Chapter III

Descriptions of the Selected Data Initiatives

Overview

This chapter provides a detailed description of each of the 10 selected initiatives. The initiatives are *not* ranked according to priority. Rather, each initiative appears in the same order as in chapter II (i.e., by the method used to create new or improve existing information resources).

For each initiative, a detailed summary is provided that includes the following sections:

- **Initiative Summary.** A brief summary of the initiative.
- **Background.** Background information on the need for conducting the initiative, including information from the previous literature (and the informant interviews) that outlines current gaps in the knowledge of food assistance and nutrition programs.
- **Methodology.** The current availability of the data to carry out the initiative, additional required data, and data collection methods.
- **Advantages and Limitations.** The advantages (e.g., list of research questions that could be addressed) and limitations for food assistance research.
- **Feasibility Issues.** Any feasibility issues that need to addressed before moving forward with the initiative (e.g., confidentiality), including initial suggestions to address these issues.
- **Potential Cost.** The major costs associated with the initiative, to both USDA and other organizations.
- **Criteria Summary.** The initiative in relation to the nine evaluation criteria in chapter II.

1. Micro-Level Database of Food Stamp Program Participation Records

Initiative Summary

This initiative proposes to develop a new micro-level administrative FSP database that tracks FSP participants in every State. The FSP Quality Control (QC) System data represent the largest nationwide effort to collect demographic and program information on FSP participants. The new database would significantly expand the QC system by obtaining information on all FSP participants from State administrative files and adding this information to a national database.

The initiative would address a major limitation of current surveys (e.g., SIPP) and administrative records (e.g., the QC system), which provide limited information on the dynamics of FSP participation. Researchers could use a full administrative file of FSP participants to examine dynamic transitions at the State and national level. This file would contain all FSP households rather than a small-percentage subsample like the QC data, allowing researchers to link it to other administrative or survey datasets without losing observations for nonsampled cases. This would be an important gain when linking to surveys that are small subsets of the universe.

The proposed database would include information on all FSP participants from State administrative files for a specified month. The differences in State administrative databases and technologies mean this database could start by including information from selected States and evolve as more States develop the technology to produce and store large data files. USDA could collect administrative data on FSP households for a 1-month period and include an agreed-upon minimum set of data elements that use consistent definitions in all 50 States.

USDA will need to consider several major cost issues when evaluating this initiative. Significant technical issues associated with developing, creating, storing, and monitoring data could make this implementation very expensive. In addition, it will likely place an additional burden on State administrators, which could influence the feasibility of this task. The existence of a full administrative file for WIC and recent findings by Maxfield et al. (1999) on creating a similar FSP database for administrative purposes, however, suggest that developing this initiative is feasible.

Background

The FNS-388 and FNS-388A forms provide two sources of FSP administrative data at the national level. These forms provide the most current information available on FSP participation, but they are limited for detailed research because they do not include any information on individual characteristics. Researchers therefore generally restrict their use of these files to examining broad trends in the FSP.

The forms were designed many years ago, when the FSP's emphasis and operations were quite different, and States' level of automation was much more limited. The FNS-388 form requires States to report FSP participation by individuals, participation by

households, and benefits issued each month. The same information must be provided twice a year (in January and July) for each "project area" (typically a county) on the FNS-388A form. Although States have to submit this data to FNS monthly, there is only one data item on household characteristics and it is only reported in 2 months each year. The one non-FSP data item that is obtained over time is whether each FSP household also receives cash assistance from Temporary Assistance for Needy Families (TANF)/Aid to Families with Dependent Children (AFDC), Supplemental Security Income (SSI) or a general assistance program. Participants who receive such assistance are categorized as "Public Assistance (PA)" households. Households where at least one member does not receive such assistance are characterized as "No Assistance (NA)" households.

The annual FSP QC sample addresses the limitations of the aggregate administrative files by providing an administrative survey of a representative sample of approximately 50,000 FSP households.[3] The QC data includes detailed demographic, economic and FSP eligibility information for a sample that is stratified by month and State. The primary objective of the database is to assess the accuracy of eligibility determinations and benefit calculations for each State's FSP caseload.

According to Brinkley and Ewell (2000), State QC reviewers first gather administrative information from the selected household's case file and then visit the participant to determine if the household received the appropriate level of FSP benefits. States that have fewer than 10,000 FSP participants have to pull and review at least 300 cases. States with more than 60,000 participants have to pull at least 2,400 cases monthly.[4] The reviewers then send the case information to FNS' national computer center where they are entered into the QC system database. FNS regional offices then conduct a Federal re-review of a subsample of the original State sample. This information is used to determine whether each State's computer error rate is correct or whether it needs adjusting. During the key stakeholder interviews, State agency officials said this process is quite burdensome in administering the program.

While the primary objective of the QC system is to assess the accuracy of eligibility determinations and benefit calculations, researchers also use it as a source of detailed information on FSP participant characteristics. Researchers have used these data to publish annual reports on the characteristics of FSP households, as well as to develop estimates for FNS' QC microsimulation model (Rosso and Fowler, 2000; Brinkley and Ewell, 2000).

The QC system has several limitations for research. First, while it allows for comparisons over time of the characteristics of FSP households nationwide, it does not provide longitudinal information on individual households with which one can study program dynamics (outside of historical information on FSP participation trends) or study changes

[3] It also samples a somewhat smaller number of households who experienced a denial or termination from FSP.

[4] If the average monthly caseload is between 10,000 and 60,000, the standard minimum sample size is derived by the following formula: 300 + 0.042 (the average monthly caseload – 10,000)

in the circumstances of individuals or households. Second, the QC system contains a very limited sample size for important subpopulations, such as immigrants, the elderly or childless adults subject to the FSP time limit. Its sample size precludes analysis of differences among all the States. There is a lag in reporting in the QC system for approximately 9 months after the close of a fiscal year.

A full FSP administrative file would address some of the limitations of the QC data, as well as those of the aggregated data from the FNS-388 and FNS-388A forms. First, the file would provide longitudinal information on a very large number of FSP participants– the entire universe rather than just a sample. Researchers could use this information to examine dynamic changes in the FSP program, including changes in benefit levels and transitions onto and off the program. Second, because a full file would include all participants, researchers could perform analyses on small subpopulations of interest, such as those mentioned previously. Third, unlike the QC data and general surveys, the full sample would not be subject to sampling and respondent error (including nonresponse).

A recent study of tracking administrative outcomes across States for the FSP by Maxfield et al. (1999) suggests that creating a full FSP administrative file is feasible. The primary purpose of that study was to examine whether it would be feasible to create a national administrative database for FSP and TANF that would document interstate duplicate participation and track time limit requirements.[5] They found two potential options that could make a national database feasible. The first involved the creation of an "eligibility database," in which a central facility transmits case records from all States to individual States to identify duplicate records. The system would also track program participation history to monitor time limits. The second is a "one-stop database" that extends the eligibility databases and links to several other Federal and State databases to allow direct verification of income, assets and employment. The Maxfield et al. report found that creating either option for just the FSP program is feasible, though it would be more cost-effective to coordinate joint efforts with FSP and TANF.

Methodology

The primary purpose of this initiative is to develop a limited research database that includes all FSP participants during a particular month each calendar year. The initiative would incorporate key elements of the information collected in the QC reviews, as well as focus on client outcomes and year–to–year participation dynamics, and include information on specific subpopulations (e.g., immigrants).

We do not propose the development of a continuous FSP administrative file covering every month of the year, for three reasons. First, and most importantly, without a compelling administrative purpose behind the initiative, downsizing the State burden at least initially seems like a good strategy. Second, the creation of a smaller file for research purposes would be less costly for USDA and easier to manage than a file that collected data for all 12 months. In addition, the proposed file would put fewer burdens

[5] The study added to a previous report from the Department of Health and Human Services (1997) to build a national administrative database of records from mean-tested transfer programs.

on States than the larger file proposed by Maxfield et al. Third, the development of the smaller file could be seen as a first step in creating the larger file suggested by Maxfield et al. As States become more accustomed to the reporting requirements of the smaller files, they can begin to adjust their data reporting systems to eventually produce a data reporting system suggested by Maxfield et al.

The lessons learned during the creation of a full administrative file for WIC should be beneficial to building a similar database for the FSP. The main issue that arose during the creation of the WIC database was the definition of the "minimum data" set (i.e., the minimum number of data elements to be provided by each State) primarily because there is variation in the data elements across State systems. To address this issue, FNS created a minimum dataset with a common set of data elements standardized across States. FNS also created an optional dataset for States that wanted to report more information. Data are reported centrally for just a single month each year, forming a series of cross-sectional point–in–time databases.

FNS maintains an aggregate WIC file and State agencies control individual level data that are available upon request. This aggregate file, referred to as the WIC Participant Characteristics (PC) database, provides an overview of the characteristics of WIC participants in a single point-in-time "snapshot." State WIC files linked to these records provide longitudinal data on the included individuals: demographic characteristics, eligibility criteria, and benefit levels. Some States have more extensive files on their clients, including referrals, family identifiers, and documentation of such services as immunization status. Researchers can access State administrative files (with the State's permission) to track client services, and conduct longitudinal research by linking the mother to the child.

The process for developing the FSP administrative initiative could mirror that for the WIC database. Federal officials, State program administrators, local program administrators and researchers would collaborate to develop the framework, minimum dataset requirements, data collection procedures, and confidentiality agreements. In the WIC program, the participatory process worked well by involving the State WIC directors from the start and should be replicated in this effort. In total, the WIC data system took approximately 2 years to construct.

The first step would be to develop a framework for the FSP database. Similar to the WIC database, the most efficient method would likely gather data from a specific month each calendar year. While it is possible to collect information from multiple months, the size of State FSP files could be quite large, which would significantly increase the cost associated with processing and storing the data.

The second step would be to develop a "minimum dataset." Similar to WIC, State FSP agencies collect varying levels of details on participants. The choice of data elements depends upon two important criteria. First, USDA would select data elements of importance to research questions that are available in all States (or can be developed in all States). USDA would need to consider the reliability of elements in administering

programs because certain elements may be more reliable than others. For example, if administrators do not need to track certain variables—such as education—to administer their programs, then the variables are more likely to have missing or incorrect values. The second criterion is that the data would need to be comparable across States (and localities). During interviews with key stakeholders, State agency officials said the variability of State data systems might make it difficult to develop standardized systems by which new data could be collected. Because each State has its own system, using different data definitions and data elements, it may be difficult to compare data among States. While some of these issues can be identified by fully understanding the program rules and databases in each State, in some cases USDA may need to make a special request to States to add comparable variables to their systems.

Several potential variables could be included in a "minimum dataset," including:

- a case identifier (e.g., TANF/FSP/Medicaid case number)
- Social Security Numbers of case members (SSNs)
- FSP benefit amount
- net income
- children in household
- elderly person in household
- person with a disability in household.

It may also be desirable to develop "optional" datasets for States that have more detailed information in their administrative systems. As was true with the WIC program, certain data elements may exist that are not essential to the main objectives of the system. These optional data elements allow for some State or local programs to contribute to a rich database that can be used to compare outcome or participation dynamics. Potential examples of "optional variables" include those that identify participation in other programs, such as TANF.

The third step would be to develop a method for data collection. Maxfield et al., (1999) found that States had the capacity to produce the types of FSP administrative files necessary for this initiative on an ongoing basis. They surveyed officials about the capabilities of producing similar types of records and found that administrators could prepare and send an extract of their case records with relative ease to a central facility. USDA currently is conducting a comprehensive survey of the computer capabilities of all State FSPs. The survey results should more information regarding the feasibility of creating this research database in all States.

The micro-level State records would be stored at a centralized location. Records could be kept "active" for, say, 5 years, and then archived off the system. A more ambitious effort would have FNS house all of the micro-level data in one location.

The database could start with a selected number of States that have the capability to produce a minimum dataset and then be extended to larger samples as States move forward with the technology to produce data extracts. The Health Care Financing Administration (now the Centers for Medicare and Medicaid Services), Department of

Health and Human Services (HHS), used a similar method to create the State Medicaid Research File (SMRF), which now contains individual level administrative records from every State's Medicaid program.

The final step would be to develop data access, confidentiality and security protocols. According to Maxfield et al. (1999), privacy concerns fall into three categories. The first is that State agencies own FSP records and, the records may not be disclosed to another public or private entity without the State's permission. This problem can be solved by executing agreements between FSP State agencies and USDA.

The second issue is the potential for unauthorized physical or electronic access to the national client database. This threat to security could be controlled by building a secured computer system, similar to that used by the U.S. Census Bureau and Social Security Administration to store their restricted data extracts. The third issue deals with authorized, but "unintended uses" of the data. This risk is the most difficult to control. The Census Bureau attempts to control this by reviewing all final data products that use the restricted data. To alleviate any confidentiality concerns, Census limits output in tables to include at least 50 observations, and encourages projects to use multivariate analyses, rather than descriptive analyses, in presenting the results.

A more ambitious initiative would build on this database by negotiating agreements with State agencies to provide supplemental records from other months. Obviously, if all States provided this information for every month, the proposed file here would be identical to that from Maxfield et al. Because of differences in State systems and technology, however, it is likely that many States will not be able to produce and/or archive a full file for all months. Nonetheless, it would be important to have access to such records, even if it is limited to a select number of States. For example, researchers interested in understanding the dynamic program participation trends and conducting longitudinal analyses may wish to access a file that includes all months.

Advantages and Limitations

A full file of FSP administrative data would significantly expand research opportunities for USDA and potentially create a new avenue for program administrators to identify duplicative payments and to monitor a time limit on Food Stamp Program participation that applies to able-bodied adults without dependents who do not meet certain work requirements. First, the data would provide detailed and reliable program information about large samples of FSP participants and, unlike survey data, are not susceptible to nonresponse or respondent error. Second, the size of the database would allow for detailed subgroup analysis both across and within States. Third, once the database is established, updating the files with additional years of data is a relatively low-cost venture, particularly when compared to the costs of constructing longitudinal files from multiple waves of survey interviews.

If monthly State databases could be added, researchers could track outcomes of program participants longitudinally both within years and across years. Fourth, because this file would contain all FSP households rather than a small percentage subsample like the QC

data, researchers could link it to other administrative or survey datasets without losing observations for nonsampled cases. This would be a gain of particular importance when linking to surveys that themselves are small datasets of the sample.

Researchers could use the files to address several questions related to FSP participation, including:

- How do the program characteristics of FSP participants vary across and within States?
- What are the program characteristics of specific subgroups of FSP participants?
- Has the composition of FSP participants changed since the implementation of welfare reform in ways that differ across States related to their reform measures?

There are three important limitations of these files. First, unlike survey data, the FSP files would contain relatively limited information on important background characteristics, such as education. This limitation could be addressed for a small (but representative) subset of the cases in the file by linking the data to other existing State or national surveys (e.g., CPS or SIPP). Second, researchers could only track program participants while they are participating in the FSP program. This is a relatively minor issue, given that nonparticipation is an important outcome in its own right signaled by the lack of inclusion in program files from any State. Finally, these files do not include information on long-term nonparticipants.

Feasibility Issues

The biggest feasibility issue will be in developing a minimum dataset that includes a comparable set of data elements across disparate State systems. During interviews with key stakeholders, State agency officials said the variability of State data systems might make it difficult to develop standardized systems by which new data could be collected. A second important issue will be in storing and monitoring the data. The data files will likely be quite large, and may require additional funding for States to provide and transmit the data to a centralized location. This issue would become even more important if USDA pursued the creation of a joint administrative-research file centrally, which would require significantly more Federal computing resources for space to store data on individual FSP participants from all months.

The existence of the WIC administrative files coupled with the findings by Maxfield et al., (1999) suggest that the initiative should be feasible. Key informant interviews conducted with State FSP directors suggest that there is significant interest in developing an alternative to the FSP QC system that focuses on program participation and outcomes. This option could become particularly appealing to USDA if it simultaneously allowed for a reduction in the effort spent on QC. For example, if USDA could develop a new method of auditing State files using the new administrative files for quality control purposes (e.g., auditing a randomly selected, representative subset of State records), USDA might be able to significantly cut the effort spent in collecting QC data each year.

Potential Cost of the Initiative

USDA will need to consider the following costs when developing a plan for this initiative and considering its overall desirability. Costs to the Federal Government include the following:

- Data Access Costs: According to Maxfield et al., (1999), these costs should be relatively minimal given that State agencies should be able to produce reliable extracts.
- Storage Costs: The storage costs include establishing a data warehouse to store the data.
- Data Update: The data updates should be relatively inexpensive after establishing a minimum dataset.
- Monitoring: USDA will need to monitor data usage to protect the original confidentiality agreements.

States also will have the following costs:

- Data Manipulation: The programming time and costs to manipulate the data will be substantial, as programmers from States will need to generate data elements for the minimum dataset. The file must also be converted into a format usable by researchers.
- Data Storage: States may need to develop data storage systems to transmit the data each year to a centralized location.
- Data Transmission: The data will need to be sent in some form to a centralized location.

While the startup costs of creating a linked data systems are high, updating the files on a regular basis with new information should not be as burdensome, as administrative data are a relatively low-cost source of information whose structure tends to remain constant over time and whose processing can become routinized over time with each new installment of data.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
X			X	X	X	X	X	X	Moderate	Substantial	Moderate

Notes on Individual Criteria:

(2) Could be extended in the future to include a continuous administrative file (similar to Maxfield et al.) that administrators could use the database to track duplicative payments across States.

(3) Provides detailed information on participant characteristics, including characteristics of subgroups (e.g., people with disabilities, immigrants, Able-Bodied Adult Without Dependents (ABAWDs), dynamic transitions onto and off of the FSP; and differences across States in participant characteristics and outcomes.

(5) Would include a minimum dataset from 1 month for all 50 States, though some States may provide more extensive information from other months.

(6) Would be updated on annual basis.

(7) Has some moderate feasibility issues related to 1) developing a minimum dataset that includes a comparable set of data elements across disparate State systems and 2) storing and monitoring the data

(8) While the costs of creating the file are substantial, updating the file should not be costly

2. Building Aggregated Administrative Statistics from Local Agency Records on the National School Lunch Program

Initiative Summary

This initiative seeks to capitalize on information routinely collected by local agencies and aggregated across individual program participants but not systematically compiled for research use as a database of local-level observations. Often these data become part of higher-order aggregates reported and analyzed as State totals, but the details–easily accessed at the prior step–are then lost. A "data capture" of these local-level aggregates could produce important benefits for all of the Department's food assistance and nutrition programs. This concept should be implemented first for the National School Lunch Program (NSLP) and the School Breakfast Program (SBP) because the FSP and WIC programs have already developed more evolved systems of capturing aggregated data from their State administrative databases.

The focus of this initiative is to create a database containing information at the school district level about the NSLP and SBP. The database would contain school district-level information about children who qualify for free and reduced-price meals, as well as the number of meals served to all children in the district. Many of these data are already collected by local school districts and reported to their State Departments of Education. These data are then aggregated to the State level to produce the federally mandated monthly and annual *Reports of School Program Operation*s for each State.

A national database that contains data at the school district level could be used by researchers and program administrators to compare and analyze data across both State and local geographic areas. The portion of the database containing information about children participating in the NSLP/SBP would portray the distribution of free or reduced-price meal recipients in a variety of dimensions, including their demographic characteristics (e.g., percent age 7, percent Hispanic) and eligibility categories, all in summary form for school districts as a whole. School district-level information would also be included such as name, geographic location, size (number of students), and meals served broken out by paid, free, and reduced-price.

The initiative proposes that district-by-district data of this sort be collected for 1 month each year corresponding to the timing of the States' annual reporting requirement to FNS. This will create a point-in-time view of program characteristics and participation of low-income children, providing comparable information that highlights differences across school districts. The database would be stored at USDA as an electronic file using methods similar to those currently used to store data on other agriculture programs. The entire database will be available to Federal program administrators. For confidentiality reasons, it may be preferable for only a subset of the data to be made available to the general research community for confidentiality reasons. It should be both technologically feasible and relatively simple to construct the database for limited access.

The initiative will be conducted by completing some preliminary activities prior to developing the actual database. First, an inventory will be conducted to determine the information currently collected by local school districts administering NSL and NSB programs. These data would be categorized into demographics of the children certified for free and reduced-price meals, the meals served, and any other important areas specific to the NSLP/SBP. The initiative will explore whether nutritional program data currently collected at the district level could be combined with school district demographic information to facilitate comparisons of program participation among districts of comparable size and demographics.

A second step will be an assessment of the potential for using such aggregated information in national research. Included in this assessment will be the development of criteria for creating an accessible national database, the limitations on the amount of data that will be made available to persons outside the Federal Government, and methods by which these data could be accessed.

The actual implementation of the initiative will involve developing the database structure, implementing a test design, and populating the database. In particular, the database will need to be designed as an analytic file that links participant data to local agencies. Consistent data definitions for client and agency data will need to be developed, and agreement will need to be reached with State officials on how the data will be collected. Local district data should be collected by States, but before being aggregated as is currently done, States would transfer the data to the national database in the format reported by the local school districts and a common format established nationally.

A final key step will be to encourage States or localities that do not now provide such measures on a routine basis to do so for the sake of the national database. Emphasis on the benefits accruing to State and local agencies, including examples from other States, will be essential to "selling" participation in the initiative to currently non-reporting jurisdictions.

B. Background

Over the years, a number of national databases have been developed to track participation in food assistance programs (e.g., the WIC program participant characteristics database and the Food Stamp Quality Control System). These national databases rely on State program administrators to collect data from local programs delivering services and to aggregate the information into a format prescribed by USDA. In general, these national databases tend to focus on client and program participation issues, benefit distribution, and client demographics.

However, State and local food assistance programs often provide far more details in their data than ever are reported at the national level. For example, data regarding the number of school meals served in the NSLP and SBP are collected first for individual schools, then reported to school districts, passed on to State agencies as district-wide aggregates, and finally aggregated into a single State total reported to the Food and Nutrition Service, USDA. By the time the information is collected and aggregated at the Federal level, it is

impossible to compare demographics or participation rates between school districts within a State, let alone among school districts of similar size located in different States. There is also potentially some important program descriptive and cost information collected locally but never aggregated or reviewed at either the State or Federal level.

This phenomenon occurs in the WIC program and the Food Stamp Program as well. For example, local WIC agencies collect a variety of nutrition, medical, and referral data that are recorded in charts and in many cases reported in State data collection systems, but these data never reach the Federal level. The WIC program does collect and aggregate some useful information about participant program characteristics and benefit delivery through its Participant Program Characteristics database. With regard to the FSP, each month States provide aggregated data to FNS on participants and benefits issued. Rather than requiring the States to provide more detailed aggregated data to FNS or to maintain the local program area-level identifiers in their monthly FSP reports (Form FNS-388), a separate data development initiative would collect and store local-level data in a national database, thus permitting USDA and researchers to track not only aggregate program characteristics, but also program participation patterns across States and localities in relation to other State and local variables.

Having better access to State and local data is important if research is to be expanded beyond simple information about program participation and benefit delivery. Issues related to the effectiveness of program services can only be examined if more and better data are available. The report of the Advisory Panel on the Research Uses of Administrative Data cited the problem of limited access to local data, noting that "program administrators and policymakers will need reliable State and local data if, among other things, they are to summarize program operations; determine who is being served by which programs; who is being underserved; who is not being served; and how services can best be targeted to those in greatest need …"(Hotz et al., 1998). In addition, the panel noted that national survey research could not adequately monitor the diverse local programs currently being funded by State and local governments. To better understand the dynamics of food assistance programs at the local level, as well as to better understand how well clients are being served, better access to local data seems critical.

C. Methodology

The methodology for conducting this initiative in the area of the NSLP/SBP consists of the preliminary steps previously described, and the actual building and populating of the database. A description of the specific methodology follows:

1. ***Conduct a preliminary assessment of the actual amount of useful information currently being collected by local School Lunch/School Breakfast providers or that is available in local records but not reported to State agencies.*** Because States may have different administrative and reporting requirements for local school districts, an assessment of all State programs will be necessary. A number of steps would be involved in conducting this inventory, which will include:

a. ***Collect and analyze State-level data collection instruments to determine which data are collected by State program administrators but not reported to the Federal government in disaggregated form***. The initiative proposes to examine the actual data collection instruments and methods used by State programs to collect client eligibility, participation, and meal delivery information at the school district level. In addition, the initiative proposes to collect information on how State program administrators identify local program demographic information, including any information that would identify relative program size and geographic location. Once these instruments are collected, it will be possible to identify the extent to which different data elements are collected consistently in most or all States.

b. ***Conduct an inventory of local program data***. Most of the data needed for this initiative will be reported to States, but many States grant flexibility to their local school districts as to what types of administrative and program data are collected. The second step will be to identify which data might be collected by these local school districts on a routine basis, but are not reported to State school food service program administrators. To this end, data reporting will be examined for a sample of school districts within each State that allows the type of flexibility described.

