Integration of a State specific sampling database with the NASS List Sampling Frame: a Case Study of the California Fruit Acreage Database

Stan Hoge, National Agricultural Statistics Service, Washington, D.C.

Abstract

Recently NASS has moved to centralize the national List Frame as part of an operational efficiency plan. There have been issues with agency efforts to standardize metadata and maintenance procedures in relation to Field Office maintained administrative and/or standalone databases and how this information can be effectively utilized in a centralized environment. This paper discusses issues specific to the California Fruit and Nut Acreage Database and how to maximize the use of this valuable data. The paper explains how the California Field Office has established the Fruit and Nut Acreage Database utilizing data through detailed county assessor information. These data are unique to California in that each land parcel has extensive information on crop(s), variety, year planted, planting system(s), tree or vine spacing , the number of trees or vines and the parcel acreage. Each parcel has information on the owner and/or operator which is used to link to the NASS List Frame. The List Frame is NASS's primary sampling frame for state and national surveys and for the Census of Agriculture although the CA FO also utilizes its Fruit and Nut Database for sampling state and locally funded Fruit/Nut and Objective Measurement Surveys So, the Fruit and Nut Database is a separate stand alone database that in most cases is used as a sampling frame and it is maintained in parallel to the national List Frame. The issues discussed in the paper are where existing maintenance and metadata procedures conflict with new standards being established for the national List Frame.

Keywords: Assessor Parcel Number, Operating Arrangement

1. Purpose

The National Agricultural Statistics Service (NASS) is in the process of a major reorganization which will involve the centralization of processes and the regionalization of NASS's 45 Field Offices (FO) to nine regional centers. Currently, the California FO maintains a separate Fruit Acreage Database (FAD) that is a critically important tool to NASS in providing accurate and timely fruit and nut production statistics. The records in this database are land based parcels containing detailed commodity acreage data reported by growers. As NASS moves towards regionalization, one of the challenges is to integrate state administrative record systems into a centralized approach.

This paper documents the development, maintenance, and use of the California FO Fruit Acreage Database (FAD) and how the database and the associated applications could be integrated into a more centralized processing environment in line with the NASS reorganization.

2. Description and Background of the Fruit Acreage Database

California is the largest agricultural state in the nation as measured by value of sales of agricultural products and leads the nation in agricultural diversity with more than 400 commodities harvested each year. California's agricultural abundance accounts for nearly half of the fruits, nuts, and vegetables grown in the U.S. Fruit and nut farms account for just over fifty percent of all farms in the state and more than one-third of the state's total cash income. As part of the agriculture estimating program,

USDA's NASS California Field Office maintains the Fruit Acreage Database (FAD), a database of orchards, vineyards and fields planted in fruit and nut crops throughout the state.

The FAD is organized into land ownership units, identified as area parcels, then further subdivided into blocks. Each block is a reported area of land identified by a unique combination of land management, crop type and variety, planting year, and planting system of the trees or vines. In addition to these attributes, the FAD includes a link to the NASS national List Frame database by operator identification number. . The FAD includes data for approximately 58,000 parcels and 224,000 blocks.

Funding in support of the database is provided by a series of longstanding cooperative agreements, some dating back to the 1960s, with state government and agriculture industry groups. Each year, these agreements provide funding for detailed annual acreage reports for grapes, almonds, walnuts, dried plums, and citrus. In addition, the database provides a data rich sampling universe for state and industry funded early season production forecasts for raisin grapes, almonds, walnuts, and oranges. These forecasts are based on fruit and nut counts from randomly selected trees within sampled area blocks. These are among the most important reports prepared by the CA FO, as they serve an integral role in the development of commodity marketing decisions impacting both domestic and foreign markets.

3. History and Development

The origins of the fruit acreage database date back to the late 1930s. In order to meet the growing and specific needs of California's agriculture industry, fruit and nut acreage estimates were prepared jointly by the NASS California Field Office (then known as the California Crop & Livestock Reporting Service), the California Department of Food and Agriculture, and the county Agricultural Commissioner offices. Beginning in 1948, complete enumerations of commercial parcel level acreages were made in selected counties on a rotational basis. Funding was provided under the Federal Research and Marketing Act of 1946 with state matching funds.

