discussions within an international audience. To provide a forum for such a discussion, the UNECE organized jointly with Eurostat, a Work Session on Electronic Data Reporting in Geneva, Switzerland in February 2002. The goal of the work session was to share experiences and approaches to EDR from around the world with a view to assisting those who would like to venture into this area.

What is Electronic Data Reporting? There are different definitions but a simple approach could be to consider it as direct raw data collection from respondents without the interference of an interviewer (e.g., using electronic questionnaires for self-interviewing, touch-tone data entry or computer-assisted interviewing). An advanced version of this could be the extraction of data from the respondent's information systems.

The current issue of the Statistical Journal looks at the various aspects of EDR, taking into consideration its management and organization, the requirements for security and confidentiality, and the related metadata, conceptual models and standards. In addition, when analysing its application in practice, an important aspect is users' acceptance of the EDR option. Examining users' experiences with EDR provides a good basis for improving this data collection method.

The first part of the publication focuses on management and organizational issues in developing and offering EDR to respondents. Questions are considered such as the integration of EDR with existing data collection methods, strategies to address operational resources needed to deliver the EDR options, and strategies for migrating research into production. The Eurostat paper serves as a good introduction to the topic by defining EDR and related terms and providing an overview of the relevant standards. The paper from Statistics Canada describes issues encountered while introducing EDR, such as technological and respondent relations difficulties, as well as the integration of the data with existing collection modes. Several factors affect the adoption of EDR as a viable option by respondents, such as the availability of a help-desk and efficient communication with the statistical office. The paper from the Statistical Office of the Republic of Slovenia discusses several management issues, such as the implications of outsourcing when applications must be maintained afterwards, and migrating research into a production environment.

Management commitment to the new environment is an important factor for success and changes to the organization's culture are necessary. The paper by the United States Bureau of Labor Statistics documents well the detailed organizational structure set up to allow for EDR as a collection option. The key issues taken into account when considering a centralized versus a decentralized environment for EDR are discussed, as well as the advantages and disadvantages to both approaches. The Statistics Norway paper describes new approaches taken towards data collection, including a web-based solution and the importance of integration with usual collection methods. The paper from Statistics Netherlands details the concept of the "Internet-enabled self interview", ways to promote this option with respondents and to choose the right EDR method for the circumstances.

The second part of the publication considers the different security solutions deployed by statistical organizations and the development of new guidelines and policies on security, confidentiality and privacy. The strict policies of statistical offices related to confidentiality have an impact on the flexibility, user-friendliness and overall acceptance of EDR by respondents. The paper from the Israeli Central Bureau of Statistics looks at the solution of data collection through a National Data Collection Center. Internet is used as a medium for data transfer. This incurs changes in the data collection strategy and allows the better implementation of the data security policy. The paper from Statistics Finland includes an overview of their data collection activities, and how EDR fits into that picture. Their security issues encountered as well as their experience in implementing EDR and a discussion of the use of PKI for the future are also covered.

The next two papers look at metadata issues and uses of emerging technologies such as XML or XBRL. These new developments allow automated extraction of information from respondent systems which can have a significant impact on the consistency and quality of statistics. The Statistics Austria paper presents the principles of e-Quest - a metadata-based system for electronic raw data collection. It is a metadata driven generic solution that can be used for any survey. Started as a research project in 1998, it is now implemented and is being used by over a thousand respondents. The paper from the Australian Bureau of Statistics considers the potential of XBRL for electronic data capture. XBRL is often seen as the industry accepted business reporting language and it could be expected that the financial information in XBRL format could be used for multiple purposes, including statistical ones. The paper concludes that this type of by-product extraction from businesses computer systems is by far not a solution for all data collection problems, but it would allow for a significant reduction of the respondent burden.

As evidenced by the National Statistical Service of the Republic of Armenia, the necessary preconditions for EDR are also being determined in the Commonwealth of Independent States (CIS) countries. Their paper covers the development of a metadata system in Armenian statistics, involvement in international metadata systems, and other development projects.

As you will see by reading these papers, issues related to this new emerging world of EDR are plentiful. Confidentiality, security and privacy-related concerns still play a very important role when trying to implement EDR as a data collection option. The need for organizational change (if not the structure but the culture) to accommodate EDR must be considered. The integration of EDR data into existing data collection environments in a seamless manner must also be emphasized. And, as experience grows with the increasing offer of EDR, more lessons will be learned related to respondents' reactions in terms of burden, ease of use, and overall acceptance of EDR as a collection option.