Information from each selected school district will be collected by asking local officials to provide copies of the data elements contained in their local records. These records can then be compared to the data collection instruments provided by the State agency. Data that are collected and not reported, but that are important for national research, can then be identified. The survey will also determine which of these variables are regularly aggregated to the school district level, and on what periodicity.

c. ***Develop a matrix of data available but not currently collected***. The final step in the process will be to develop a data matrix that displays the data available from each State and/or school district in local units but not passed on to the Federal Government in that form. Data will be classified as to whether they are currently being reported by local districts to the State, whether they are compiled by most districts by not passed on to the State, or whether they simply exist at the micro- (i.e., participant) level in a local record. This matrix will contain information on the data elements available, the extent to which each data element can be reported, and the value of that data element to national research objectives. Separate matrices will be developed for data on children receiving free or reduced-price meals, the meals served, and the local school district demographic data. The extent to which these data are consistent across States and local programs will also be assessed.

2. ***The next preliminary step will be to prepare a national picture of which data are currently available, and the extent to which these data can be used to enhance national***

research priorities. It will be very important not to simply collect data because they are available, but to set the proper context by identifying how collecting these data at the local program level will enhance program research efforts. Once a complete database matrix is developed for the NSLP/SBP, it will be important to conduct an evaluation of whether these data are valuable to program administrators or researchers in achieving national research goals. It must be remembered that some of the data collected by local programs may serve a limited purpose related to local program administration, and may not be useful or important in conducting research on national issues. Therefore, it would be important for USDA to conduct an assessment of the data elements available, in order to limit the overall size of the data file to those most important.

It is not likely that all State or local programs will collect the same or similar data, so it will be important to examine the similarities and differences in the types of data collected and any variations in the data definitions across States. Once the inventory is complete, data elements will be analyzed to determine the extent to which they are consistent and are collected and available in electronic format. From this information, profiles of datasets can be developed to begin the process of designing the potential database.

3. ***Develop and populate the database.*** The main focus of the initiative will be to design and populate the database. One of the first steps in developing the database will be to construct some specific parameters for how the database will be structured, the platform that will be used for storage and manipulation of the data, and how the database will be populated. A committee of representatives from the national Association of School Food Service Administrators (ASFSA), State and local programs, and researchers should be assembled to work with ERS, FNS, and any potential contractor to develop these parameters. This process of including State and local program representatives and other stakeholders in the development of national databases has worked well in the past, and will provide Federal officials with valuable input on the difficulties that might be involved in collecting these data, as well as their potential uses.

   Once the parameters for the database have been established, USDA should examine how best to collect and store the data. Data can be collected at the same time the State agency must prepare its report of annual participation. It is envisioned that the system will need to be able to accept data electronically in a variety of formats and platforms. Because of the need for flexibility, the database will need to be constructed in a way that translates data provided in different formats into a single, consistent structure.

   USDA will need to decide if it wants to contract out the development and maintenance of the database or use internal resources within the Department to create the database. In either case, because data will likely be collected in a variety of formats, consideration should be given to examining Internet interface technologies that can transpose data shipped in different formats into a common database structure. To use an Internet interface, specific guidelines for data formatting, data definitions, and data quality will need to be produced.

   The frequency of data collection is also an important issue. Because these files will potentially be very large, it is recommended that USDA collect these data only once a

year to create a point-in–time picture of each food assistance program, similar to how the WIC Participant Characteristic database is now compiled. Depending on laws and customs adopted in individual localities across the country, steps may need to be taken to ensure the security and confidentiality of the data aggregates for individual local programs prior to any release of the data to researchers. Researchers should be able to access the database, either through Internet access or through a specific request procedure developed by USDA.

Advantages and Limitations

A number of advantages would come from having local- and State-level food assistance program data available in a single database.

1. ***Access to data in order to compare participation and outcome dynamics between State and local school district programs of similar size and demographics.*** During key informant interviews conducted for this project, representatives from NSLP/SBP noted that it would be valuable to compare program participation trends across local school districts with similar demographics and characteristics, but not located within the same State. For example, it would be useful to have the ability to compare the number of free and reduced-price meals served and the rates of school lunch and school breakfast participation in large school districts located in inner cities across various States.

2. ***Data on the number of meals served and program participation among children eligible for free and reduced-priced meals at the local* level *will be useful in determining how to better serve clients.*** The number of school meals served, broken out as free, reduced-price, and paid, will be available at an aggregated level for each school district. This will allow researchers to examine the extent to which school districts of similar size and demographics are serving these client groups as well as the proportion of children enrolled in local schools who eat meals funded by the NSLP and SBP. These outcomes then can be examined in relation to school district policies, procedures, or outreach methods. After examining the outcomes from other State and local programs around the country, local school districts then can contact the other programs to discuss why some s are more successful than others.

3. ***Better access to trend analysis to analyze program evolution.*** By having an expanded database that includes information at the school district level in a single month each year, State and local administrators will better be able to determine the long-range effect of policy changes on program participation patterns

4. ***Data can be merged with survey and administrative data.*** Researchers could merge the aggregated file with individual records to examine correlations between broad trends in the NSLP/SBP at the local level and individual participant characteristics. In addition, data could be merged with Census data to examine

trends in program participation based upon community demographics and income levels of school district residents.

5. ***NSLP/SBP data from local areas can be compared with data from other food assistance programs.*** Data from local school districts can be identified by a specific geographic unit (e.g. counties), allowing those data to be compared with data from other food assistance programs serving the same geographic areas. This type of comparison would allow program administrators and researchers to examine whether programs are reaching those in need of services.

In addition to the advantages discussed previously, aggregated data collected from local school districts would be valuable to program researchers. The data could be used to answer some basic questions about participation trends in local school district programs. Some examples of the types of research questions that this initiative could address include:

1. To what extent do school districts of similar size and demographic characteristics succeed in enrolling low-income clients into the free and reduced-price school meals programs?

2. What are the characteristics of school districts that serve a high percentage of free and reduced-price meals to students as part of their school meals programs?

3. How does enrollment in the free and reduced-price school meals programs compare with enrollment in other food assistance programs within specific geographic areas?

4. What trends in NSLP/SBP participation occur over time? How do these trends vary by type of school district?

This proposal has a number of limitations, including:

1. There will be no detailed information on individual participants to allow for tracking of participation dynamics and specific client outcomes over time. The database is aggregated at the local school district level, making significantly more information available to researchers than in the current files that have been aggregated to the State or Federal level. However, the database by itself cannot be used to track individual participant outcomes.

2. By limiting the data collection period to a 1-month observation each year, data will be limited in its ability to detect short-term trends. While the database will be able to provide a very complete picture of program participation during the 1-month data collection period each year, it will not be able to track changes on a month-to-month basis. This will limit the usefulness of the database in tracking short-lived fluctuations in participation. But the data could be merged with individual-level participant data for this purpose.

3. Because the data are collected only once a year, it will take several years for the database to accumulate enough time periods to conduct any but the most rudimentary trend analyses.

Feasibility Issues

Deciding which data elements to include in the database will likely be a complex process requiring a significant amount of time. FNS discovered in developing the WIC characteristic database that an inclusive process involving input from State and local program administrators is a time-consuming process. It will be very important to limit the database to those elements considered to be key ones by the government, constituents, and researchers. There may be much disagreement among these groups about which data elements are most important. A clear, time-limited process should be used to identify the elements to be included.

Populating the database also may be difficult. It is likely that even if such an initiative is undertaken, some State and local programs may not wish to participate. It is unlikely that USDA will want to force them to participate, so convincing State and local programs to provide data will be a major issue. There may be less resistance to developing the means by which these data can be reported and included in the database than to entirely new data collection initiatives since the focus of this initiative is to create a database populated with data that are already collected. New Internet-based interfaces will make it easier for State and local agencies to load data into the national database. By using an inclusive process, many of the concerns of program administrators can be dealt with early on in the process.

There may be local school districts without the computer equipment, software, or skills to collect and distribute data in an electronic format. This may be especially true in rural areas. But most agencies are required to provide participation data in an electronic form now, and given the increased use of computer technology in local schools, this should not be a major issue for long.

Potential Cost-effectiveness of the Initiative

To conduct the data inventory and create the national databases, the main cost centers for this project would be:

- **Conducting the data element inventory of the NSLP/SBP**. In order to determine which data are currently being collected and not reported, a sample of data collection forms and local school district records will need to be collected and analyzed. This analysis must be conducted to determine the extent to which data elements are available across all types of local programs, and the extent to which data elements are comparable across States. While the initial review will not be a large undertaking, later stages of data collection and analysis will require a strict quality control system to ensure consistence of data elements among the States and localities contributing to the database. This will be a major undertaking if many of the roughly 10,000 school districts take part.

- **Bringing together program administrators, researchers, and government officials to develop the parameters for the database**.

- **Developing and programming the database**, including the cost of creating the Internet interfaces for data collection and the accessibility protocols for accessing the data.

- **Quality control to insure the accuracy of the data collected and entered into the database.** This will include programmer time to develop, edit, and update protocols for the database. The initial quality control system will likely be costly, as edits will need to be conducted to determine that data elements reported are comparable. Once the initial verification of data elements is complete, ongoing quality control will be less costly and more routine.

- **Data distribution, including the creation of Web-based access files or the creation of CD-ROM files for distribution.**

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
		X	X	X	X		P	X	Moderate	Moderate/ Substantial	None

Notes on Individual Criteria:

(2) This would be of value to researchers and program administrators interested in understanding caseload trends across districts of similar size and demographics.

(4) This would provide information to examine participation trends of students over time, including in the context of program interventions.

(5) The scope of the initiative would only be limited by costs and school districts' willingness to provide data.

(8) The large number of school districts involved could create a substantial data collection cost.

3. Matching State WIC Program Administrative Data with Point-of-Sale Grocery Store Transaction Data

Initiative Summary

The U.S. Department of Agriculture provides billions of dollars to States for WIC participants to purchase authorized foods. While the general food categories in which participants are allowed to purchase their foods are established by Federal regulations, very little is known about which specific foods are purchased, or if participants purchase all of the foods to which they are entitled.

This initiative proposes to match a single State's WIC program administrative records with point-of-sale transaction records at grocery stores where participants purchase their foods. The feasibility of using this technology was identified in an FNS-funded feasibility study released in 1998 (Bell et al., 1997), but specific applications of this technology to program administration and evaluation have not been explored. This initiative proposes to utilize this technology to examine food purchasing patterns of WIC participants as they relate to the nutritional goals established by the program.

Background

The WIC program uses the bulk of its funds to pay for food products purchased by participants. For most of the program's 27-year history, little has been actually known about which specific foods are purchased by participants. While the WIC program controls the general category of foods that participants can purchase, it does allow participants to choose which specific foods they will purchase within the approved WIC food categories.

The Food and Nutrition Service has a strong interest in examining which foods WIC participants purchase. Specific areas of interest include:

- ***Assessing program compliance***. WIC program officials must deal with significant program compliance issues when it comes to purchasing authorized foods. The program has made considerable efforts in the past to examine if participants use their WIC fiscal instruments (FIs) to purchase unauthorized foods, nonfood items, or to trade their checks for cash (Bell et al., 2001). Program officials also want to make sure that retail grocers do not overcharge the program by charging for foods a participant may be authorized to purchase, but decides not to.

- ***Examining WIC participant food purchasing patterns***. The WIC food package is designed to be a food prescription that will assist the participant in addressing an identified nutritional risk factor. It is very important, therefore, that the participants purchase and use all of the WIC foods prescribed. The WIC program combines a number of foods on one FI, so program officials do not know if a person purchases all of the foods they are authorized to

purchase. For example, if milk, eggs and cheese are authorized on the WIC FI, and the person only purchases the milk and eggs, program administrators do not know that the participant has not purchased the cheese. It is important for program nutrition staff to know this information so they can counsel participants to use all of their food prescription.

- ***Evaluating nutrition education efforts***. One of the main benefits of the WIC program is its delivery of nutrition education to participants. The nutrition education efforts are usually targeted either to making sure the participants purchase all of their WIC foods, or select foods that will help address an identified nutritional or health risk (such as obesity). In the past, the only way to evaluate whether or not the nutrition education efforts were effective was to ask the participants which foods they purchased or ate. Self-reporting creates a strong, possibly inaccurate bias in favor of nutrition education efforts. Participants may understand which foods they *should* buy, and may report that they actually did purchase the foods, in an effort to please the nutrition professional asking the questions when, in fact, they did not purchase the foods.

With the advent of new technologies, it has become possible to track the food purchasing patterns of grocery-store customers. The introduction of Universal Product Code (UPC) and Local Product Code (LPC) scanning systems was one of the most significant developments in store technology. . The systems allow cashiers to simply wave product codes over a scanning device, and the product and price are recorded into the store's database. Grocery stores and food manufacturers became aware of the potential commercial use of this technology to track customer food purchasing patterns, and began developing additional systems to take advantage of this technology as a marketing tool. Systems are now in place to track the purchase of specific foods, and to print out store "cents off" coupons for specific products. Stores have o developed "frequent shopper" programs, which give shoppers discounts for using a scannable card, which allows the store to record an individual's food purchasing patterns. The data then can be used by store managers to identify trends in purchasing and to measure the success of product advertisements.

Supermarket scanning systems include a feature known as a payment tender, which is used to identify the method of payment. Tender systems identify whether a shopper paid cash, used a check, or used an automatic teller machine (ATM) or credit card, as well as whether the customer used food stamps or WIC fiscal instruments (FIs). The great majority of scanning systems used today have the tender feature, which means the store database can link the tender code to the products purchased to identify products as having been purchased with WIC FIs.

WIC transaction records have been successfully linked to WIC program participation records. A study by Health Systems Research, Inc. (HSR) released in 1998 examined the feasibility of linking WIC transaction data to participant demographic data. The study was able to identify a methodology that used the stores' tender systems to record the WIC

check number (the same way they record personal checks) into their scanning system, which allowed the transaction database to be linked to the program administrative database.

Methodology

Although this initiative's main focus is the matching of point-of-sale transaction data with State administrative records, several preliminary activities must take place in order to accomplish this goal. The proposed methodology outlines the preliminary steps that must be undertaken to make practical use of the data, and then discusses the actual system that will be used to match the data.

1. Preliminary activities needed to implement the initiative

a. ***Identify and select a State to participate in the initiative***. In its initial application, the initiative should be undertaken in just one State. The first step is to identify a State in which to conduct the initiative. States that do not use retail food delivery systems for their WIC programs–such as Mississippi and Vermont, which use direct food delivery systems–would be excluded from the initiative. In addition, because of unique food delivery issues, Alaska, Hawaii, and tribal WIC State agencies also would be excluded. ERS can work with the FNS regional offices to identify the interest level of specific States.

b. ***Identify the range and scope of stores capable of scanning food purchases***. Once a State is identified for the project, the geographic area of consideration within the State needs to be selected. In order to select a geographic area with a significant number of WIC clients who shop at stores that have scanning technology, an assessment first must be made of which stores have scanning equipment. (Assessment of the size of the WIC clientele in those stores is discussed in the next item.) To do this, an analysis of the State's authorized WIC vendor databases will be conducted. State WIC officials collect a significant amount of demographic information about grocery stores on an annual basis prior to authorizing the stores to participate in the program. Among the data collected from grocery stores is a determination as to whether or not they have scanning capabilities. Using these databases, one can create a file of all authorized stores that use scanning equipment.

c. ***Determine the number of WIC participants who patronize stores with scanning technology.*** The next step is to determine the percentage of WIC participants who use those stores. The State WIC program maintains records of how many WIC participants in a given state use each grocery store every month. By merging the file of stores with scanning equipment with the file of the number of transactions by store, a picture of the number and percentage of all WIC participants using stores with scanning systems can be developed.

d. ***Develop the "universe" of stores that can be included in the initiative***. Once the total number of participants using stores with scanning equipment is determined, geographic profiles of areas where WIC participants shop at stores with scanning equipment can be developed. A geographic area should be selected that has a large percentage of its WIC population using stores with scanning equipment. This will minimize the number of WIC participants shopping at stores without scanning equipment. If possible, the area also should be chosen to be fairly representative of the State as a whole in terms of the types of families served by WIC.

e. ***Develop the scope of UPC/LPC codes to be collected***. Once the extent to which participants shop at stores with scanning equipment has been determined, the next step will be to examine how the technology can be targeted to specific program issues. One obvious method is to limit the UPC/LPC codes read to those related to specific WIC-approved foods (additional codes would not be a problem were they to be provided).

2. **Collecting and matching WIC transaction data with program client demographic files.** A number of steps will be needed to implement this stage of the initiative once the preliminary activities are complete.

a. ***Recruiting stores to participate in the initiative.*** Stores are under no obligation to participate in this initiative. One of the main activities will be to convince store management to allow the government to have access to their UPC/LPC databases. In addition, store managers must be willing to enter the WIC FI number into the database. By limiting the geographic area to be examined, only a limited number of stores will need to be approached. In order to ensure that the recruitment effort is successful, alternative geographic areas will be identified as a backup.

b. ***Training store personnel***. It will be necessary to either train store personnel on how to enter the WIC FI number into the transaction database or to provide materials to training officers employed by the stores. All personnel who handle transactions will need to be trained. Training materials and quick reference guides will be developed for cashiers to have at store registers.

c. ***Collecting data***. Data detailing purchases over a 2-month period from all WIC transactions will be collected from each store in the sample. The data often are processed weekly by stores, so access to the data should be relatively quick. Data can be transmitted electronically from the store to either USDA or to USDA's contractor representative. Once the data are received, they will need to be matched with the UPC/LPC identification files so records of the foods that were purchased can be created. The

UPC/LPC files contain the product name, date purchased, quantity, price and size.

Once these data are obtained, the client demographic file (the Electronic Benefit Transfer (EBT) file for FSP participants, and the reconciled WIC transaction file from WIC agencies) will need to be acquired from the State agency in order to match the administrative data with the UPC/LPC transaction files. Complete administrative data from State agencies are often available within a 3-month period after the date of benefit issuance. Using the links created in the UPC/LPC database, the data are matched, and an analytic database is then created with the State administrative file linked to the foods purchased. These data then are analyzed to determine food purchasing patterns.

Advantages and Limitations

A number of advantages to expanding this technology exist. These include:

1. ***Enhancing the ability of State program officials to examine purchasing patterns relative to specific program objectives***. By showing how the technology can be used to monitor important aspects of WIC program service delivery, the technology becomes useful to both Federal and State program administrators. The technology is simple to use, which will allow State and local program administrators to make decisions to use the technology–once developed–to study community-based or Statewide food purchasing issues.

2. ***Eliminating the bias of self-reporting of food purchases when evaluating nutrition education activities***. As noted earlier, the nutrition education efforts of WIC program nutritionists tend to focus on making sure the client understands the value of purchasing and eating the prescribed foods, and directing clients to select foods designed to address a specific nutritional risk. It is important to know whether these goals are achieved. Much of the nutrition data gathered for this purposes is based upon self-reporting of what foods are purchased. Obtaining vendor records of foods purchased through the program will go a long way in determining whether self-reporting is accurate, and may, in fact, be used to eventually replace self-reporting for items covered by the WIC program. Having information about which WIC foods are purchased also will help local nutritionists modify their approaches to conducting nutrition education related to WIC food purchases, if necessary.

3. ***Providing a deterrent to program abuse if store personnel know that purchasing patterns can be examined***. If store owners know data can be collected and analyzed to determine if proper foods are being purchased, they will be less likely to allow participants to substitute unapproved items for authorized foods.

4. ***Providing an effective tool for program administrators to monitor client purchasing trends***. Because the data can be collected over time, patterns of food purchasing can be monitored.

There are some limitations to the expanded use of this technology, including:.

1. The methodology cannot be applied to stores that lack scanning systems. In the case of the WIC program, this will eliminate stores that are more likely to be out of compliance with food purchasing regulations. The number of stores not using scanning systems is small and continuing to drop (Bell et al., 2001).

2. Rural areas may be underrepresented, because many small rural stores do not have scanning equipment. However, they also are likely to serve small numbers of clients.

3. In addition to UPCs, local product codes (LPCs) are used to identify products sold either as store brands, or products such as deli cheeses that are sliced to order. The codes used for these products are established by individual stores, and therefore will vary among stores and from region to region. It will be important to include LPCs in the database. Extensive research will be required to identify the full range of WIC-approved foods with LPCs at each store included in the project.

4. The methodology cannot provide a complete picture of all foods purchased by WIC participants. Nutrition professionals would find it extremely useful to know the foods WIC participants purchase in addition to those provided by the program. The store's UPC/LPC database captures all foods purchased by WIC participants when they use their WIC checks, so this information can be included in the database. But purchases made by WIC participants when they are not using their WIC checks are not recorded in a manner that identifies their purchases as being made by a WIC participant. This means that only a limited amount of data is available for each WIC participant.

   This disadvantage can be reduced if a store has a frequent shopper program that tracks all purchases by an individual or family. Identifying information from the WIC UPC/LPC file can be linked with frequent shopper data to create a more complete profile of WIC participant shopping patterns. However, stores located in low-income areas may be less likely to have these types of programs, so the extent to which this approach can be used is unknown.

Feasibility Issues

While both studies cited earlier show the technology to be feasible, there are a few feasibility issues that must be examined, including:

1. It may be difficult to create complete individual food purchasing pattern files if a person shops at one store with scanning equipment and another store without it. This is particularly an issue for the WIC program, where milk might be purchased at a small neighborhood convenience store, but the family does most of its shopping at a larger store. If all WIC participant records are linked to the stores where products are tracked, these purchase amounts can be subtracted out of their total redeemed amount to determine the extent of this problem.

2. Store owners may be reluctant to cooperate and provide the scanned data to State or Federal officials. The WIC feasibility study was successful in convincing stores to provide data, but in that case, there was no threat the data would be used for program compliance. Store owners may be reluctant to cooperate if they know the data may be used for program compliance. In order to convince store owners this project is worthwhile, it will be important to focus on how this technology will improve program outcomes and operations from the stores' point of view.

3. Confidentiality issues that deal with matching transaction data with purchasing data will need to be examined. If only State or Federal officials use the technology, the confidentiality issues will be limited. But if the technology is used by nongovernment researchers, there may be significant confidentiality issues. In the feasibility study discussed earlier, client identification information was stripped from the files prior to analysis. However, clients may not be happy to learn the government is in a position to use personal identifiers to examine what they are purchasing even if that step is never taken. To minimize this perception, client identifiers should be stripped from all data records immediately after scanner and program data are linked.

4. Cashier errors made when keying WIC FI numbers into the tender system was one of the key issues in examining food purchasing patterns of WIC participants identified in the HSR feasibility study. This suggests that another pilot project that is designed to examine the potential to barcode WIC FIs be undertaken. Currently, WIC FIs have a micro-encoded line on the bottom that displays the check number, so banks can read the FI. The FI then can be matched with State records. The WIC FI number could be added there as well. As noted earlier, to track purchasing patterns of participants at the grocery store, the FI number needs to be entered into the store's UPC database. It is inefficient for store clerks to punch in FI numbers, and errors can occur during the procedure. It would be better to barcode the FI number on the WIC check so that it could be scanned along with the check number. This way the FI number would be in the database linked to the purchase. The data then could be matched to the State participants' file.

5. Local Product Codes (LPCs), reflecting products sold as a store brand, or products such as baked goods, deli items, etc., vary by store and by region. Incorporating LPCs into the UPC database is possible, but a cumbersome task. The difficulty of this task may be reduced by examining only certain targeted UPCs and LPCs.

Potential Cost of the Initiative

- Based upon information provided in the feasibility study, this technology is likely to be cost-effective. By limiting the geographic area and the range of UPC/LPC codes to those used for WIC, at least in the initial rollout, costs of managing the data file should be minimal Problems could occur if the database gets so large that it becomes expensive to maintain and analyze.

- A second cost factor will be the cost of obtaining participant transaction and demographic data from State agencies. This will be a relatively minor cost, as State programs prepare reports using these data on a monthly basis, and will only have to copy the files.

