

Profiles in Success of Statistical Uses of Administrative Data

**Mark Prell, Howard Bradsher-Fredrick, Carol Comisarow, Stephen Cornman,
Christine Cox, Mark Denbaly, Rochelle Wilkie Martinez, William Sabol, Michelle Vile**

Abstract. Administrative records have a variety of statistical uses. An administrative data source can be linked either to survey records or to other administrative records to support formal evaluation of a public program or simply to provide an analysis that informs policy development. Administrative records provide sample frames for surveys. Comparisons of survey and administrative data provide insights that can improve survey design. Often, administrative records can serve these statistical uses only if at least one entity (i.e., a federal, state or academic organization) has access to all of the relevant files in order to link them together to create the research file or summary statistics. However, sharing administrative data poses risks of disclosure of personally identifiable information and raises issues of privacy, confidentiality and informed consent. Data-sharing projects also entail legal, financial and technical challenges. Other challenges include identifying the feasible and appropriate scope of a project, as well as managing both internal processes and interagency relations. In 2008, a cross-agency team of researchers under the sponsorship of the Federal Committee on Statistical Methodology conducted the project “Profiles in Success of Statistical Uses of Administrative Data.” The project examined seven successful data-sharing arrangements between various federal, state and academic entities and conducted conference calls with the projects’ original participants. They discussed data-sharing challenges and how challenges were met. This report conveys lessons that can help facilitate future data-sharing agreements involving the statistical use of administrative data.

Key words: administrative records; data linking; data matching; program evaluation; data sharing

April 2009

Executive Summary

Administrative data have a variety of statistical uses which often are met only if federal, state or academic partners succeed in developing a data-sharing agreement. In 2008, a cross-agency team of researchers under the sponsorship of the Federal Committee on Statistical Methodology conducted the project “Profiles in Success of Statistical Uses of Administrative Data.” The project examined seven successful data-sharing arrangements between various federal, state and academic entities. It conducted focus groups, via conference calls, with the projects’ original participants who discussed the challenges they faced in developing a cross-agency data-sharing agreement (e.g., a Memorandum of Understanding or MOU) and how those challenges were met. The report considers the statistical uses of administrative records at a broad level. It then describes the design and protocol of the “Profiles in Success” project. The report explains the topics of the seven data-sharing projects chosen for study and how each project bridged data gaps with administrative data.

The “Profiles in Success” project identified six core challenges: (1) project definition and development (i.e., when two or more agencies that are potential partners first begin to define a project that is desirable and feasible); (2) financial challenges; (3) legal challenges; (4) technical challenges; (5) managing internal processes; and (6) managing interagency relations. Focus groups provided insights on these challenges and lessons learned on factors that contributed to the success of their projects. Based on the seven case studies, the “Profiles in Success” research team distilled and synthesized these factors into four “elements of success.”

The findings of the report are contained in four “elements of success”: (1) vision and support by agency leadership; (2) narrow but flexible goals; (3) infrastructure; and (4) mutual interest. These elements can help facilitate future data-sharing agreements involving the statistical use of administrative data. Each of these elements can benefit many projects, including more familiar or routine ones. However, the elements have special nuances—and may be particularly critical—for the success of developing a cross-agency sharing of administrative data. The report not only describes each of these elements but provides detailed evidence of how the successful projects exemplify the four elements of success in actual practice. The major findings, without the full project details, are provided below followed by brief descriptions of the seven projects.

A companion to this report is “*Model Data Use Agreement for Sharing Administrative Records: Recommended Components Annotated With Examples*” (Cornman, Stephen, Howard Bradsher-Fredrick, Carol Comisarow, Christine Cox, Mark Denbaly, Rochelle Wilkie Martinez, Mark Prell, William Sabol, and Michelle Vile). The companion report, in draft form, focuses on the *MOU components* (e.g., parties to the agreement, legal authority, consideration, costs and reimbursement) that together compose a legal agreement which applies to all projects that share administrative records. In contrast, the “Profiles in Success” report examines challenges and strategies for managing the *MOU process* that develops projects and their data-sharing agreements.

Four Elements of Success for Data-Sharing Agreements.

