INTEGRATING AGRICULTURAL DATA FOR ANALYSIS AND PUBLIC USE

Mark Elward, Statistics Canada and Theresa Holland, United States Department of Agriculture
Mark Elward, Agriculture Division, Statistics Canada, Ottawa, Canada, K1A OT6, elwamar@statcan.ca
Theresa Holland, National Agricultural Statistical Service, USDA, Washington, DC 20250, tholland@nass.usda.gov

ABSTRACT

Agricultural economists have been integrating agricultural data since the conception of supply and demand theory. Consequently, integrated data programs often date back to the development of the statistical programs themselves because they rely on integrating agricultural data into structured formats.

This paper will look at integrating agriculture and food data from three perspectives: traditional statistics, structured accounts, and surveys. As the topic is broad, the paper will approach the subject from a highly summarized perspective.

Traditionally, the integration of agriculture and food statistics has been an integral part of many agricultural statistical programs. Commodities have long been the focus of supply-disposition tables (balance sheets) used to analyze data, as a tool to set estimates, or to display information. In many cases, the commodity data flow to the economic data set so that estimates such as farm income or value added can be derived. Also, commodity data from different sources, such as censuses versus seasonal production surveys, must often be integrated into a statistical program so that the public data do not differ because of different methods.

One way of integrating data is to develop a series of structured accounts that are logically integrated and provide more information as a result of the integration. A brief look at an integrated system of accounts including value-added accounts, balance sheets, cash flow accounts and producer income accounts will be presented.

Integrating data from different sources that use different concepts and methods often has some serious challenges that can be met, in part, by integrating surveys. The paper will review the experiences of USDA’s National Agricultural Statistical Service with the Agricultural Resource Management Study and of Statistics Canada in relation to the Unified Enterprise Survey.

The authors hope that examples from the Canadian and United States experience will illustrate the value of integrating agriculture and food statistics.

Key Words: agricultural accounts, integrate, supply-disposition, balance sheet, survey

1. INTRODUCTION

Integrating agricultural data from different sources must have been one of modern man’s earliest mathematical endeavors. One can imagine administrators in earlier times trying to determine if their food supplies were sufficient by obtaining data on supplies, such as food in storage or production, and demand, such as number of mouths to be fed.

The integration of agriculture data today may be more sophisticated, but traditional types of integration are similar to earlier times. More modern forms of integration rely on modern statistical programs to both obtain the data and structure the data for analytical or public use.

Statistical organizations often develop analytical tables as tools to integrate data, assess the quality of survey and administrative data alike, and prepare official estimates for public release. These working data sets normally contain considerable detail – survey results, ratios, check data, previously published data and data prior to release. Analytical tables often contain more detail than is practical to publish, confidential data, and data that are not designed for release.

Benefits

There are many advantages to integrating data. Fundamentally, some data sets simply cannot be produced without integration. It is only by integrating data from different sources that agricultural statistical organizations can produce estimates such as supply and demand, farm income or value added for agriculture. Ultimately, agricultural data are often used for national accounts, a highly integrated framework, to produce indicators such as gross domestic product.
In addition, it can be much more efficient to integrate data from different sources, thereby reducing both costs and the burden on survey respondents. Administrative data and data obtained as a result of governmental regulatory activities are routinely used in agricultural statistical programs in conjunction with data from surveys.

Further, integrating data can lead to higher quality information. Timeliness and quality are two statistical needs that often conflict – the faster the estimates are produced, the higher the risk of error. Agricultural information, due to the continual demand for and perishable nature of food, must often be produced in a very timely manner. Routinely forecasts or estimates based on expert judgment or trend analysis are used. As time goes on, better quality data become available from sample surveys, census surveys or administrative sources. These data can be integrated and used to fine-tune earlier estimates. Even when particular estimates are survey-based, analysts normally use supply-disposition balance sheets to verify the reasonableness of the survey data and, if necessary, adjust the estimates prior to publication. What the analyst is really doing is providing a more reliable estimate than the survey results alone would indicate.

Statistical organizations often integrate data at the analytical and output level to ensure that only one number is published for each variable of interest. Users can be baffled by more than one estimate for the same item from the same statistical agency.

Limitations

While data integration has many strengths and is an integral part of agricultural statistics programs around the world, it also has some limitations. One problem associated with integrating data is that sources of data are often numerous and the concepts, definitions, methods and data limitations associated with each source can be different. Statistical organizations can determine the concepts and methods used when surveying for information; however, their ability to influence the concepts and methods underlying administrative data are usually limited. When sources are merged, the differences can lead to quality concerns because of inconsistent data. No data set is perfect. If anything is wrong, it is important to be consistently wrong because data users typically work with both level and change.

Another problem is that while data integration may allow the calculation of information that would not be available otherwise, the results may be flawed, particularly if the derived data are calculated residually. A residual derived from an integrated data set will contain statistical error. For example, in a grain crop balance sheet, “feed, waste and dockage” actually represents “feed, waste and dockage, plus statistical error”. If the data used to derive this estimate are from the same source, odds are that a bias in the supply would be partially offset by a bias in the disposition. When the sources of supply and disposition statistics are different, the odds of offsetting biases are much smaller, leading to a less accurate residually derived estimate.

This problem may be magnified if the calculated residual is considerably smaller than the data used for the calculation. For example, in a farm income balance sheet, a small change or a small error in receipts, expenses or inventories can lead to a large change or a large error when deriving net income.

2. TRADITIONAL DATA INTEGRATION

Modern statistical organizations often integrate data in traditional ways – ways similar to those used by early economists or the way the organization has been integrating data for decades. The best example of traditional data integration is the supply-disposition balance sheet. This tool is essentially a simple commodity-level model. Stock and flow components of the balance sheet usually come from different sources.

A typical supply-disposition balance sheet will have certain basic elements. “Total supply” will be the sum of stocks (inventories) at the beginning of a period and flows into the sector during the period, while “total disposition” will be the sum of flows out of the sector during the period and stocks at the end of the period. Theoretically, total supply will equal total disposition. (For some commodities, beginning stocks are either close to zero or unknown, so the supply-disposition will rely solely on flows.)

Several challenges arise when working with balance sheets. One is to estimate for one or more important items that are not measured directly (e.g. feed, waste and dockage), usually by residually deriving the data element(s). Another is distributing error when the total supplies do not equal total disposition. A third challenge is assessing the quality of the
balance sheet overall. These challenges are difficult to overcome when integrating data from numerous sources with different concepts, methods and data limitations. Agricultural analysts use a combination of science and judgement, developed from years of experience, when working with supply-disposition tables.

There are numerous types of supply-disposition balance sheets displaying different data sets and used for different purposes. Analytical balance sheets may contain items that are low in quality while the public data set might aggregate items of varying quality into one data point. Analytical balance sheets might also use a two-column approach, one for previously published data and the other for revisions.

Balance sheets can be used to depict data at different geographical levels – sub-provincial, provincial or state, regional, national, global, etc. For example, on a monthly basis, the U.S. Department of Agriculture’s World Agricultural Outlook Board publishes a report containing supply and demand estimates for various commodities for the world and major producing countries. (USDA 1999a)

Sometimes, a balance sheet will be used to depict farm-level or commercial-level activity, or farm and commercial levels may be combined. In the latter supply-disposition table, often used at the national level, producer deliveries from the farm to the commercial level cancel because they represent a flow within the defined sector.

<table>
<thead>
<tr>
<th>Table 1. Grain commodity supply and disposition table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Beginning stocks</strong></td>
</tr>
<tr>
<td>Stocks on farms</td>
</tr>
<tr>
<td>Commercial stocks</td>
</tr>
<tr>
<td><strong>Total beginning stocks</strong></td>
</tr>
<tr>
<td><strong>Flows in</strong></td>
</tr>
<tr>
<td>Production</td>
</tr>
<tr>
<td>Imports</td>
</tr>
<tr>
<td>Producer deliveries</td>
</tr>
<tr>
<td><strong>Total flows in</strong></td>
</tr>
<tr>
<td><strong>Total supplies</strong></td>
</tr>
<tr>
<td><strong>Flows out</strong></td>
</tr>
<tr>
<td>Producer deliveries</td>
</tr>
<tr>
<td>Grain exports</td>
</tr>
<tr>
<td>Product exports</td>
</tr>
<tr>
<td>Seed requirements</td>
</tr>
<tr>
<td>Feed, waste &amp; dockage</td>
</tr>
<tr>
<td>Human food</td>
</tr>
<tr>
<td>Industrial use</td>
</tr>
<tr>
<td>Loss in handling</td>
</tr>
<tr>
<td><strong>Total flows out</strong></td>
</tr>
<tr>
<td><strong>Ending stocks</strong></td>
</tr>
<tr>
<td>Stocks on farms</td>
</tr>
<tr>
<td>Commercial stocks</td>
</tr>
<tr>
<td><strong>Total ending stocks</strong></td>
</tr>
<tr>
<td><strong>Total disposition</strong></td>
</tr>
</tbody>
</table>

Table 1 is a sample analytical balance sheet that displays an aggregation over geographical regions and sector levels. For practical purposes, statistical agencies would normally display the different types independently in a time-series format. Also, terms like “flows in” would not be used explicitly. Flows within the defined sector, such as farm-to-farm sales within a defined geographical region, are often not displayed in public releases because they are not required to balance supply and disposition, and because the data either do not exist or are of poor quality.

Time can also be used to depict supply and disposition. Balance sheets can be produced for any time period needed – crop year, calendar year, quarter, month, year-to-date, etc. Consequently, supply and disposition can be compared over time, to other geographical levels or aggregated with other supply-disposition balance sheets to provide additional information.
The supply-disposition balance sheet is a powerful tool. In fact, it is so powerful and in such common use that it is an important tool used in price determination. When statistical organizations release data related to the production of agricultural commodities, there can be an immediate market response in terms of the price of that commodity. But the response is truly more because the supply and disposition relationship as depicted in the balance sheet has changed. Production goes up, supplies are greater and there is downward pressure on prices.