In the mid-1970s, the paper based records system was converted to an agency-wide NASS mainframe computer system. Maintenance and updates to the area parcel records were completed within a controlled mainframe processing environment. The data were only available for review on microfiche and consequently only as current as the last microfiche print. This advancement coincided with the initial development of the NASS List Frame and first enterprise wide mail and maintenance system. In 1998, with advances in the development of local area networks (LAN), the mainframe processing environment was converted to a local in-house IT environment.

Under the new processing environment, the database is managed in Visual Fox Pro. Daily and continuous access to the FAD is provided for multiple authorized users. Users are able to add, delete, or edit the data from their workstations. Changes and updates are captured in a history file for later review if needed. The updated information is immediately available for review in real time. In 2002, the three major application components of the fruit acreage database (Lookup, Update, and Processing) were completed by CA FO IT staff. All end user surveys and projects are now conducted using these system database applications. They are menu driven utilities that escort the user through the desired tasks based on the specific input from the user. The FAD also provides a link to the NASS list frame for each record. The ability to connect individual parcels and blocks of land to records on the NASS list frame allows the CA FO to customize data collection instruments and other data collection processes for individual producers. More than 14,000 contacts with farmers are made each year as a result of the surveys drawn from the Fruit Acreage Database.

4. Primary Functions of the Fruit Acreage Database

Much of the California Fruit and Nut statistical program depends directly on the data provided on the FAD. Acreage information from the FAD is utilized for sampling fruit and nut acreage surveys. Variety, location, and year planted information are used to select samples for objective measurement studies for almonds, walnuts, separate varieties of oranges and raisin grapes. Up-to-date operator reported information on the FAD allow for the direct publication of acreage by variety and year planted. In addition, the FAD is also used administratively during the review and analysis of the Census of Agriculture data, and for Fruit acreage surveys in determining non-bearing fruit acres and to provide reference data for various other NASS surveys including the Fruit Chemical Use Survey and the Agricultural Resource Management Survey.

5. The Relationship between the Fruit Acreage Database, the NASS List Frame and ParcelQuest

Each record on the FAD represents a land based parcel and each individual parcel aligns with a county assessor parcel number (APN). Each parcel record includes detailed information on the owner, operator, type of crops, variety of crop, the year the tree or vines were planted, spacing, number of trees or vines and the parcel acreage. This data is reported by growers throughout the year via state surveys. Owner and Operator information are pulled from the NASS List Frame on a daily basis to update the FAD. The link between the parcel on the FAD and the NASS List Frame is the unique Person Operation Identification number (POID). The Operator's POID is the unique reporting unit identifier for most NASS surveys and all national program list samples. This amount of detailed parcel by parcel coverage for the entire state is unique to California and offers an expanded resource to enhance the statistical program for California fruit and nut production.

One use of having the FAD parcel records linked to the list frame by the POID allows for acreage on the Fruit Acreage Database to be summed to the commodity level and input into the NASS generalized sampling system (GENESIS). With over half of all active California farms growing fruit and/or nuts, the FAD provides an excellent source for maintaining acreage data on the list frame. The detailed parcel

