

Distr.
GENERAL

CES/SEM.47/16
30 January 2002

ENGLISH ONLY

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Seminar on Integrated Statistical
Information Systems and Related Matters (ISIS 2002)**
(17-19 April 2002, Geneva, Switzerland)

Topic II: Secure communications and data confidentiality

**DESIGNING A SECURE DATA COLLECTION, SUMMARY, AND PUBLICATION SYSTEM
FOR ONE-HOUR TURNAROUND**

Contributed paper

Submitted by the National Agricultural Statistics Service, United States¹

I. INTRODUCTION

1. In the past 2 years, the Livestock and Grain Market News Branch of the Agricultural Marketing Service (AMS) of the United States Department of Agriculture has been charged with designing and implementing a Mandatory Price Reporting system which requires large livestock slaughter plant sites to report all of their transactions within designated time periods and the AMS to summarize and quickly release reports based on those transactions, with many reports required 1 hour after the data are received.

2. This paper will briefly discuss the market conditions which led to such strict reporting and publication requirements and then outline the basic approach that AMS has taken to fully implement the new law. The paper is intended to be instructional and to form a starting point for decision making if another statistical organization would be faced with similar timing and security requirements.

II. BACKGROUND OF THE NEW LAW

3. The impetus for the new law was growing concerns about the effect of concentration of companies buying from producers. By 2000, over 90 percent of the Federally inspected slaughter of cattle, hogs, and sheep in the country was performed in approximately 114 slaughter plants. Many of the 114 plants were owned by 4 companies, which slaughtered about 80 percent of all fed cattle, 55 percent of all hogs, and about 80 percent of all lambs. In addition, livestock production itself was becoming very concentrated. About 116 feedlots accounted for 40 percent of all fed cattle, with the remaining 60 percent coming from over 97,000

¹ Prepared by Rich Allen (rallen@nass.usda.gov).

feedlots. About 45 percent of the total hog inventory was owned by 110 operations, with 77,150 operations making up the other 55 percent.

4. Besides concerns about concentration itself, marketing procedures in the United States had rapidly changed. Much of the meat trade had shifted from shipping carcasses for future cutting into retail cuts to an approach of cutting and shipping boxed products such as loins. Slaughter and shipping operations are most efficient when fairly standard sized and quality animals are being slaughtered. To assure consistent qualities and supplies, livestock packers increasingly entered into private marketing arrangements with producers. Those arrangements such as formula pricing, forward contracting, and other incentive agreements usually offered some premium to operators who could produce a consistent supply of desired quality animals. In addition, some packers owned some of their own animals through subsidiary operations.

5. The AMS had a long-standing Livestock Market Price Reporting service. It was on a voluntary basis but AMS had taken many important steps to produce quality price information such as verifying details of sales transactions with both buyers and sellers before compiling price reports. However, the special purchase arrangements were private transactions and not included in AMS reports. Since many of those special arrangements had traditionally been on some basis to the open market transactions there were originally not too many concerns from producers about the lack of price information. However, purchases of hogs for slaughter shifted from 80 percent or so open market to 20 percent in just a few years and there was likely less of a tie to open market prices. Small operators, who did not produce the volume to qualify for such special arrangements, felt that they were not privy to true price information.

6. The United States Congress expressed its concerns about the lack of full price information and the possible impacts of concentration by passing the Livestock Mandatory Reporting Act of 1999. The Act required all livestock operations above certain size limits to report the details of all purchases to AMS and established detailed reporting requirements such as swine purchases three times a day (purchases up to 10:00 a.m., from 10:00 a.m. to 2:00 p.m., and total daily purchases by 7:00 a.m. the next morning), cattle purchases twice a day, lamb purchases once a day, and similar requirements for meat sales. There were many other technical requirements such as each slaughter plant reporting lowest and highest net prices in addition to average prices, plus carcass weights and quality measures of the animals slaughtered.

III. BASIC SYSTEM REQUIREMENTS

7. The AMS was charged to set up a system to receive and validate all data reported by the companies and to issue statistical summaries within an hour. Some 91 different reports were designated to be created by AMS, many containing information which had not previously been available. AMS had to develop a system which was simple enough to implement that all affected plants could be trained to use the system, which had sufficient safeguards to protect integrity of the data being transmitted, and which allowed for quick data review, validation, and summary. Data also had to be efficiently and securely archived since the law also established an ongoing audit requirement of all reporting companies. And, even though the law mandated reporting, AMS was to develop procedures to protect the confidentiality of all reporting entities.

8. It was clear that AMS needed an approach which was always available, which had strong systems integrity, and which could capture and protect proprietary data. It also required the capability for mathematic manipulations to prepare the summary reports, many of which involved combining data elements through some modeling operations. A few common electronic systems came to mind as models to consider. The systems used for electronic banking and automated transactions by individuals at ATM machines have to be available 24 hours a day and maintain detailed, secure databases. However, with the exception of multiple checking, savings, etc., accounts linked to the same individual, electronic banking systems do not need the

capability to link data across accounts to prepare timely reports. The systems used for financial stock markets have been written to constantly update indices and other models whenever transactions occur. However, those systems do not have to consider confidentiality factors before issuing new reports. The conclusion was that systems such as electronic banking and tracking financial market trading helped describe features needed in the mandatory reporting system but no off-the-shelf system was presently available.