The sophistication of supply-disposition balance sheets can range from “back of the matchbox” calculations to highly developed models for commodities such as hogs that use a balance sheet in conjunction with survey data and biological constraints, distributing “residual error” based on the quality of the input data. Even so, all such analytical tables must incorporate the fundamentals that underlie all balance sheets.

One important feature of most supply-disposition tables is that they usually integrate data from many different sources, each with their own unique concepts, methods and data limitations. For the analyst who will likely work with survey data and administrative data, determining the error and attributing it appropriately is virtually impossible. Consequently, the analyst must decide whether to display the error, hide the error, or distribute the error. This is done using reasonable judgement based on experience and some knowledge of the inputs. Analysts usually have an indication of the quality of survey data, but they have less information and control related to the quality of administrative data. The fact that the different components of the integrated statement have different data limitations is a problem in itself, particularly when an item is being residually derived. And, the smaller the residual, the more serious the problem. If the errors were somewhat consistent, then at least the calculation of a residual would not compound the error. If, however, the errors are not consistent and there is no compensating error, the quality of any residual calculation can be dubious at best.

The limitations of integrated data are a major concern to analysts working with economic data. For agriculture, net farm income is an important measure, which is calculated residually by subtracting expenses from receipts. The sources for the receipt and expense data vary considerably. Receipts typically are based on producer deliveries or slaughter data obtained from administrative sources, and prices obtained from administrative data and price surveys. Expenses, on the other hand, often use survey or tax data. The value of inventory change is normally based on survey and balance sheet data and associated with average prices. One can quickly see that because this integrated data set uses data from different sources that any net number derived could have severe data limitations.

The severity of the data limitations is a function of the size of the error in one component not offset by error in another component and the size of the residual itself. If the residual is small, as net income can be, then the error may be magnified. Time is also a factor. Data users demand timely estimates, if not forecasts, and the result is that the components are often revised. A small revision to a large component can have a large impact on the calculated residual. The ratio to net income is regularly 5 to 1 for receipts and 4 to 1 for expenses. These ratios can climb much higher if net income is particularly low – just when analysts and politicians alike are keeping a close eye on farm income. If the receipts are later revised in one direction and the expenses in another, a large change in the net income estimate is likely.

Table 2 shows an example, based on real data, in which receipts were increased by 5 percent and expenses lowered by 5 percent. As a result, net income changed dramatically. Cash income, a smaller residual, changed by 42 percent, while realized and total net income basically doubled. This illustrates how the calculation of net income can be very sensitive to statistical error. This problem is exacerbated when working at smaller geographical levels, such as a province or state, where some of the input data will be of lower quality than similar data at the national level.

<table>
<thead>
<tr>
<th>Table 2. Net income - what happens when data change?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Base</strong></td>
</tr>
<tr>
<td>______________________</td>
</tr>
<tr>
<td>Total cash receipts</td>
</tr>
<tr>
<td>Operating expenses</td>
</tr>
<tr>
<td><strong>Net cash income</strong></td>
</tr>
<tr>
<td>Income in Kind</td>
</tr>
<tr>
<td>Depreciation</td>
</tr>
<tr>
<td><strong>Realized net income</strong></td>
</tr>
<tr>
<td>Value of inventory change</td>
</tr>
<tr>
<td><strong>Total net income</strong></td>
</tr>
</tbody>
</table>
National accounting data is possible only by integrating data from many different sources. Without integrating data, it would be impossible to produce such fundamentally important indicators as gross domestic product (GDP). Consequently, national accounts have the same quality limitations as any other integrated data set. Even though GDP is calculated as a residual, the result is proportionately larger than net income making this indicator much less sensitive to error. Although GDP can be calculated in several ways, operationally it is often calculated in two different ways, then the data are “confronted” and the difference is distributed. Currently, Statistics Canada has underway a project to put all business survey data under one umbrella with similar concepts, methods and data limitations. (The Unified Enterprise Survey will be discussed later.) A unified survey approach is possible in Canada, partially because of the centralized nature of the statistical system. Conversely, it would be more of a challenge in the United States where the statistical system is decentralized.

3. INTEGRATED AGRICULTURAL ACCOUNTS

Over the past decade, the United States and Canada have been working to improve their agricultural financial data systems by developing integrated sets of financial accounts. Conceptually integrating accounts allows a more thorough analysis of the agriculture sector that would not be possible otherwise. In addition, like the supply-disposition tables or the net farm income statements, the accounts integrate data from a multitude of different sources. The integrated accounts should result in better information and, thus, facilitate better decisions by decision-makers.

Canada

The Agriculture Division of Statistics Canada has partially developed a set of Integrated Financial Accounts for Agriculture that include:

- Value-added,
- Balance sheet,
- Cash flow,
- Capital flow,
- Farm capital value, and
- Total income of farm operators.

The existing financial data system was developed in the 1930s and 1940s in response to pressures for financial information. Early agricultural policy concentrated on research and development in crop production and animal husbandry and on improving methods of marketing grain. With the onset of the Depression, there was a shift to policy designed to improve the income position of farmers and to increase their market power through centralized marketing. During World War II, production maximization was emphasized. In the postwar period, policies concentrating on productivity, marketing, price stability and the adequacy of incomes were developed. Recently, agriculture policy initiatives have expanded to include goals such as stable food prices with adequate supplies, stable and fair producer returns, the quality of rural life, conservation of soil and other resources, environmental impacts, rural-urban balances, trade agreements, the impact of international price wars and the fulfillment of international food supply responsibilities.

The financial accounting system for the Canadian agricultural production sector has become somewhat obsolete or inadequate as the information requirements of public and private decision-makers and the structure of the sector have changed over time. The accounts establish valid linkages between sector incomes and those of the farm operator family “institutional” group. The business orientation of these accounts should deter the use of business information for drawing welfare implications of people involved in the sector. The welfare policy analyses would be better served by institutional accounts, such as the total income of a farm operator’s account.

The following is a brief description of each of the six accounts that will display the integration of data both within the six accounts and from a multitude of different data sources. The accounts themselves are aggregate accounts and do not display all of the commodity detail that some users require, as this information is available elsewhere. For example, aggregated farm cash receipts are used as opposed to commodity level receipts.

The Value-Added Account (or Production Account) measures the value of economic production of goods and services from farm establishments. Conceptually, as with all of the integrated accounts, it is based on the establishment concept
and displays the information on a calendar year basis, by province. Economic production is defined as the production of goods and services that are exchanged for money in the market economy.

The format of the production account is designed to display the sources of economic production and distribute the incomes earned in this production to the factors of production. The sources of production fall into four main categories: sales of agricultural products (including farm to farm sales), sales on non-agricultural products (woodlot production, recreation), other sources (program payments, rent), and own account production uses (value of inventory change). The sum of the four sources of income results in the gross output of the agricultural industry.

The allocation of the income falls into four categories including: product expenses (intermediate inputs), indirect taxes, depreciation, and the distribution of net value added to labor and capital. Both gross and net value added are displayed. The net value added is distributed to wages (non-family), rent to non-operators, interest, corporate profits, and unincorporated operator family returns. From the family returns, family wages are removed to derive unincorporated operator returns, which represents a return to the unpaid labor and equity capital of the unincorporated farm operators.

The Canadian Farm Balance Sheets have been designed to display assets such as current assets, machinery and real estate, liabilities on a current and long term basis, and equity. Provincial-level data are shown as of December 31 for any given year. Four different balance sheets, in actuality, are used to separate farm businesses from farm operator households and from non-operator landlords. The published data focuses on the set that most closely reflects the assets employed in the production of agricultural products.

Of interest from an integration perspective is that the balance sheet also includes a set of financial ratios, some of which rely on data from the Value-Added Account. The liquidity and solvency ratios are independent; however, the profitability (return on equity) and efficiency ratios (interest coverage) use data from different statements that can only be relied on if the data have similar conceptual frameworks.

The Cash Flow Account summarizes cash flows in and out of farm businesses over a calendar year. Cash flows from transactions and cash flows from changes in levels of debt outstanding or net capital expenditures are accounted for. Also, the accounts include specific ratios such as loans outstanding to cash income. Both the account itself and the ratios rely on integrating data from the Value-Added Account and the Balance Sheet.

The Capital Flow Account is designed to explain changes to the balance sheet and the Farm Capital Value Account displays capital assets located on farm establishments regardless of ownership. Clearly these accounts must be integrated with the balance sheet.

The sixth account in the series focuses on the farmer by providing a measure of the total income of farm operators, regardless of the source of that income. This account is yet to be developed, but would provide new information for policymakers based on integrating data from the Value-Added Account with other sources that depict off-farm income.

The strength then is in the whole. The integrated accounts provide a set of financial data answering many of the questions related to both farming as a business and the people involved in farming.

United States

The Economic Research Service (ERS) of the U.S. Department of Agriculture (USDA) is responsible for developing and maintaining a similar set of agricultural accounts for the United States (USDA 1999b, 1999d). Among the accounts developed annually are:

- Value Added/Farm Income – value added, net farm income, returns to operators and net cash income,
- Farm Sector Balance Sheet – assets, debts and equity,
- Capital Expenditures and Replacement Costs,
- Cash Receipts for Agricultural Commodities, and
- Cost of production for selected crop and livestock commodities.

These data for the farm sector are integrated into the National Income and Product Accounts of the United States developed by the Bureau of Economic Analysis of the U.S. Department of Commerce (Seskin & Parker 1998).
The development of the U.S. agricultural accounts dates back to the early part of this century. The content, format and frequency have evolved over the years in response to changing policy needs from the Depression Era, to World War II and the postwar era, to the changing farm economy of the 1970s, 1980s and 1990s. ERS recently switched to a detailed value-added format for presenting farm income data to make it easier to determine which factors are driving changes and trends in farm income.

Like the Canadian accounts, the various U.S. accounts are highly integrated, sharing input data sources and derived values from one account to another. Much of the data used in these accounts come from surveys of farmers conducted by USDA’s National Agricultural Statistics Service (NASS), including the Agricultural Resource Management Study (described below), various crop, livestock and prices surveys, and the quinquennial Census of Agriculture. Administrative data from various other USDA and other Federal agencies are also used.