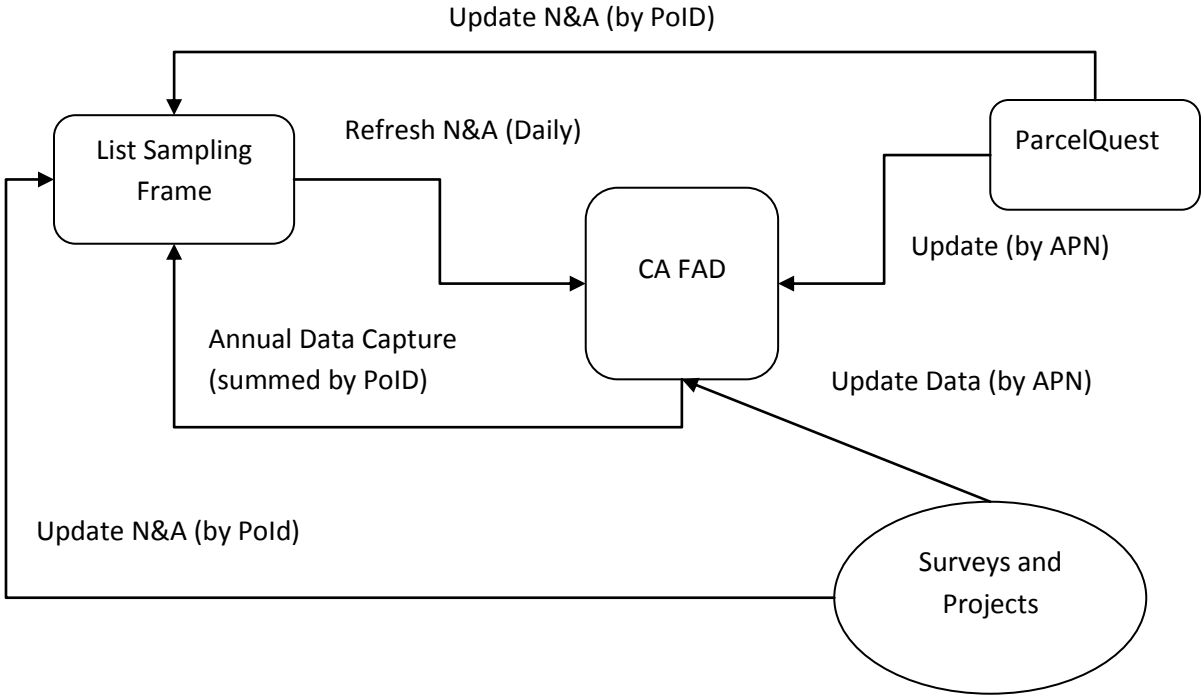

Figure 1

information on the FAD is not maintained on the list frame. Operators on the NASS List Frame may be associated with multiple parcels on the FAD.

The FAD, in conjunction with ParcelQuest, is an effective source for list building and maintenance. ParcelQuest is a proprietary electronic database of California County Assessor parcel records and is used to track land ownership changes through manual lookups by APN. Land-based acreage surveys are conducted throughout the year, so new operators and changes in operators are discovered and verified through both the survey and the ParcelQuest on a nearly continuous basis.

A combination of previously reported data on FAD and current parcel data on ParcelQuest are used to help minimize duplication on the NASS list frame. Because both are land-based, the connection of a single operator to a specific piece of land can be established. An example where this sort of relationship is useful occurred during the 2007 Census of Agriculture. Review of FAD and ParcelQuest information provided an explanation of the difference in acres of navel oranges reported on the Census with the acres of navel oranges reported on the annual Fruit Acreage Survey and stored on the FAD. Reporting discrepancies and possible duplication were effectively identified, researched and addressed.

Name and address changes identified from returned fruit and nut survey forms are used to update the NASS list frame through the usual protocol for maintaining these attributes. Names and addresses are maintained only in the NASS List Frame. The list frame name and address information is then posted to the FAD at the beginning of each workday using the POID (see Figure 1). The FAD is a database of land units or parcels which are linked to the list frame for all information about the operation. In this way, the FAD behaves as a discovery tool for keeping the list frame names and addresses current for fruit and nut operators and the associated operating arrangements. Currently, name and address changes are processed through the Enhanced List Maintenance Operations (ELMO), which is the system of applications for the maintenance of the list frame. In this way as ELMO transactions are received, the List Frame is in sync with the FAD. Timeliness of the updates for the contact information and other associated operation changes are critical for the follow-up mailings and phone calls. Timing is also important for Objective Measurement Sampling and preparing permission agreements.

6. Established procedures from Fruit Acreage Database to the List Frame

Data in the FAD are captured once per year for input into the NASS generalized sample select system (GENESIS). Individual parcel data are edited and updated continuously in the Fruit Acreage Database from annual fruit acreage surveys, but "captured" once per year during annual sample selection process. This capture of reported data represents a snapshot in time or what NASS refers to as a frozen frame. This is consistent with standardized procedures for all other surveys conducted by NASS. Each parcel in the FAD represents a specific piece of land which has a unique Assessor Parcel Number (APN) assigned by the County Assessor Office. The APN helps guarantee that each piece of land is reported only once in the FAD. At the time of data capture, each parcel's fruit acreage is summed by commodity to a total for each Operator. Operator POIDs are added and maintained in the list frame. Transaction records are then created in the standard NASS format for submission into GENESIS, summed by commodity to the Operator POID, and uploaded for processing. Transactions from the FAD are then evaluated during the GENESIS process against all other data captures of fruit acreage surveys (including Fruit Chemical Use Survey, ARMS, Organics, state Non-Citrus Survey, and other surveys conducted during the past year).