(1) Vision and Support by Agency Leadership. Agency leadership is the element that assembles, organizes, prompts, motivates and facilitates all of the other elements of success. Projects on familiar topics may require little more than tacit approval by agency leadership. In contrast, cross-agency data-sharing projects can require significant involvement by agency leadership. One function of leadership is to provide, or sometimes to coalesce, a vision for the direction of the agency’s work on a data-sharing activity. This vision is needed to address cultural barriers that may exist within an agency. Generally, such barriers include the intrinsic difficulties of innovation—of introducing anything new into an established organization. Furthermore, in a statistical agency, which may be a data-receiving partner, the use of administrative records can be a significant departure for a survey-oriented organization. Their use can even be seen as a challenge to the importance of traditional survey work. In a program agency, which may be a data-supplying partner, statistical uses of the agency’s data may be considered of secondary importance to administering the program. Such uses may even be seen as a threat to protecting administrative data from unauthorized or inappropriate use or disclosure.

Agency leadership no doubt recognizes that green-lighting project development and, later, signing an MOU are necessary for success. However, data sharing is too new and too complicated to happen on automatic pilot. An MOU may be unlikely to be drafted for signature—that is, the project will be abandoned along the way at the development stage—if agency leadership and management do not provide follow-up support after the initial broad approval for development. Agency leadership can make these difficult projects a priority for the agency, assign appropriate staff, and ensure that policies and procedures are in place that are appropriate for data sharing, especially on data stewardship issues.

(2) Narrow but Flexible Goals. The second “element of success” gives equal weight to two important characteristics of project goals: narrow and flexible. Narrow goals can be conducive to cross-agency discussions that take into account both technical feasibility and data stewardship issues. A large relational database can involve hundreds of millions of records and dozens of variables. For success in developing an MOU that can be signed, it can be important for the MOU to specify goals narrowly—sometimes down to the level of which particular fields in a database will be shared between agencies and how those fields will be used. Narrow goals are also helpful because they facilitate cross-agency discussion of data stewardship issues. In general, the data stewardship issues of privacy, confidentiality and legal authority for sharing data can depend on what datasets are involved, which subset of variables from the datasets will be used, how they will be used, and the physical arrangements and procedural safeguards that are in place for data security.

Even though it is beneficial for goals to be narrow enough to discuss fruitfully in cross-agency discussions, goals must also be flexible—that is, to change as a result of those discussions. These changes can take two forms: evolving goals, meaning that initial goals change before any project output is reached, and added goals, meaning that a successful project can do more things as time goes by. In the successful data-sharing projects under study, when some initial goal encountered a barrier or challenge that proved to be difficult or impossible to meet, the goal shifted to adapt. If instead specific or narrow goals had been cast in concrete too early, a project

may have found that there were insurmountable barriers for those particular goals. The ability to allow for goals to evolve can be critical for new projects. Flexibility also permits goals to be added over time as new uses of data come to be identified as a by-product of the two agencies working together. The art of developing and managing these projects involves, in part, creating cross-agency relationships and environments of mutual trust and support. In such environments, undetected opportunities become detectable. Goals that are flexible can result in expanded activities as initial success begets more success.

Goals changed and goals evolved in each of the seven case studies. These projects make significant contributions even though the final form of the project may not have quite been the original goal or the original means, or had additional goals added with time. The visions behind the projects—the benefits of using administrative data or linking it to survey data—were motivation to keep going and craft projects that were both profound and feasible.

(3) Infrastructure. The third element of success, “infrastructure,” has two components: staffing, and policies and procedures. Cross-agency projects benefit from people who are results-oriented, supportive of the project’s goals, experienced with the data, and able to work cooperatively with people in their agency and the partner agency. The quantity of staff time required was frequently a concern for the seven data-sharing projects examined for this report. The projects were consistently understaffed because the cumulative hours of staff time required were underestimated. Moreover, the number of months it took to complete the MOU, funding and technical file preparation processes was underestimated. The explanation for systematic underestimation of staffing requirements seems to be that these data-sharing projects are examples of innovation. The projects were “first-generation” projects and simply so new that the full extent of the necessary time was hard to anticipate. As a result of these projects, though, there is now more information for future projects to better estimate and even to reduce the time required.

The second component of infrastructure is the importance of having appropriate policies and procedures in place to support data-sharing activities. For laws and regulations that are established outside the agency, the agency may be responsible for providing or obtaining a legal interpretation of how the laws and regulations on privacy, confidentiality, and data stewardship apply to its activities. An agency may need to create, implement and manage the process by which the MOU review will occur. Management may not know *a priori* the relevant laws and regulations, the technical requirements, or the data-sharing project’s specific goals. Protecting data integrity and the confidentiality is at stake. Ensuring that an agency’s policies and procedures are conducive to data sharing is a management activity of paramount importance.