A review of the individual U.S. accounts will not be made as this information overlaps with the content of the Canadian accounts discussed above.

4. SURVEYS

Statistical agencies have turned to surveys to provide integrated or unified information in order to improve the quality of the data and the efficiency of the survey process. Quality improvements can be made by standardizing the concepts, methods and data limitations allowing logical linkages that may be weak when working with several independent programs. In some cases, eliminating duplication of contacts and/or questions can ease respondent burden.

Even so, survey integration has some limitations. Respondent burden may be increased if more questions are asked of more respondents. Integrated programs are often large programs and timeliness may be compromised. Although size can lead to efficiency gains, it can also lead to a larger, more complex process, which can be costly from a management perspective. Many statistical agencies have met with success by focusing surveys on target groups (e.g. crops versus livestock). This is because the design, process and output can cater to the exact needs of the specific target group, efficiently producing reliable and very timely statistics.

4.1 Agriculture Resource Management Study

USDA’s Agricultural Resource Management Study (ARMS) is designed to provide data on farm finances, production practices and resource use. The data collected by the National Agricultural Statistics Service are used by the Economic Research Service in various economic and environmental analyses, including the agricultural accounts described above. The ARMS as we know it today evolved from the integration of various surveys over the years.

*Farm Production Expenditures Surveys (FPES)* began in the mid-1950s, and have been conducted annually since 1972. These surveys have provided information on what farmers spend for production items. This information is used to calculate prices paid index weights, and to estimate net farm income and general costs of production for crop and livestock commodities.

Annual cost of production estimates for wheat, feed grains, cotton and milk were mandated by the U.S. Congress in 1973, leading to the development of specific *Cost of Production Surveys (COPS)* for these commodities. Over the years, other crop and livestock commodities have been added to the COPS program to satisfy various farm program and other analysis needs. Due to survey costs and respondent burden considerations, cost of production data for specific commodities are obtained on a rotating basis, providing benchmark data for a particular commodity only once every 4 to 6 years. In the intervening years, ERS uses production and price data from other NASS surveys to update cost estimates.

In 1984, the FPES and COPS were integrated. The resulting *Farm Costs and Returns Surveys (FCRS)* collected farm financial information for all farms in the sample, while obtaining costs of production for subsamples of farms having the targeted commodities for the particular year. Increasing survey efficiency and data quality were the goals of this integration. Combining samples would make more reports available for estimates. Eliminating duplication of questions asked in both surveys would mean more consistent data. Response rates would be improved as respondent burden was reduced by eliminating the potential for multiple contacts to collect similar information. In addition to production
expenditure and costs of production information, the new survey design provided regional-level data on chemical use and production practices.

Information on crop production practices has been collected under several guises over the years. Since the mid-1960s, fertilizer data has been collected in conjunction with objective crop measurement surveys. In the late-1980s, questions on seeding rates, tillage practices and pesticide use were added to these surveys. In 1990, as a result of concerns about pesticides and plant nutrient levels in water, a Congressional initiative to enhance water quality led to the implementation of a separate Cropping Practices Survey (CPS). One goal of the initiative was to establish a database linking fertilizer and pesticide use with farming practices, soil characteristics, water quality and farm expenditures. The CPS was designed to collect information from farmers on fertilizer and pesticide quantities and application rates, target pests, nutrient management and farming practices (seeding, tillage systems, crop rotation, etc.) for selected field crops. The design provided state-level chemical use and production practice data, but did not have an economic component due to cost and respondent burden constraints.

To establish the link between farm financial information and field crop chemical use and production practices, the FCRS and CPS were integrated into the Agricultural Resource Management Study (ARMS) in 1996. As a result, ERS can now conduct economic analyses relating field crop chemical use, production practices, and cost of production (USDA 1999c). In the integration process, efficiency of data collection was improved by eliminating duplication of chemical use and production practices information that had been obtained on both surveys. The feature of rotating targeted cost of production commodities every 4 to 6 years was retained in the integrated design.

The ARMS is conducted in three phases. ARMS Phase I involves screening the sample of farms selected from NASS’s list sampling frame. The sample is selected during the spring of the year in such a way to ensure coverage of farms likely to have the particular year’s targeted cost of production commodities, as well as farms in general. The Phase I screening determines farm status (operating or not), target commodity status (have this year or not) and farm value of sales. This information is then used to select samples for the next two phases.

ARMS Phase II focuses on the targeted commodities. During the fall of the year, data on chemical use, production practices, variable costs of producing the particular commodity, and quantity produced are obtained from a sample of farmers identified as having the commodity during Phase I. In ARMS Phase III, information on farm economics and risk management practices are obtained from all Phase II respondents and a general farm sample. The link between agricultural resource use and farm finances comes from the farms in both Phase II and Phase III.

The screening phase has allowed the reduction of sample sizes required and costs per usable report in Phases II and III. In addition, the survey design allows further integration to satisfy other data needs that require linkage to the environmental and economic data ARMS provides. For example, in the year 2000, the Agricultural Economic and Land Ownership Survey, a follow-on survey to the 1997 Census of Agriculture, will be conducted in conjunction with ARMS Phase III.

Data from the ARMS are used to generate a variety of reports. Among them are Agricultural Income and Finance Situation and Outlook (USDA 1999b) and Farm Business Economics Report (USDA 1999d), ERS publications relating to the agricultural accounts mentioned previously, as well as ERS’s environmental report, Agricultural Resources and Environmental Indicators (USDA 1999c). NASS and ERS jointly produce publications on Farm Production Expenditures (USDA 1999e) and Agricultural Chemical Usage - Field Crops (USDA 1999a). While all of these reports make extensive use of ARMS data, they also take the concept of integration even further by incorporating data from various other NASS surveys (June Agricultural Survey, Fall Area Survey, Census of Agriculture, Fruit and Vegetable Chemical Use Surveys, Farm and Ranch Irrigation Survey, prices received and prices paid surveys, among others), and other USDA and other Federal administrative sources.

### 4.2 Unified Enterprise Survey

On a much broader level, Statistics Canada is looking to integrate all of their economic surveys across all business sectors. The Project to Improve Provincial Economic Statistics (PIPS) is a major undertaking that will rely, in part, on integrating data. The improvements will result because of four important thrusts.
1. Coverage: All industries will be covered by the Project, except where an industry is not significant in a particular province. The industries are defined according to the new North American Industrial Classification System (NAICS), which will also be used throughout Statistics Canada wherever business statistics are found.

2. Reliability: For each significant industry in each province, sufficient sample sizes or adequate coverage from administrative records will ensure reliable data.

3. Consistency: A consistent approach must be used for all industries and provinces, with similar statistical units, terminology, questionnaires, survey methodologies, and procedures for capturing and processing data.

4. Coherence: Information on what businesses sell (supply-side data) must be confronted by and balanced against information on what consumers spend (demand-side data). In addition, data collected from establishments and data collected from their parent enterprises must be internally consistent and comprehensive, and not duplicated.

Under the PIPES umbrella, the Unified Enterprise Survey is crucially important, as it will be used to integrate many of the 200 separate business surveys into a single master survey program. The survey aims to ensure Statistics Canada receives consistent and integrated data from many types of surveys and sizes of businesses, with enough detail to produce accurate provincial statistics.

The Business Register is the survey frame for all of Statistics Canada’s business surveys. As part of the Project, the frame will include all Canadian businesses, be they incorporated or unincorporated, employers or non-employers. Further, Revenue Canada recently implemented a new single business number that is now incorporated into the Business Register. This number makes it possible to assemble statistical data from Revenue Canada, including tax records and payroll deduction data. The Business Register has developed a new system to track all contact made with responding businesses.

The Unified Enterprise Survey (UES) will rely on obtaining data from two different sources, a sample survey and tax data. In the future, the UES will utilize the General Index of Financial Information that will contain financial statement detail for all incorporated businesses in Canada from their income tax files. This means that tax data for all enterprises that file a tax return will be available in a standardized format. Also, in the future, the data will be available on a NAICS basis.

The sample survey portion of the program focuses on collecting data from the enterprise rather than the establishment. Enterprises are divided into two types – complex and simple. Complex enterprises are those having more than one company (legal entity) or having establishments in more than one province or industry. Simple enterprises have just one legal entity and usually one establishment. If there is more than one establishment, they must be in the same province and industry.

The UES includes three main parts. First is a survey of the 8,000 complex enterprises of significant size. Income statement, balance sheet and structural data are collected on this survey. Second is a survey of the establishments within these complex enterprises requesting information related to production, sales, employment, etc. These data are used to generate value-added by industry and province. The third part is a sample survey of simple enterprises collecting data that is similar to part 2. The survey data are used in conjunction with tax data to produce estimates.

The first value-added account developed was for aquaculture. This account relied solely on data from the UES, proving that the concept behind the survey works. The fact that a structured and detailed account could be built in the first place is an important step forward. More so, there is real strength in terms of quality as the entire account was built from one survey program – the UES. This means that the concepts, methods and data limitations are similar for all of the components making the integrated account much more reliable, particularly when deriving residually calculated estimates such as value added. Further, the chances are less that a revision will have a huge impact on the residual, which is proportionately larger. Revisions will likely be made to inputs and outputs at the same time and relationships will be respected.

This concept of integrating data with similar concepts, methods and data limitations can be carried to the next step – to the national accounts. The idea of a single master survey providing the backbone for economic statistics and integrating

999
into a structured system will provide a data set that is not only reliable at the industry level but is also reliable when data are aggregated. Standardizing the concepts, methods and data limitations that underpin the myriad of statistics used for national accounting purposes is a long-term goal at Statistics Canada. No doubt, the road to arrive at this statistical utopia will be rocky and the reality something less than ideal, but the promise of such an improved system will hopefully make the work worthwhile.