The Fruit Acreage Database is completely integrated with the NASS program. Because all names, addresses, and POIDs come from the List Frame, all list maintenance decisions are made under the standard NASS list maintenance procedures – whether the initial change is coming from a NASS program survey or a State Fruit Acreage program survey. This is similar in concept to existing Multiple Frame maintenance for the NASS annual survey program. The Multiple Frame is used in the national survey program where a second independent Area Frame is sampled and surveyed and the results are used to

measure for the incompleteness of the list frame. The Area Frame Survey is conducted in June and as mentioned previously the List Frame samples are selected from the "frozen frame" close to the June reference date for a comparison to the Area Frame as it existed in June. One major difference is that the list and area frames for the Multiple Frame are independent and the list and FAD are not.

The FAD is similar to an Area Survey under a multiple frame design where one Operator (for one tract of land) can operate in multiple area samples or segments. Similarly in a FAD sample one List Frame Operator can operate multiple parcels in the Fruit Acreage Database. The definition of an Operator is the same for the NASS program as it is for the Fruit Acreage Database. Using the List Frame POID as the "link" between List Frame and the Fruit Acreage Database ensures that no matter what changes are made to the NASS program, the Fruit Acreage Database will mirror those changes and remain in sync and operational.

Through the integration of FAD with the List Frame, samples from the FAD are able to be processed to produce listings, analysis, extract samples and label questionnaires. Surveys originating from the FAD and processed through these applications are also able to handle special data collection issues in a similar manner to other NASS national program surveys. An example of one special handling situation is for Management Companies. There are many fruit/nut operations in California that are operated by Management Companies. Management Companies are agricultural businesses and do not own or lease any land. Having these Management Companies identified is extremely helpful in data collection. Management Companies usually have all operational information that Owners do not. Respondent burden is reduced in the case of Management Companies because enumerators can make one contact with a Management Company that can report for dozens of individual operations. On ELMO and the Fruit Acreage Database, Management Companies are identified under a special Active Status code (AS=12). List frame records identified as Management Companies are not sampled in standard NASS Surveys. In most cases, comments on ELMO identify what Management Company an operation uses and contact information for them. As far as data capture from the Fruit Acreage Database, data is summed by commodity to the POID and update transactions are created for both the Management Company and for operations that use Management Companies.

7. Discussion on the Benefits and Limitations of the FAD

The key benefit of the FAD is the detailed level of information at the parcel/block or field level. The detailed information is essential for objective measurement studies and tracking parcel level specific crop characteristics. Conceptually, the two frames are different. The NASS List Frame is not designed to maintain this level of detailed information. The NASS List Frame is maintained at the operator level, while the FAD is maintained on a land-based level this difference can be an issue if the operator is not identified correctly because the operator is the link between the FAD and the NASS List Frame. For this reason, the CA FO maintains operator names and addresses only on the NASS List Frame through ELMO applications. The ability to link parcel/block records to ParcelQuest through the APN makes it possible to utilize GIS information to map samples for enumeration and verify new or updated county assessor information. Through ParcelQuest the FAD can fully utilized county assessor and the associated GIS information.

A fundamental factor that makes the FAD possible is the stability of fruit and nut acreage., Once tress/vines are established the planted acres are more or less permanent year to year. This allows for accurate detailed record level information that changes very little over time. Crops that are rotated on an annual cycle would be difficult to maintain. Tracking permanent plantings allows the sampling of specific trees/vines on an Objective Measurement Study or Fruit Acreage Survey.

The long term success of the existing cooperative agreements and the resulting program of surveys, objective studies and state projects provide a consistent source of high quality reported data for updating and maintaining the fruit and nut operators/operations on both the FAD and the NASS List Frame.

Although the FAD only covers fruit and nut acreage in the state and it is unique to California, it does enhance the national program. The FAD is used to track ownership changes, remove duplication from the list, and estimate bearing and nonbearing acreage for the Census of Agriculture, Fruit Chemical Use, and other national surveys. The FAD is a product of a positive cooperative relationship with state and industry groups.