(4) Mutual Interest. To reach successful conclusion, data-sharing arrangements benefit each partner to the project. In particular, since statistical agencies’ benefits are likely to be one-way receipt of a new data file, it is critical to identify benefits to the data-supplying agency. These benefits must outweigh and address any perceived risk. Each agency has its own set of statutory and policy requirements to protect the confidentiality of its own data. Each agency has its own legal, policy and cultural environment. Developing an awareness and sensitivity to each other’s policies and rules is achieved through cross-agency negotiation and discussion. Together, the

partners can discern if an MOU is possible that can accomplish something that serves the mutual interest of both agencies.

Project Descriptions. The “Profiles in Success” research team selected the seven case studies to achieve diversity on several project dimensions, including: the “type of partnership” based on federal, state, and academic partners; the “unit of observation” based on whether the data involved persons or business establishments; the “type of project” for example, whether it was one-time research or intended for production on an ongoing basis; and the “use of the project” for example, survey improvement, and program evaluation. The seven case studies were:

- SNACC project, which merged survey data from the Census Bureau and the National Center for Health Statistics with administrative records on Medicaid from the Centers for Medicare and Medicaid Services. SNACC was designed to address the longstanding concern about the range of estimates of Medicaid enrollees reported across government surveys and the Medicaid program records.
- Longitudinal Employer-Household Dynamics (LEHD) project, which is a federal-state partnership currently involving the Census Bureau, 49 States and the District of Columbia. LEHD combines federal and state administrative data on employers and employees with censuses and surveys to inform such issues as: the dynamics of labor markets at the state and sub-state level, based on LEHD’s component Local Employment Dynamics (LED) program; immigrants; and transportation.
- Food Assistance project, which involved three research centers and four states that had collaborated on six distinct studies using administrative records on food stamps and/or cash welfare. They studied: differences between survey and administrative data on receipt of government assistance; various outcomes (such as employment); and issues about survey design and methodology.
- Education and Incarceration project, which involved academic researchers who used state-level data to study education or incarceration. Researchers used administrative records to help conduct studies examining student achievement, teacher quality, school effectiveness, and education programs and the effects of earning a GED while in prison.
- NASS/FSA project, which documented existing and potential data-sharing efforts between the National Agricultural Statistical Service (NASS) and the Farm Service Agency (FSA), as well as improved access between the agencies, including NASS access to FSA micro-level data and FSA access to unpublished estimates by NASS.
- Industry Coding project, which enables the Census Bureau to send a file of Employer Identification Numbers to the Bureau of Labor Statistics and receive, in return, the results from matching the file against the BLS’ Business Establishment List, thereby improving the quality of industry coding on the business register, lowering costs, and lowering respondent burden.

- NTID/SSA/Cornell project, which is a partnership between the National Technical Institute for the Deaf (NTID), the Social Security Administration (SSA), and Cornell University that linked NTID application and enrollment data with SSA data to examine employment, earnings, and disability program participation. The project can be considered an example of using administrative data for program evaluation.

Profiles in Success of Statistical Uses of Administrative Data

Mark Prell, Howard Bradsher-Fredrick, Carol Comisarow, Stephen Cornman,
Christine Cox, Mark Denbaly, Rochelle Wilkie Martinez, William Sabol, Michelle Vile¹

1. Introduction

Administrative records can be used for a variety of statistical purposes. A report by the Committee of National Statistics (CNSTAT) of the National Academies states, “*Statistical purposes* include description, evaluation, analysis, inference, and research [1, p. 2, emphasis in original]. To contrast “statistical purposes” with other non-statistical purposes, the report continues:

For these [statistical] purposes, a statistical agency may collect data from individuals, establishments, or other organizations directly, or it may obtain data from administrative records, but it does not do so for administrative, regulatory, or law enforcement purposes. Statistical purposes relate to descriptions of groups and exclude any interest in or identification of an individual person or economic unit. The data are used solely to describe and analyze statistical patterns, trends, and relationships involving groups of persons or other units. [1, p. 2]

In many cases, administrative data collected by one agency are useful for another agency’s research and statistical purposes; here, a variety of organizational units are called “agencies,” including federal agencies, bureaus or centers; academic and non-profit research institutions; and entities within State governments. A data-sharing agreement, or Memorandum of Understanding (MOU), is the legal instrument by which one agency obtains and uses administrative records from another. This report examines case studies conducted in 2008 that examined several data-sharing projects and their MOUs. The case studies identified challenges associated with developing and executing the MOUs, efforts undertaken to overcome the challenges, and the lessons learned for future data-sharing MOUs.