5. SUMMARY

This paper has looked at only a few of the ways statistical organizations integrate agricultural data to provide relevant information that would not be available otherwise, to add value to survey and administrative data, to present data in a way that is meaningful to users, and to improve program efficiency. Traditional balance sheets provide a way to analyze production and disposition of agricultural commodities. Integrated financial accounts allow thorough analysis of the economic health of the farm sector and agriculture’s contribution to the national economy. Survey integration can lead to better quality and more consistent data, and may create more statistically sound links between databases used for agricultural, environmental and economic analyses. Certainly, there are many other ways that data integration helps statistical organizations meet their goals.

As ever-developing technologies allow easier and faster access to vast amounts of information, data users are demanding more and better quality data. National data sets are being integrated into international data sets in order to understand the complexities of global economies. Users then are becoming more sophisticated in their analyses and the use of integrated data. Consequently, statistical agencies will need to provide increasingly complex data sets. Not only will they need to understand the relationships between the data they produce and other data sets in the public domain, but they will also need to work towards standardizing concepts, definitions and methods.
REFERENCES


Statistics Canada (1999a), Agriculture Economic Statistics, Cat. No 21-603, Agriculture Division, Ottawa, Canada.

Statistics Canada (1999b), Cereals and Oilseeds Review, Cat. No 22-007, Agriculture Division, Ottawa, Canada.

Statistics Canada (1999c), Livestock Statistics, Cat. No 23-603, Agriculture Division, Ottawa, Canada.

Statistics Canada (1999d), Food Consumption in Canada, Cat. No 32-230, Agriculture Division, Ottawa, Canada.

Statistics Canada (1998), PIPES Information Package, Technical Series Number 19, Ottawa, Canada.


THE AGRICULTURE AND FOOD INFORMATION SYSTEM IN SLOVENIA: ITS TRANSFORMATION AND DEVELOPMENT

Irena Orešnik, Mateja Rojc, Metka Zaletel, Simona Dernulc, Helena Puc, Sluga Marija, Joja Krznar, Angela Juvanc
Statistical Office of the Republic of Slovenia
Vožarski pot 12; 1000 Ljubljana, Republic of Slovenia

ABSTRACT

In 1993, the first steps were taken in order to bring the Slovenian agriculture statistics in line with the national needs and also to harmonize it with EU standards. Therefore, on the basis of a set of pilot surveys, a new programme of agricultural statistics was introduced, aiming to meet the above stated objectives. Upon having selected the most appropriate method to be used in view of the preparation of a list frame (i.e., farm register), the most objective and cost effective methods of data collection were determined. As Slovenia is a country with relatively undeveloped administrative structures and a high share of farms, producing only for their own consumption, the situation was such that agriculture statistics was in fact compelled to make use of a system, mainly based on statistical surveys.

Recently the Integrated Administrative and Control System (ICAS) has been introduced in Slovenia. This information system is to be used for the support of subsidy payments and market interventions in the frame of the Common Agriculture Policy (CAP). It goes without saying, that upon full implementation of IACS, some of the statistical surveys will be discontinued. And the agricultural statistics will become an integral part of the national food and agriculture information system, which on the one hand will have to cover the requirements for individual data in the frame of administrative procedures and on the other hand satisfy the needs for aggregate statistical data.

Key Words: agriculture statistics, farm register, data collection methods, administrative data, data dissemination.

1 Introduction

Before gaining its independence, in Slovenia 93% of all agricultural land was privately owned or leased and was cultivated by a large number of small farms. Although state farms cultivated only 7% of agricultural land, they have been favoured by agricultural policy. After having gained independence in 1991, the position of family farms in the Slovenian agricultural policy has changed; their importance increased.

However, the structural changes after 1991 in Slovenia were only minor ones, mainly to be seen in the decrease in the number of very small farms of up to two hectares. Therefore the needs for changes in agriculture statistics in Slovenia derived mainly from changes in agriculture policy, the EU accession process and unreliable methods used in the past.

The introduction of a new program of agricultural statistics was based on a set of pilot surveys, by which the optimum methods were selected and the necessary budget estimated. The objective of the pilot surveys was to define the appropriate method for the preparation of a list (farm register) and to find the most objective and cost effective methods of data collection. In the following chapters the activities concerning the introduction of the new program of statistical surveys are described.

The EU standards were the reference to be met, but in real life the situation in the Slovene structure of agriculture as well as the existing system of agricultural statistics surveys were hardly comparable to those of the EU.

The EU legislation and gentlemen's agreements define the characteristics to be surveyed, the reference time or period, the coverage and the reliability standards. The method of data collection and the sample design are within the decision and responsibility of each country. The EU legislation and gentlemen's agreements define the structural surveys, the land use surveys, the livestock number surveys, the agricultural production and yield surveys, the agricultural output and input prices and the methods of calculation of food balances and agricultural accounts.

The farm structure survey has to be carried out every two to three years on a sample base and every ten years as a census of all agricultural holdings. Every five years countries are obliged to carry out the survey on orchards and every ten years the survey on vineyard plantations. The latest data can be obtained from the administrative register.
of vineyards, but only under the condition that the data meet certain quality criteria. The number of cattle has to be surveyed in June and December, the number of pigs in April, August and December and the number of other livestock each year or every second year in December. Each year, data have to be collected on areas sown and on crop production. Monthly data on slaughtering have to be collected as well as monthly data on prices of agricultural outputs and inputs. Data on milk production and processing have to be provided partly on monthly basis and partly on an annual base. Food balances and agricultural accounts have to be produced annually.

The old system of agricultural statistics in Slovenia was designed and its implementation co-ordinated by the Federal Statistical Office of Yugoslavia. The Slovene Statistical Office had a position of a regional office, whose responsibilities were predominantly directed to data collection and data processing. Since 1991, however, the Slovene Statistical Office is a national statistical office and it took over the responsibilities of methods. Lack of experience in view of methodological work had an inhibiting influence on the introduction of necessary changes in statistics.

In Slovenia, prior to 1991 all data on state farms were surveyed annually or more frequently by means of a census. The structure of private individual farms was surveyed only once in every ten years, as a supplement to the census of population. The latest such census was carried out in 1991. Data on livestock numbers were collected annually by personal interviewing on a sample of farms, selected from the census list. The sample remained unchanged for ten years (panel). The data on crop areas and production were not reliable. Areas under crops were never collected from the farmers, even during a census. They were estimated by employees of the Agriculture Advisory Service. The estimates were based on land cadastre data on land use. The land cadastre information was not regularly updated and it had overestimated the agriculture area, what resulted in the overestimation of both: areas and production of important arable crops.

The relatively bad administrative structures and a high share of farms, producing only for their own consumption, forced the development of agriculture statistics to remain based on pure statistical surveys. The use of data from existing administrative sources is limited only to the use of updating the farm list.

For the future, we are anticipating several changes in the direction of more intensive use of the administrative data, thus replacing the statistical surveys. We anticipate that the main administrative source which we shall be able to use is the Integrated Administrative and Control System (IACS).

2 Preparation of the Farm Register

The construction of the farm register was one of our priority tasks, undertaken in order to obtain a frame for the agriculture sample surveys and for the agriculture census.

In Slovenia the administrative farm register does not exist. The existing census 91 list of farms was without data on areas of crops and even that was not updated. Since some structural changes happened, and the list was not updated, it was clear that the changes were not covered in the census list. The introduction of sample surveys, based on census 91 list of farms, would not provide reliable results. There were two possibilities, either to carry out the census as soon as possible or to update the census list with the existing statistical and administrative data bases. It was decided that we should start first with the compilation of a farm register from the existing data bases.

In order to construct the farm register with the necessary characteristics, the available statistical and administrative sources were studied in detail. The technical characteristics as well as the content and reliability of the data were examined and an appropriate strategy for the farm register construction and updating was chosen.

The most important database on farms in Slovenia is the list of farms, obtained in the 1991 Census of the Population, Households, Dwellings and Agricultural Holdings. The data on livestock numbers could be obtained from different administrative and statistical data sources, whereas the data on the size of the farms and their land use could be obtained from the Census list or from the Land cadastre. Technically it was possible to match those data bases by personal identification numbers (PIN) or by codes of addresses (RTE), which existed in the data bases.

If more data existed on the same characteristic, the most reliable data were used for the construction of the register.
The updated frame was very complete in respect of coverage of the farms. The weak point of the frame was the non-
existence of data on areas of crops by the farmers and inclusion of the farms no longer in operation, the so called
dead farms. These two deficiencies caused problems in the accurate estimation of the size of the sample by which
reliable data on areas of crops on national level could be obtained. Nevertheless these deficiencies did not affect the
estimates based on the samples selected from the new register.

The new register was first used in the pilot survey in 1995 and became fully operational in 1997. In 1998 the
Ministry of agriculture food and forestry (MAFF) started to register the subside payments on the level of individual
farmers. In order to obtain the subsidies farmers had to provide the data on land use and animals reared on the farm.
In 1999 in the data base of MAFF existed already most of the farms of economical importance, which is between 50
and 60% of total population of farms exceeding the EU threshold.

This new source of data became very important for the updating of the list frame. At the end of the year 1999, the
existing list frame was updated with that data base. The constructed frame was sent to the employees of the
Agriculture Advisory Service so that they could delete the dead farms and add the missing farms. Each employee
obtained only data for the farms for which he is responsible in his/her regular work of giving advice to the farmers.
The relatively clean list is going to be used in the Census of agriculture in June 2000.

All the updates and the management of the register was carried out using the SAS language. This is an excellent
tool for all such data management, but proved to be very dependent on having skilled employees to manage this
work. In the beginning of 1998, it was decided that a more independent system of updating should be established.
Therefore a project was prepared, with the aim to migrate the list frame into Oracle and to establish the automatic
updating of the frame with data from statistical surveys and with available administrative data bases.

3 Selection of data collection methods

In order to obtain the appropriate data collection method for data on farm structure and for data on areas of arable
crops, the pilot postal survey on farm structure was carried out. In parallel, the quality of data bases for the updating
of farm register was evaluated.

Slovene farms are small and difficult to access. Due to fragmentation of the land in different parcels and also to low
specialisation of the farms, it is possible to obtain reliable estimates only with large samples. Therefore, data
collection by face-to-face interview is very demanding, time consuming and expensive. A cheaper data collection
method is the postal survey.