- The third cost factor will be the personnel time necessary to create the data files and analyze the data. This will involve programmer staff as well as analytic staff. In the prior feasibility studies, standard statistical software (SAS) was used for data analysis, and was determined to work for this project. So no special programming will be required.

- One other cost factor that must be considered is the cost of training store clerks to enter WIC check numbers into the scanned database. Many larger stores have training systems set up to train new cashiers or to train cashiers in new checkout procedures, potentially holding down costs. There will be some cost involved at the store level for entering the WIC FI number into the system, but because most stores currently enter personal check numbers into their systems, this cost should not be significant.

- The only other cost that might be associated with the initiative is the purchase of data from the stores. It is not likely stores simply will give their data files to government officials or researchers without charging for copying the file. In the WIC food purchasing study, the cost of obtaining data from 10 chain stores for a 4-month period was $1,000. This would extrapolate to around $100 per store per 2-month data collection period.

This initiative would provide an important database for research related to the WIC program. The database would be a valuable resource for researchers and administrators, allowing them to: enhance the ability of State program officials to examine WIC program participant food purchasing patterns relative to specific program objectives; eliminate the bias of self-reporting of food purchases when evaluating nutrition education activities; provide a deterrent to program abuse if store personnel know that purchasing patterns can be examined; and provide an effective tool for program administrators to monitor client purchasing trends.

Both researchers and program administrators could use the database to examine food purchasing behavior, to evaluate nutrition education efforts, and to assess program compliance as those elements relate to client well-being and program administration. This database should support research that is initially limited to select States or localities, but

with significant efforts, could become national in scope. The database could be updated by FNS on a regular basis for continuing research by researchers and program administrators.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
	X		X	X	X	P	P	X	Moderate	Substantial	Moderate

Notes on Individual Criteria:

(2) Allows program administrators and researchers to examine purchasing trends, estimate the targeted impact of eliminating or adding foods to the approved WIC food list, and evaluate nutrition education interventions.

(3) Allows researchers to examine basic program outcomes related to whether or not WIC participants purchase all of their WIC foods, and whether they take advantage of nutrition education. Also allows officials to examine if stores are in compliance with Federal regulations.

(4) Could provide information on participation dynamics if WIC records are linked across months (years) and/or to other data sources.

(5). Could become national in scope, depending on the extent to which WIC participants shop at stores with scanning equipment.

(6). Allows for ongoing research on food purchasing patterns in response to State agency efforts to modify their food packages, increase or modify nutrition education efforts, or expand grocer compliance efforts.

(7). Limited by the extent to which stores will be willing to 1) provide data, and 2) have their cashiers key in the WIC check number.

(8). Implementing the initiative will be expensive at first, as store personnel are trained, databases created, and systems for merging and editing files identified. Once in place, cost of maintaining the system would be moderate.

(9). May be objections from WIC clients to their purchases being monitored, even though clients will not be identified.

4. Using the Internet to Collect Program Data from State and/or Local Agencies

Initiative Summary

This initiative proposes to use surveys administered from a central website location to collect ongoing program data from State and/or local agencies. Currently, if USDA wants to collect survey information from State or local agencies, it must either mail out survey forms to potential respondents or contract with a survey company to conduct telephone interviews. Even simple surveys, such as asking State WIC directors about changes they might be considering to food packages, or asking local WIC agencies about nutrition education plans, require a significant amount of effort and cost. Advances in Web technologies and security, as well as the prevalent access to Web browsers by government offices make Internet-based data collection a feasible alternative to the more traditional mail, telephone, and in-person data collection methods.

State and local agency personnel would access a central secure website to complete ongoing surveys, provide program administrative data, and/or access the results of these data collection efforts. Most State and local program administrators interviewed as part of the key informant interview process indicated that sufficient data to support program administration and research currently is being collected at the local level. But many of these data are not reported to State or Federal officials because the mechanism to report them does not exist. Federal WIC officials noted during the key informant interviews that an initial review has been conducted by USDA Information Technology Staff on developing an Intranet capacity that would link the Federal office with State WIC agencies. There has been very little done in the way of examining the use of the Internet as a survey tool. This initiative offers one approach to addressing this need. The proposed system is generally applicable across a broad range of USDA program areas. This discussion focuses on how it might be implemented within the WIC program.

Use of Internet technology would allow data to be collected, analyzed, and dispersed without going through the cumbersome processes typically associated with data collection methodologies such as regular mail or e-mail surveys, which might require production of hard copy, multiple mailings, and data receipt and data entry activities. The use of Netscape and Explorer browsers is widespread among business users, and most computer programs developed to collect information are supported by these browsers. Thus, the Internet represents an ideal medium for providing nearly universal access to program data and resources from almost any geographic location.

Background

Hotz et al., (1998) found that policy makers and program administrators in the future will require more and better data sources to adequately monitor program operations and evaluate program outcomes. Their report said greater efforts are needed to ensure that comparable and high-quality data are gathered across States, and across agencies within States. The data development initiative seeks to address this need by using the Internet to host surveys and computer-assisted data entry programs to collect administrative and

survey-type data in a uniform way. Data can be collected, cleaned, and posted back to a central website location quickly, enabling more timely use of the data by interested (and authorized) USDA agency and program staff. The ability to obtain expedited data will allow USDA to better deal with current policy and program issues and also may broaden the types of research it conducts.

RTI has developed project-specific websites and implemented data collection and reporting systems using the Internet for the U.S. Department of Education (National Postsecondary Student Aid Study, and the Integrated Postsecondary Education Data System); the U.S. Department of Health and Human Services' Substance Abuse and Mental Health Services Administration (the National Household Surveys on Drug Abuse); the Office of Indian Education (Office of Indian Education Study); the North Carolina Employment Security Commission/Department of Labor (Occupational Information Network (O*NET)); and confidential commercial clients.

Methodology

The initiative proposes to build a central website and associated data collection and reporting system that would allow USDA to use the Internet for the transfer of data files, the collection and dissemination of survey or program administrative data, and as a communication resource for USDA agencies and personnel. Both the amount and type of data that could be included on this kind of website are wide-ranging. The initiative proposes focusing on the use of the website to field surveys or computer assisted data entry programs, but the website could be expanded to serve as a repository of different kinds of program data, information resources, survey data, budget reports and analyses, contact lists, and protocol information.

The first step in building a system and website is to hire one or more contractors to build the website, establish the site's security, develop access levels, and develop a data collection implementation plan. A contractor could administer short surveys on topics of interest to USDA or use the website as a low-cost way to collect program administrative data. For example, surveys could be administered to WIC agency directors on hours of operation, outreach efforts, plans for expanding clinic sites, and nutrition education activities. State WIC directors could be asked such questions as plans for changing or limiting the WIC food package in the coming year, activities conducted related to the training of WIC vendors, or activities related to the coordination of WIC and immunization programs.

The following issues will need to be addressed before a system is designed to collect, analyze, and report survey or program data:

Access Issues. Several different kinds of access issues exist—which agencies will have access to the system, which staff within these agencies will have access to the surveys or data posted on the website, and what kind of modem or browser will be most compatible with the kinds of instruments deployed on the website. At a most basic level, the system will be limited to those with access to the Internet. The program codes should strive for "platform independence," meaning that the code used to program a survey or data entry

program should *not* be dependent on viewing by any particular browser. As long as the user has Explorer or Netscape versions 3.0 or higher, the user's ability to view and respond to surveys should be a minor issue. A larger issue for USDA to consider in terms of system management is whether to permit all agencies to have access to the system or to confine the initiative to a single agency, as a pilot study of the system.

System Design. Once USDA determines the target audience or target group of programs that the system should be designed to serve, the Department will need to acquire a contractor (or contractors) to design and build the website. In addition to the design of the Web page, there are issues associated with building and designing a secure website, including security, access, how the surveys or computer assisted data entry programs will be designed and implemented, and how results will be analyzed and reported.

Survey or Computer Assisted Data Entry Program Design. Any survey or data entry program designed for deployment on the Internet must be designed to function with multiple browsers. At a minimum, surveys should be designed to be viewed using either Netscape or Explorer at versions 3.0 or higher. If a user has a nonstandard browser, such as Lynx, a text-only browser, then surveys or data entry programs that use graphics of any kind will not function at an optimal level. Before designing a survey or data entry program, USDA needs to find out first how widespread the more standard Netscape or Explorer browsers are among the regional offices, and whether nonstandard browsers may be a problem.

Speed. Anyone who accesses the surveys or data entry programs will need to go through an Internet service provider (ISP) to access the website. Organizations like USDA often act as their own ISP, rather than having employees dial to an ISP like America Online, to gain access to the Internet. If any USDA offices must dial an ISP using a 56K modem rather than obtaining access through high-speed telephone lines, the rate at which the survey questions are delivered and the data are entered could be slowed down. For data entry programs where error checking is done at the "user end" rather than at the central server site, dialing through a 56K modem and having data checked before being sent could also slow down the application.

Security. The system must be designed to be accessible through the Internet but protected by an NT security system using user identification and password protection. USDA may also want to explore the option of using Secure Sockets Layer (SSL), which is an information technology that provides encryption between the user's Web browser and a Web server. All data sent over an encrypted SSL connection are protected with a mechanism that detects tampering. This is the type of encryption system that is used when credit card information is sent over the Web. SSL technology will cause a 10–15 percent decrease in the speed with which data are transmitted, so USDA may want to consider selectively applying SSL to the website, such as only on pages where data are entered.

Exhibits 1 and 2 are Internet pages extracted from surveys conducted by Research Triangle Institute (RTI). They are examples of Web pages used to collect data over the

Internet. Exhibit 1 is a set of questions from an employee satisfaction survey, including a question that permits open-ended text. Exhibit 2 reflects more of a computer-assisted data entry approach, where data can be keyed in a uniform manner, despite differences in how individual schools may maintain these kinds of records. Using either of these approaches, the resulting data file can be easily filtered (i.e., subset or sorted) or analyzed to prepare summary-level or unit-level reports.

For the website to function not only as a location where data are collected but also as a location from which data are disseminated, it could be designed with several layers. The top-level page of the website could be for "public use," where data or surveys available to all USDA personnel might be posted.

Subsequent layers of the website would be accessible depending on the level of user access. By clicking on one link from the main page, a Food Stamp Program director could access a special survey targeted just to Food Stamp Program directors, as well as deeper levels for site-specific data. By clicking on another link, WIC directors might have access to a data file of WIC applicants by region. Typically, for a website such as this, as data are received and processed, the website is refreshed daily. It is made available 24 hours a day, 7 days a week. If technical support were required, technical support staff would need to be accessible via an e-mail link available from every page of the website.

An example of how a contractor could implement the Internet-based system could include a survey of all WIC agency directors on issues related to clinic operating hours and staffing patterns. To start with, the contractor would send a message explaining the purpose of the survey to an e-mail list of WIC agency directors. The website address (URL) would be included within the body of the e-mail, and the agency directors would be asked to click on the website address to participate in the survey. Upon clicking on the website address, the director would be taken directly to the centralized website. As mentioned earlier, the website may have a "public use" component, where data are posted for viewing by all USDA personnel. There also would be a special log-on box for those with access to the particular survey. In such a case, the directors would log on using a password and an ID and thus gain access to a separate page within the website. The WIC survey would appear on that page. In this way, only those with appropriate access would be able to participate in the survey.

Exhibit 1—Example of Web-based employee satisfaction survey questions

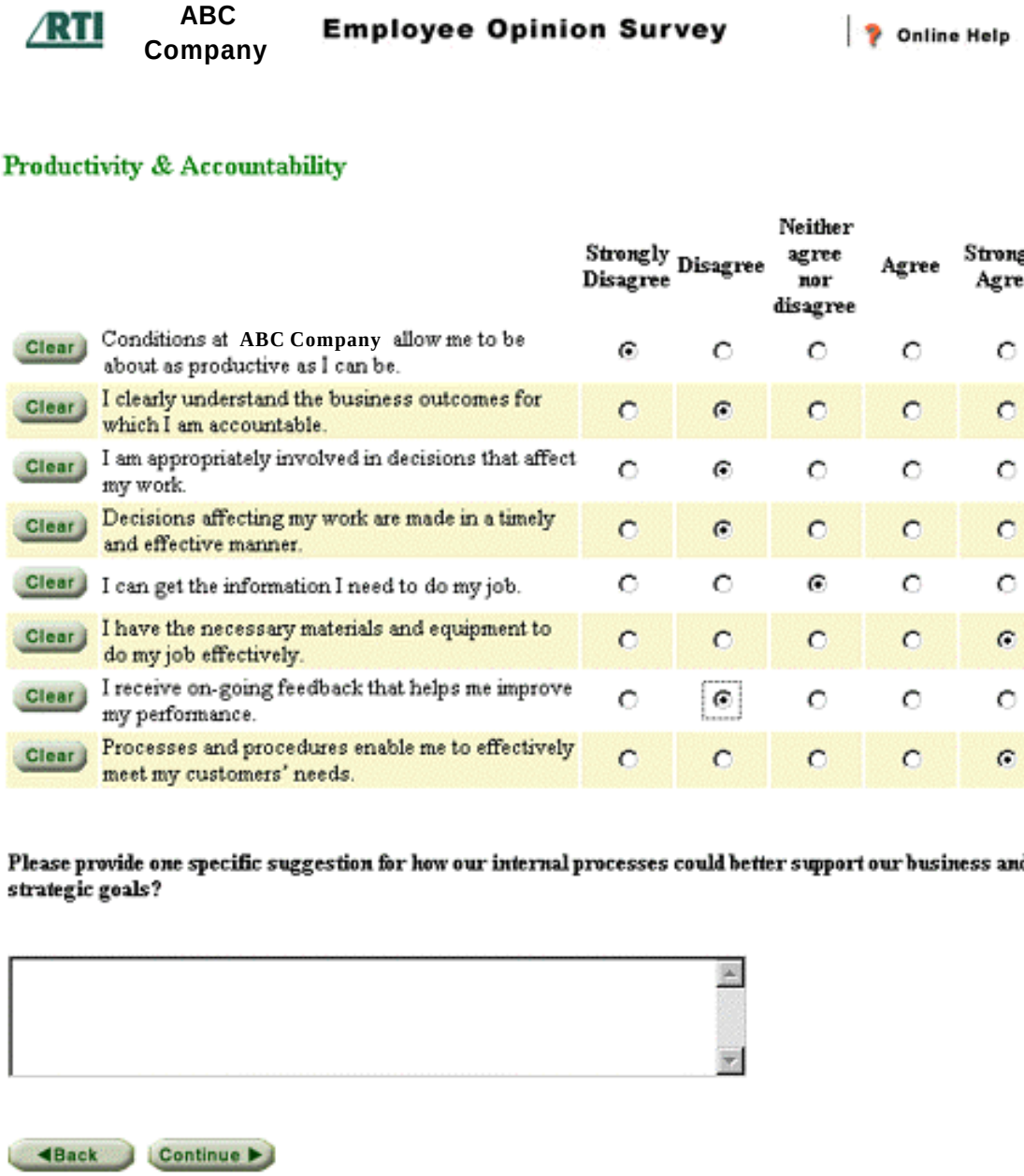


Exhibit 2—Example of Web-based computer-assisted data entry program

Period: 2000-2001 Data Collection
District: A
School: District A Elem. School

Kentucky Data Collection Project

Worksheet 1 Services to Students	**Worksheet 2** Community Involvement	**Worksheet 3** Law Violations
Worksheet 6 Board Violations	**Worksheet 9** Victims and Offenders	**Worksheet 10** Firearms Expulsions
Worksheet 11 School Security Measures	**Worksheet 12** Suspensions by Transportation Code	**Worksheet 13** Expulsions by Transportation Code

School Worksheet #4: Incidence of Law Violations

Incident Counts

7. Please report the number of incidents involving students and non-students in each of these categories.

				# of Incidents Resulting In:			
Incident Type	**Total**	**# of Weapons Related Incidents**	**# of Gang Related Incidents**	**Calls to Police**	**Arrests**	**Charges Against Offenders**	**Civil Damages Pursued by Victims**
Arson	0	0	0	0	0	0	0
Weapon - A. Handgun	0	0	0	0	0	0	0

Continue | Go Back | Save and Exit

After completing the survey, the WIC director would exit the page, and the data would then be uploaded automatically to the survey database. Should the WIC director have to exit before finishing the survey, the data could be saved so that when the director re-enters the survey the survey would begin on the last page before the breakoff. Previous responses will be stored so that if a respondent wants to back up, the earlier responses would be available for review. As with any survey, all respondents will have the ability to modify answers, clear questions they have completed, and return to the survey anytime until the they choose to click the "submit" button. Once a survey is submitted, the respondent will no longer have access to it, as only one response would be allowed per agency. As the data are processed from each respondent, the system can determine who still has not responded, allowing for "reminder" e-mails to be sent to those who still have to provide data or complete a survey. The notification process can be automated if desired.

Once the survey data collection period is closed, the survey would no longer be active on the Web. Data from all respondents then would be cleaned and reported in an agreed-upon report format, or made available on the website to directors (and others) as a spreadsheet, text, or database file. The advantage of the website, and the strength of this type of data collection initiative, is that summary-level or detailed reports can be designed, displayed, or disseminated almost automatically once the data are collected. Reports can be designed for various levels, depending on the audience Exhibit 3 displays a "mock" high-level report that is geared to a broad audience and is designed to allow the reader to click on various links within the report to access data files or related tables. These types of reports can be automatically generated, depending on the tables or text required.

A website can be deployed either within the protected firewall of USDA, or, if preferred, an equally secure site can be deployed outside the firewall. Operationally, and from a user's perspective, there is no difference in either the logon operation or the level of security. A contractor might create and deploy a website on a Windows/NT server. It would be accessible through the Internet but protected by an NT security system using user ID and password protection. Security of the data and the website contents is assured using NT security, which incorporates folder-level locking on files on the server. Differing levels of security can be established for personnel working on any given survey and/or for other USDA personnel. To ensure that this security is not breached, the website should not be registered with any search engine or search service and the URL address should only be given to authorized personnel.

Data from surveys or other sources can be stored in a SQL Server database maintained either internally at USDA or by a contractor (SQL Server is a Microsoft database and data analysis program). In either case, the database should be backed up nightly, and data can be maintained indefinitely so that it can support USDA for ongoing reporting and analysis. One possibility for processing and disseminating data is for the contractor to provide USDA a weekly data feed of all survey responses or program data. Depending on confidentiality requirements, individual identifiers can be removed prior to giving the data to USDA. As mentioned earlier, the data can be delivered in various formats or the contractor can create a set of summary reports with data collected in the survey. Whenever standard reporting spreadsheet templates can be utilized, report production can be more easily automated. All reports that are generated should be done with an understanding that if confidentiality of the data is an issue, all identifiers should be removed such that individual agency or program directors (individuals or offices) cannot be linked to any particular set of responses.

Exhibit 3—Example of a "mock" report with embedded Web links

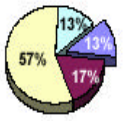
Spring 2001

PERSPECTIVES

Report on the People, Progress, and Performance of Global Corporation

13% of current clients expect legislation to begin significantly affecting prices in the **first** quarter of this year.

When do you think prices will begin to change?



INSIDE PERSPECTIVES

STAFF VIEWS

- Key stakeholders reveal their major concerns
- Management and Staff rate the top executive
- What benefits are important to you?

CLIENT SIDE

- Why they are turning to our competitors
- Top 3 reasons women select our products
- How advertising influences the decision to buy

Survey Participation

Click a category for detailed demographics about each group.

Employees	972
➢ Management	195
➢ Staff	777
Key Clients	83

Global Climate Remains Warm

Spring 2001 Client/Staff Survey Results

Fair Fares - A question of Cost

When client purchasing representatives were asked to communicate their reasons for choosing/not choosing specific Global products, we found that 73% felt the maintenance costs were too high.

Global Consulting fees, seem to be in line with the competition, as 80% of those surveyed felt the fees were appropriate.

Click each percentage to view a Regional breakdown.

How do you rate Global's fees for:	Appropriate	Too High
Maintenance	27%	**73%**
Consulting	80%	**20%**
Initial Setup	60%	**40%**
Upgrade Activities	55%	**45%**

The Power of Influence

Major factors influencing the decision to partner with Global include:

- **42%** – Referral by an existing Global client
- **32%** – Industry specific need
- **12%** – Advertisement in trade publication
- **5%** – Professional workshop

Major factors influencing the decision **NOT** to partner with global:

- **38%** – Perceived expense of the business relationship over time
- **25%** – Actual expenses
- **10%** – Lack of knowledge about Global service line
- **10%** – Long term contract

Click here to find out how you can become a marketing resource for Global Corp.

Presidential Performance

Global management & staff rated their top executive on four key issues.

Global Corporation President Ian Sparks, III

Advantages and Limitations

There are several reasons to consider the use of a centralized website to field surveys, collect program administrative data, and ultimately disperse data for use by State, local, and Federal agency personnel. First, the kinds of analyses supported by this initiative are limited only by the kinds of data available at a State or local level. Although State and local agencies may have different automated systems to support their programs, the use of a centralized website to host a computer-assisted data entry program can allow the uniform collection of data, facilitating comparisons at either the local or aggregate level.

Second, with Internet-based data collection, response time is shortened. For paper and pencil surveys, the forms need to be designed, printed, mailed, and returned. The surveys then need to be keyed in, edited, and cleaned, most of which is eliminated by online systems. Thus, timely analyses of time-sensitive issues are often delayed. With Internet-based data collection, USDA could post a survey on its website, limit the survey response period to a week or two, and then analyze the survey results. For time-sensitive research questions, such as how a program is adapting to a new legislative directive (e.g., various welfare reform changes), this permits a much quicker analysis of the policy or program options.

Conducting a pilot program within one food assistance program is a cost-effective and manageable way to implement an Internet-based data collection system. In the future, the initiative could be expanded to include many of the food assistance and nutrition programs. Potential research questions that could be addressed include:

- The Food Stamp Program could use the Internet to survey State FSP administrators on changes being made or planned in outreach efforts.
- The child nutrition programs could use the technology to survey local school districts on issues related to transportation and participation in the school breakfast program, private vendor access to campuses, and limitations on servings for the free and reduced-price school lunch program.

The benefits of using a Web-based data collection from a centralized website include:

- Access. The hardware and software that are required are standard for almost every office site in the country. As long as an office has access to the Internet and uses Netscape or Explorer at version 3.0 or higher, there are minimal barriers to that office participating in a survey or answering other requests for data. Rather, barriers to survey participation will depend more on an agency's degree of willingness to participate in the data collection activity and on whether an agency has the desired data.

- Multi-Use System. A centralized website offers the opportunity not only to collect data but to post it back on the website in files that can be subset, sorted, analyzed, or otherwise displayed to all (authorized) interested parties.

- Timely Analyses. Data are collected and fed instantaneously into a database and can be analyzed in real time for results (and response rates, if desired), enabling analysts to examine data within a narrower timeframe. Similarly, there is no transcription or

data entry necessary as there is with a mail survey, so the error rate is reduced and data are available more quickly.

- Better Quality Data. Validation checks are programmed directly into the instrument, which allows the programmer to specify what answers are permissible and in what format. This ensures that the survey will be completed, and that the responses to questions will be usable. There is very little data cleanup required because there are fewer opportunities for skipped questions, multiple answers to single-answer queries, or ambiguous responses.

One potential limitation of this kind of data collection approach is that its success depends on how willing State and local agencies are to reply positively to requests for data. For this reason, USDA may want to consider the use of incentives, such as a small cash payment, gift, or coupon, to encourage participation. Another limitation is the extra effort that may be required on the part of State and local agencies to produce data that "fits" the requests being fielded on the website. For example, if a request is posted that asks WIC directors for information on the number of WIC applicants by demographic status, and this information is not typically captured at the local level, then it may require more work for the agency to produce the required statistics. This is primarily a survey design consideration. USDA would need to ensure that the data requests mirror the way data are typically captured within any given agency, or at the least, allow some variation in the range of acceptable responses. Finally, as with any survey involving more than nine individuals, USDA would need to obtain Office of Management and Budget (OMB) clearance. Since these surveys will be done electronically, they may be more likely to gain quicker approval than surveys that require filling out a hard copy form. If similar kinds of data are going to be collected on an ongoing basis, USDA might want to work with OMB to obtain a bundled clearance, submitting multiple data collection requests in a single clearance package.