The case studies were conducted by a cross-agency team of researchers in the project “Profiles in Success of Statistical Uses of Administrative Data” (hereafter, the “Profiles” project). The team members belonged to the MOU subcommittee of the Statistical Uses of Administrative Data Subcommittee, which in turn was convened under the auspices of the Federal Committee on Statistical Methodology. The project examined seven successful data-sharing arrangements between various federal, state and academic entities and conducted focus groups via conference calls with the projects’ original participants. They discussed data-sharing challenges and how challenges were met. This report conveys major lessons that can be learned from the seven case studies that can help facilitate future data-sharing agreements involving the statistical use of administrative data. Four elements of success, together with specific examples from the case

¹ This paper represents part of the ongoing work of an interagency subcommittee under the auspices of the Federal Committee on Statistical Methodology (FCSM). The views expressed represent the individual authors and not necessarily the full FCSM or the agencies at which the authors are employed, which are: Bureau of Justice Statistics (William Sabol), Economic Research Service (Mark Denbaly and Mark Prell), Energy Information Administration (Howard Bradsher-Fredrick), Office of Management and Budget (Rochelle Wilkie Martinez) National Center for Education Statistics (Stephen Cornman), National Center for Health Statistics (Christine Cox), U.S. Census Bureau (Carol Comisarow and Michelle Vile).

studies, are identified: vision and support by agency leadership, narrow but flexible goals, the importance of infrastructure in terms of staff and processes, and mutual interest.

The report's distillation of the experiences of the "first generation" of data-sharing projects is meant to make future data-sharing projects easier to develop. While these types of projects are not routine, especially for agencies that have not done any data-sharing previously, the benefits of data-sharing can merit the cross-agency efforts such projects require. Data-sharing can be time-consuming and resource-intensive. At the same time, though, it can be a *relatively* low-cost route to certain goals. For example, the costs of collecting certain administrative data and survey data may have already been incurred by two agencies. The additional cost of merging the two datasets can be less than the cost of new primary data collection.

It is important to recognize that this report adopts the term "agency" strictly for brevity. It is used to avoid repeatedly listing the many types of federal, state and academic entities that the case studies examined. As the case studies themselves demonstrate, state agencies and independent researchers in academic institutions or research centers are deeply involved with using administrative data. They envision and develop projects of statistical uses of administrative data. They conduct data analyses. They can be the entrepreneurial change agents who establish relationships with a government agency, fellow academicians, or research centers to form partnerships for data sharing and analysis. As used in this report, then, the term "agency" should not be interpreted as synonymous with "federal agency." Indeed, it could be a federal (or state) agency that is responding to overtures from an independent researcher. In such a case, "vision and support by agency leadership"—one of the elements of success—may be referring to vision and leadership of a professor, a dean, or a researcher or director at a research center. Leadership and other talents are found in each of the many types of organizations that the case studies examined.

A view that underlies this report is that the life-cycle of a data-sharing projects (or interagency projects in general) has two broad phases: development and execution. A development phase of interagency negotiations precedes—and hopefully culminates in—the signing of an MOU by the project's partners. A project milestone is reached when the MOU is signed, after which the project enters its execution phase. In this two-phase view, a signed MOU represents a *mid-point* for a project—not its *starting point*.

Those who are new to data-sharing activities may be prone to thinking of a project as "starting" once an MOU is "done and out of the way" so that the "real work" can begin. Such a view might not deny that projects have two phases, but it more or less considers the only worthwhile phase to be the execution phase. In this view, jumping to hammering out a draft MOU is thought of as the quickest way to get a project done. This jump-to-the-MOU approach is understandable, but unhelpful. It is understandable because, indeed, a signed MOU is the legal instrument by which a project is executed and project results are finally achieved. It is unhelpful, however, because impatience and a preoccupation with reaching the execution phase can be self-defeating. The jump-to-the-MOU approach tends to give short shrift to the effort, care, and time needed to think through and develop a viable project and its MOU, which can result in a draft MOU that stands little chance of passing review and being signed (by *either* agency).

A companion to this report is “*Model Data Use Agreement for Sharing Administrative Records: Recommended Components Annotated With Examples*” (Cornman, Stephen, Howard Bradsher-Fredrick, Carol Comisarow, Christine Cox, Mark Denbaly, Rochelle Wilkie Martinez, William Sabol, and Michelle Vile, 2009.) The companion report focuses on the *MOU components* (e.g., parties to the agreement, legal authority, costs and reimbursement) that together compose a legal instrument without which no project is executed. In contrast, the “Profiles in Success” report examines challenges and strategies for managing the *MOU process* that develops projects and their data-sharing agreements.