Postal surveys of farmers had not yet been undertaken in Slovenia. To obtain information on the suitability of this
method, the pilot postal survey on farm structure was carried out in June 1995.

A sample of about 500 farms was prepared (out of 120.000). A questionnaire with questions on farm structure,
including areas of crops, was prepared and designed to suit postal data collection. The survey was carried out
according to the rules of the "Dillman's Total Survey Design".

The response rate was less than 40%. Part of the non-response was due to frame errors, but it seems that the main
problem was that the questionnaire was too long and too complicated to be answered without any help of the
interviewers.

For all sampled farms, census data (1991) and cadastre data on land owned by their members were obtained as well.

Since data on land use by farms were different in those three data sets, the quality of the pilot survey data on land
use as well as the quality of the data in the census and cadastre were tested by re-interviewing the respondents in
August 1995. All their fields were visited, their area was measured and land use was defined. The pilot, census and
cadastre data were compared with the data obtained by field measurements. Comparisons were done on an aggregate
level for the comparable land use categories.
The conclusions of the June pilot survey can be summarised as follows:

- In Slovene circumstances, the farm structure survey has to be carried out by face to face interviews. The response rate in postal surveys is too low and the quality of the responses insufficient.
- The data on the size of the farms and on land use were most reliable in the census 1991 data base. This database could be complemented by data from the cadaster, from which only addresses of potential farms and their size could be taken.
- The cadaster data on land use were out of date. Since data on crop areas and production were based on cadaster data, they were not reliable. The methodology had to be changed.

Due to budget limitations, the introduction of more frequent surveys on livestock numbers could be possible only if data were collected by post. The postal method was tested. The response rate was in all tests above 70% and the quality of the responses, measured by consistency in time, was sufficient. The results of the pilot surveys were judged as acceptable and the method became operational in 1997.

The census of orchards has not been carried out in Slovenia yet. A pilot census was carried out in 1996. The updates of the list of fruit producers as well as interviewing were carried out by the Agricultural Extension Service, i.e. agriculture specialists. Based on the results of the pilot census, the methodology was finalised and the census was carried out in 1997. The use of interviewers from the Agriculture Extension Service proved to be very good, so the same method of updating the lists and interviewing is planned for the horticultural and vineyard census in 2000. Both censuses will be carried out immediately after the General Census of Agriculture, which will be used to introduce the final updates of the list of the farms which have to be surveyed in the horticulture and vineyard census.

4 New program of agricultural statistical surveys and related problems

On the basis of the pilot surveys, it is possible to compile an adequate farm register from existing data bases, by which the data collection from farmers by means of sample or census surveys is possible. The most appropriate methods of data collection were also determined.

National needs, EU standards and the results of pilot surveys were the background for the preparation of a new program of statistical surveys on agriculture structure and production until year 2003. The budget was also calculated and had to be increased by more than 100% comparing to the situation before 1997. The main reasons for such high costs were:

- large number of small, non-specialised farms have to be surveyed by large samples;
- due to unfavourable age and education level of holders, most of the data have to be collected by face to face interviews.

In the year 1997 part of the new, EU harmonised, program was already realised. Livestock was surveyed in EU reference periods, by post. The orchard census was carried out. The farm structure and areas of arable crops were surveyed by Sample Farm Structure Survey. In order to be able to recalculate the time series for data on areas of arable crops, the areas were estimated also by experts, using the old methodology. Use of expert estimates, based on outdated land cadstre data, resulted in serious errors in estimates of areas and consequently in errors in estimates of the production of most widespread arable crops. Data on potatoes, wheat, grain maize and silage maize were seriously overestimated, area of potatoes for more than 100%.

The approach of carrying out the estimation of arable crops areas by experts in parallel to data collection from farmers in the Farm Structure Survey enable recalculation of time series without losing the data on changes in the last year. The Farm Structure Survey data on crop areas were taken as the correct data for 1997. The annual changes estimated by experts were taken to recalculate the data for the previous years. Through the introduction of the new surveys, most of the urgent needs were met. Except for the orchard survey, the estimates were provided at national level only. To cover the needs for data at local level, the agriculture census is planned to be carried out in this year. By means of the census, the farm register will be updated. The elimination of dead farms from the register and improvement of the data on farm characteristics will reduce the size of samples and consequently the costs of sample surveys after the census will be reduced.
As already mentioned, the new program is very costly. In addition it is also a heavy burden for the farmers and an increase in non-response is anticipated.

5 Future development of the Agricultural statistics systems

When building the Slovene agriculture and food information system, we are going to concentrate mainly on two aspects:
- use of available administrative data for the needs of statistics
- disseminating the agricultural data in time in a user friendly manner and supplementing the data from statistical surveys with the data from administrative and other sources in order to provide a broad picture of agriculture in one dissemination data base.

5.1 Use of the administrative data for the needs of agricultural statistics

The volume of the existing administrative data defines the selection of the appropriate sample frame and the ways of its updating. On the other hand it defines the number and the content of the statistical surveys which need to be carried out in the country in order to fulfil the national requirements and international requirements on data on agriculture and food sector.

The implementation of the common agriculture policy in EU requires a large amount of timely and accurate information on the level of the individual farmer. The Integrated Administrative and Control System is indispensable in managing the subsidy payments to the farmers. Slovenia as an accession country is obliged to establish such an information system, which has to include the:
- register of farms
- register of vineyards
- register of olive oil cultivation
- register of livestock
- register of milk quotas
- annual collection of the data from the farmers at least on the production of products which are the subject of the common agriculture policy.

Through a step by step introduction of IACS in Slovenia, as statisticians we are faced with the duplication of collection of some characteristics by some of the farmers. Some of the data collected by IACS are of a good quality on an individual level, but in statistical terms they are incomplete. IACS does not include all of the data needed for statistical purposes and does not cover all the farms. Data collected within the IACS cover mainly the products which are the subject of subsidising and only those farmers who ask for the subsidies. Farmers providing data for IACS feel that additional statistical data collection is an unnecessary burden. On the other hand, users are becoming confused by two different data for the same products.

The situation asks for proper co-ordination, which is somehow limited by tight timetables for implementation of the Acquis Communautaire (EU regulations) in both domains and also by the need of reliable, representative statistical data for the negotiations with the EU.

For the time being, the stage of development of IACS in Slovenia does not allow much use of its data for statistical estimates. It can make more harm than help to the statistical system. The data collected from the farmers by administration can be effectively used only in preparation of the farm lists, used as a sampling frame and as a frame for the censuses.

We expect many changes in agricultural statistics towards more regular use of administrative data for updating the farm lists (sampling frame) and also as a replacement of certain statistical surveys. In Slovenia substantial differences in comparison with the existing system could be expected in vineyard statistics, milk statistics, livestock numbers and crop area statistics.

One of the instruments for managing quotas is the register of vineyards and a system of annual reports of production by wine producers. Data are reported to those administrative bodies, responsible to the Ministry of Agriculture.
It is assumed that upon full introduction of vineyard registers and EU harmonised annual reports on production, the statistical surveys on vineyards will no longer be required. Administrative data will be used to cover statistical needs.

Milk production is managed in the EU by quotas as well. For its administration, many of the data are collected on the level of the individual farmers, including details on protein and fat content. It means that upon the introduction of a system of quotas in Slovenia, the statistical data on milk could be obtained from this information system.

In Slovenia the preparation of a daily updated database of livestock is under way. The system of statistical surveys on livestock numbers is well established in Slovenia. Nevertheless, when the database on livestock becomes operational, the need for multi-annual surveys of livestock numbers should be reviewed and some of the surveys even discontinued.

With the introduction of IACS in Slovenia, the data on crop areas will be available also in the frame of IACS. If the data from IACS could gain the appropriate coverage, then it would be worthwhile to have close co-operation between the statistical offices and MAFF in the preparation of the IACS questionnaire on areas of crops in order to cover in the questionnaire not only for those crops which are the subject of CAP. In this way the response burden on farmers would be reduced and IACS data could be used for statistical estimates on areas for all crops.

Beside IACS, there are also other administrative data, such as tax records, which can be used for evaluation of the income and social situation of the farms. These topics are covered also with statistical surveys which are not treated as agriculture statistics surveys, such as the Household budget survey. Combining these data with the data from agricultural surveys can provide new information, without additional surveying and can be another step towards reduction of the burden on the respondents. Some of the agriculture questions can be added to other statistical surveys, e.g. in the population census certain filter questions can be added to obtain the farm list.

5.2 Timely, user friendly data dissemination

In the last five years, the changes in the Slovene agricultural statistics were so large and fast that it was difficult to manage adequate data dissemination. The list of the characteristics, the definitions and the data collection methods have changed. Time series were broken and a lot of time was spent to recalculate the data. Some of the data were missing and in general they were published late.

In the year 2000, the whole system will became rather stable again. In addition, an enormous quantity of data will be obtained in the year 2000 - with the census of agriculture, census of vineyards and census of horticulture.

To make the best use of the collected data to the benefit of the user, the new project on data dissemination is proposed. Its aim is to establish the flexible data tabulation system for the censuses carried out in the year 2000 and to establish a user friendly data base of aggregate data from all agriculture statistical surveys. The data base has to include also some important administrative data, relevant for policy analysis, all to provide the analysts with a sound data base for policy and other analysis.

Within the project also the data dissemination policy will be decided upon. It will be decided what will be available free of charge, what will have to be paid and for what fees. The strategy on putting data on a web side will be decided as well as the whole publication plan will be reviewed in line with the new situation in respect to data needs and in respect to the new technical possibilities of data dissemination.