Feasibility Issues

One of the most important issues to consider when designing surveys or attempting to collect administrative data is how the confidentiality of the data will be maintained, and how the privacy of program participants (if participant-level data are collected) will be addressed. There are several options for USDA to consider. One option is for USDA to house the website on its own server and handle all survey deployments or all deployments of computer-assisted data entry programs. With this option, USDA personnel would be responsible for controlling the security access to the website, managing the various data collections, and ultimately preparing reports and files to post back to the website. A disadvantage to this option is that it may raise concerns about the confidentiality of the data, as the officers and directors who are asked to provide data may want assurances that their data will not be linked directly to them, or their programs.

Other issues to consider is whether to allow all State and local agencies to access and use this website, and if not, how to determine and regulate appropriate access levels. As mentioned earlier, one option is to consider a design that has a top-level "public use" page where program and survey data that are acceptable for widespread dissemination could be posted. More restrictive access might be considered, depending on the kind of

data that has been collected. USDA may also want to consider having all data transmissions encrypted as they are sent over the Internet, which would provide an additional layer of assurance and confidentiality to the State and local agencies.

An additional consideration is whether State and local agencies will "buy in" to this method of providing data and how best to obtain their cooperation. Will participation in surveys or being asked to provide data in a way that differs from how the local agency typically stores the data be perceived as an administrative burden? Will the ability to access the resulting database be a sufficient incentive to overcome this kind of "barrier"?

Another issue to consider is how the website should be designed and maintained. An outside contractor could potentially design and build the website, including implementing security control as needed. USDA might want to use more than one contractor, such as a website developer, a data collection organization, and an analysis organization (for the design and analysis associated with reporting the program or survey data). USDA could potentially design, develop, and manage the website and data collection efforts internally.

Potential Cost of the Initiative

The costs associated with establishing a centralized website will be based on the number and types of surveys or computer-assisted data entry programs that are deployed on the website; the design and maintenance of the website; and the number and types of reports that are desired. The more the reports can be automated, the less labor will be required to produce the reports. Should USDA prefer that only the raw data be made available–without creating reports–then costs would go down. USDA will need to consider the following costs when developing a plan for this initiative and considering its overall desirability:

- Website Design and Maintenance Costs. The costs associated with designing and maintaining the website are reflective of the complexity of the website design and how frequently the design might change. These costs will primarily be labor costs.
- Survey or Computer-assisted Data Entry Program Development and Implementation Costs. Costs will reflect the number and types of surveys or computer-assisted data entry programs that are deployed on the website. Costs also might include managing the data collection effort, development of the "sample" to be surveyed, e-mail notification to sample members and later to nonrespondents, use of incentives, and operation of a Help Desk during data collection.
- Data Processing Costs. Costs will reflect the degree to which survey or program data can be collected in a uniform manner. The more "cleaning" the data may require, the more costly the data processing will be. If someone needs to review survey data to ensure that no individual names are mentioned or that no ID numbers are included with the data that are uploaded to the website, labor costs will be increased. These costs also will include developing electronic codebooks for the user community.
- Report Design and Dissemination Costs. The number and types of reports that are desired will affect costs. The more the reports can be automated, the less labor will be

required to produce them. Should USDA prefer that only the raw data be made available—without creating reports—then that would decrease costs.

- Security Design and Maintenance Costs. The number of different access levels needed will affect the initiative's cost. The more complicated the access structure, the more time will be required to track password changes, allow new employees to have access, and disable access for former employees. These will be primarily labor costs.
- Storage Costs. Data storage costs will vary depending on the amount of data stored. Storage space is typically "leased" on a gigabyte per month rate (e.g., $90/GB/month) and likely would be charged this way should USDA use a contractor to deploy the website. Storage costs are not likely to a large component of the initiative cost. Rather, labor costs, including data collection management, analysis, and reporting, and overall website management, will be the primary expenses in implementing this initiative.

The initiative compares favorably with more traditional data collection methods such as mail or in-person surveys, because production, mail-out, and receipt costs are lower. Further, there are no data transcription costs using computer-assisted data entry programs.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
X	X	X	X	X	X	P	X	X	Low	Moderate	Moderate

Notes on Individual Criteria:

(2) The proposed website would be a valuable resource to administrators at the Federal, State, and local levels, allowing them to share information about their respective programs through survey data or through the collection of standardized program data.

(6) The website could be updated when new surveys or requests for program data are fielded, analyzed, and posted on the website.

(7) The technology is not a major barrier because most offices have access to the Internet.

(8) The cost of labor to create and maintain the system should be moderate. Survey and reporting design costs, data processing costs, and storage costs, are expected to be high at startup, but to diminish significantly after implementation.

(9) Client privacy is not a concern if the initiative is used strictly for administrative purposes.

5. Using a Probability-Based Web-Enabled Panel to Collect Data from Low-Income Families through the Internet

Initiative Summary

This initiative proposes to use an existing panel of families equipped with Web-enabled interactive TV to collect data on food shopping patterns, the effect of a mother's labor-market on food preparation and the provision of meals, knowledge of nutrition, effectiveness of nutrition education messages, participation in food assistance programs, and other issues relevant to low-income families (both program participants and nonparticipants).

The base sample of participants is available from Knowledge Networks, Inc., a strategic partner of RTI, founded in 1998 by two Stanford University professors, Norman Nie and Douglas Rivers.[6] The company is headquartered in Menlo Park, CA and provides Web-based survey data collection using a scientifically representative sample of households equipped with Web-enabled interactive TV. KNI supplies volunteer participating households with an interactive TV device and free access to the Internet for up to 3 years.

Conducting periodic surveys of a nationally representative sample of low-income families generally would reveal important information not just about program participants but also about the pool from which food and nutrition program participants are drawn. USDA could use these surveys to gather information on reasons for not participating in and reasons for leaving programs, as well as family and economic dynamics that immediately proceed, accompany, or follow program entry and exit.

The Knowledge Networks panel is the only probability-based Web-enabled panel currently in existence in the United States. So the use of the panel would require a contractual arrangement with Knowledge Networks, Inc.[7] The creation of a new panel would be too costly to develop solely for the use of USDA. Building onto the existing panel with a targeted sample of low-income households might be a cost-effective alternative to obtain new information from Web-based surveys. This option is discussed in more detail later.

Background

The use of an existing panel to field short surveys presents some unique opportunities for USDA to collect data from a targeted, nationally representative sample of low-income families. Unlike telephone surveys or paper and pencil surveys, the use of Web-enabled surveys permits the incorporation of multi-media, making it an ideal technology for testing visual items, such as nutrition education messages or strategies. The use of a panel

[6] More information about KNI can be found at www.knowledgenetworks.com/index2.html

[7] RTI's strategic partnership with Knowledge Networks, Inc., gives RTI the first right of opportunity to use the Web-enabled panel for any competitive government procurements for which the Web-enabled panel is an appropriate methodology.

sample also lends itself well to longitudinal survey designs, allowing USDA to measure knowledge, behavior, and attitudes over time.

In a recent article, Couper (2000) categorizes the different types of Web surveys currently in practice (table 3). While many of the approaches listed by Couper rely on volunteers with Internet access, the Knowledge Networks panel is an example of what he refers to as Category 8, which is the only approach that allows generalization beyond the current population of Internet users (Couper, 2000). This approach solves two of the major problems of Web surveys—that of coverage bias and lack of browser compatibility. Providing recruited sample members with Web access solves the coverage concern to a large extent. To address the second concern of browser compatibility, every panel member is provided with the identical equipment to view the surveys, allowing the survey instruments to be delivered in a consistent way to all panel members.

Table 3—Types of Web surveys[8]

Nonprobability methods	Probability-based methods
1. Polls as entertainment	4. Intercept surveys
2. Unrestricted self-selected surveys	5. List-based samples
3. Volunteer opt-in panels	6. Web option in mixed-mode surveys
	7. Prerecruited panels of Internet users
	8. Prerecruited panels of full population

Knowledge Networks has recruited and maintains a Web-enabled panel of 125,000 sample members. The panel is recruited using stratified random-digit-dial (RDD) telephone sampling. As part of a household's agreement to participate in the panel, it is provided with an interactive TV device and free access to the Internet for 2 to 3 years. This innovative sampling methodology distinguishes the probability-based Web-enabled panel from other Web panels that are essentially nonrandom samples of households that already have Internet access.

The probability-based RDD sample is drawn from all 10-digit telephone numbers in the United States. Households that do not have a telephone are not covered (approximately 6 percent of U.S. households), nor are households in areas where Web-TV is not available (approximately 13 percent of U.S. households). The sample is purged for numbers outside the Internet provider's range for obvious nonworking numbers and businesses. All numbers that pass the purging process are sent to a commercial reverse address-matching service. All matched numbers receive an advance letter and incentive information, and all numbers passing the purge process are sent to a telephone interviewing organization for recruitment. A random 50 percent subsample of the unmatched numbers also is included in the final sample sent for recruitment.

[8] Couper, Mick (2000). "Web Surveys: A Review of Issues and Approaches," *Public Opinion Quarterly*, Winter 2000, p.477.

RTI has used the Knowledge Networks panel for several recent studies. Most recently, RTI and Knowledge Networks conducted the Survey on Health and Aging[9] in August and September 2000, with more than 6,300 panel members aged 18 or older. The survey completion rate was 78 percent, meaning that 78 percent of eligible households completed the survey during the data collection period. RTI also conducted two surveys for the Bureau of Transportation Statistics in 2000 on airline satisfaction. One survey was conducted with a sample of 4,795 adults aged 18 or older, who were asked to complete a 35-minute questionnaire. Respondents were offered a $10 incentive; the survey completion rate was almost 71 percent. For the second survey, a total of 1,546 panel members were sampled for the study and asked to complete a 20-minute questionnaire via interactive TV without an incentive. The survey completion rate was 78 percent.

For any individual survey, a sample is drawn from the panel using a stratified probabilities proportional to size (PPS) design in which the measures of "size" are essentially poststratification weights that make the panel resemble the U.S. population along the following sociodemographic dimensions: gender, age, race/ethnicity, region, metropolitan status, and education. These measures also include adjustments for sources of unequal probabilities of selection in the panel creation State. These sources include multiple telephone lines per household, some geographic oversampling, and subsampling of households without an address match, which are mostly unlisted numbers.

When a household agrees to participate in the panel, it is provided with free hardware (an Internet appliance that connects to a television), free Web access, free password-protected e-mail accounts for each household member age 13 and older, ongoing technical support, and an incentive program to encourage continued participation. A condition for participation in the panel is that the household must own a telephone and a television set. Upon recruitment, potential participants are informed that the expected length of panel tenure is 2 to 3 years and that they will be expected to answer short questionnaires on a weekly basis for that time period. Knowledge Networks is responsible for recruiting individuals into the panel, instructing participants on how to install the appliance that provides access to the Internet and to the surveys, and providing ongoing technical support as needed. Occasionally, panel participants drop out of the panel. Attrition is just less than 1 percent (personal communication with Lisa Thalji, RTI, Jan. 19, 2001). Since panel participation is restricted to 2 to 3 years and the panel is in its infancy (i.e., recruitment began in November 1999), attrition has not proven to be an issue at this point.

The Knowledge Networks panel reflects the broad diversity and key demographic dimensions of the U.S. population and tracks the U.S. population closely on age, race, Hispanic ethnicity, geographical region, employment status, and other demographic elements. The differences that do exist are small and are corrected statistically in survey data (i.e., by nonresponse adjustments). There is a modest under-representation of Hispanics, due primarily to the fact that households where Spanish is spoken exclusively are currently not in the Web-enabled panel. In addition, panel members tend to report

[9] The study results and methodology may be found at http://www.rti-knowledgenetworks.org/.

slightly higher household incomes and a higher education status than the U.S. population. The most important factor in this difference is the exclusion of nontelephone (predominantly low-income) households from the panel.

An example of the growing interest in this new technology is the nearly a dozen papers based on Knowledge Network's (formerly called InterSurvey) data that were presented at the annual meeting of the American Association of Public Opinion Research (AAPOR) in May 2000 (table 4).

Table 4—Recent presentations on Knowledge Networks Web-based data collection: American Association of Public Opinion Research (AAPOR), May 2000

Rivers, D. (Knowledge Networks). Probability-based Web Surveys – an Overview
Dennis, J.M. de Rouvaray, C, and Couper, M. (University of Michigan) Questionnaire Design and Respondent Training for Probability-based Web Surveys
Frankovic, K. (CBS News). Internet Panel Response to the 'State of the Union' Address: An Experiment
Greenberg, A. (Kennedy School of Govt.) and Bocian, M. Uncertainty in Web-based Polling
Krotki, K.P. (Knowledge Networks). Sampling and Weighting for Web Surveys
McCready, W., (Northern Illinois University) and Tortora, R. (Gallup). A National Random Sample Comparison Between CATI and Web TV
McLaughlin, T. (Cyber Dialogue). Customer Database Research: Guidelines for Complete and Ethical Data Collection
Nie, N. (Stanford), Price, V., and McCready, W.C. A Preliminary Examination of a Joint Survey/Online Discussion Method for Studying Electoral Attitudes
Price, V. (Annenberg). Opinion Expression and Opinion Gathering on the Web
Rainie, L. (Pew Trust). The Pew Internet and American Life Project
Terhianian, G. (Harris). How to Produce Credible, Trustworthy Information through Internet-based Survey Research

Methodology

This initiative proposes to use the Knowledge Networks panel to institute a survey related to food and nutrition (or other topics of interest to USDA) with low-income families on the panel. As with any survey tool, this initiative presents an opportunity for USDA to collect data on attitudes, knowledge, and behavior patterns. The difference is that this information will be collected using Web-enabled interactive TV, rather than through slower and more costly telephone or in-person surveys. The types of questions that can be included on Web surveys are comparable to those administered via mail or phone but with some additional features. A Web survey allows for the incorporation of skip patterns, as a survey administered using computer-assisted telephone interviewing (CATI), it also permits the incorporation of graphics and audio.

The combination of panel structure and multimedia access might lend itself well to a nutrition education evaluation. For example, one could randomly assign panel members to treatment or control groups and give "pre" and "post" food intake surveys to both groups. The treatment group could be given a nutrition education multimedia tutorial on the joint USDA-HHS Food Guide Pyramid during the intervening period.

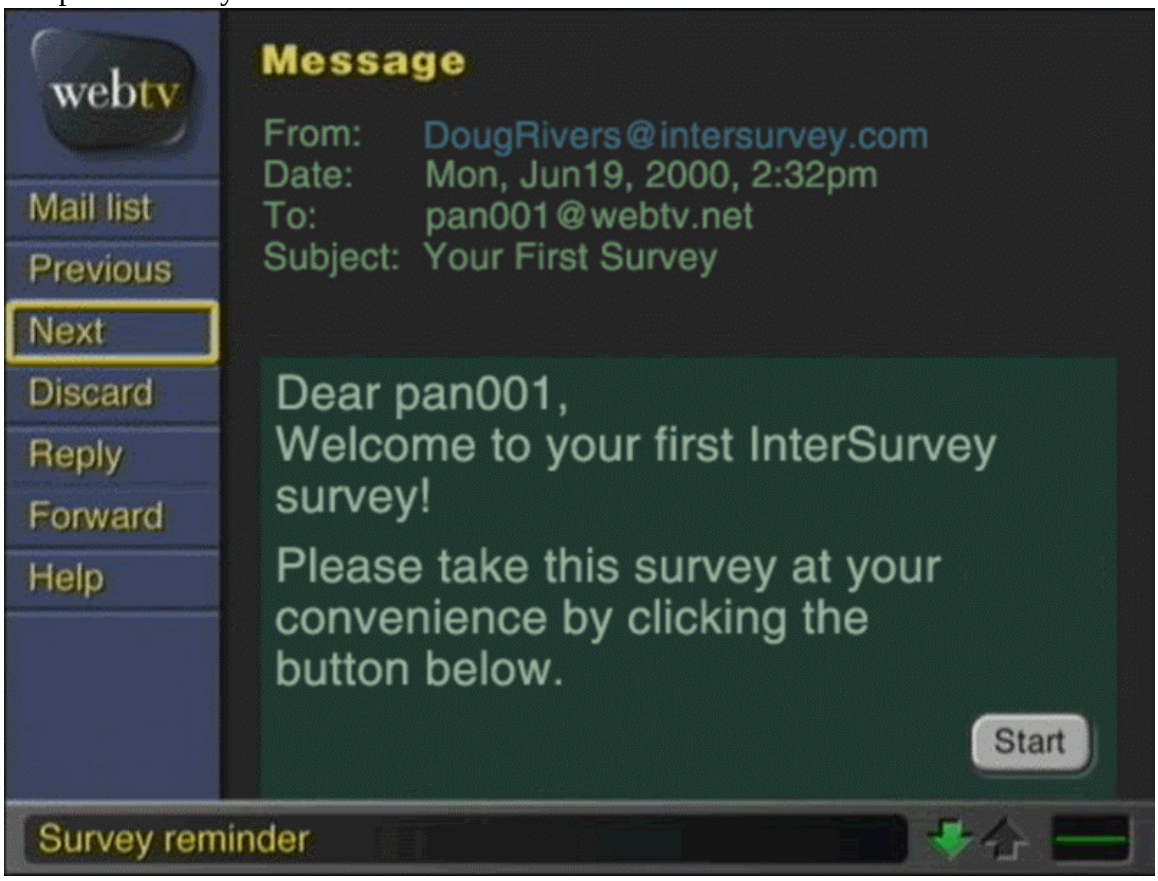
A further application of this technology could utilize brief screeners, that queries a sample of panel members about the occurrence or nonoccurrence of a particular event, such as whether the panel member experienced any of a series of symptoms of food insecurity or whether or not the panel member applied for food stamps. Given a positive response to the screener, the panel member would then receive a longer questionnaire. No further demands would be made on households that did not experience the event.

- Surveys could cover a broad range of issues for the FSP, and, if sufficient sample sizes exist for the other programs, for WIC and the NSLP. Examples of research questions that could be addressed are provided in the Advantages section.

It is important to note that while the data collection methodology is different from a more traditional mail, in-person, or telephone survey, this initiative still proposes using a survey to collect data from individuals. In this respect, sample design, questionnaire design, nonresponse analysis, and weighting will all play an important role in any study that uses this technology.

For panel participants who are also participants in USDA programs, various administrative data could also be linked to the survey data, where available. Survey data might be combined with food-purchasing data obtained from grocery stores. At this time, data is not available on what percentage of low-income participants in the Web-enabled panel are currently receiving food stamps, WIC, or other USDA program assistance.
To initiate a survey, the contractor would send an e-mail message to selected panel members who satisfy the screening criteria for the particular study. The e-mail message informs the recipient that a survey is waiting for him or her. Exhibit 4 provides an example of an e-mail message informing a panel member that a survey is waiting. The participant clicks on a button within the e-mail to start the multimedia questionnaire and responds to the survey questions by clicking on the desired response. Surveys longer than 15 minutes are broken into segments and administered incrementally to avoid overburdening the respondent. This is important because short surveys are significantly more likely to be completed and submitted promptly by respondents. In general, nonrespondents to surveys are sent up to two e-mail reminders to complete the survey; telephone reminder calls are also an option.

Exhibit 4—Example of an e-mail sent to a panel member, inviting the member to complete a survey



The project team suggests this type of data collection methodology be a pilot program first, using either a regional or a small national sample. To date, neither Knowledge Networks nor RTI have used the panel to field surveys specifically targeted to low-income populations. A pilot study is recommended that either uses the existing panel alone or supplements the existing panel with an area frame sampling methodology to specifically target and recruit low-income households into the panel. The latter option is discussed in more detail in the Feasibility section of this report.

The findings also need to be compared to results from other national studies within comparable sociodemographic groups. The biggest challenge to the generalizability of survey findings will come from the initial recruitment to the sample. The restriction of the national sample frame to households with telephones may mean the poorest of food stamp participants will not be part of the Panel. CPS data (in 1998) indicate that 94 percent of all U.S. households have telephones. This rate of telephone coverage by income is shown in Table 5:

Table 5—Telephone coverage of U.S. households

Characteristic	Percent with telephones
Total U.S.	94.1
Race/ethnicity:	
White	95.1
Hispanic or Latino	88.8
Black	88.1
Family income:	
$30,000 or more	98.0
$20,000–29,999	96.4
$15,000–19,999	93.8
$10,000–14,999	91.2
$ 7,500–9,888	88.7
$ 5,000–7,499	82.4
Under $5,000	75.7
Source: Current Population Survey, 1998	

A related issue concerns the extent to which families living in nontelephone households may vary from families living in households that do have telephones. The National Health Interview Survey (NHIS) (1985–86) suggests that with variables such as the percentage of people who were overweight, this difference can be relatively small (23.7 vs. 24 percent). On the other hand, regarding other behaviors, such as exercise, it appears more likely that nontelephone households are less likely to engage in a health behavior (32.4 vs. 40.5 percent) (unpublished RTI analysis). Accordingly, USDA would need to be cautious about these kinds of differences in interpreting survey results.

At the conclusion of the data collection period, USDA could either obtain the raw survey data and weights to analyze and report the data or work with a contractor to have the data analyzed and reported in an agreed-upon format. The file could then be made publicly available for research.

Advantages and Limitations

This type of probability-based Web-enabled panel survey offers several important benefits for data collection:

- Speed. By using the Internet, interviewing and data collection speed is greatly accelerated. Survey fieldwork can be completed in days, instead of weeks or months. Sample size does not determine the length of the field period.

- Sample Quality. Since the panel is recruited using RDD sampling techniques, it represents a true, scientific, population-based sample. Data collected from the Web-

enabled panel are more reliable than data collected via Web surveys that use self-selection or quota sampling.

- Multimedia interviewing. This technology is ideal for administering surveys that use audio, video, and three-dimensional (3D) graphics in the questionnaire. The richer content broadens the types of questions that can be asked. From the respondent's point of view, the inclusion of video, audio, and 3D graphics in the questionnaire makes the survey experience much more engaging and less burdensome than conventional telephone interviews.

- Low respondent burden. Panel members complete no more than one short survey per week over a 2- to 3-year period. Surveys are self-administered in panel members' homes at their convenience.

- Reduced Screening Costs. Information on each panel member's household composition and income are maintained, permitting a more targeted survey approach to families with particular income or family characteristics. These data also need not be collected with each survey.

- Cost. Web-enabled panel surveys provide sample quality and interviewing capability on par with the best face-to-face surveys at costs equal to or below telephone surveys of high quality. Costs will vary depending on the complexity of the survey and the required sample.

Research questions that could be addressed using this technology include:

- To what extent have food assistance programs adapted methods to increase participation among low-income families? Among low-income working families? Do clients find these efforts satisfactory?
- What kinds of shopping strategies are used in households where resources are limited?
- What kinds of nutritional messages work best for this population, using print, audio, or television? How can USDA tailor nutrition education strategies to meet the needs of low-income families?
- What is the level of knowledge among these families about USDA food assistance programs and HHS Medicaid programs, and what are the factors associated with higher awareness of these programs? Does awareness differ by urban/rural status, family composition, age, cultural background, and economic circumstance?
- (For low-income working families) How does a parent's work (or more working hours) affect the provision of meals in a family, including food preparation, meals outside the home, shopping patterns, etc.?
- What is the nutritional intake of children in low-income families? Do factors other than income and time play a role, such as peer-group socialization or heightened stress?

- What kind of nonfood assistance do low-income families need? What about access to preventive healthcare?

As discussed earlier, a limitation to the use of this panel is the extent to which the survey findings will be generalizable to the population of low-income households. Other issues are whether attrition among low-income families will become a factor as the Panel ages over time, and how successful the contractor and USDA can be at achieving high response rates. These latter two issues are discussed in detail in the section that follows.

Feasibility Issues

The panel exists and surveys are being fielded using this panel. In this respect, there are no barriers to USDA using the panel. USDA would have to work with a contractor and with Knowledge Networks to ensure that an appropriate survey instrument and sampling strategy are developed before implementing any data collection activities with panel members.

As with any survey of more than nine individuals, USDA would have to obtain OMB clearance to field surveys with this panel. If similar kinds of data are going to be collected on an ongoing basis, USDA might want to work with OMB to obtain a bundled clearance, submitting multiple data collection requests in a single clearance package.

Another issue to think about related to feasibility over the long term is whether this relatively "new" panel (i.e., less than 3 years in operation) will experience higher attrition rates as panel participants reach their 2- to 3-year limit on the panel. Higher attrition, if it is associated with certain demographic characteristics, could potentially provide a less stable basis for data collection (Dennis, 2001).