The next section examines how administrative data can be used for statistical purposes and certain strengths and limitations of both administrative data and survey data. It concludes by reviewing some of the broad challenges involved when developing MOUs to share administrative data for matching to survey data or for other statistical purposes. The third section describes how the Profiles project was conducted. The fourth section describes the case studies, the issues that motivated them, and the roles played by administrative records. The fifth section identifies emerging themes that arose from the Profiles project—specifically, the four elements of success that can be critical for developing a data-sharing MOU. A concluding section summarizes major lessons of the Profiles project.

Section 2. Statistical Uses of Administrative Data

Summary. In some instances administrative records can approximate a census, enabling them to serve as the basis of sample frames for surveys. Combining administrative and survey data can be fruitful for econometric analysis or program evaluation that uses the distinct strength of each data source. A strength of a program’s administrative records is that their data are highly accurate on program participation (“recipient” vs. “non-recipient”) and a participant’s level of program benefits; surveys do collect such data but face sampling error, coverage error, non-response, and measurement error. A strength of surveys is their ability to collect more complete information on the characteristics, attitudes and behaviors of respondents and to include data on non-participants, making survey data relatively more complete. Sharing data poses risks of disclosure of personally identifiable information and raises issues of privacy, confidentiality and informed consent. Completing a successful MOU between agencies requires careful attention to legal and procedural issues that can differ between the agencies due to their different missions, statutes and regulations.

Two characteristics of administrative data strengthen their use for statistical purposes. First, in terms of inclusion of the population universe, in some instances administrative records can approximate a census. It is this property of some administrative files that makes them the basis of sample frames for surveys.

A second characteristic of administrative records is that they contain with considerable precision a listing of the program’s participants and the benefit amount paid, if any, to each participant. Administrative records may have relatively better quality for at least some of the data they contain. While surveys can solicit information on program participation and benefit levels, survey data may be subject to more error on this information than the program’s own administrative data due to sampling error, coverage error, non-response, and measurement error. For the binary variable of participant/non-participant, measurement error is tantamount to a

misclassification error, that is, a data record mistakenly indicating that an actual participant is a non-participant or (perhaps more rarely) that a non-participant is a participant. Such misclassification would affect survey-based estimates of participation rates and of aggregate or average benefit receipt. Misclassification affects econometric analyses and program evaluations that contrast program outcomes between participants and non-participants. Estimated coefficients on the “participation” variable are biased when participation is measured with error. Misclassification would also affect studies that estimate the effects of other factors that influence the household’s or producer’s participation decision using participation as a dependent variable. In contrast, administrative data are much less subject to measurement error.

Despite these strengths, administrative records sometimes have limited information on the personal or household characteristics of those in the database. In contrast, a survey can have more complete information on the characteristics, attitudes and behaviors of respondents. Furthermore, administrative databases contain no information at all on non-participants whereas a survey captures both participants and non-participants. Thus, survey data can be relatively more complete.

Combining administrative and survey data can be fruitful for using the distinct strength of each data source—and the administrative data, which are already in a database, can be utilized at a marginal cost that is low compared to a new data collection. Projects that match an administrative data source to survey data or to other administrative records are well positioned to conduct such work as: (1) a descriptive comparison between characteristics of participants and non-participants, (2) evaluation of program outcomes (by comparing experiences of participants and non-participants), (3) assembling a data file of households that participate in more than one program (by linking multiple sets of administrative records) to examine cross-program experiences and effects, or (4) comparing survey and administrative data that should “match” to garner insights that can improve both survey design and administrative recordkeeping. In an era of heightened interest in program performance and accountability, as reflected in the President’s Management Agenda and the processes based on the Program Assessment Rating Tool (PART), the possibilities of expanding the use of administrative records for statistical purposes is promising.

At the same time, however, concerns about disclosures of private data, whether deliberate or inadvertent, may be greater than ever. The activity of matching administrative and survey data, or of simply sharing administrative data even if no linking occurs, involves data-sharing arrangements between two or more agencies. Sharing data poses risks of disclosure of personally identifiable information and raises issues of privacy, confidentiality and informed consent. Officials responsible for data stewardship must be satisfied that risks will be successfully managed and related issues are addressed before they agree to share data. Completing a successful Memorandum of Understanding (MOU) between agencies requires careful attention to legal and procedural issues that can differ between the agencies due to their different missions, statutes and regulations.

Besides the legal issues, data-sharing projects involve financial and technical challenges. Other challenges include defining a project’s scope jointly with the other agencies involved and managing interagency relations. Managing an agency’s internal processes is another challenge.