An important tool in data dissemination of agricultural statistics is the classification of agricultural products and land use. In the process of large changes it was impossible to introduce the stable classification of agricultural products. In addition, there is no world standard or at least EU classification of agricultural products. In 1993, the CPA (classification of products by activity) was accepted as a standard in the EU. Based on that, a detailed classification of industrial products was prepared (PRODCOM), but nothing similar occurred in the domain of agriculture. Nevertheless, we have decided to prepare the Slovene classification of agricultural products, based on CPA. This classification will be used in the dissemination data base. No big problems are envisaged concerning the data entered in the data base for the years after 2000, when all the characteristics will be completely harmonised with the
EU. Several problems are expected in view of inclusion of the old data. Some data were already recalculated, while some data are comparable only on higher levels of aggregation. In some cases, the common base can be found only by totals.

6 Conclusions

By finding the adequate data sources for preparation and updating of the list of the farms in Slovenia, the grounds were established for the implementation of the set of agricultural statistical surveys by which data were collected, satisfying the EU and national requirements. By pilot surveys three different data collection methods were selected: postal surveys for agricultural enterprises and co-operatives and for surveys such as livestock number survey where the questionnaires are easy to fill in. The method of personal interviewing by hired interviewers was selected in case of larger and more complicated surveys on land use (areas of crops, sown). In the case of very specialised surveys, such as those on orchards, vineyards or horticulture surveys, the personal interviews, carried out by agriculture experts, employees of Agriculture Advisory Service, proved to be the most reliable method.

Once the methods were decided, the new EU harmonised program of agricultural statistics surveys was implemented. In the year when a particular new survey was introduced, the double surveys of the same characteristics were carried out in order to be able to recalculate the time series. In year 2000 the first round of the new program will be completed by carrying out the general census of agriculture and the specialised survey of vineyards and horticulture. All characteristics will be harmonised with those of the EU. This will enable also the introduction of new the classification of agricultural products and the new classification of land use. It will be based on the EU standards in product classifications; i.e. CPA (classification of products by activity.) This will simplify our future work on enhancing data dissemination tools and procedures.

The possibility of using more administrative data for the statistical estimates in the future is a challenge for the Slovene statisticians. We would have to be able to judge the quality of the administrative data and to decide whether they are of acceptable quality to be taken as official statistics.

In addition to the influence of IACS, changes in the agriculture sector, such as the decrease in the number of self-sufficient farms and the stabilisation of market channels, will reduce the burden of statisticians in respect of data collection. Especially the quality of data on production of meat and crops will be improved by parallel simplification of the system of surveys.

Nevertheless, administrative data deal with the full coverage surveys. The data entry, data editing and data processing are generally longer than in case of smaller sample surveys. By replacing the traditional statistical surveys with the administrative data, the statisticians will have to be in a position to provide early estimates or even forecasts, based on small samples and using fast methods of data collection such us CATI (Computer Aided Telephone Interviewing).

One has to be aware of the particularity of the situation in Europe. Namely, the information system IACS is supporting highly subsidised agriculture in the EU. Two questions have to be raised: how long will this source of administrative data exist and how often will it be changed. In case one decides to use data of IACS for the statistics, one has to be aware that the priority of IACS is to support the administrative needs. On the other hand, the role of statistics is to measure the phenomena in time and time series have to be continued. Therefore the possible use of IACS data for statistical estimates has to be seen also from the perspective of time series continuation.
7 References


Central Statistics Office of Ireland (1993), General description of the implementation of an EC Community supported programme to restructure Irish agricultural statistics including the establishment of a comprehensive farm register.

Council Directive 76/625/EEC of 20 July 1976 concerning the statistical surveys to be carried out by the Member States in order to determine the production potential of plantations of certain species of fruit trees.


Commission Decision 89/651/EEC of 26 October 1989 relating to the definitions of the characteristics and the list of agricultural products for the surveys on the structure of agricultural holdings in the period from 1988 to 1997.

Council Regulation 90/837/EEC of 26 March 1991 concerning the statistical information to be supplied by the Member States on cereals.

Council Regulation 93/959/EEC of 5 April 1993 concerning the statistical information to be supplied by the Member States on crop products other than cereals.


Dillman, D.A. (1978), Mail and Telephone Surveys; The Total Design Method, John Willey and sons.


Orešnik, I., Zaletel, M., Dernulc, S.(1999), Community Typology for Agricultural Holdings, Special Workshop for Candidate Countries, Eurostat, Luxembourg.

Orešnik, I. (1999), The Adaptation of Potential New Member Countries to the EU Statistical System - the Point of View of One of Potential Member, Meeting on Food and Agricultural Statistics in Europe, Meeting on food and agriculture statistics in Europe, Geneva.

DATA INTEGRATION AND ITS ROLE IN THE DEVELOPMENT OF BETTER AGRICULTURE AND FOOD INFORMATION SYSTEMS

by

L. Kabat, L. Naiken and P. Narain

Abstract

Recently agricultural activity has undergone considerable change. Today agriculture involves a large number of institutions and people in an inter-sectoral scenario. While making policies for growth of agricultural production it needs to consider, within the process of overall economic development, the question of its sustainability over a long period of time takes into consideration the social, economic and environmental issues. This requires an information system containing data which has been collected using different methods. This paper presents three aspects of data integration in this context, namely integration of socio-economic data collected in statistical inquiries to derive secondary statistics, integration of socio-economic data collected in statistical inquiries with the information collected from administrative sources, and integration of physical and monetary data.

Key words: statistical inquiries, administrative sources, secondary statistics, physical and monetary data

Introduction

1. Agriculture plays an important role in the economic development of any country. It is a traditional occupation which constitutes a significant share in either the national income or in foreign trade for many developed and developing countries. In many countries transformation of agricultural products is also an important component of manufacturing industries. Recently agricultural activity has undergone considerable change. Today agriculture involves a large number of institutions and people in an inter-sectoral scenario. Governments are seeking to acquire food and fibre for the country taking into account their domestic production, export/import possibilities as well as the status of foreign exchange. While making policies in this context, it needs to consider the nutritional status of the population and the conditions of its weaker sections and minorities. Another important aspect of this sector is that it heavily relies on natural resources. Issues of sustainability and environment are now directly and indirectly linked with the decision making process. The economic and social importance of agriculture coupled with the sector’s dependence and impact on the environment highlight the need for a well organised information system indicating its linkage with other sectors to support decision making. This paper presents a global picture of data needs for a statistical information system for the use of food and agricultural decision makers and goes into some of the important aspects of data integration.

Information system needed by the food and agricultural decision makers

2. In common parlance information is “details” or “knowledge” of any specific activity/event or situation which is not necessarily qualified by its coverage or reliability. However today specialised information systems are being created for targeted audiences presenting structural information. According to Chambers English Dictionary the science of information relates to “processing and communication of data for transfer of knowledge” and system means “anything formed of parts placed together or adjusted into a regular and connected whole”. Taking the two things together an Information System is a group of different databases and other documentation put together in a systematic manner for transfer of knowledge and communication of data on any subject. Such a system is generally created with a view to:

a. identify and analyse problems,
b. set priorities and formulate policies and program, and
c. monitor and evaluate policies and program performance.

3. An information system can be either in the form of some textual information including data that are static in nature or bibliographic information and meta databases or consisting of a set of statistical databases. The paper
concentrates basically on the last type of the system which contain statistical data. Statistical data are numbers with known concepts, coverage and reliability. These are operationally meaningful through the levels of aggregation, valuation, basic units of count and weighting schemes used to define them. These characteristics are determined by some theoretical background and with some analytical aim in mind. These data are provided either in the form of a table in a publication or in a magnetic media from which a specific selection can be made. In such cases when it is available through any magnetic media generally it can be downloaded to the user's PC for further analysis. In general a typical information system would have the following structure:

![Figure 1: An overview of Information System](image)

4. The structure of the information system presented above makes it clear that the process of data integration is part of the system. This is necessary for giving a closer look at some issues, which are of social, economic or scientific nature and therefore require data originating from different sources. In decision making process there are occasions where one need to derive certain ratios/indicators such as share of female labour force in total labour force or any other ratio which can be derived using data collected/generated within one source for that matter. The present paper treat all such matters as part of primary data processing activity and does not include them while discussing issues relating to data integration. Some of the very common examples requiring data integration are creation of secondary statistics such as index numbers, national income, food availability, etc.

5. Such a system would not only be useful to monitor the current trends of growth of agricultural production within the process of overall economic development but also for relating various data sets for use of policy makers in taking decision on sound use of agricultural inputs and technology of production which takes into consideration the social, economic and environmental issues and long-term sustainability. Such an information system that could provide support to decision makers in the formulation of policies requires large amount of statistical data relating to (a) economic forces (e.g. production, input, prices, wages, food consumption, macro-economic aggregates, aids and assistance) (b) human resources (population, labour force) (c) natural resources (soil, water, climate). Data on these factors are collected using various methods and techniques that may also sometime need special high-tech scientific
methodology and instruments. In order to make the information system user friendly it is necessary to supplement the system with derived indicators and secondary statistics like domestic product, food availability. One of the common constraints for creation of such a system is that the various data components used in the system have different concepts, coverage and reliability. The statistician as a producer and manager of such an information system is faced with the challenge to provide a comprehensive, reliable and consistent picture through a set of “statistical data”. To accomplish this task it is essential to integrate various data sets. The task of integrating various sets of data depends on the sources which have been used for collection of the data.

Main sources for core data

6. The statistical system of any country has usually been developed over the years primarily to meet the needs of the country’s administrative set-up. In any statistical system including an agricultural statistical system, part of the data originates from information contained in different administrative (official) returns which individuals and corporations submit to the government while a part of data are collected through sample surveys and censuses organised by the statistical counter-part of the administration to meet special needs. As mentioned earlier, agriculture being heavily dependent on the natural resource base, also requires data on natural resources which have a significant impact on environment. With the technological breakthrough in the field of information science and growing information need of policy makers to manage the fragile ecosystem, satellite based technologies like remote sensing techniques form another important source of data which may be different from other socio-economic data sets. Thus there are three main kinds of sources which provide core data for developing an information system. Each of these sources has different role in developing a statistical system as well as has their own limitations. It is therefore essential to analyse and adjust the databases to create an integrate system. In the following discussion we describe roles and limitations of these three kinds of sources.