Finally, USDA must always consider the issue of response rates. With this type of survey, a panel response rate is computed. The panel response rate takes into account four different stages of potential nonresponse. The more intense the nonresponse follow-up efforts are, the greater the potential for reducing nonresponse at each of these stages.

Potential places for nonresponse include:

- not responding to the initial (RDD sample) telephone recruitment offer
- not installing the Internet appliance once recruited
- not completing the first "profile" survey, which is required before any further surveys can be completed
- not responding to any individual survey (such as one that USDA might field).

While experience has shown that response rates to individual survey requests are more than acceptable, nonresponse at the initial recruitment stage is potentially more of a concern. Currently, the cooperation rate at the initial stage of panel recruitment (i.e.,

responding to the initial telephone recruitment offer) is 56.2 percent, reflecting the number who agree to participate when called. If one computes an AAPOR response rate (AAPOR response rate No. 3) which includes those telephone numbers for which no determination of household status can be made, the rate drops to 53 percent. While response rates at the initial recruitment stage are comparable to those observed on RDD surveys, this is just one of the four response rates that is computed to determine the overall panel response rate for any given survey. For this reason, serious thought should be given to the kinds of nonresponse followup that might be possible with any given survey.

As mentioned earlier, an option for USDA to consider with respect to implementing a survey targeted at low-income families, is the possibility of creating a special "custom" panel, to use to supplement the current panel. RTI has developed an area frame sampling methodology (i.e., drawing a national sample of households) to supplement the existing Web-enabled panel. The advantage to building a special panel is that it helps address some of the coverage bias that results from using an RDD-developed sample frame. This could be particularly beneficial in targeting low-income families. If USDA chooses to implement a special custom panel, RTI would be responsible for recruiting households into the panel using in-person recruiting and also would be responsible for installing the Internet appliance.

Potential Cost of the Initiative

USDA will need to consider several costs when developing a plan for this initiative and considering its overall desirability. Costs to the Federal Government include the following:

- Use of the Web-enabled Panel. RTI's strategic alliance with Knowledge Networks offers USDA the option of using a probability-based Web-enabled panel to implement surveys. The costs will depend on the number and types of surveys, the sample design, and the degree to which USDA requires any data processing, analysis, or weighting of the survey results. Since Knowledge Networks is the only supplier of a probability-based Web-enabled panel, any government agency using this methodology by necessity will have to work with a single supplier.

- Survey Design and Data Collection Management Costs. Survey costs will depend on the complexity of the instrument, the sample required, and the management of the data collection effort. A pilot survey with a smaller sample would be less costly than a more complex full-scale survey. Similarly, using the existing panel will be less costly than recruiting and building a supplemental panel. These will be labor costs primarily, as surveys administered over the Web do not require the "production costs" typically associated with mail or field surveys.

- Data Analysis and Reporting Costs. Depending on the kind of analysis and reporting required, USDA might choose to handle this internally or work with an outside contractor. Computing panel response rates can be fairly complex, so someone with

knowledge of survey response rates and the sampling methodology would need to work with USDA on survey results to ensure accurate response rate computations.

Panel designs have several economic advantages over one-shot survey designs. First, in a panel, the costs of hardware and recruitment can be amortized over the life of the panel. This permits more expensive sampling techniques to achieve higher response rates without having to pass recruitment costs on to clients in a "lump sum." Second, profile information on panel members can be collected just once, and then used in conjunction with data collected in later waves. In one-shot designs, demographic data must be collected with every survey.

Internet panels also have advantages over conventional panels. Recontacting costs are much lower than in a telephone panel because e-mail is cheap to send and does not require the respondent to be at home when a recontact is attempted.

Although the enrollment, maintenance, and participation incentive costs are substantial, on an individual study basis, the methodology is typically far less expensive than most telephone survey methods. As with any survey, the costs vary depending on the target population, the number of completed surveys required, and the length of the questionnaire.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
X	P	P	X	X	X	X	X	X	Moderate	Moderate	None

Notes on Individual Criteria:

(1) Because of sample size issues, this initiative may be limited for the WIC and NSLP.

(2) Researchers and administrators could use the surveys from the Web-enabled panel to examine such issues as FSP participation, income changes, family status, knowledge of nutrition, and reasons for nonparticipation in food assistance and nonfood assistance programs.

(3) Outcome information is available for a large sample of low-income families.

(4) Some information on participation dynamics could be provided if the surveys followed individuals longitudinally.

(5) Research could be supported that is either national in scope or limited to particular States/localities.

(6) Web-enabled panel is available for use, and surveys could be fielded on a regular basis for continuing research by USDA.

(7) Feasibility concerns include the ability to obtain a sufficiently high response rate and the ability to target a sufficient number of low-income families.

(8) Cost issues involve the development of periodic surveys, possible recruitment of a supplemental “low-income” sample of households, and data analysis.

6. GIS Internet Map Server (IMS) Applications for Project Management and Data Analysis

Initiative Summary

Key informant interviews with Federal, State and local program administrators and researchers, carried out as part of this data initiatives project, indicated that geomapping technology is perceived as a valuable tool for examining program outcomes, conducting needs assessments, and targeting specific populations for outreach.

The initiative will explore the use of GIS Internet Map Server technology for food assistance programs. Automated methods of geocoding State and national data from food assistance programs and linking the geocoded data to Census 2000 and vital records data would be discussed. Demographic characteristics or birthweight, which are important to the management and analysis of nutrition services, are examples of useful records data. The methodology for implementing GIS/IMS is discussed, along with examples of potential USDA applications.

Background

A geographic information system (GIS) is a geographic database management system that provides users with a powerful set of tools for viewing and analyzing geographic (i.e. map) data and performing spatial analysis. The geographic data in a GIS consist of a series of (spatially referenced) map layers that contain information about features that are located in specific locations. These could be: 1) census tract boundaries with sociodemographic variables collected by the U.S. Census; 2) WIC clinic locations and associated information, such as hours of operation or capacity; or 3) ZIP code centroids or boundaries with data on low birthweight births or Medicaid-eligible populations.

Because data in a GIS are spatially referenced, there exists tremendous potential for linking WIC and other USDA data with Census 2000 data, vital records, economic and other health outcomes data. Census 2000 data of particular relevance to WIC programs would include demographic data, such as populations of women of child-bearing age or populations of children under 5 years old, or economic data, such as the number of children under 5 years living at or below the poverty level. Vital records data contain information about the number of live births, behavior and risk factors of mothers, and birth outcomes, such as low birthweight, birth defects, and congenital anomalies. Methods for linking these data with food program data are discussed below.

GIS functions have the potential to be extremely valuable to the decision-making and program administration responsibilities of food and nutrition program managers and analysts. The most widely recognized and commonly used function of a GIS is map production. Maps provide the viewer with a visual picture of the distribution of features across a geographic area and the relationships among them. Other commonly used GIS functions include geocoding, distance computations, spatial queries, buffering techniques, and overlay analysis (Vine et. al., 1997).

To date, GIS has been underutilized in the social sciences (except, in the area of demography) and there is little documentation of its use for social services, especially applications developed for direct practice or administrative purposes (Queralt and Witte 1998). This situation is changing, however, prompted by The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (the "Welfare Reform Act"). Some GIS applications have been developed in response to program changes resulting from welfare reform legislation. GIS was used in North Carolina to respond to changes in the Child and Adult Food Program and an application was developed to automate eligibility determination of Family Day Care Homes, based on geographic location and census block group child poverty status (Hanchette, 1999).

For many years, the use of information technology in the health sciences lagged behind its use in other fields. But in recent years, there has been a rapid proliferation of GIS use for health applications. The contribution of information technology to public health is now widely recognized and has led to the emergence of the field of public health informatics. Public health informatics encompasses GIS technology and has been defined as "… the application of information science and technology to public health practice and research" (O'Carroll, 1997). Much of what has been written about GIS in public health informatics (e.g. Hanchette, 2001) also applies to its use in food assistance and nutrition programs.

Ten years ago, geographic information systems usually resided on powerful Unix workstations and required a long learning period for their use. Digital spatial data were not widely available and about 90 percent of the cost of GIS implementation comprised the development of map datasets. Over the past decade, GIS technology has become increasingly based on personal computers; the costs have continually decreased; and most GIS software interfaces are very user-friendly, with menu, tool and button interfaces. Perhaps even more important is the fact that a wide array of digital spatial datasets is now available, at little or no cost. The most widely used have been the U.S. Census TIGER/Line files, which include census boundaries, streets, major hydrology, and landmarks. Even with these changes, the acquisition and use of GIS by non-GIS users has been somewhat intimidating.

In the past 3 to 4 years, GIS technology has been revolutionized through the development of IMS technology, which provides users with access to mapping capabilities and spatial queries through the use of a Web browser such as Netscape or Internet Explorer. These require initial setup and programming (and some level of subsequent maintenance) by GIS personnel, including geocoding data (such as food assistance and nutrition program data) for the development of map layers. Once developed, they allow users to view and analyze spatial data and they require no GIS software purchases or data storage and little or no training for the non-GIS user. These capabilities can be password-protected for limited access, or made available to the public. Their potential of IMS technology for management and decision-making for health and social services programs is nearly untapped.

Three examples of IMS applications for public health demonstrate the utility of this technology, its ease of use, and its potential for food and nutrition services: 1) the Geographic Analysis Tool for Health and Environmental Research (GATHER), was developed by the Agency for Toxic Substances and Disease Registry (ATSDR), Centers for Disease Control and Prevention, HHS, to provide access to spatial data for analysis of public/environmental health issues (http://gis.cdc.gov/atsdr/); 2) the National Cancer Institute (NCI), National Institutes of Health, HHS, has an IMS application that allows users to produce customized maps of cancer mortality rates during 1950–1994 for more than 40 cancers (http://www.nci.nih.gov/atlasplus/); and 3) an interactive mapping application is available for *Women and Heart Disease: An Atlas of Racial and Ethnic Disparities in Mortality*, developed by the Office for Social Environment and Health Research at West Virginia University and the Cardiovascular Health Branch at CDC (http://www.cdc.gov/nccdphp/cvd/womensatlas/Statemaps.htm).

Methodology

There are three major tasks involved in implementing IMS for food assistance and nutrition programs: 1) to determine the GIS agency or organization that will be responsible for the geocoding of USDA data and the implementation and maintenance of IMS applications; 2) to acquire and develop spatial data for mapping and analysis; and 3) to develop and deploy the IMS application itself. The second requires the most resources. In addition to these three tasks, methods of linking USDA data to other data sources are discussed.

Identify Agency or Organization Responsible for GIS/IMS Development

IMS applications can be developed for any level of geography. They could be developed for the USDA, to collect data for all States, or, they could be developed by State and local agencies for more specific purposes. The decision about where to house the GIS/IMS operations will depend on the existence and location of GIS operations in food assistance and nutrition programs and their supporting agencies. The startup costs of IMS can be hefty and include the purchase of a server, and IMS software and deployment licenses. For agencies without GIS, it might be cost-effective to outsource IMS operations to an organization that is already set up to provide these services.

Acquisition and Development of Spatial Data

Many of the GIS datasets developed for health services or other applications would be important for food assistance and nutrition programs. There are a number of basic map layers, available at the State and/or national level, that are important to the development of nearly any GIS application in the health or social services. These are sometimes generically referred to as “framework” datasets. The status and availability of many of these datasets can be determined by searching the national Geospatial Data Clearinghouse (http://www.fgdc.gov/clearinghouse), an online data catalog developed by the Federal Geographic Data Committee.

Lee and Irving (1999) have provided an excellent compilation of spatial data for health planning. Many of the datasets they have described are critical for GIS applications for food programs. Table 6 provides a list (and some of the characteristics) of spatial datasets

important for the development of GIS applications for nutrition services. Many of the datasets would need to be created from food and nutrition program data through the use of geocoding.

Geocoding is a critical function of any data development efforts that involve the use of food and nutrition services or other computer files developed by State or Federal agencies. Geocoding is the process linking a record in a non-spatial database to a geographic feature or location in the map layer through a geographic identifier (i.e. geocode) such as State, county, ZIP code, census tract, or street address. These identifiers must be contained as a field in both the database record and the map layer information. Common identifiers, such as State, county and census units, are usually represented as Federal Information Processing Standard (FIPS) codes. For example, Somerset County, PA, has a FIPS code of 42111 (42 = Pennsylvania; 111 = Somerset County). Many State and local agencies use their own city or county coding systems, so it is critical to have an understanding of which coding system is in use prior to geocoding.

Table 6—GIS datasets for inclusion in IMS

Theme/database name	Source	Geographic coverage	Availability	Comments
TIGER/Line files	U.S. census	U.S.		
Sociodemographic and economic data	U.S. census	U.S.		Linked to TIGER Line files via FIPS codes
Food program service providers, e.g. WIC clinics, Family Day Care Homes	State government agencies	State	Varies by State	Computer databases imported to GIS and geocoded by county, ZIP code or street address
Food program participants, e.g. WIC participants	State government agencies	State	Varies by State	Computer databases imported to GIS and geocoded. Data may be aggregated by county or ZIP code.
Medicaid eligibles and participants	State government agencies	State	Varies by State	Computer databases imported to GIS and geocoded. Data may be aggregated by county or ZIP code.
Perinatal indicators	National Center for Health Statistics; State government agencies	State/National		Better geographic resolution with State data

A common method of geocoding address data is through address matching, which involves matching a numbered street address in a database to street network information in the map layer. This works well in urban areas, where the majority of residents have actual street addresses, but less well in rural areas where residents have post office boxes or rural route mailing addresses. The development of enhanced 911 (E-911) systems is changing this.

Geocoding provides: 1) the ability to link to other spatial data; 2) mapping capabilities; and 3) spatial data for use in other data initiatives. A major first step in developing GIS and IMS applications for food and nutrition services is to develop processes for the routine geocoding of USDA databases. Some States, such as South Carolina, have developed automated methods of geocoding vital records data (Laymon, 1999). The North Carolina Central Cancer Registry geocodes its registry data on a quarterly basis. These routine and automated geocoding methods could be developed to "spatially enable" State and national data from food assistance programs.

Confidentiality and privacy issues must be considered because address information, especially numbered street addresses, can contain identifiers. This is not as major an issue when data are of restricted use (internal to a program), but must be considered when data are distributed externally, or maps are published. Two common methods of addressing confidentiality of spatial data are: 1) to aggregate street address data to an area unit such as a census tract, county or ZIP code; and 2) to develop procedures to randomly displace geographic coordinates of point data (Armstrong et. al., 1999).

Geocoding has an additional, often overlooked, benefit. It can be used in the data quality assurance process as a way of flagging records that lack correspondence among geographic identifiers. For example, a single record should have city, ZIP code, and county correspondence. Lookup tables can be set up to flag records that don't. Since geography is often a basis for summarizing statistics (e.g. county-level mortality rates), assigning observations to an actual geographic location provides a means of validating information in certain data fields.

Developing, acquiring and processing spatial data are ongoing processes and, when developing funding mechanisms for GIS/IMS applications, funds for ongoing data maintenance should not be ignored.

Development and Deployment of Internet Map Server Applications

With today's IMS software, map server applications can be developed and deployed in a matter of days, once the geographic data have been developed or acquired. Of course, additional programming and applications development is needed for more customized applications. Out-of-the-box IMS functions include mapping; display features such as zooming in or out, panning, and symbol selection; spatial queries; feature selection; distance measurement; and buffering. Figure 1 shows a screen from RTI's Venipuncture Project Management Internet Map Server, which was developed with Redlands, CA-based Environmental Systems Research Institute's ArcIMS software. It is a password-protected application developed by RTI's GIS Program for RTI epidemiologists who need spatial information for project management and decision-making.

The Venipuncture project protocol requires blood samples for all study participants. Project managers need to make important decisions about the placement of field offices and the hiring of phlebotomists and subcontractors to draw blood. One of the logistical requirements is that patients live within 60 miles of the nurse/phlebotomist office. In

addition to the management of field staff, project managers are interested in the completion rates (percentage of study population with blood drawn) by primary sampling units (PSUs).

The map in figure 1 shows study participants (red dots) and phlebotomists (green and blue dots). Framework data, such as State boundaries are included, and county boundaries are drawn and labeled at larger map scales (see figure 2). Primary sampling units, drawn in pink, consist of aggregations of ZIP codes. The application was being used by Venipuncture project managers and their staff within hours of its completion, with no training other than a "cheat sheet" explaining the tool bar functions. Project staff have used it to: 1) add primary sampling units to the study; 2) allocate patients to phlebotomists, by using the buffer function to determine which patients lived within 60 miles of a specific field office (figure 2); 3) set up additional phlebotomy field offices in areas where patients weren't covered; and 4) track progress of PSU completion.

Figure 1—Venipuncture Internet Map Server, RTI

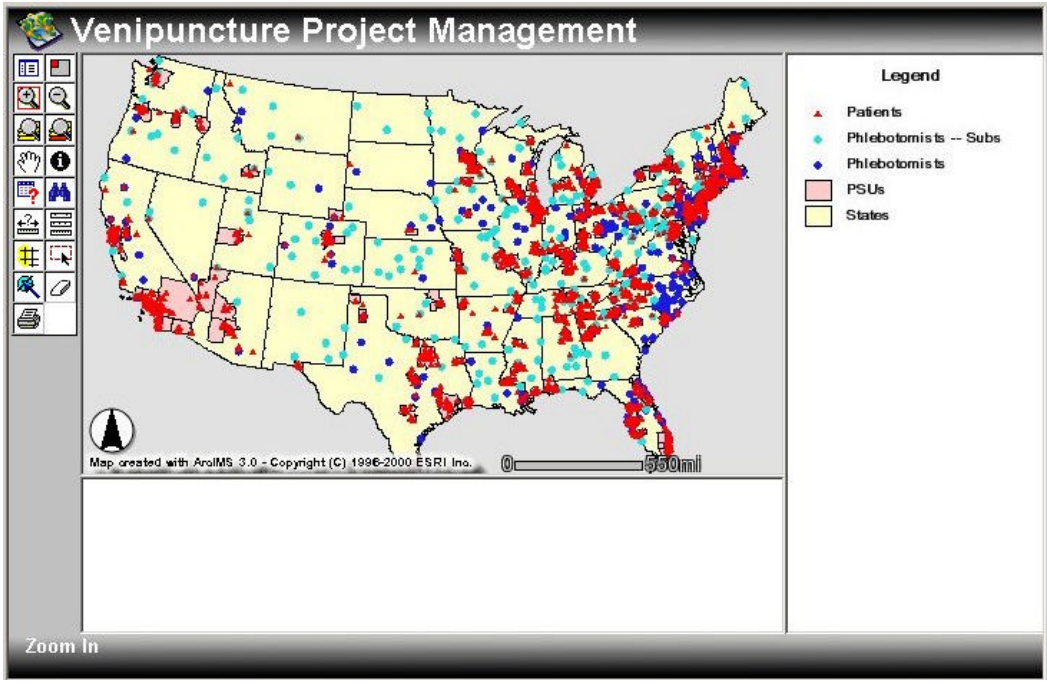
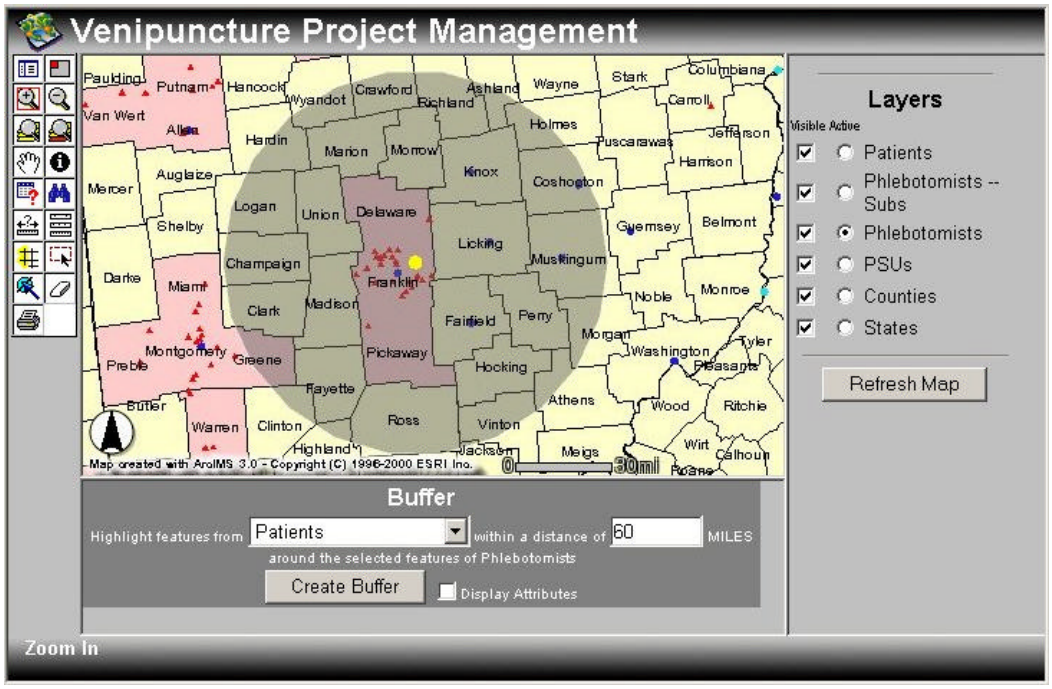


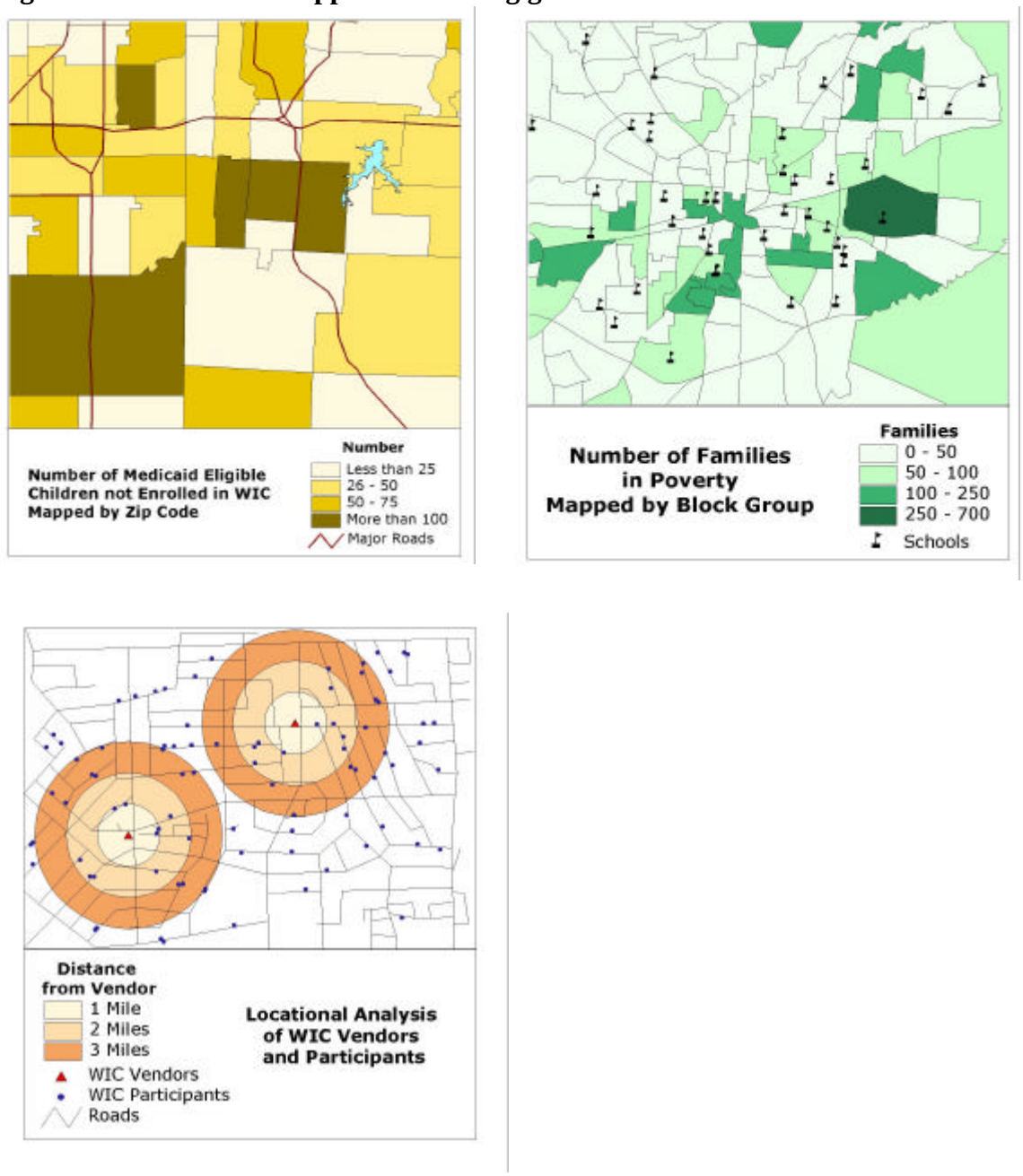
Figure 2—Study participants within 60 miles of field office



Many of the IMS functions used in the Venipuncture project, and standard to IMS applications, would be extremely valuable for the management and analysis of services provided by food assistance and nutrition programs. The maps in figure 3, while produced with standard GIS software, could easily have been produced with IMS software and demonstrate potential data and applications that could used by food program administrators. These could include map-based eligibility determinations for nutrition assistance program participants and locational analysis for the siting of new services.