Sharing administrative records is an innovative practice, especially in an agency's first data-sharing project. Innovation requires change or, rather, innovation *is* change. As always, some aspects of change can be accomplished through the agency's pre-existing internal processes. But for a data-sharing project, unlike more routine activities, some aspects of change will involve modifying or expanding the agency processes themselves—in these cases, the change occurs first *to* the processes and then *through* the new processes.

3. Design and Protocol of the Profiles project

Summary. The Profiles project is a project about projects. Its focus leans more towards “how” successful data-sharing projects were developed. The research team of the Profiles project developed a list of candidate “successful” projects for selection. To study projects of different types, seven case studies were chosen to achieve diversity on several dimensions. The nature of success—the particular milestones or outcomes that a project passes to be deemed “successful”—differed from project to project. Information was gathered through conference calls with projects’ original participants. The protocol of the conference calls consisted of 14 questions under four headings: Project Goals, MOU specifics, Challenges and Barriers, Outcomes and Lessons. The six items on Challenges and Barriers are the core questions of the Profiles project that involve: (1) project definition and development (i.e., when two agencies that are potential partners first begin to define a project that is desirable and feasible); (2) financial challenges; (3) legal challenges; (4) technical challenges; (5) managing internal processes; and (6) managing interagency relations. Information on these challenges and how they were met can be very hard to gather from professional publications. Yet insights and lessons learned on these six core questions are potentially very useful to other agencies developing or contemplating data-sharing projects.

The Profiles project is a project *about* projects. Its focus leans more towards “how” successful data-sharing projects were developed and less towards “what” the projects were studying. The topics of the projects are of interest primarily for understanding the motivation and context for the data-sharing arrangements. The Profiles project examined partnerships among federal agencies, state agencies, and academic researchers in a university or other non-governmental research center. For brevity, this report refers to any partner as an “agency.”

The Profiles project examined only projects that have already achieved success. It did not examine data-sharing efforts that were abandoned or others that could be considered “ongoing” but not yet successful. Thus, a limitation of the Profiles project is that it cannot answer the penetrating question “Why do some projects succeed while others are abandoned?” Instead, the narrower—but still informative—questions posed to project participants were about the challenges they faced, how challenges were overcome, and what the lessons are for future projects.

The research team of the Profiles project developed a list of candidate “successful” projects for selection as a case study. The list reflected the knowledge and expertise of the team's members, as well as input from the larger subcommittee of which the research team was a part. Case studies were chosen from the larger list to achieve diversity on the basis of six criteria, such as “type of partnership” (combinations of federal, state, and academic partners); “unit” (person or establishment); “type” (one-time research or intended for production on an ongoing basis); and

“use” (e.g., support research for policy issues, survey improvement, and evaluation). This report provides a brief project description for each case study to provide the context in which the data-sharing partnership was formed. The descriptions draw from various sources, including responses provided in focus groups. For each description, particular attention is paid to the administrative records involved and their use.

The projects included in this research are each successful, but they are successful in different ways. That is, the *nature* of success—the particular milestones or outcomes that a project passes to be deemed “successful” by its participants—differed from project to project. One project was successful once administrative and survey records were merged, essentially on a one-time basis, and estimates from the resulting data file provided answers to the policy and research issues that motivated the project. In another case, a project was successful once a system and procedures were in place to collect and merge administrative and survey records on a periodic basis and the first (of a series) of data products was publicly released. In yet another case, aspects of success included interagency access on an ongoing basis to program data or to unpublished statistical estimates. The factor responsible for this project-specific definition of success is the underlying diversity in the uses of administrative records for statistical purposes.

Each case study was conducted in three stages: organizational, conference call, follow-up. In the organizational stage, a member of the Profiles project research team arranged a conference call with a project’s participants and obtained certain types of background information that would free the call to discuss participants’ insights rather than easily accessible facts. The questions to be answered in this first stage are displayed in Table 1.

Table 1
A. Does the project have a standardized name that would be understood by its agency partners to distinguish and identify the project’s activities?
B. Who were the participating organizations and what was the main contribution of each in terms of data and/or funding?
C. What specific data and reference periods were involved with the project?
D. What agencies developed new MOUs or Interagency Agreements specifically for the project?
E. What are the citations of for the legislative and regulatory requirements that each agency faced in the course of developing MOUs?
F. Who would we contact to obtain copies of the MOUs?

Across the projects, the first stage was broadly successful, though some portions were collected during the conference call. In the end, focus group participants were not pushed to provide exact citations for legislative and regulatory requirements in Question E. One or more MOUs were obtained for each of the seven case studies. The second stage was the conference call itself. As

displayed in Table 2, the protocol of the conference calls consisted of 14 questions under four headings: Project Goals, MOU specifics, Challenges and Barriers, Outcomes and Lessons.