IV. SYSTEM DESIGN APPROACH

9. AMS did a thorough, far reaching evaluation of data processing and security alternatives and interfaces. Over 35 different laws or regulations were identified which needed to be followed to ensure that proper protocols and security measures were employed. AMS specifically identified the sensitivity of the system regarding the all important aspects of confidentiality (assuring that sensitive data can be accessed only by authorized individuals and not revealed to other individuals or the public), integrity (protecting all data from improper modification through computer viruses or fraudulent changes), and availability (ensuring that data, applications, systems, and networks are accessible whenever authorized users need them). Since there were no existing government facilities and systems which could meet the unique needs of this new program, it was necessary to use competitive contract procurement for the communication systems, hardware, and programming required.

10. Another key challenge for AMS was the communication with and training of the companies and plants which were affected by the new law. It was vital to assure the companies that their data would be protected along with the identity of the company. It was helpful to learn of the company concerns about setting up new reporting procedures, hiring new technical personnel in many cases, and adjusting internal company communications to compile data fast enough to meet the reporting requirements. The constant communication was helpful in alleviating many of the fears and concerns about the new procedures.

V. THE OPERATING SYSTEM

11. The system design uses the Internet functionality for internal communications and for the capability for effective transmission of reports from the system. A browser-enabled application has been established on a commercially acquired, redundant NT platform. Communications have been established through the commercial company's World Wide Web Infrastructure. Authorized market news reporters and auditors are able to connect through a U. S. Department of Agriculture router. Each affected slaughter company has selected an Internet Service Provider for their communications to the system. All system accesses are through specific Public Key Infrastructure certificate security implementation. All users are granted rights only to those files and features of the system appropriate for their functioning. For example, companies can load data or files to the system but not retrieve any information. The numbers of market news reporters with access rights are strictly controlled and auditors are provided access only to the information needed for a particular audit that they are to conduct.

12. All user accesses to the system are also controlled through the use of unique identification numbers and passwords. Companies can submit data through direct data entry or by uploading files using File Transfer Protocol over a Secure Hypertext Transfer Protocol connection. All transmitted data are encrypted for added security assurance.

VI. DATA HANDLING, EDIT, SUMMARIZATION, AND RETENTION

13. In addition to providing a secure data transmission system to ensure confidentiality and integrity, AMS needed a very efficient data handling system to meet the 1-hour turnaround requirement for summary and publication. One key feature is that three types of databases are employed. Data from companies are

loaded into what has been named the quarantine Web server database and archived. Data remain in that database until reviewed by reporters for consistency and accuracy. When a data set for a particular report has been accepted, reporters begin the report generation. Routines in the report subsystem extract the needed data items to create the current report for publication. All original and corrected data files are stored in a permanent database. This is needed since companies must retain all reported data for at least 2 years for possible audit purposes and to provide a base for further analyses of this dynamic industry.

VII. SYSTEM PERFORMANCE

14. As with many major computer application systems, it took longer to implement than originally hoped. However, it was vital to be sure the confidentiality and integrity safeguards were well tested. One unusual aspect in the testing of the new system was that the new law did not permit AMS to be collecting voluntary and mandatory data at the same time. This required AMS to prepare its own test data for both program development and quasi-operational testing. It also meant that AMS could not continue most of the voluntary reports that were formerly issued in order to give the industry parallel reports for evaluation. Some new reports specified by the legislation could not be generated when the system began operation since it was necessary to collect and evaluate live data before finalizing the calculations and report formats. They were brought on line as quickly as they could be finalized and tested. Only one problem has occurred with the new program. A wrong assumption in the testing of the most complicated report of the system (which calculates carcass level equivalent value from the prices of various meat cuts) resulted in incorrect calculations when the system was implemented. When the error was noted, AMS quickly fixed the formulas and issued a corrected set of data.

VIII. CONFIDENTIALITY

15. One aspect of the system that may be of particular interest to statisticians is the confidentiality handling of the new program. The law required AMS to ensure that the identity of reporting firms was held confidential as well as the proprietary data being reported. One step that AMS has taken to provide extra confidentiality is to never indicate how many firms are included in any total which is published. AMS started with a very strict 3/60 rule for each reporting period requiring at least 3 firms to be reporting, with no firm having more than 60 percent of any aggregate that was to be reported. This meant that many data cells had to be suppressed within reports that were released and some reports could not be issued at all. Many industry participants were upset about the lack of data and they often assumed that even more data were being suppressed since AMS did not distinguish between blank and suppressed data cells. In many cases, National reports during the day could be released but many planned regional reports could not because of the 3/60 requirement. Analysis of reported data from the first 60 days of operation of the new system demonstrated that reporting patterns in most regions were very random. For the most common purchasing arrangements such as negotiated or formula purchases there were often many (10 or so) active companies but perhaps only one or two were successful in purchases during a particular time period. It was not the case that certain companies were always active on particular days of a week, for example. Also, even if only one company bought during a particular time period, other companies would not know which company or even that it was only one company. This led to an intriguing new confidentiality rule referred to as 3/70/20. This rule allows publication of data if an average of at least 3 companies had been active daily in the past 60 days, no company had 70 percent or more of the volume of purchases during that time, and, if days had occurred in which only one company purchased, that no specific company would have been exposed more than 20 percent of the time. This rule was widely discussed with the industry and presented to major U. S. statistical agencies for their information before it was implemented. Using this approach has reduced the percentage of hog reports not issued from about 30 percent to less than 3 percent and all cattle reports have been able to be issued except for daily forward contract sales.