Other potential applications might involve the location of grocery stores and food sources in poor neighborhoods. FSP and/or WIC participation rates could be shown in the background as area data, with FSP and WIC retailers and other food sources shown as point data. If food and nutrition services data were linked to geographic units, such as county or ZIP code, then GIS/IMS users could make their own queries and produce maps most suited to their analyses. Spatial queries could show areas where the proportion of all children who receive free school lunches was greater than a specified number or the ratio of the FSP population to persons in poverty was above or below a specified number. The *USDA Strategic Plan 2000-2005* includes a number of key outcome measures for reducing hunger and improving nutrition among children and low-income populations. Spatial data can be used to determine areas that fall below the targeted outcomes.

Figure 3—Potential GIS applications using geocoded USDA data



Data Linkages

Geographic identifiers provide a means of linking USDA data to other types of data, such as Census 2000 socioeconomic and income data, vital records and Medicaid data. In North Carolina, data on Medicaid eligible populations has been linked to WIC data, by ZIP code, to determine target populations for WIC programs. In this case, tabulations of Medicaid eligible women and children were made by ZIP code, using State-maintained Medicaid databases. These numbers were linked to tabulations of WIC participants, by ZIP code, to determine numbers and percentages of Medicaid women and children NOT

enrolled in WIC. Twenty-five North Carolina counties had high numbers of Medicaid eligibles not enrolled in WIC. ZIP codes areas within these counties were then targeted for higher enrollment.

Address geocoding results in a map, with residences or facilities pinpointed, as shown in the WIC participant map in figure 3. A GIS function, point-in-polygon analysis, can be used to determine in which polygon, or geographic area, a point lies. For instance, each of the WIC participant locations in the figure 3 map can be overlaid on a census tract map layer to determine which census tract the participant resides in. Census demographic and economic variables (e.g. number of children under 5 years of age; percentage of children under 5 years old living below the poverty level) for that census tract can then be applied to those participants. Hence, census data have been linked to WIC participant data through geography.

This method was used in North Carolina to determine reimbursement levels for Family Day Care Homes (Hanchette, 1999). Address data for more than 3,600 Family Day Care Homes (FDCH) were geocoded, then overlaid on a map layer of census block groups. The USDA had requested a special run from the U.S. Census to compute the percentage of children 0-12 years of age who were at or below 185 percent of the Federal poverty threshold. After overlaying the FDCH data on the census data (providing a direct link to the census data), spatial queries were made to determine which FDCH homes were located in poverty areas and automatically qualified for a higher level of reimbursement.

These same methods can be used to link WIC (or other USDA) data to vital records and other geographically referenced health outcomes data. Vital records data can be geocoded and aggregated by geographic units, such as census tract or ZIP code. Vital records contain information, such as low birthweight, that can be used for WIC targeting and screening efforts. A combination of census and vital records information, linked by geographic unit, may identify census tracts or ZIP codes with high numbers of low-birthweight babies and high poverty. The areas could be targeted for intervention, and the outcomes of intervention (the aim is to improve infant health), and tracked over time.

Advantages and Limitations

The advantages of GIS/IMS technology are many and include the ability to visualize information about food and nutrition services, map and analyze outcome measures such as those developed for the *USDA Strategic Plan 2000-2005* and make decisions about program implementation, management and administration. In North Carolina, the Childhood Nutrition Services Section (which administers the WIC and Family Day Care Homes programs) has used GIS to: 1) examine WIC vendor and participant locations; 2) develop means testing for Family Day Care Homes; and 3) map linked WIC/Medicaid eligible data, by ZIP code, to identify areas for targeting services. Additional advantages of automated geocoding, such as quality control, have been discussed.

Some of the limitations of using GIS technology and implementing IMS applications for USDA include: 1) restrictions due to lack of data or poor data quality; 2) erroneous interpretations of data by untrained users; and 3) production of misleading maps, due to

poor cartographic design. An additional limitation is the confidentiality of participant data used to serve the system. Internet map servers are automatically available for public access. Restricted access can readily applied and tightly controlled by login/password and system security functions. There may be circumstances under which program managers need to be able to visualize client locations and access client attribute information without a heavy investment in GIS software and training. IMS applications can meet these needs by making the appropriate tools available on the Internet.

.

Research questions that could be addressed by GIS/IMS technology include:

- Which areas of a State or county are experiencing a drop in FSP participation rates?
- What is the average (or minimum) distance that WIC or FSP participants have to travel to a service provider/food retailer?
- What percentage of eligible children in a high-poverty area participate in the National School Lunch Program or the School Breakfast Program?
- Where are farmers' markets located in relation to low-income populations?
- What is the spatial distribution of counties, or ZIP codes, with low scores on key outcome measures?
- Where are areas of high birth rates and high poverty located, so that better nutrition intervention program planning and administration can be targeted toward them?

Feasibility Issues

The use of GIS technology for the mapping and spatial analysis of nutrition services in North Carolina has demonstrated the feasibility of its use and its effectiveness in enhanced decision-making and program administration. The development of IMS applications for other applications has demonstrated its utility. However, each State has its own set of food and nutrition programs, organizational structure, budget processes and data collection and management systems and the feasibility of GIS/IMS implementation must be examined in these contexts. Key informant interviews with WIC program managers indicated there was some concern that State and local data systems were not sophisticated enough to utilize Internet mapping technologies. State and local systems must be examined on a case basis.

GIS is a much sought-after technology and comes with its own politics. Where GIS fits into an organization's structure and the funding and prioritization of projects all have political implications.

Potential Cost of the Initiative

The potential cost of this initiative depends entirely on the number of map server applications that will be developed, where these applications will be housed and who will be responsible for their development, and the accuracy of food and nutrition data that is collected by State agencies and the USDA. A pilot program with IMS technology should be applied specifically in one or two States. Costs would involve:

- geocoding USDA databases
- acquiring and processing spatial data for the IMS application
- purchase and installation of server and IMS software (unless this is outsourced)
- development of IMS application
- ongoing maintenance/update of USDA spatial data.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
X	X	X	X	X	X	P	P	P	Low	Moderate	Moderate

Notes on Individual Criteria:

(2) Administrators could examine a variety of program issues and outcomes, such as vendor and participant locations; map to linked WIC (or other program)/Medicaid eligibility data by ZIP code; and validate the elimination of food vendors from the program due to vendor fraud.

(5) While it is possible to use this technology on a national level, it will likely be limited to State/local analyses because each State program has its own organizational structure, budget processes and data collection that will affect implementation.

(7) The only feasibility issues include initial set up and programming (and some level of subsequent maintenance) of the technology by GIS personnel, which are relatively standard procedures to implement.

(8) IMS requires no GIS software purchases or data storage and little or no training for the non-GIS user.

7. Extension of State Projects Linking Administrative Data Across Programs and Over Time to Food and Nutrition Topics

Initiative Summary

This initiative would assess whether the Food Stamp Program data contained in databases linking participant-level information across multiple public assistance programs can be used in their current form, or with slight modification, for future empirical research.[10] The FSP is a logical focus when seeking new uses for State-developed "master files" of this sort related to food and nutrition policy. A number of such databases already exist due to the historical linkages between the FSP and the AFDC/TANF and Medicaid programs (UC-Data, 1999). Further linkages to WIC and School Lunch files may also be identified based on proposed work by Abt (2001), which concentrates on a few individual State datasets in isolation. If compiled across States, and possibly enhanced, researchers could use these same files to examine State trends in food assistance and nutrition programs in a variety of settings and make informative comparisons across States.

The initiative builds on a recommendation by Hotz et al., (1998) to foster the development of administrative data files for research purposes. Hotz et al., suggest that State and Federal agencies develop permanent, ongoing data capacities involving micro-level data from multiple administrative systems. They suggested that a centralized repository of information on administrative data efforts be established (and funded). FSP administrative records appear in many of the studies reviewed by Hotz et al., which puts USDA in a strong position to develop such a repository. USDA also may have the strongest interest in the range of individuals and families contained in such files, since its programs—particularly the FSP—serve a larger portion of low-income Americans than any other parts of the "safety net."

In this role, USDA could work as a broker, helping State agencies develop protocols that would allow researchers to access existing linked data files from numerous States for the purposes of food and nutrition research. After securing the approval of participating States, USDA could use the protocols to grant access to researchers at USDA or under contract to USDA interested in examining issues related to food assistance and nutrition programs in multiple States.

After establishing a brokerage agreement, USDA would store the data (or the rights to the data) at a centralized location. States that agree to participate would provide USDA with extracts either on a periodic basis or on an "as needed" basis. State agencies could provide periodic extracts that would be stored and archived at a centralized secured facility at USDA. Alternatively, USDA could request data on an as-needed basis from State agencies.[11]

---

[10] Several projects also include survey components that might be useful in future projects related to food and nutrition.

[11] This second option is less expensive because USDA would not need to create a centralized data warehouse.

USDA-approved researchers (i.e., those who satisfy the requirements of the broker agreement between USDA and State agencies) could then order the data extracts from USDA. USDA could provide these extracts in various media, including CD-ROMs, and through the Internet. Alternatively, USDA could require that researchers work at a "secured site" to ensure the confidentiality of the data. These "secured sites" could be restricted to only USDA facilities, or expanded to contractors and/or universities who meet SSA data security requirements. For example, Cornell University has a restricted data access center on its campus to process restricted data files from various government agencies.

The development of this centralized system would significantly expand the use of administrative data for research on nutrition-related topics. Without this type of system, researchers need to contact various State agencies and/or independent contractors to obtain State administrative data. Given that States often have different protocols for data access, it is not surprising that most researchers constrain their current analyses to only one State or county. A common protocol would also allow USDA to more efficiently plan future research projects that use administrative data without the costs and uncertainty of obtaining access to individual State databases. USDA could review its inventory of State databases in the centralized system to identify potential internal or external research projects.

Background

Throughout the 1990s, there was a major devolution of responsibility for social programs from the Federal government to the States. These changes started with early State waiver efforts involving AFDC, Medicaid, and FSP which allowed States to make changes to their income support programs (e.g., work requirements, time limits) by relaxing Federal requirements. They continued on a national basis with the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) in 1996. One of the biggest changes under PRWORA was the conversion of the AFDC program from an open-ended entitlement program into TANF, an appropriated block grant program. . PRWORA also reduced funding for FSP, tightened child eligibility for Supplemental Security Income, narrowed program eligibility for legal immigrants, targeted funding to family day care homes under the Child and Adult Care Food Program, and enacted reforms in the child care programs and in the Child Support Enforcement Program.

In response to the changes, several Federal and State agencies funded projects to examine a variety of issues related to welfare reform. For example, the Administration for Children and Families (ACF) in HHS funded projects in 17 States to study the implementation and effectiveness of State welfare demonstrations that began operating prior to the implementation of PRWORA.[12] More recent studies include 14 welfare

[12] All of the States selected had policies that promoted self-sufficiency (e.g., time limited assistance, strong work requirements, strong sanctions for non-participation), though there were some significant variations across State policies. These States included Arizona, Connecticut, Florida, Illinois, Indiana, Iowa, Maryland, Minnesota, Nebraska, New Hampshire, North Carolina, North Dakota, Ohio, Texas, Vermont, Virginia, and Wisconsin (Stapleton et al., 1999).

leavers projects funded by the HHS Assistant Secretary for Planning and Evaluation designed to provide information on the short- and long-term outcomes of welfare reform on individuals and families who leave TANF.[13]

Of particular relevance are the ERS-funded studies that examined the circumstances of FSP leavers. Looking at the linkage between FSP and other programs (e.g., TANF), ERS wanted to evaluate whether the new welfare reform rules were effective in moving families toward self-sufficiency and, more specifically, whether the FSP was still meeting the needs of low-income families. Most of the welfare and FSP leaver studies relied on linked administrative data extracts from several programs. In some cases, State agencies and independent researchers produced one-time data extracts to complete the proposed research project. In other cases, these entities established ongoing extracts that were updated on a periodic basis.

At the same time, advancements in the use of automated systems in social services increased the capacity of administrators to produce and store information about their programs (UC-Data, 1999). These advancements have helped policy makers and researchers meet the growing demand for administrative data for research purposes at the State and local levels.

The proliferation of State projects in addition to technological advancement by State agencies has allowed the agencies to produce linked data that has resulted in the creation of linked administrative data extracts in several States. UC-Data (1999) found more than 100 administrative data extracts in a recent survey of 26 States, many of which included links across several programs. They found 80 percent of those States had major external research projects or databases in development. Several other States also have created similar linked databases but full information is not available on State files not covered by the survey.

The information included in the linked data extracts varies significantly. Some State databases include extensive linkages across programs. UC-Data (1999) documented linkages among records from the Food Stamp, AFDC/TANF, Medicaid, Foster Care, Child Care, Job Opportunities and Basic Skills Training (JOBS), Child Support, and Unemployment Insurance (UI) programs in selected States such as California and Illinois. The linkages in other States were much less extensive, generally focusing on basic linkages across the Food Stamp, AFDC/TANF and Medicaid programs.

The structure of the files also differs across States (and projects). Some data extracts are cross-sectional and focus primarily on producing unduplicated counts of program participants. Other extracts, particularly those from the aforementioned evaluations, are longitudinal. Researchers and administrators can use them to track transitions over time But the longitudinal extracts tend to cover only a specialized subset of assistance

[13] The State and counties for the studies included: Arizona, Cuyahoga Co.–Ohio, District of Columbia, Florida, Georgia, Illinois, Los Angeles–California, Massachusetts, Missouri, New York, several other California counties, South Carolina, Washington, and Wisconsin.

recipients, such as those effected by waiver provisions applied in selected counties to certain categories of families.

Recently, Abt (2001) extended the work by UC-Data by examining the potential for using administrative data from the WIC and child nutrition programs. They are surveying State agencies and school food authorities to inventory their system capabilities, data sharing arrangements, record linkages, and "best practices." In addition, they are testing the feasibility of linking data from multiple food assistance programs in select States.

There has been no coordinated effort to organize and transform the information in existing records into research-ready databases for future empirical research. One major issue is that each State has its own set of protocols to access the administrative databases and, in some States, obtaining this access may involve contacting multiple State agencies. While the Abt project provides a detailed inventory of capabilities in a large number of States, it does not establish brokerage agreements for researchers to use these data.

Because of these issues, it is difficult to track the development of various State data resources. Consequently, researchers may be missing valuable research opportunities and/or duplicating previous efforts because of a lack of a centralized system.

In a report on the research uses of administrative data, Hotz et al. (1998) made several recommendations to develop administrative databases for future research. Of particular relevance is the role of government agencies (and their contractors) in coordinating these efforts. Specifically, Hotz et al. noted that past collaborations between one or more State agencies and outside academic or independent research groups were the keys to developing successful and ongoing data for future research. A larger Federal role could contribute significantly to this momentum.

Methodology

The methodological approach builds on Hotz's suggestion to create a centralized system of data access. A major advantage of coordinating this effort through USDA is that most States already link FSP administrative data to records from AFDC/TANF and Medicaid using integrated database systems. According to Abt (2001), 38 States use these systems, which are generally referred to as the Family Assistance Management Information System (FAMIS).[14] Presumably USDA, with support from contractors, researchers, and other government agencies could gain access to its own administrative records and play a key role in the coordination effort.[15]

The resulting system would serve as a coordinating mechanism for monitoring the State data resources. Specifically, USDA would broker agreements to use the State data,

[14] In most States, the FSP, AFDC/TANF, and Medicaid programs share a mainframe computer system and, in many cases, a common "program identifier" to identify individuals participating across each of these programs.

[15] A coordinated effort between the Federal agencies responsible for the AFDC/TANF and Medicaid programs (i.e., HHS' Office of Family Assistance and HHS' Centers for Medicare and Medicaid Services–formerly the Health Care Financing Administration) and ERS is an option.

compile an ongoing summary of the available State data resources, and make recommendations to coordinate efforts across States that would enhance comparability of systems. A nondirective role for USDA along these lines could provide substantial value added without treading on State and existing USDA prerogatives.

The first step would be to identify which databases are most promising for future research. While several potential databases exist, not all are in a format that would benefit the research interests of USDA. USDA could use several criteria to select these databases, including:

- *Success in previous projects:* Most of the aforementioned State welfare evaluations included detailed assessments that provide insights on the quality of the data used in their analysis.

- *Frequency of update:* The development of complex relational databases has improved a State's abilities to update and store data about program participants on an ongoing basis, thus providing a rich potential source of longitudinal data.

- *Linkages to other programs*: Linkages to other social welfare programs determine the value of administrative files for examining a variety of cross-program issues. Of particular interest would be linkages to UI wage records that allow researchers to potentially examine earnings patterns of FSP participants, before, during, and after their receipt of FSP benefits.

- *Type of State welfare program*: Databases from States with "interesting policy changes" surrounding food assistance and related programs could provide the most valuable information for research, including indications of the possible effects of policy innovations in the FSP program in other States. For example, many of the studies done by HHS' Administration for Children and Families included experimental designs that USDA could build off to identify policy impacts in the FSP.

- *Availability of linked survey information*: Survey data can provide detailed background information that is not available in administrative records, thereby enhancing the information from the administrative extracts. Stapleton et al. (1999) documented the availability of several surveys linked to administrative databases.

Several existing databases meet one or more of those criteria, include those in California, Illinois, Massachusetts, Texas and Oregon (Hotz et al., 1998). In each of these States, researchers, contractors, and administrators have developed relationships to use administrative databases on an ongoing basis. Such databases would be a logical starting point for negotiating agreements, given their successful use in previous research projects. It is likely that the Abt (2001) study will also identify other very promising States.

USDA would need to take three important "operational steps" to establish relationships with State agencies to obtain access to these databases.[16] (In many cases, USDA may be able to rely on contractors who have used these data in State-sponsored projects to assist in these operational steps.) First, USDA should review the original interagency agreements from the initial file linkage and determine if new agreements are necessary to obtain access for USDA and its researchers. While these agreements are important, ideally most State databases will already have an established set of protocols across agencies for data access. It is likely that the findings from the Abt project will significantly enhance these efforts. Second, USDA would negotiate agreements with one or more agencies that would allow the agencies to retain adequate control over the data, and outline any new burdens that this initiative might impose on administrators. These protocols would State how the data will be used (e.g., by which agencies) and where it will be stored. Issues related to data storage are discussed in more detail, because they are critical both in establishing and using the data on and ongoing basis. Finally, protocols would be necessary to ensure the confidentiality of the data. These protocols would outline how to satisfy the confidentiality restrictions in accessing the data and disseminating the results.[17]

After obtaining access to the data, USDA could hire researchers to assist State agencies to standardize the record formats for research purposes. It is likely that these formats vary significantly across States, depending on the State mainframe system used to process the data. Manpower Demonstration Research Corporation (MDRC) has developed a standard relational database for this purpose, called jemis, developed with Oracle software, which USDA could use to read these formats (Freedman, 1997). Jemis is written in a database language that converts non-standard source data into standard analysis files and has several other beneficial features for working with linked data sources. For example, MDRC has used jemis in multi-State evaluations to update links across multiple files and individuals, match records, and perform data quality checks.

USDA, with assistance from researchers, would then document the data elements in these records. It will be critical for USDA to have a full understanding of the data's historical development, State practices for overwriting, purging, and archiving data, and program rules for the documentation. In addition, it will be important to understand the obstacles that other researchers faced in using previous State extracts, including potential issues

---

[16] This list is similar to the suggestions made by Hotz, et al. to establish protocols for using administrative data.

[17] According to Maxfield et al., (1999), privacy concerns fall into three categories. "The first is that FSP … [and other linked data] are owned by the State agencies providing benefits to the case member and may not be disclosed to an agency of another State or Federal agency. A solution to this problem is to execute privacy agreements between State agencies and the sponsoring Federal agency. The second category of privacy concerns is unauthorized physical or electronic access to the national client database. This threat to security could be controlled by building physical, electronic, and procedural safeguards into both the centralized component and the State agency component of the database. The third category of privacy concern is the threat of authorized but unintended uses of the data. This risk to client privacy is the most difficult to control. The wide variation among State privacy laws suggests that this risk is minimized by construing the national client database as a Federal system of records subject to the Privacy Act and the Computer Matching and Privacy Act (pg. XIV)."

related to data integrity (e.g., missing data, duplicative observations). Ideally, USDA would develop documentation for these procedures based on previous reports, but there may be some information gaps that require further investigation into State policies and practices.

USDA could use this documentation to assess whether the selected State databases were ready for cross-State analyses. In making the comparisons, USDA could draw on the experience of researchers familiar with methods to assess the comparability of data, as well as administrators who have a detailed understanding of the data elements and program operations. Based on the findings, USDA could make suggestions to the participating States to add or modify data elements that may improve comparability across States.

The final step would be to establish a method for researchers to access the data on an ongoing basis from a centralized location.[18] Of particular importance is the location that researchers use to access the linked databases. One option is to establish a secured site at USDA, or with one of its contractors, to manipulate State extracts. USDA could use the sites to store data on an "as needed" basis. Researchers would then go to this site to process the data. USDA may be able to allow contractors to take the data off the secured site by stripping important confidential identifiers, such as Social Security Numbers, from the extracts, depending on the terms established with State agencies supplying the data. The site location and the ability to process the data will depend on the arrangements for the protocols.

Advantages and Limitations

The extension of projects using State administrative data would significantly expand research opportunities for USDA. These data provide detailed and reliable program information on large samples of FSP participants, and, unlike survey data, are not susceptible to nonresponse or respondent error. In addition, the size of the databases allows for detailed subgroup analysis both across and within States. Finally, many of these databases included several periods of information that will allow researchers to track long-term outcomes of program participants longitudinally.

Researchers could use the files to address several questions related to FSP participation, including:

- What are the participation rates of FSP participants in other State welfare programs?
- Do these rates vary across or within States? Over time?
- What are the program participation patterns of short and long-term FSP participants who participate in other State programs?

[18] A more ambitious option is to structure a data warehouse to process, clean, and store the State extracts on an ongoing basis. This warehouse would give USDA more control over the databases, but it would likely involve costly infrastructure and significant programmer time.

Administrators could use this information to better understand the multiple program participation patterns in their own State. Administrators could use this information to identify whether TANF recipients in certain areas are more likely to receive FSP benefits than other areas. This information could be helpful in targeting outreach efforts both within and across States.

It is also possible that these administrative files could be linked to other surveys, such as the CPS and SIPP. To establish these types of linkages, however, USDA would need to obtain unique identifier, such as Social Security Numbers, which may make negotiating protocols to use the data more burdensome.

There are two important limitations of these files. First, researchers could only track program participants while they are participating in the State programs. In one sense, this is not a problem since nonparticipation is an important outcome in its own right, signaled by the lack of inclusion in program files. But researchers will not be able to distinguish between participants who left the program and those who moved from the State–a problem faced by all prior uses of administrative data and not considered crippling by most researchers. Second, these files include no information on long-term nonparticipants. Unfortunately, only surveys could provide this type of information because non-participants by definition never encounter the FSP program, and therefore there is no way to track these individuals.

Feasibility Issues

To obtain the data extracts, USDA will need to show that the match would be beneficial to State agencies. This should not be a major concern because administrators could use this information to better understand the multiple program participation patterns in their own State.