Table 2
<p>Project Goals</p> <ol style="list-style-type: none"> 1. In your words, what was the project’s overall goal that the MOU served, and what were agency-specific goals and possible secondary objectives? 2. Did goals change as discussions between the agencies progressed, or perhaps as more partner agencies became involved? <p>MOU specifics</p> <ol style="list-style-type: none"> 3. What is a list of the Divisions or offices or official positions within each agency that scoped the project and created and approved the MOU? 4. How did previous or ongoing relationships between the partnering agencies influence the new MOU or Interagency Agreement? 5. Who were champions at the outset of the project? 6. Who propelled the project to completion once an MOU was in place and ready for execution? <p>Challenges and Barriers.</p> <p>Any project involves a variety of challenges and barriers that can be grouped in various categories. We are using six groupings: project definition and development; financial; legal; technical, internal processes (internal to your agency); interagency relations.</p> <ol style="list-style-type: none"> 7. What challenges in <u>project definition and development</u> were faced and how were they overcome? 8. What <u>financial</u> challenges were faced and how were they overcome? 9. What <u>legal</u> challenges were faced and how were they overcome? 10. What <u>technical</u> challenges were faced and how were they overcome? 11. What challenges involving <u>internal processes</u> (internal to your agency) were faced and how were they overcome? 12. What challenges involving <u>interagency relations</u> were faced and how were they overcome? <p>Outcomes and Lessons</p> <ol style="list-style-type: none"> 13. What specific outcomes—which survey fielded from a sample frame, which research database, which government report(s)—have resulted from the project? 14. Are there general observations or “lessons learned” that you would like to share with other agencies interested in statistical uses of administrative data?

The type of information gathered by the six questions in the Challenges and Barriers section can be very hard to gather from professional publications that do not typically include the behind-the-scenes stories and impressions. Yet, answers to these types of questions are potentially very useful to other agencies developing or contemplating data-sharing projects. The Challenges and Barriers questions can be considered the core of the Profiles project.

Participants themselves may use different languages or frameworks for thinking about their work and the problems they face. For example, when staff at two agencies are not reaching agreement about a project description and scope suitable for putting into a draft MOU, is there a barrier in “project development” (a stages-of-projects framework) or is it a “legal” barrier (a functional framework) or an “interagency relations” problem (an organizational behavior framework)? Any given barrier might be thought of and described differently by different persons. By building in a variety of frameworks into the six core questions, the Profiles project was designed to elicit helpful insights from participants regardless of which language or framework they used to describe the challenges they experienced.

4. Case Study Descriptions: Bridging Data Gaps with Administrative Data

Describing the goals and activities of the seven projects provides the context for the participants’ experiences and highlights the critical role played by administrative records in bridging data gaps. In summary, the overarching purposes behind the seven case studies include supporting research for policy issues, survey improvement, serving as enumeration list and sample frames, and program evaluation.

Case 1. SNACC. The project is commonly referred to as the SNACC project, an acronym based on the project’s partners: SHADAC (State Health Access Data Assistance Center, University of Minnesota); NCHS (National Center for Health Statistics); ASPE (Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services); CMS (Centers for Medicare and Medicaid Services); U.S. Bureau of the Census. The SNACC project involved a number of funding sources, data files, and activities spread across several phases of the project. The project benefited from funding provided by ASPE and by the Robert Wood Johnson (RWJ) Foundation, both of which provided support to SHADAC at various times. The RWJ funding of December 2005 provided impetus.

The SNACC project reached its first milestone in January 2006 with an interagency agreement (IAA) between Census and CMS, which authorized Census to re-use certain CMS data and to acquire new state-level data from CMS that selected states would provide to CMS. The project has invited seven States to provide state-specific data for Medicaid and State Children’s Health Insurance Program (SCHIP): Florida, Maryland, California, Pennsylvania, New Jersey, Louisiana, and Minnesota (Research Project to Understand the Medicaid Undercount, 2008).

The SNACC project is also known as the “Medicaid Undercount Project,” which was designed to address the longstanding concern about the range of estimates of Medicaid enrollees reported across government surveys and the Medicaid program records. For example, the Congressional Research Service (CRS) reported that a 2000 Current Population Survey (CPS) estimate of about 29.5 million enrollees is below administrative figures of about 40.5 million enrollees or more.