ParcelQuest also enhances the national program, It allows research of new addresses for post office returns, identification of tract operators on the June Area Survey, and additional tracking capabilities for ownership changes. However, access to statewide county assessor information through ParcelQuest is not available in every state

A factor to consider is that resources are required to keep the FAD in sync with the NASS List Frame. As NASS moves towards a regional or centralized environment, the agency will need to address what resources will be available for maintaining the FAD. Regardless of how and where the list frame maintenance is conducted, the CA FO will continue to have primary responsibility for the FAD, state fruit acreage program, and the related surveys. Under an enterprise wide infrastructure, the FAD will have to be integrated into the centralized system.

Finally, the FAD system was developed in 1998 and written in Visual Basic FoxPro. NASS is currently phasing out FoxPro. The current FAD system will be rewritten in an agency supported software. It is important that any new system be a coordinated effort with the California FO and in line with the agency's reorganization goals.

8. FAD Issues to be addressed during the NASS reorganization

Under the NASS reorganization there are specific plans to improve operational efficiencies. These plans include the regionalizing of the list frame activities and transitioning toward a centralized database processing environment. To reach that goal, NASS will have to overcome many barriers so that all software tools and processes operate efficiently in that new enterprise wide environment. While obtaining a centralized environment is a major benefit of this plan, there are several Field Offices with legacy systems that will need to be evaluated in the reorganization process.

The FAD is not the only state level database system currently administered at the Field Office level in NASS. It is the most extensive and probably has the longest history of development and support from the state government and industry groups. There is no other FO maintained database in NASS that seamlessly provides auxiliary grower information for national surveys sampling and the Census of Agriculture. The way the FAD is linked to the agency's list sampling frame should be examined by NASS and possibly the USDA as way to better share available information without affecting the operational integrity or structure of national programs.

One of the most challenging issues concerning the FAD is the coordination by the California Field Office with the new Frames Maintenance Group at the National Operations Center on updating the NASS List Frame. Timing of updates to the FAD from the List Frame is critical. For Fruit and Nut Acreage Surveys updates to owner/operator records must be current for follow-up phoning and second mailing. List Frame must also be current and in sync with the FAD for sampling and permission agreements for the Objective Measurement Studies. A new update and comment application has been developed that will address this concern

9. About the FAD in the Future

NASS is in the process of developing an Administrative Data Repository (ADR) which is a collection of standalone systems and outside administrative data sources. The FAD, is an auxiliary sampling frame consisting of grower reported survey data and plans are for the FAD to reside in the ADR (see Figure 2). Currently administrative information from the USDA's Comprehensive Information Management System (CIMS) and remote sensing data from NASS's Research and Development Division reside in the ADR with plans to expand to include other State/National data now outside of the enterprise architecture. As the ADR is developed it will have a layer of NASS metadata to allow the sharing of available data in the ADR and data services infrastructure to allow access to various enterprise services. In the case of the