(a) Administrative Sources

7. Administrative set-up of each government department, depending upon their functions, regularly collects information on different aspects for meeting their own day-to-day needs. Most common example of this type of information is the information on ownership of land, market arrivals of agricultural commodities, collection of direct and indirect taxes, quantum and value of goods and services exported and imported, birth & death, prices of different commodities in different type of markets, number of different types of vehicles on road, number of residential and non-residential buildings, etc. In most of the countries with centrally planned economies (like Eastern European Countries) a large amount of information on economic activities is kept in administrative registers could also be included in this category. The basic objective of collecting such information is to monitor the tax collection system or making economic and financial policies for the country. Some of the basic characteristics of this source of information are:

(a) Cost of data collection is relatively negligible as the information flows through its normal administrative channel. As this information is collected using normal administrative system, the collected information also contains information from remote and inaccessible areas.

(b) Theoretically administrative data should be more reliable than the data collected in censuses and surveys. This is mainly because (a) the format for collection of information is fairly known to respondent, and (b) the process of collection normally has a strong legal backing and consequently the system permits the person responsible for collection of information to review and verify accuracy by going through basic documents. However, at times two sets of data, although collected and compiled using same concepts, may result into different estimates. Sometimes the reasons for these discrepancies are known (e.g. time or place of recording), but sometimes it may not be feasible to reconcile or explain these differences.

(c) The objective of collecting administrative information is different from the data collection process adopted in statistical inquires. Administrative information may or may not have well defined concepts and definitions.

8. Making effective use of administrative data for statistical purposes is a challenge. The statistical system’s needs are not normally high on the priority list of most administrative information system developers. The latter are relatively rigid about the content and methods of collecting information. Often, it is not feasible to incorporate additional information in the “form (administrative return)” for adjusting concepts and definition of the information to meet the statistical needs. In most of the cases in the countries with centrally planned economies sometime, users had reservations about its quality. Never the less administrative information has numerous uses in developing a sound statistical information system:
The information could be used for creation and maintenance of a frame for collection of statistical information and as an aid for improving sampling design. Information could be used to make clusters or for stratification of “the population” for creating homogeneous groups.

The information could also be used for validating data collected from other sources.

The information could also be used for making assumptions for imputing missing data either collected in a survey/census or supplement the data collected in a survey/census.

The information could also be used to reduce respondent burden.

9. However, to make this information useful for developing a statistical information system it is necessary to (a) make adjustment in the concepts and definitions using information from other sources or making assumptions, (b) re-tabulate the information according to the coverage required under the information system, and (c) compile averages to present the true picture by making the series continuous and giving appropriate weighting system.

(b) **Statistical Censuses and Surveys**

10. Statistical censuses and surveys have long been used for collection of data on various aspects. These statistical inquiries are conducted to collect various data on population, labour force, nutritional status, poverty, manufacturing establishments, research and development activities, etc. Traditionally the census is a means to collect data with comprehensive coverage (complete enumeration) while survey techniques have been developed to reduce cost and time in collection of data. Agricultural censuses and surveys have long been used for collection of data on the use of agricultural land and labour, agricultural production, agricultural inputs, livestock, agricultural machinery, cost of production, farm management practices, etc. FAO every tenth year provide guidelines for developing agricultural census programs by making suggestions for improving coverage and concepts to be used by countries for achieving international harmonisation of census results. These guidelines also help in establishing linkages with data being collected and used by meeting various needs. The FAO has also developed special handbooks to deal with data collection on specific issues such as labour force, post harvest losses, agricultural production, agricultural prices paid and received by farmers, aqua-culture, etc. Sufficient literature is also available on using holding or household as the sampling frame or creation of sampling frame based on remote sensing data. Apart from agricultural censuses and surveys there are number of such sources which provide direct/indirect information on matters related to food and agriculture which could be used for developing an information system. Some of the well known issues for collecting data through these statistical inquiries are:

- The collected agricultural data can be incomplete in terms of (a) the range of commodities covered (for example, in many cases only cash crops or large farms are covered only), (b) coverage of the nation can be partial (sometimes parts of the country are excluded from the national statistical reporting system), and (c) to maintain cost or for other practical reasons some time limits are imposed in these inquiries in terms of volume or value of the output, number of livestock, size of holding, number of trees, or area to be covered etc.
- Data collected in various surveys/censuses follow different sampling frame or do not relate to same population, making any integrated analysis difficult.
- Concepts and definitions used in different statistical inquiries are very often different.
- Sometimes there are aggregation issues due to differences in level of non-response and precision.
- Even when data are available, their reliability is often questionable due to presence of sampling and non-sampling errors.

(c) **Data collection using satellite system and remote sensing techniques**

11. In late seventies and early eighties sustainability of agriculture was considered as one of the important area for decision making. It was thought that agriculture may be treated as an industry rather than as a traditional occupation. Agricultural practices may be developed to maximise the socio-economic benefits and should aim at (a) raising the level of farm productivity which could be sustained or enhanced over a period of time, (b) minimising adverse impacts on the natural resources, and (c) reducing use of agri-chemicals which may effect quality of soil or ground water. This aim needed an information system which could take into account quality of land along with risks associated with the vagaries of climate. Since quality of land and soil are closely related to various acts of human activities and impact of climate, the agricultural information system need data not only on the principal forms of soil degradation (water erosion, wind erosion, soil acidification, salinisation, soil structure, etc.) but also data on host of
variables such as emission of greenhouse gases, temperature, rainfall, cloud cover, sunshine hours. These sets of data which are highly scientific in nature are generally collected using satellite system and remote sensing techniques. It is evident that these data are different from other socio-economic databases in terms of unit of measurement as well as area to which they are related. In spite of this basic difference in the nature of such databases, it is useful to have an integrated view to make plans and take policy decisions.

Data integration for creation of an information system

12. Looking at the nature of the data originating from the three types of sources and the needs of decision makers it is obvious that it is essential to integrate these data sets. Data integration issues could be broadly divided into three classes: (a) integration of socio-economic data collected in statistical inquiries to derive secondary statistics (e.g. GDP, food availability), (b) integration of socio-economic data collected in statistical inquiries with the information collected from administrative sources, and (c) integration of physical and monetary data (e.g. satellite accounts) to study issues relating to nutritional status, environment, etc. Some of the important aspects of these three data integration processes are given below:

\( \text{(a) Integration of socio-economic data collected in statistical inquiries to derive secondary statistics} \)

13. Derived secondary statistics such as index number of prices and production, macro-economic aggregates, per capita total consumption, etc. which are constructed using a set of ‘statistical data’, have well defined concept and coverage. These statistics require data on large number of items which normally cannot be collected in one inquiry. Thus while deriving such secondary statistics it is essential to integrate socio-economic data collected in different statistical inquiries. For example, national accounts (or input-output table from which national account originates) is one tool which facilitate this process of data integration. There are two important characteristics of these data sets which need to be considered for integration of such data sets: (a) For many items data may be available from more than one source requiring decision to choose among the alternative sources, while for some items no data may exist, and (b) these derived statistics are normally produced as time series, but some of the essential data may be becoming available only at certain period.

14. In cases where more than one set of data exist, one has to compare alternative sets of data for (a) coverage (geographical, items included), (b) concept and definition, (c) time period to which the data relates, (d) sampling frame (if applicable) (e) responding unit, and (f) possible sources of sampling and non-sampling errors, and select those sets of data which are more closely related to the concept and coverage of the targeted statistics and identify the gaps which need to be filled by imputation through integration of information from other sources. One of the simple examples of such data integration is preparing estimates of agricultural production. Some of the obvious questions are:

- Whether the data cover the total geographical area of the country and all items (crop, livestock),
- How agriculture has been defined, e.g. taking only crops and livestock or products and by-products are also included or if it includes standing crops,
- If the data are related to agricultural year/calendar year/financial year,
- If the data cover all holdings or only households or only large farms

15. Depending upon such information one has to look for some other suitable sources which provide information for imputing the missing links. Similar analysis is also required to locate gaps when data on some part/item are not available. However in such cases one also need to look for sources for providing information on ‘proxy variables’\(^2\) which meet some of the criterions as indicated in the above paragraph, (e.g. coverage, concept, definition, time period, sampling frame, responding unit, etc.). In cases where only periodic estimates are available, generation of regular time series also requires similar information on proxy variables which can present true movements over time. It is feasible to understand

\(^2\) In the absence of complete data for the compilation of various flows or the details of disaggregation sometimes use is made of data available from pilot studies, case studies or other sources giving partial information. The most common method for using such information is to work out some required ratios and estimate the population value using a variable for which information is available for the population as well as from the study. These ratios are known as "Structural Ratios".
reliability and margin of errors while using national accounting\(^3\) structure to integrate different data sets originating from different sources. This is mainly because national accounting structure is based on a double entry system as well as identities based on economic theory. “The central framework of the SNA presents a number of characteristics which give it the advantages of an integrated accounting structure. It is exhaustive and consistent within the boundary of the economic activities it covers; that is to say, each unit, transaction, product and purpose is given a place, and only one, in the classifications and accounts of the System. Moreover, the set of concepts adopted by the System is fully coherent. (SNA, 1993)” Integrating the results from all kinds of censuses and surveys into a consistent framework may increase both their relevance and their reliability. Integration of basic data enhances the possibility of more policy issues being monitored and analysed inter-relatedly.

(b) Integration of socio-economic data collected in statistical inquiries with the information collected from administrative sources

16. To minimise the cost of collecting data many countries employ a mix of data compiled from administrative sources and the data collected in statistical inquiries. The degree to which this approach is adopted depends on a number of factors such as awareness of the statistician about the information being collected by their administrative counter parts, storage and retrieval system for maintaining information, importance and frequency of the information being collected, confidentiality provision and co-operation among participating agencies. The process of integration also depends basically on the ultimate objective of the final product. One of the most useful illustrations of such an integration is to derive per capita food availability to assess the nutritional status at national, regional and global levels. Normally a large amount of data would be required for estimating per capita food consumption. However, by integrating statistical data on agriculture production (food crops, livestock and fishery products) with administrative information on items such as foreign trade, non-food uses of food products FAO derives per capita apparent consumption for various countries and published in the form of Food Balance Sheets\(^4\). In spite of their limitations, food balance sheets represent the only comprehensive source of world-wide information on food supplies. Their primary contribution is to provide statistical information quantifying world supplies, as well as to reflect changes that have occurred in per caput dietary energy supplies and food supplies patterns. The national aggregates are suitable for assessing the overall shortages and surpluses in the food supply of the country.