There are two major feasibility issues in creating data access and generating comparable estimates across States. For data access, USDA will need to develop protocols to allow "USDA-approved" researchers equal access to each State's database. It is possible, for example, that some States may agree to provide access to data at a centralized location and others restrict use only to local agency offices (and their contractors). For comparability, it will be important to standardize procedures across States to produce consistent extracts. During interviews with key stakeholders, State agency officials noted that the variability of State data systems might make this difficult. Because each State has its own system, using different data definitions and data elements, it may be difficult to compare data between States. For example, a $200 AFDC/TANF payment may represent an actual payment in some States and a scheduled payment before AFDC/TANF income offsets (e.g., for work or other income) in other States. While some of these issues can be identified by fully understanding the program rules and databases in each State, in some cases, USDA may need to make a special request to States to add comparable variables to their systems. It is not clear whether this would be feasible for programs outside the purview of the USDA.

Despite these issues, there seems to be a significant demand for starting a centralized system for obtaining administrative data. USDA could be in an excellent position to coordinate these efforts, particularly given the existing linkages to the FSP, as technological advances continue to improve the ability of administrators to produce data for research purposes.

Potential Cost of the Initiative

USDA will need to consider certain costs in developing this initiative and considering its overall desirability. Costs to the Federal government include the following:

- Data Review: The labor time to review the large number of existing data sources should not be that large, given the previous efforts by Hotz et al. (1998) and UC-Data (1999).
- Data Access: Data access costs may be minimized by building on previous agreements by other contractors to use the data, though, in some cases, negotiations may be necessary to extract additional data.
- Data Protocols: Legal counsel may be necessary to establish and monitor the use of the linked databases on an ongoing basis. These costs could be relatively high for because the Department USDA will be charged with protecting data confidentiality.
- Data Documentation and Comparability: Creating comparable files will likely represent the highest costs of this initiative. According to Freedman (1997), these costs can be quite substantial, particularly in developing the human capital to thoroughly understand all of the details of each administrative system. In some cases, however, the costs of this exercise may be minimized by previous efforts in the field to create comparable data. One of the goals of the 14 welfare leaver projects conducted under the auspices of the HHS' Office of the Assistant Secretary for Planning and Evaluation was to create comparable datasets across each site. USDA may be able to piggyback on some of these earlier efforts.
- Storage Costs: The State administrative research files are likely to be very large and costly to maintain in terms of labor and computer hardware. Proper storage facilities will need to be developed to store and manipulate the administrative data files and protect their confidentiality for the extracts. If USDA chooses to house all of the data rather than periodic extracts, the costs of the initiative increases significantly because each State administrative file is likely to be quite large.
- Monitoring: USDA will need to monitor data usage to protect the original confidentiality agreements and gauge the value of the initiative on an ongoing basis.

It is important to note that each of the preceding costs will rise with increases in the number of State FSP administrative data sources and years of data. While the startup costs of creating a linked data systems are high, updating the files on a regular basis with new information should not be as burdensome, as administrative data are a relatively low-cost source of information.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
X	P	P	X	X	X	X		X	Moderate	Moderate	Moderate

Notes on Individual Criteria:

(2) Information is provided to examine cross-State and longitudinal trends in multiple program participation by FSP participants. Administrators could use this information to better understand the multiple program participation patterns in their own State.

(5) Linked databases may only be available in select States and localities.

(6) This should be available from States for continuing research on an ongoing basis.

(7) The largest issue is in making the data comparable across States.

(8) Costs are limited by reliance on existing State-level efforts to develop linked databases.

8. Micro-Matching of SIPP and CPS Records to Food Stamp Administrative Records

Initiative Summary

This initiative will address the inadequacies of current data by linking the SIPP and the CPS data to State FSP administrative records. Most surveys only track respondents for a limited number of periods, which makes their use in research of long-term trends limited. While administrative data sources generally track large samples of participants over multiple periods, these data include limited background information on participants. The linkage of survey and administrative data sources would combine the strengths of survey data, which provide detailed background information on individual characteristics, and administrative data, which follow participants over multiple periods. Researchers could use the linked data source to address research questions related to FSP program dynamics of FSP participants and nonparticipants.

There are three primary reasons to use SIPP and CPS records for the linkage. First, unlike most surveys, both the SIPP and CPS collect information on the respondent's Social Security Number (SSN), name, and address. The existence of these unique identifiers, which are generally not included in most surveys, will significantly enhance the linkage to administrative records, which usually include information on SSNs, as well as name and address. Second, researchers have used both the SIPP and CPS extensively in past research. Consequently, improving these data should enhance options for future FSP research. Finally, unlike most surveys, several years of SIPP and CPS data from the 1980s and 1990s are available for the match.

The availability of FSP administrative records varies by State, so the linkage would likely be limited to a select number of States. It is important to note, however, that the creation of a national FSP administrative database could significantly expand this initiative in the future years (see Micro-level FSP Administrative Data Initiative).

In the past, the Census Bureau has created research files that link SIPP survey records from the 1984 SIPP to FSP administrative files from Florida, Pennsylvania, and Wisconsin (Marquis and Moore, 1990). The primary purpose of that effort was to examine response error issues in the SIPP.

This initiative would build on the Marquis and Moore project by linking multiple panels of the SIPP, as well as the CPS, to several years of FSP administrative data from various States, not just Florida, Pennsylvania and Wisconsin. The methodology for linking the records would be the same as that used in the 1984 study. Specifically, the linkage would link common identifiers that exist in the SIPP, CPS and FSP administrative records, such as Social Security Numbers and names, confirming that the right match took place by examining confirmatory variables (e.g., date of birth, demographic characteristics).

Despite the research advantages, USDA will need to account for several major cost issues in evaluating this initiative. Specifically, significant technical and legal issues associated

with developing, creating, storing, and monitoring matched data could make implementation very expensive.

Background

The SIPP is a nationally representative dataset that provides detailed longitudinal information on households, families, and individuals over approximately a 32-month period.[19] It oversamples low-income populations to provide added information on issues concerning poverty, income support programs, and the low-wage sector of the work force. SIPP panels are available for each year from 1984 through 1993 and then again in 1996. Each SIPP panel surveys up to 20,000 households (approximately 50,000 sample members), though the sample size of each panel varies. [20]

The CPS is a monthly survey of a nationally representative sample of the civilian noninstitutionalized United States population. The annual March Demographic Supplement, the primary data proposed for the match, contain detailed questions about household composition, employment, and sources of income. The CPS sample size in most recent years is in excess of 55,000 households and 150,000 individuals. Unlike the SIPP, each March CPS represents a separate cross-section of the United States. The March CPS is available on an annual basis.

Several studies have used survey data from the SIPP and CPS to examine a variety of FSP issues. The primary advantage of these databases is that they provide detailed individual survey information on FSP participants and non-participants. For example, researchers used the SIPP and CPS to examine the relationship between various factors (e.g., business cycles, individual characteristics) and FSP participation decisions (Martini and Allen, 1993; Yelowitz, 1995; Gleason, Schochet, and Moffitt, 1998; Castner, 2000). In addition, Gleason et al. (1998) used the panel nature of the SIPP to examine dynamic changes in FSP participation over a relatively short period.

The surveys have two major limitations. First, they do not allow researchers to examine long-term patterns (e.g., those lasting at least 3 years) of FSP participation. While some researchers used the SIPP to examine multi-period earnings (Gleason et al., 1998), their analyses only provide information on earnings outcomes over the life of the SIPP panel. Second, these surveys tend to undercount the number of FSP participants, as well as participants in other means-tested transfer programs, due to under-reporting of benefit receipt among survey respondents (U.S. Bureau of the Census, 1999).

To address these limitations, researchers have turned to administrative files to supplement their survey analyses. Castner (2000) uses a combination of Food Stamp Quality Control (QC) and CPS data to estimate FSP participation rates. Specifically, she uses CPS data to estimate the size of the population potentially eligible for the FSP, and the QC data to estimate the actual size of the population receiving FSP benefits.

---

[19] The 1992, 1993 and 1996 SIPP panels include a longer sample frame.

[20] The 1984, 1990, 1992, 1993, and 1996 SIPP panels include at least 20,000 household observations. The remaining panels have fewer than 20,000 household observations.

Some studies have also used administrative files in lieu of survey records, because these files include large samples of program participants over several periods. Recent studies, such as welfare leaver studies sponsored by HHS, rely on the large multi-period samples available in administrative records to track transitions into and out of welfare programs (including FSP).

Administrative files also have significant limitations. In general, these files contain no information on nonparticipants and relatively limited information on participants' characteristics (e.g., there is nothing in the administrative files on education). In general, information from administrative records will only be reliable if it is necessary to administer the programs.

The original match of the 1984 SIPP to FSP records included a limited extract of administrative State FSP records linked to the first two waves of SIPP interviews, to address issues of underreporting FSP participation in the SIPP survey. Bollinger and David (2000) used these records to compare SIPP interview responses to Food Stamp questions to FSP administrative reports. While they found that most survey responses corresponded with the FSP administrative cases, there were several disparities.

This initiative would build on the 1984 SIPP-FSP match in four ways. First, the match would be extended to CPS data, which includes significantly larger samples relative to the SIPP for State level analyses. Second, recent advances in the use and storage of administrative data should allow for the linkage of several years of administrative files. Third, because more States use administrative files for research now, additional States should be available for the linkage. Finally, the initiative would repeat the link for several rounds of the SIPP and CPS, which researchers could use to track trends across different cohorts of survey respondents.

The proposed initiative might also be able to build on existing files of CPS and SIPP data that have already been linked to Social Security Administrative (SSA) earnings and program participation data. Researchers have used various extracts of matched SIPP and CPS data on a periodic basis to examine issues related to SSA programs (Stapleton et al., 1999). It is possible that these files could be combined with the proposed matched FSP files, which would allow researchers to examine dynamic FSP program participation and earnings trends for several cohorts of survey respondents.

Unfortunately, obtaining the restricted files has become problematic for a recent project that proposed to use matched SIPP-SSA data because of data restrictions.[21] Therefore,

[21] Recently, the Lewin Group proposed to use the matched Census files in a research project to examine dynamic earnings patterns of different cohorts of FSP participants. The project proposed to follow SSA earnings histories for different cohorts of FSP participants from the 1990 and 1996 SIPP panels. To obtain access to these files, researchers must submit proposals for approval by the Census Bureau and Internal Revenue Service. Because of the confidentiality of earnings records, the access to the data is restricted. Unfortunately, obtaining access to these data has proven more difficult than originally anticipated. For a copy of the guidelines in obtaining these data, see http://www.ces.census.gov/download.php?document=50.

this linkage is not proposed here, despite its obvious desirability. Of course, the value of a Census to FSP administrative match becomes even more important given the lack of information on program dynamics from other sources.

Methodology

The Census Bureau, in cooperation with USDA and State agencies, would match the FSP records with the CPS and SIPP surveys using a probabilistic record linking methodology.[22] This methodology is widely recognized as one of the most reliable methods to match data records. Census could implement the methodology using a commercial software program (e.g., Automatch) that calculates probabilistic matches.[23]

Probabilistic record linkage assumes that no exact match between fields common to the source databases will link a person with complete confidence. Instead, probabilistic record linkage calculates the likelihood that two records belong to the same person, by matching together as many pieces of identifying information as possible. This approach seeks to limit the probability of false links and nonlinks in the data matching procedure. The precision of the match improves with the addition of common data elements that uniquely identify each individual. Hence, information on data elements, such as SSN and name, would significantly improve the precision of the match.

The matching fields would include all of the data elements that are common to the FSP and survey files. These fields would likely include SSN, name, address, and demographic information. The SIPP and CPS surveys include information on all of the matching fields. In general, most States also include all of the data elements for the matching fields in FSP records. UC-Data (1999) found 95 percent of the States in its survey linked FSP with TANF and/or Medicaid using some combination of the matching elements.

Past linkages suggest that the SIPP surveys gather reliable information on the core matching elements. As part of the ongoing SIPP program, the Bureau of the Census and SSA validate SSNs for SIPP sample members in the course of normal survey operations. Census also attempts to locate SSNs for persons for whom an SSN is not reported in the survey (except for persons refusing to provide their SSN). According to Hu et al. (1997), this process resulted in a "validated" SSN for approximately 90 percent of original sample members age 18 or older and for about 80 percent of persons under the age of 18 in the 1990 SIPP panel.

Since many State FSP administrative databases link to other systems using several of the matching identifiers suggested above, the matching elements in these databases are expected to be very reliable (UC-Data, 1999). A current study by Abt (2001) should provide some insights on the reliability of using identifiers from FSP administrative

---

[22] This is the same methodology used to generate the 1984 SIPP-FSP link.

[23] A simple alternative would be to link FSP and Census records using just a single unique identifier, such as a SSN. This simple approach, however, might generate a large number of "false links" (links between records for different people) or false nonlinks (links of two records for a single person that are missed) because of missing data, transcription problems or other errors with the SSN identifier on the FSP and/or Census datasets.

records to link to other systems. Abt will test the feasibility and accuracy of linking data from different nutrition assistance programs, including the FSP, to obtain unduplicated counts of clients in a limited number of States.

An added challenge in linking records is the variation in available matching variables in different State FSP administrative systems. Federal auditing provides an incentive for each State agency to keep accurate, machine-readable records in administering the FSP, but differences across State program rules and administration might have significant impact on the information record in each State's administrative files. It would be important to understand whether such differences will affect the precision of the probabilistic match, perhaps making matches in some States too unreliable to use, particularly if the administrative importance of certain matching variables, such as the SSN, varies across States.[24] In addition, it would be important to understand the comparability of any data elements *not* used in record matching but available for analysis when linked data are later used in research (e.g., benefit amount, household composition) across FSP administrative databases, particularly those related to FSP eligibility.[25]

To better coordinate the data elements used for probabilistic matching, Census and USDA may need to develop a standardized method for developing, storing, and updating FSP administrative databases when converting them to research use. This coordination could include designing record formats to facilitate research; developing a variable imputation procedure for any missing fields; standardizing fields from different systems; writing code to summarize and/or delete duplicate records; establishing eligibility links for individuals living in FSP households; and performing final verification checks.[26] The "cleaned" FSP administrative records would then be transferred to a standardized research file for the probabilistic matching to survey records.

Maxfield et al. (1999) found that States had the capacity to produce the types of FSP administrative files necessary for this linkage on an ongoing basis. They surveyed officials regarding the capability to produce similar types of records and found that administrators could prepare and send an extract of its case records with relative ease to a central facility. Some States, however, have records that are more accessible than others for various reasons (e.g., automated data systems, centralized data warehouses).

Presumably, USDA would limit the initial extracts to a few promising States. USDA could collect administrative data on FSP participation for as long as these records are available. According to UC-Data (1999), several States should have multiple years of FSP administrative records available.

---

[24] For example, while UC-Data found that almost all States include information on SSNs on their FSP administrative databases, only 6 of the 26 States used the SSN as the primary mechanism to organize their records. Presumably, the information on SSN will be more reliable in States that use this variable as their primarily identifier.

[25] For example, differences in State program administration may affect the method of recording certain data elements. Researchers will need to understand these differences when using linked files to make comparisons across States, as well as in aggregating over several States.

[26] This list is based primary on a summary from Freedman (1997).

USDA would transfer the administrative files to a secured Census site for the probabilistic match. Sworn Census agents (e.g., contractors who sign a confidentiality agreement), would then match these records.

Census and USDA would develop protocols that would allow researchers to access these data on a periodic basis. Such protocols could be similar to those already available on the Internet to use the matched SIPP-SSA and CPS-SSA records.[27]

Advantages and Limitations

The combination of survey and FSP administrative records would significantly expand research opportunities beyond that provided by survey or administration data files alone, particularly in the area of FSP dynamics. The survey information would provide detailed background information on demographic, income, health, and other program characteristics of FSP participants and non-participants. The administrative records would supplement this information by providing lifetime FSP histories for each SIPP and CPS respondent.

Researchers can use these data to observe detailed transitions of CPS and/or SIPP respondents before, during, and after their interviews. While transitions onto FSP can be observed using FSP administrative data alone, the linked survey data allow for the construction of detailed profiles of family characteristics at these transition points on FSP and non-FSP participants. Consequently, the linked data would provide detailed contextual information related to program dynamics unavailable elsewhere.

Researchers could use the linked files to address questions related to the dynamics of FSP participation, including:

- What are the characteristics of participants who cycle on and off of FSP programs? Are these characteristics related to participation in other programs?
- How do the characteristics of FSP participants compare to non-participants?
- Do FSP non-participants in a particular period participate in other periods?
- How do changes in participation dynamics vary by demographic groups and States?
- Have dynamic patterns in FSP participation changed since the passage of welfare reform?
- How do FSP participation patterns of SIPP and CPS respondents before, during, and after their interviews?

In addition, a matched file will also allow Census to update efforts to assess the reliability of reported program participation with data from recent surveys. Of particular interest could be changes in underreporting rates over time, particularly with the replacement of older survey instruments with Computer Assisted Personal Interviews (CAPI) and Computer Assisted Telephone Interviews (CATI) in the mid-1990s.

---

[27] For example, see http://www.ces.census.gov/download.php?document=50.

The linked data do have some important limitations.. First, detailed characteristics from SIPP and CPS interviews are only available over the life of the panel. Attempts to characterize 2000 FSP recipients using information from, say, the 1990 SIPP panel, will be problematic because some characteristics, such as health, income, and family status are likely to change as a person ages. Second, there might be important inconsistencies in State FSP administrative databases that affect the reliability and interpretation of certain linked elements. Third, it is possible some State administrative databases will not be available for the link. Consequently, the research using these data could be constrained in both national usefulness and its statistical precision (sample sizes from the SIPP and CPS will drop sharply due to the exclusion of whole States).

Feasibility Issues

USDA will need to show that the match would be beneficial to both the Census Bureau's data collection efforts and State agencies that hold the FSP administrative data. The match would benefit Census data collection efforts in two ways. First, Census could update the findings from Bollinger and David (2000) by matching recent surveys of the SIPP and CPS to FSP administrative records. This update would be very important because both the CPS and SIPP surveys have changed their survey instruments and data collections methods since the initial FSP match. Second, an FSP match would enable Census researchers to better understand dynamic participation patterns in their own data. They may be able to use this information to redesign the questions in their own surveys. Researchers have used previous matches between SIPP and SSA data to recommend options for changes in the SIPP questionnaire (Stapleton et al., 1999). For State agencies, USDA would need to show how the research would directly benefit their programs. Researchers could use the matched files to illustrate the dynamic program participation patterns of different subpopulations (e.g., AFDC/TANF recipients, families with children, nonparticipants). For large States, such as New York and California where the survey sample sizes in the CPS are sufficient for State analyses, researchers could also show how these patterns vary among States. Administrators could then use this information to better target their programs to specific subgroups.

To obtain State data, USDA will need to satisfy State protocols to access and distribute the merged data. Each State FSP program has its own bureaucratic structure, funding stream, and data systems, making access to each State's administrative records burdensome (UC-Data, 1999). One major issue will be ensuring the use of the data does not violate any of the State's confidentiality clauses.[28] There are also significant technical

[28] As mentioned in Initiative 1, according to Maxfield et al., (1999), privacy concerns fall into three categories. “The first is that FSP … [and other linked data] are owned by the State agencies providing benefits to the case member and may not be disclosed to an agency of another State or Federal agency. A solution to this problem is to execute privacy agreements between State agencies and the sponsoring Federal agency. The second category of privacy concerns is unauthorized physical or electronic access to the national client database. This threat to security could be controlled by building physical, electronic, and procedural safeguards into both the centralized component and the State agency component of the database. The third category of privacy concern is the threat of authorized but unintended uses of the data. This risk to client privacy is the most difficult to control. The wide variation among State privacy laws suggests that this risk is minimized by construing the national client database as a Federal system of records subject to the Privacy Act and the Computer Matching and Privacy Act. (pg. XIV).”

issues related to producing a "clean" readable State administrative file that is void of missing, inconsistent, or purged data. It is likely that a data warehouse would need to be created to store and clean the State administrative data files for the match. While creating such warehouses is feasible, it involves constructing a costly infrastructure to house the data and significant programmer time to clean the data.

Obtaining the Census data for the match should be less burdensome. To use the data for the match, USDA will need to negotiate agreements that allow its researchers (and contractors) to become sworn agents of the Census Bureau. These agreements will document how the data will be matched and distributed. In past SSA projects, we have used matched SIPP-SSA data by becoming sworn agents of the Census Bureau and processing the data at a secured site (at the SSA).

The existence of previous matches of SIPP-FSP data illustrate that generating an updated linked file is feasible. In fact, the technology, methods, and data systems to facilitate such a match have significantly improved since the original linkage of these records described in Marquis and Moore (1990). These improvements should allow programmers to link more data in an efficient process.

Potential Cost of the Initiative

USDA will need to consider two main costs in developing this initiative: the development of a plan for this initiative and the consideration of its overall desirability. Costs to the Federal Government include the following:

- Data Access Costs: According to UC-Data (1999), gaining access to data sources could require "prolonged legal negotiations with State agencies, involving drafting formal contracts, interagency agreements, or specific legislation that authorizes or requires interagency access to confidential data." However, based on the findings from Maxfield et al. (1999), USDA can minimize some of these costs by requiring all States to provide "minimum dataset" with a common set of elements.
- Storage Costs: The State administrative FSP research files necessary to fully execute the match are likely to be quite large and to require extensive data processing capability. Proper storage facilities will need to be developed to store and manipulate the administrative data files and protect their confidentiality. Many of these State files could contain in excess of a million observations that will need to be stored to generate a State file for the linkage.
- Data Manipulation: The programming time and costs to manipulate the data will be substantial, as programmers will need to become familiar with the data elements from various State systems. There also likely will be costs associated with purchasing commercial software, such as Automatch, for the data match.
- Data Update: The administrative data linked to each of the survey files will likely need to be updated on an ongoing basis. Hence, protocols will have to be established to obtain and manipulate data from various States on an ongoing basis.
- Monitoring: Census and USDA will need to monitor data usage to protect the original confidentiality agreements.

It is important to note that each of the costs will increase with increases in the number of State FSP administrative data sources and years of data included in the link. While the start up costs of creating a linked data systems are high, updating the files on a regular basis with new information should not be as burdensome, because administrative data are a relatively low-cost source of information (and the required survey data could be made available through outside funding).

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
X			X	X	X	X	P	X	Moderate	Substantial	Moderate

Notes on Individual Criteria:

(2) Researchers and administrators could use this information to better understand the heterogeneous of short- and long-term participants, as well as in identifying potential outreach efforts to assist non-participants.

(3) Information is provided on program outcomes, including detailed information on FSP participant characteristics (e.g., race, education, gender, family background, and income)

(8) Costs could be quite high unless there is a coordinated effort to make FSP administrative data consistent for all of the States.

(9) The need for personal identifiers for survey respondents (for the data linkage) makes client privacy a moderate concern. However, similar issues have been resolved in the past by other government agencies and researchers.

9. Matching WIC Administrative Records with Medicaid and Vital Records Data

Initiative Summary

The purpose of this initiative is to create a database by asking State WIC agencies to link WIC, Medicaid, and vital records data at the micro level, using Federal protocols. This nationally representative database will be rich in data on the services provided to and the health outcomes **expected** of WIC participants. State-level merged datasets of this sort would be combined into a single Federal-level database enabling both State program administrators and program researchers to do a much better job of monitoring outcomes for WIC clients and identifying gaps in services, and would help to direct resources to improving health outcomes.

The initiative proposes to create a nationally representative database composed of WIC client demographic, nutritional risk, and birth outcome data obtained from the States' WIC program files; vital records data obtained through State birth registries; and data on Medicaid participation accessed through the State's Medicaid agency. By linking WIC and vital records data, a database can be established that contains pregnancy and birth outcome data not available on the WIC record alone (such as Apgar scores, complications of delivery, etc). By linking the WIC record with vital records, the database then can be used to link the mother's record with the child's. The WIC program enrolls children separately from their mothers, so this link cannot often be established from the WIC record alone. By adding Medicaid participation to the database, the potential exists for long-term tracking of access to health care and ongoing health status (as approximated by use of Medicaid-covered services). In addition, long- term links can be established to track families who have participated in the WIC program through the births of several children.

Background

Since its inception, the WIC program has been viewed as a nutritional program designed to be an adjunct to the health care delivery system. The WIC program is the only food assistance program administered by State health departments. Over the years, local WIC programs have been encouraged (and in some cases, required) to refer low-income women and children to health care services. Among the types of services to which WIC clients have been historically referred are prenatal care, well-baby checks, the Medicaid program, and immunization programs.