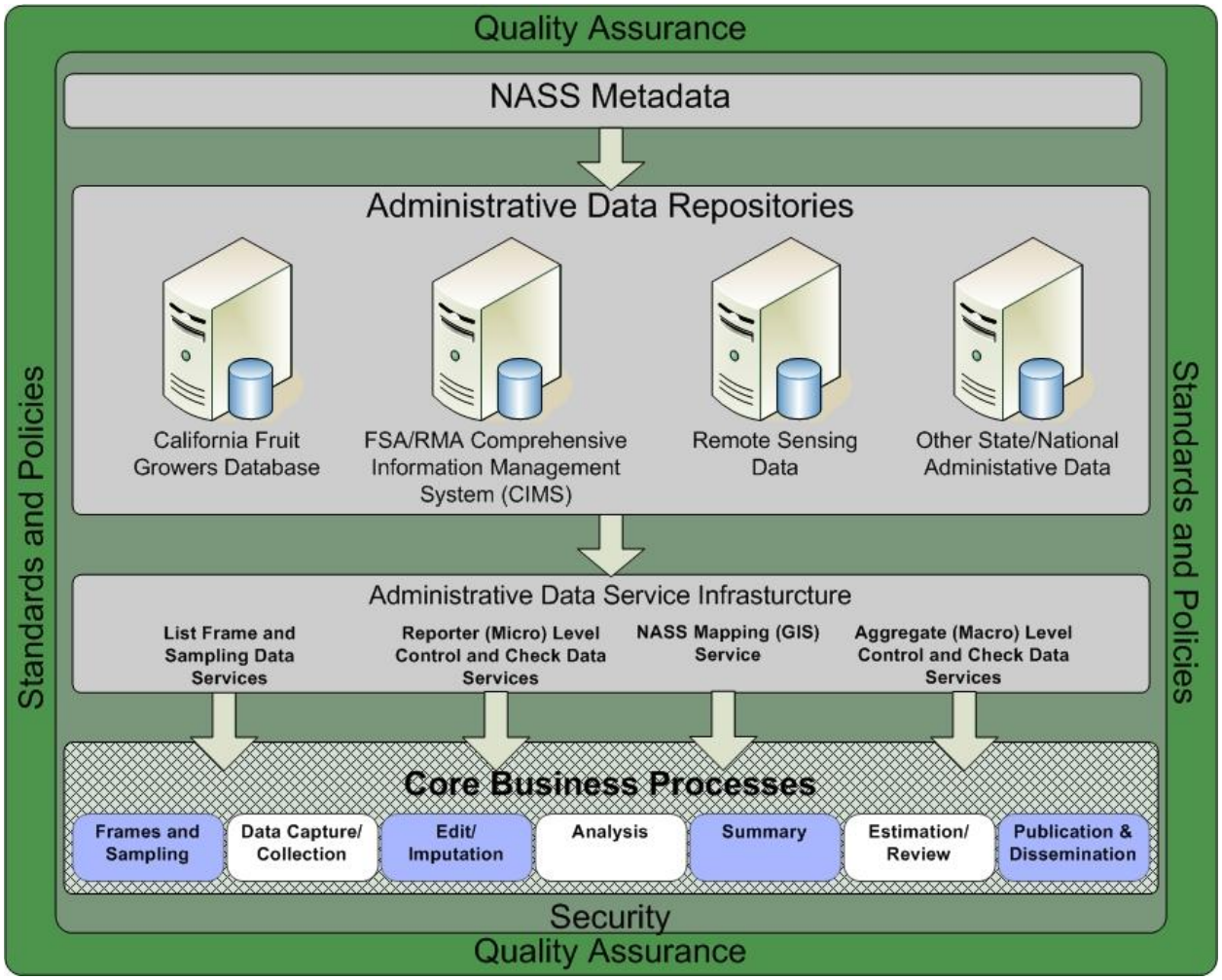

Figure 2

FAD this would be the list frame and sampling services.

FAD already is set-up with many of the metadata and administrative services outline in Figure 2 so conversion to the ADR should be transparent to the data users. Plans are to use the existing Look-up, Update and Processing modules through an Open Database Connectivity (ODBC) driver or until there is a need for need for a redesign of these modules.

10. Recommendations

The most significant challenge for the FAD during the NASS reorganization is the daily maintenance and timing. California's Regional Office and the Frames Maintenance Group should work closely to develop and document new standards for list frame maintenance at the National Processing Center and the same for the FAD. Under the reorganization the FMG has the responsibility for the for List Frame maintenance and will have to work with the California Regional Office on the timing of the NASS List Frame updates to coincide with the requirements of the fruit acreage survey program.

A second challenge is to design a new database for FAD compatible with the new agency standards and centralized reorganization. This will be accomplished while maintaining mission critical applications and functionality for California's existing Fruit and Nut program and should be transparent to the California Regional Office during the transition period. Any redesign of the FAD or its associated applications should include a concerted effort by NASS staff in headquarters, the FMG and the CA FO. The CA It staff have already developed a system of sharing data though shared metadata and NASS should explore the possibility and of expanding on the lessons of the FAD design and develop ways to share other administrative data (FSA, RMA, and NRCS). One interesting feature of the FAD is the flexibility of low granularity of design at the parcel level to share information and utilize applications on other systems.

A third challenge is to research the expansion of county assessor information (APN) into administrative database design. Although California FAD is only possible through a unique combination of cooperative agreements with state and industry and a focus specifically for fruit and nut crops, it is the application of a state county assessor data at the APN level that is the catalyst that allows the FAD to sync with other databases and utilize outside applications. Currently not many other states have a central standardized statewide database of county assessor information, like ParcelQuest, but this type of system should expand to other states as technology improves and this is something NASS/USDA should be prepared to implement.

References

ParcelQuest. (2008). *Standardizing the California County Assessor Use Codes.* Folsom, CA: www.parcelquest.com.