(c) Integration of physical and monetary data

17. The utility of national accounting framework in integrating various data sets has already been discussed. There are certain limitations in what can be accommodated in national accounting framework as the accounting structure is based on monetary system. However, in several cases additional or different requirements necessitate the development of complementary or alternative categories and concepts for collection of data. The 1993 SNA realised the need to expand the analytical capacity of national accounting for selected areas of social concern in a flexible manner, without overburdening or disrupting the central system and recommended use of satellite accounts. The satellite accounts are useful in integrating physical and monetary data. On the one hand, satellite accounts are linked with the central framework of national accounts and through them to the main body of integrated economic statistics. On the other hand, as they are more specific to a given field or topic, they are also linked to the information system specific to this field or topic. In this manner it allows for better integration of monetary and physical data.

18. In a satellite account it is desirable that the concepts, accounting structure and classifications should be tailored to the specific analytical purposes which facilitate data integration. Satellite accounts can focus on the role of people or natural resources. This calls for extra breakdown of economic sector for conceptual and numerical linkages among all kinds of related monetary and non-monetary phenomena, which may be expressed in different measurement units.

\(^3\) When a comprehensive set of accounts is prepared using a variety of sources, certain discrepancies may emerge between the two sides of an account. These discrepancies may arise because of incompatible assumptions used in integrating various data sets. The size of the discrepancies reflects the degree of consistency of derived statistics from national accounts.

\(^4\) The food balance sheet presents data on: selected food products/items in quantity terms, giving distribution of domestic supply and domestic utilisation by broad categories; total per caput supply per year in quantity terms; and per caput per day supply in terms of quantity and also in terms of caloric value, protein and fat content (calculated by applying appropriate food composition factors for all primary and processed products).
Concluding Remarks

19. To understand the dynamic nature of the economic system in terms of the social, economic and demographic characteristics of the population, it is essential to integrate information/data on various aspects. This task requires some basic considerations for collection and compilation of statistical data. Some guiding principles to achieve this in a cost-effective manner are to establish linkages between data at micro level and areas requiring decision making. This could be achieved if:

(i) The stratification for the collection of statistical data are the same as that of the administrative regions for which policies are generally framed. The coverage of both should be examined and stratification may be adjusted so that it should be possible to construct data sets for administrative regions.
(ii) The collection of statistical data should keep in view the need for uniformity with the decision making process.
(iii) The concepts and definitions used in the collection of statistical data should be examined carefully and matched with those required in the system for making administrative decisions. For example, it may be necessary to include some extra details in statistical inquiries that may have help in making adjustments to the administrative data.
(iv) The derived secondary statistics generally include some imputed values for non-market activities. The basic assumptions followed in these imputations should be examined and checked with corresponding administrative information if available.

A statistical system conceived in this way would be well suited for the integration of statistical data and administrative information for analytical purposes. Another alternative to integrate physical data with other monetary data is the use of the Geographical Information System (GIS).

55 Discussant Comments:

Mr. Rich Allen and Fred A. Vogel* have suggested, a reordering of guiding principles. The suggested order is: 1. consider the decision making process before designing the data collection, 2. concepts/definitions need to be consistent with the administrative requirements, 3. sampling should be done by administrative regions, and 4. allowances must be made for items which don’t move through markets.

Discussants have also suggested three additions to this list, as indicated below:

1. “The data requirements and statistical methods should be coordinated across the different levels of government.” In many countries the provincial governments produce statistics to satisfy their decision making requirements and do so independently of the central government. Meanwhile, the central government is also producing statistics. Confusion results when the summation of data across provinces differs in level and concept from the centrally derived data. The U.S. approach to this problem is described in Vogel (1995). A similar approach was also followed in Canada (Trant and Whitridge, 1998).

2. There should only be a single source for each set of “official” statistics.” In many countries, there is a central statistics office providing agricultural statistics. In addition, there may be a ministry of agriculture also providing agricultural statistics resulting in parallel data sources differing in quality, concepts, and actual levels. Vogel (1999) describes such a situation in China with suggestions on how to resolve the problem. This problem existed in the United States until 1997 when the Census of Agriculture was transferred from the Census Bureau to the U. S. Department of Agriculture.

3. “There should be a data dissemination policy that provides the public open and equal access to all data in the government information systems.” The paper by Oresnik, et al., provides a good overview of these issues.
References:


* Associate Administrator and Deputy Administrator for Programs and Products, respectively, National Agricultural Statistics Service, U.S. Department of Agriculture. All opinions expressed are those of the authors and not necessarily those of the U.S. Department of Agriculture.


Introduction:

The three papers complement one another and provide a balanced view of the many issues involved when integrating data for information purposes. They range from a rather formal summary paper to a practical presentation to a case study of improvements in one country. The paper by Oresnik, et al., provides an excellent overview of the approach that the Republic of Slovenia has taken and its experiences in transforming and developing a statistical system. Some suggestions based on experiences in other countries will be presented. The Elward - Holland paper uses the U.S. and Canadian data systems to describe four different approaches to data integration. This paper outlines what these two large developed countries have done in the past or are doing now, but does not address the many structural changes occurring in agriculture that deserve attention. Some questions will be raised for Elward & Holland to consider. The paper by Kabat et al., provides an excellent global perspective on requirements and integration issues faced when combining data from many sources. An especially valuable contribution is their concluding recommendations for guiding principles to follow when providing linkages between different data bases. More will be said about the guiding principles.

Transformation and Development of Information Systems:

The Republic of Slovenia has outlined a long-range goal to meet all EU information flows. They are currently evaluating the standards in light of a new government agricultural administration system for subsidies and market interventions which creates a new data source. The paper outlines various data approaches that Slovenia has tested and evaluated. For example, they learned that a postal service approach will only work with very simple, short questionnaires.

Slovenia appears to be following the doctrine of many countries in their situation which means they are adopting the use of a farm register. A farm register is a critical part of a statistical system, but not the answer to all problems. Registers are virtually impossible to be kept complete, up to date, and free of duplication and dead wood. The U.S. spent millions of dollars to develop a farm register for the 1997 Census of Agriculture. Coverage evaluation showed that over 13 percent of the farms and 2 percent of the land in farms were missed. A sound statistical system needs a process to supplement the farm register. The FAO handbook on multiple frame surveys provides an in-depth approach to supplemental farm registers. Fecso, Tortora, and Vogel (1986) describes the U.S. approach while David (1998) presents a sampling strategy for censuses and surveys.

Oresnik, et al., are to be congratulated for addressing the issue of data dissemination. This is too often overlooked by countries when developing their statistical programs. Timely, user friendly data dissemination to all interested parties should be a guiding principle.
The Changing Face of Agriculture:

The four different definitions of integration illustrated by Elward and Holland are: 1. using supply/disposition balance sheets for commodity estimates, 2. combining production data or financial accounts in order to create other accounts for publication, 3. integrating different data collections of the same population into one survey vehicle to reduce respondent burden and improve analytical capability, and 4. combining multiple surveys into one survey program for efficiency and coherence. Example 2 is particularly critical since some measures such as Gross National Product can never be surveyed directly but must be derived from other information. Some comments follow about issues for consideration by the authors.

The agricultural sector is becoming increasingly vertically integrated with one ownership entity controlling all phases from seedstock production to the final processed product. Where in the pipeline do agricultural costs and returns begin and end?

The incidence of contract arrangements is rapidly increasing. Using pork as an example, the operator provides facilities and labor. The contractor provides the animals, feed, and other supplies. When the animals are marketed, the farm operator receives a fee per head or pound gained. The profit or loss accrue to the contractor. What is Farm income? Is it the fee the farm operator received, or the cash receipts minus expenses as borne by the contractor?

The U.S. and Canada both equate one farm enterprise with one household. This is a serious flaw in farm income accounts and deserves some consideration in the future.

Guiding Principles:

The Kabat, et al., paper contains many good practical suggestions. As businesses change and social and political priorities continually evolve, data systems must be constantly re-evaluated and adjusted—which ties into comments on the U.S./Canada paper. The Guiding Principles in the paper are a valuable contribution. However, a reordering might be an improvement. To paraphrase the principles in the suggested order we have: 1. consider the decision making process before designing the data collection, 2. concepts/definitions need to be consistent with the administrative requirements, 3. sampling should be done by administrative regions, and 4. allowances must be made for items which don’t move through markets. Three suggestions for additions to their list follow:

1. “The data requirements and statistical methods should be coordinated across the different levels of government.” In many countries the provincial governments produce statistics to satisfy their decision making requirements and do so independently of the central government. Meanwhile, the central government is also producing statistics. Confusion results when the summation of data across provinces differs in level and concept from the centrally derived data. The U.S. approach to this problem is described in Vogel (1995). A similar approach was also followed in Canada (Trant and Whitridge, 1998).

2. “There should only be a single source for each set of “official” statistics.” In many countries, there is a central statistics office providing agricultural statistics. In addition, there may be a ministry of agriculture also providing agricultural statistics resulting in parallel data sources differing in quality, concepts, and actual levels. Vogel (1999) describes such a situation in China with suggestions on how to resolve the problem. This problem existed in the United States until 1997 when the Census of Agriculture was transferred from the Census Bureau to the U. S. Department of Agriculture.

3. “There should be a data dissemination policy that provides the public open and equal access to all data in the government information systems.” The paper by Oresnik, et al., provides a good overview of these issues.
**Conclusion:**

We are just scratching the surface on dealing with problems related to integrating statistical systems. The globalization of the world economies is calling for integrating statistical systems across countries. Multi national companies operate their enterprises across continents. The assignment of the resources, costs, and returns to specific countries and sectors will pose many difficult problems.

**References:**