Data from the WIC program have been linked to data from the Medicaid program in the past. In 1986 the State of Missouri examined the effect of WIC participation on the Medicaid costs related to the delivery and care of newborns by matching Medicaid claims records with WIC administrative records (Schramm, 1986). In addition, a study by Buescher et al. in 1999 examined how WIC and Medicaid data could be linked to examine birth outcomes of Medicaid clients on WIC as compared to those not enrolled in WIC (Buescher et al., 1999).

A variety of other researchers have used WIC, Medicaid, and vital records data to conduct both State-level research and research in local WIC programs to examine outcomes. Researchers have found that by linking WIC records with vital records data, mothers can be linked with their children to examine birth outcomes based on such variables as number of prenatal care visits, demographic data, income data, and pregnancy weight gain. The 1987 national WIC evaluation linked WIC records with birth outcomes in order to determine the impact of WIC participation on birth outcomes (U.S. Department of Agriculture, Food and Nutrition Service, 1987).

Each of the three databases is rich in data describing the demographic and health characteristics of women and children, their health outcomes, services provided, and participation-related information. The WIC records for pregnant women contain information on the participants' demographic characteristics, their weight gain during pregnancy, their length of WIC participation, and their nutritional risk. The birth records contain information about each woman's demographic profile, including information about where she lives, age, race of mother and father, date of last live birth, number of prenatal visits, and information about the pregnancy outcome. In addition, the birth record contains information about the baby, including birthweight, Apgar score, complications of delivery, and other health status information. The Medicaid files include similar demographic information, health care utilization information, and often include preventative health indicators, such as a child's immunization records.

There is currently no database available on the national level that can be used by researchers to examine the health outcomes of WIC participants, their children, and the links between health outcomes and participation in both WIC and Medicaid**.** All three files are rich in client-level data, which means a database combining them would allow for fuller descriptions of WIC participants and provide an ongoing capability to evaluate WIC in relation to various prenatal, postpartum, and child health program objectives. By developing and updating a database that is readily available to researchers, USDA can create a data file that can be used to examine client demographic contributions (age of mother, weight gain, use of prenatal care, breastfeeding status, race and ethnicity, etc.) to a number of health outcomes, such as mother's weight gain, birthweight of children, pregnancy complications, Apgar scores, and nutritional risks of the pregnant woman and child. Participation in the Medicaid program will help to examine such issues as access to preventative health services, immunization status, and ongoing utilization of health care services. Since nutrition and health services are complementary services that work together to improve health outcomes for infants and children, the linked data will be very useful in better understanding the relationships between maternal and child health and nutrition services, food assistance, and health outcomes.

Methodology

The methodology for linking the records from these three programs is relatively simple, and has been proven effective in a number of prior studies. Creating a national database will require the cooperation of WIC agencies in many States, as well as program

administrators in the State offices of vital records and the Medicaid program. The following steps are proposed for creating the database.

1. ***Develop a nationally representative sample of WIC participants for which the linkages will occur.*** Deciding how to construct the sample frame for the database will require considerable thought on the part of USDA officials. While developing sample frames for national studies is not new, the fact that the ultimate goal is to link the data with other participant databases may provide opportunities not commonly considered in developing samples. USDA will need to consider oversampling certain populations, such as adolescents, different ethnic groups, or persons living in rural areas. Larger samples of WIC participants may need to be drawn to account for persons who participate in WIC during pregnancy, but move to a different State and are lost to data collection prior or following delivery.

   Another key issue that will need to be decided is at what point during WIC participation the record should be entered into the database. While the database should be representative of women, infants, and children enrolled in WIC, there may need to be some consideration given to the age of the child and length of participation prior to inclusion. If a 4-year-old is added to the file, and then data are linked, there will be more data available for that child than a 2-year-old. However, if the 4-year-old is only recently enrolled in WIC, then there will be less data available for that child than for a 2-year-old who has participated in the program since birth. This may mean that the sample needs to be controlled for the child's age and length of participation.

   Once these decisions are made, the sample can be drawn. The sample frame can be constructed through the States' administrative database, which contains reports of monthly participation. Each State will be able to provide a point-in-time file containing information about each WIC participant, which then can be randomized and from which a sample can be drawn.

2. ***Create the database structure so information can be linked.*** This next step will require that the specific data elements to be contained in the database be identified, and a database be constructed in a form that allows for analysis. The database structure would be hierarchical, creating a header record that links all data back to a family unit tied to the mother. Then separate detail records are created containing information about the mother and child, including the health status of the mother and pregnancy information, birth outcome data, and information about the child. The database should be large enough to capture the essential demographic components of the WIC participant, as well as health outcome information contained in the records. The WIC record for a pregnant woman will contain both demographic information as well as health outcome information such as weight gain during pregnancy and nutritional risk factors. The birth certificate information will also contain numerous demographic data for the baby's parents, as well as information about the pregnancy. The Medicaid record

will also contain information about health care utilization such as prenatal care visits.

Consequently, the database structure will need to be carefully constructed in order to recognize duplicate information that will be available from each of the three files. In the case of a pregnant woman, it is likely that all three files will contain information regarding her age, racial or ethnic identification, weight gain during pregnancy, and other demographic and service-related factors. The database will need to be structured in such a way that duplications can be recognized and inconsistencies resolved.

Additional consideration will need to be made when linking the child's file to the mother's. The database will need to be structured in such a way that individual information about the parent and child can be extracted for research purposes. At the same time, the data file must be able to create associations between the mother's pregnancy, the birth outcome, medical participation, and ongoing WIC participation.

3. ***Create a linked WIC and vital records database.*** The next step, once the sample is drawn, is to link the WIC record with the vital records information. The actual linking of the data will need to take place at the Statelevel. In order to ensure consistency in approaches, USDA should provide States with merge protocols that will create a consistent database across States.

   The merger can be done using a number of methods. Because the data files are so rich with demographic data about the clients, identifying common demographic characteristics can create matches. As has been noted, matching WIC records with birth records has been successfully accomplished in the past. In some cases, records can be linked if the mother's Social Security Number is available. In most cases, WIC programs do not collect Social Security Numbers, so matching will need to take place using demographic and other descriptive data such as name, birthday, date of delivery, and other client demographic data.

4. ***Determine Medicaid enrollment status.*** Once a combined WIC/birth record match has been established, the next step is to determine whether the individual WIC participant also is enrolled in Medicaid**.** Matching the newly linked database to the State's Medicaid enrollment file can do this. The newly created database will then have information as to whether the participant was enrolled in Medicaid or not. This will allow for outcome comparisons of WIC clients who are enrolled in Medicaid as compared to those who are not. Some States will have information on services provided through the Medicaid program, such as the immunization status of children or the prenatal visits of the mother that can be captured and analyzed.

5. ***Remove identifying information and develop access protocols.*** The next step in creating the database will be to remove identifying information about the client and develop access protocols and timelines for updating. It is recommended that the database be updated every 2 years with new data. The updating procedure should be linked with the prior database in order to update client records, and allow for tracking of certain clients over time. By using the birth record number as an identifier, information about the delivery of pregnant women included in the prior database can be updated. The resulting file will then be available for research purposes. It is recommended that the database be maintained by and made available to researchers by USDA. This would require the placement of the data into a database accessible through the Internet, similar to other databases currently maintained by USDA.

Advantages and Limitations. The advantages arise from developing this type of database include:

1. ***The ability to track the health outcomes of WIC participants***. By linking the WIC participant's record to the birth record and Medicaid file, a data-rich source of information about the mother and child is created. One then can use this database to examine contributing factors to health outcomes, including the value of WIC participation as compared with birth outcomes of individuals not participating in WIC, the impact of Medicaid participation on WIC birth outcomes as compared to those not participating, client demographic factors that may be related to birth outcomes, and relationships between various nutritional risk factors and outcomes.

   Prior studies have shown the value of linking health outcome data with WIC records, but these studies do not track outcomes over time. By creating a national database of linked WIC and health outcome records, researchers will be able to track long-term outcomes over time. Examining WIC participation on multiple pregnancy outcomes or studying the effect of WIC participation during the postpartum period on the timing and spacing of subsequent pregnancies could be tracked through this database.

2. ***The ability to conduct trend analysis***. Such a database allows for conducting trend analysis of WIC participants over time. By updating the file every 2 years, any trends in program participation and health outcomes can be examined by comparing data from prior years, and by updating prior files.

3. ***The ability to link this file to other databases***. In addition to having a single, data-rich file, the newly created file can potentially be linked to data from other national surveys, other data files maintained by State and local governments (food stamp administrative files, TANF files, etc.), or data files maintained by Federal agencies such as CDC (nutrition surveillance files). Data from these files could be used to track outcomes beyond those reported in the WIC/Medicaid/vital records files. Through linkages with CDC's pediatric and maternal nutrition surveillance systems, additional information about the nutritional status of clients can be

obtained. Only 44 States participate in the nutrition surveillance system, so data would be limited in scope.

4. ***The ability to identify gaps in program practices.*** By linking the data from the three files, issues related to the relative health outcomes of persons of color, persons living in rural areas, or adolescents can be assessed and new program interventions can be devised to better serve these clients.

This initiative could address the following potential research questions:

- To what extent does the mother's participation in WIC affect birth outcomes?
- How does participation in both WIC and Medicaid influence birth outcomes as compared to participation in only one of the two programs?
- How do WIC and Medicaid participation during the first pregnancy affect participation during subsequent pregnancies?
- What are the long-term health impacts of WIC participation?

Limitations and potential difficulties in developing this database also need to be considered before deciding to proceed with this initiative, including:

1. If limited to a subsample for cost reasons, the database would not necessarily be representative of the individual States it contains. Because the database is using a national sample, data will not be representative of State-level participation. However, the same technology could be used by the State to match all of its WIC records with birth and Medicaid records, so analysis could be conducted at the State level.

2. Vital records data are not always released at the same time in each State, so a decision will need to be made as to which year of birth records will be included in the database in each update cycle. While most States are working to improve the timely release of birth records, the initial database may need to use birth records from different years in different States. In some cases, the birth record information may not show up for a given WIC or Medicaid client until the next 2-year update.

3. USDA may want to consider working with HHS' Maternal and Child Health Bureau (MCHB) on this initiative. The MCHB has a great deal of interest in collecting data that can be used for analyzing birth outcomes and program participation. While the idea of cooperation between the two departments is reasonable, the resulting need to coordinate may slow the process down.

4. Many State Medicaid programs use managed care plans for delivering health services, so there will be limited encounter data available on individual client medical visits.

Feasibility Issues

As was noted earlier, the technology to implement this initiative is relatively simple. A few feasibility issues must be examined.

1. The initial matching effort, and subsequent updates, will require the cooperation of State programs from various agencies. There will be cooperation between the WIC program and the vital records office, but there may be a reluctance on the part of the Medicaid program to share its records. As with any initiative of this nature, the advantages of participating in the development of the database will need to be presented. Assistance with any problems may be available from the MCHB.

2. Maintaining the updated files will be important. This will mean that the database will periodically need to be changes to accommodate any changes in the standard birth record reporting system, changes in WIC data collected that may be required by new Federal rules, or changes made in the Medicaid program.

Potential Cost of the Initiative

- The cost of State agencies providing the data every 2 years, which should be minimal, since data that already exists in State administrative files will be used.

- One of the key cost considerations will be the development and populating of the database. Programming time, along with the time needed to create file structures and interfaces, could entail substantial personnel costs.

- It is envisioned that USDA will maintain the database. This will involve the cost of quality control of the data, developing access and confidentiality restrictions, and data storage. In addition, there will be costs involved in preparing and editing the data and its documentation for final distribution, creating update protocols for entering new data, and developing access protocols. The documentation workload may be particularly high, with data coming in from multiple systems in all 50 States plus the District of Columbia.

- There will be costs involved in maintaining data security and monitoring confidentiality agreements.

- USDA will likely wish to use the database for its own analyses. In this case, there are costs in data manipulation and software costs.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
	X		X	X	X	X	P	X	Moderate/ Substantial	Substantial	Moderate

Notes on Individual Criteria:

(2) Program administrators can use this information to examine the effect of WIC participation on client outcomes, examine whether WIC participation creates Medicaid cost savings, and track how well WIC clients are referred to Medicaid.

(7) States must agree to participate in the initiative and be willing to link the databases, and data would need to be available on an ongoing basis.

10. Linking Data on Students' School Performance With Administrative Records on NSLP Program Participation

Initiative Summary

This initiative will address the inadequacies of current data on the impact of the school-based nutrition assistance programs by linking data on student participation in the National School Lunch Program (NSLP)[29] with aggregate student-level data on academic achievement and other school-level performance indicators. This linkage would enable researchers to address several critically important research questions related to SLP participation, academic achievement, and other indicators of students' school performance (e.g., attendance, grade retention, drop outs, disciplinary actions).

Despite the research advantages, the USDA will need to account for several feasibility issues in evaluating this initiative. Specifically, information on academic outcomes cannot be combined for all States (potentially requiring analysis by individual States or groups of States), and there is a mismatch between the grade levels for which student academic outcomes are available and for which school-level NSLP participation data can be obtained.

Background

The NSLP. The NSLP is second only to the Food Stamp Program in the USDA's "safety net" for low-income Americans. Targeted specifically at school-age children and their families, the NSLP is the oldest of all of the nutrition assistance programs, dating back to 1946. The NSLP provides Federal subsidies for school lunches served to school children without regard to family income, but higher subsidies are provided for free lunches served to the most poor students (children at or below 130 percent of poverty level), lower subsidies are provided for reduced-price lunches served to less poor children (130–185 percent poverty level), and the lowest subsidies are provided for paid meals served to all other children. As of 1999–2000, reimbursement rates varied from a high of $2.10 per meal for free meals in schools where most meals served were either free or reduced-price, to about $0.20 per meal in schools where fewer than 60 percent of meals qualify for either the free or reduced-price subsidy. Schools also receive USDA-donated commodities (e.g., dairy products, pork, beef) for use in preparing school meals.

Essentially all public schools participate in the NSLP and about 83 percent of all public and private schools combined participate in the program, which serves more than 96,000 schools and institutions. The program is available to 92 percent of all school-age children (Burghardt et al., 1993; Burghardt and Devaney, 1995), accounting for more than 27 million children on any given school day. The total number of lunches served per year is about 4.5 billion; about half of the meals served are free and about eight percent are reduced-price. About 57 percent of all children in schools that offer the NSLP regularly

[29] This initiative is focused on the NSLP because of this program's large national public school coverage. Extending it to include the School Breakfast Program (that serves about 66,000 schools vs. nearly 100,000 for the NSLP) would only involve the minimal additional cost of obtaining the added SBP school-level meal participation data.

participate in the program (FNS, 1998), but participation does vary with household income, age, gender, and geographic region of the country. The poorest students are more likely to participate, and elementary school students are more likely to participate than middle or high school students are.

What Do We Know About The Effect of The NSLP? Because the NSLP was established to "…. safeguard the health and well being of the Nation's children and to encourage the domestic consumption of nutritious agricultural commodities and other foods …" it is not surprising that "the most widely studied outcome of participation in the NSLP is nutrient intake" with a few studies focusing on children's nutrition and health status and household food expenditures (Fox and Hamilton, 2000). What is surprising, however, is that the same authors reported that "virtually nothing is known about the effects of the NSLP on school performance or behavior." The only study identified was based on data from the late 1970's (Gretzen and Vermeersch, 1980) which found no positive effects of participation in the free school lunch program on school attendance or cognitive performance. However, as Fox and Hamilton (2000) note, this study has "serious limitations" that greatly reduce its value.

There is, therefore, inadequate information available about the extent to which participation in the NSLP is related to higher academic performance[30]. Given the fact that this program is based in schools, targets school-age children, and that there has, in recent years, been an increasing national, State, and local focus on improving academic achievement (especially for poor children) this is an important data niche that should be targeted by USDA.

Recent School Reform Movement Makes Student Data Widely Available. Although individual student data would be the ideal for this data initiative, issues of confidentiality and logistical feasibility make this a less attractive option (at least now). Fortunately, recent State and district efforts to increase school accountability make aggregate-level data available that can readily be linked with school-level administrative data on NSLP participation.

The most important school reform movement in the United States today is "systemic" or "standards-based" reform. It now dominates education policy in nearly every State. It focuses attention at the school level rather than the individual students and is the basis for essentially all Federal policy-making targeted at K-12 schools. The impetus for reforms came not from national leadership but from the States, especially the highly influential report by the National Commission on Excellence in Education, *The Nation at Risk* (1983), which galvanized the education community around the goal of combating the "rising tide of mediocrity" that was purported to be destroying U.S. schools and placing American children at risk of falling behind in the global marketplace. By the end of the

[30] The SBP has received more attention, but generally using very small samples of schools. According to Fox and Hamilton (2000) five studies have examined the effect of the SBP on students' academic performance. The authors conclude that "the evidence is mixed" but is "suggestive … of positive impacts of the SBP." But, these studies must be used with caution because of their small size—2 were based on 1 school, 1 on 2 schools, 1 on 3 schools, and 1 on 16 schools.

1980s, visionary governors were rapidly moving ahead with education reform culminating in the 1989 "education summit" that created the first national education goals that were subsequently incorporated into most of the Federal programs for elementary and secondary education.

These changes represented dramatic departures from previous Federal education policy with a shift to the use of Federal funds to encourage States to make broader changes in school systems. Specifically, States were encouraged to develop content and performance standards in core subject areas, and to align their entire educational systems–including assessment, curriculum, instruction, professional development, and parental and community involvement–around these standards.

As of the 2000–2001 school year nearly all States have developed curriculum standards, and most have in place student assessments that are aligned with these new academic standards (Quality Counts, 2001). More importantly, 45 out of 50 States[31] now make available school-level "report cards" that provide aggregate grade-specific data on student test scores (usually in reading and math and sometimes in other subjects), and often for other key school-level indicators of student performance (attendance, dropouts/school completion, disciplinary actions) (Quality Counts, 2001).

Methodology

This change in the availability of data on educational outcomes allows, for the first time, the large-scale linkage of aggregate school-level data on NSLP participation with aggregate school-level indicators of student academic performance. To implement this data initiative, then, requires two sources of data–school-level student outcome data and school-level NSLP data–and the ability to link the two information sources for individual schools.

With regard to student outcomes, aggregate school-level data are, as noted above, available for most public schools[32] in nearly every State in the Nation. In most cases, the information is easily obtained directly from State Departments of Education (or in a few instances from school districts), and can typically be accessed online via existing websites. Annual data for all 1,300 schools in Maryland on the Maryland School Performance Assessment Program (MSPAP) are available electronically for 1993–2000, and data on the Standards of Learning for all schools in Virginia can be currently downloaded for 1998–2000. This electronic availability of school-level outcome data makes retrieval a relatively easy, low-cost process.

With regard to data on NSLP participation, it is recommended that USDA work with the FNS Regional Offices to acquire aggregate school-level data on actual meals served (by subsidy category) directly from State program directors or, where necessary, from school

[31] The exceptions are Idaho, Mississippi, Montana, North Dakota, and Utah (planned for 2003).
[32] These data are not available for private schools, which are typically excluded from State accountability systems.

food authorities (i.e., usually school districts) for the construction of this data initiative[33]. [If data on SBP participation were also desirable, it could be obtained from the same data sources.] Aggregate meal counts could be for a single target month, a combination of months, or for the entire year—the only requirement, of course, is that the same basis be used for all schools (or at least so that calculations can be done to put the data on the same basis).

The final component, data linking, can be easily done using the existing National Center for Education Statistics (NCES) school codes that uniquely identify every public school (these codes are widely used to apply for various types of Federal funding, as well). Where NCES codes are not available, alphanumeric matching of school information with existing NCES universe files has been found to yield a very high linkage rate (see Puma et al., 2000).

Once this linked school-level data base has been constructed, the types of analyses that could be done are most likely to involve the use of ordinary least squares (OLS) regression to estimate the effect of NSLP participation rates (actually, school means) on a variety of school-level indicators, especially academic achievement scores (again, school means), controlling for a variety of school (size, grade level, poverty concentration) and community characteristics (urban, suburban, or rural as well as regional location). Once multiple years of data are available (and this information would be available now for a large number of States), this same analysis approach could be elaborated by taking into account changes over time. The data also can be easily linked to the NCES annual Common Core of Data for schools, thereby allowing the analysis of a wide range of school characteristics.

Advantages and Limitations

The combination of State accountability data on student academic achievement (and other indicators of school performance) and NSLP administrative records would significantly expand research opportunities related to the school-based nutrition assistance programs. Researchers could use the linked files to address several questions related to the NSLP, including:

- Is higher NSLP participation positively related to average student academic achievement?

[33] An option worth exploring as part of this initiative is the use of counts of eligible students instead of actual meals served, as a less expensive "proxy" measure. School "report cards" often include information on the percentage of enrolled students who are eligible for free and reduced-price meals, and the annual school-level Common Core of Data available from the National Center for Education Statistics reports the same information for every public school in the Nation. These data would provide a simple, low-cost way to obtain an indicator of the NSLP program's "penetration," But the extent to which these data agree with actual FNS administrative records on program participation (i.e., meals served) is not known. A subsequent analysis could test this relationship and possibly suggest a multivariate "adjustment" that can be used for later analysis, and subsequent updating of the data base (e.g., collecting actual meal counts in alternative years and using proxy measures based on eligibles for the intervening years).

- Is higher NSLP participation positively related to other indicators of students' school performance (e.g., attendance, grade retention, dropouts, disciplinary actions)?

- How do these relationships vary by different school characteristics (e.g., enrollment size, grade level, concentration of student poverty, urbanity, region, etc.)?

The linked data has some important limitations. One issue that arises with student test data is the variation in the specific tests that are used in the different States and the inability to place these different measures of student's academic achievement on a comparable scale for the purpose of pooling the data into a single national dataset. If States use the same standardized test, the data can be combined into multi-State units. Even if the standardized tests are not used, the same statistical analyses can be done by State (or groups of States) and then combined using estimated effect sizes and established procedures for meta-analysis (Hedges and Olkin, 1985).

A second issue that arises with regard to the school-level outcome data is the fact that States do not typically test all students but selectively test specific grades representing the three major levels of elementary and secondary schooling, most often grades 4, 8 and 11. Therefore, the school-level indicator for academic achievement will have to represent a single grade. Other indicators are, however, generally measured for the entire school. It is not possible to obtain grade-specific participation information about the NSLP data. Those data will have to reflect the program data for the entire school, causing a mismatch that will have to be kept in mind when reporting the results of these analyses.

Feasibility Issues

In addition to the data limitations, the only other identifiable feasibility issue is related to the need to gain the cooperation the FNS Regional offices and State program directors to obtain school-level NSLP (and possibly SBP) participation data. This is not expected to be a serious problem. It is one that is likely to be easily overcome. It is important to note that student confidentiality is not an issue because the proposed data are aggregated at the school level.

Potential Cost of the Initiative

USDA will need to consider the following costs in developing this initiative:

- Data Access Costs: As noted, most of the school-level student outcome data can be readily downloaded from existing State (or district) websites. Direct collection of data will be required in probably only a very few instances (alternatively, the data initiative could initially be restricted to the majority of States where the data can be obtained electronically). Collecting NSLP data may involve somewhat higher, but still modest, costs.

- Storage Costs: The research file, since it would be at the school level rather than at the individual student level, would not impose significant data storage costs.

- Data Manipulation: The programming time and costs to manipulate the data will be modest, again due to the relatively small size of the database.

- Data Update: The school outcome and NSLP administrative data, and their linkage, likely will need to be updated on an annual basis. Protocols will have to be established to obtain and manipulate data on an ongoing basis.

Criteria Summary

Criteria											
Program area			Value		Information				Implementation challenges		
1			2		3	4	5	6	7	8	9
FSP	WIC	School Lunch	Program Administration	Research	Program Outcomes	Participation dynamics	National scope	Continuing research	Feasibility	Cost	Client privacy
		X	X	X	X		P	X	Low	Low	None

Notes on Individual Criteria:

(3) This provides information a variety of academic outcomes including school variables (size, grade level, poverty concentration) and community characteristics (urban and regional location)

(4) Because the data are aggregated, they will not include information on dynamic program participation patterns.

(5) This would cover virtually every State but certain analyses will have to be done on groups of States because information on test outcomes cannot be combined for all States and a mismatch exists among the grade levels for which student academic outcomes exist.

(8) Most of the data are readily available and the resulting database would be small.

(9) Client privacy is not a concern because the data are aggregated.