(The CPS estimate cited here of 29.5 million is the “high” estimate, which CRS describes as “the estimate most commonly used in Census Bureau publications” while the administrative figure cited of 40.5 million is a “low” figure (CRS, 2005, p. 3); using a “high” administrative figure of 44.7 million results in an even greater difference between program and survey figures.) Other surveys besides CPS also result in Medicaid enrollment estimates that are below administrative figures, although the magnitudes depend on which survey is considered. Estimates of Medicaid enrollment can, in turn, affect estimates of the number of Americans who are uninsured, which is a major health policy issue.

Members of the SNACC conference call included staff who were at ASPE, Census, CMS, NCHS, and SHADAC during the project. They identified that the project’s overarching goal was to understand the “reasons” or “sources” or “causes” of the Medicaid undercount. The focus group also provided the context for that goal:

Children and insurance were a key issue. The undercount was vexing. How many kids were involved? What would be the cost to care for them? Those questions were being asked by the administration and by policy staff.

A related program issue of policy importance was that allocations across states of federal funds to support states’ SCHIP programs depended, in part, on the estimated number of uninsured low-income children in a state based on CPS data (Congressional Research Service, 2005).

By linking administrative records of actual Medicaid enrollees with survey data, the SNACC project estimated the size of the Medicaid undercount and better understood the reasons for the discrepancies.

Case 2. LEHD. This project is Census’ Longitudinal Employer-Household Dynamics (LEHD) program, a federal-state partnership currently involving 49 States and the District of Columbia. LEHD combines federal and state administrative data on employers and employees with censuses and surveys while protecting the confidentiality of people and firms that provide the data. LEHD data and analyses inform such issues as: the dynamics of labor markets at the state and sub-state level, based on LEHD’s component Local Employment Dynamics (LED) program; immigrants; and transportation.

Firms offering income-earning opportunities and workers supplying time, effort and skills interact with one another in labor markets. These interactions culminate in employment contracts, written or verbal, by which both sides of the market agree on jobs and acceptable wage rates. Often, data are collected from one side or another of the labor market. The origin of the LEHD project was a vision of the benefits to be gained from an integrated approach to combining data on employers and employees. Members of the LEHD focus group were at Census when LEHD was developed. They recalled:

At Census, the overall goal of the project was to get rid of “stovepipes” of data. The Economic and Demographic Directorates [which focus on business and household data, respectively] each build assets and approaches to working statistical operations. They wanted to look at both sides of the labor force.

In the LEHD project, as in several others examined by the Profiles project, more specific goals of how to implement this vision evolved over the life of the project. An initial approach involved linking W-2 records to survey records. When that approach did not prove workable, LEHD's second approach, developed in consultation with many states, arranged separate partnerships between Census and each participating state to obtain data from the state's Unemployment Insurance (UI) system. LEHD links household-based survey data on employees to the UI administrative records which originate with firms.

Case 3. Food Assistance. The Profiles project conducted a conference call with researchers from Census, the Chapin Hall Center for Children (of the University of Chicago), the Jacob France Institute (of the University of Baltimore), and the Ray Marshall Center (of the University of Texas). Through various partnerships amongst them (supported by various funding sources), these researchers have worked on a half-dozen projects usually involving administrative data from the Food Stamp Program (hence, the Profiles title "Food Assistance" for this case). Two projects compare Maryland administrative records on receipt of food stamps and of public assistance with self-reported survey information of receipt. A third project uses administrative data from Maryland, Illinois and Texas to analyze eligibility for a program of child-care subsidies, receipt of subsidies, and employment outcomes among low-income families in the three states. A fourth project examines food stamp eligibility using administrative data from Illinois, Maryland and Minnesota. A fifth project focuses on Illinois and Maryland to examine food stamp eligibility and participation. A sixth project examined child care subsidies, cash assistance, and food stamps using administrative data from Illinois and Texas and survey data. Members of the Food Assistance focus group recalled:

[For the first project] the goal was to examine reasons for discrepancy of food stamp receipt between Maryland administrative records and reported by households on Census Bureau's Supplementary Survey 2001 (SS01) . . . Other projects blossomed from the initial food stamp project . . . [For the fifth project] Do the two sources of data agree or disagree? The second goal is to describe who participates and how participation varies by characteristics of households . . . The [SIPP match] will help identify differences between SIPP and administrative records, help understand the seam bias problem in SIPP, and will be used to assess quality of event history data collection.

Goals for these six Food Assistance projects had much in common. They match administrative and survey data for individual households, within States, to study: program eligibility and participation, differences in the data on receipt of government assistance, various outcomes (such as employment), and issues about survey design and methodology.

Of the seven cases examined by the Profiles project, four included Census as a partner agency. While Census may be considered by some to have an orientation towards censuses and surveys—the agency's core mission—in recent years it has been active in data-sharing projects that use administrative data.

This new approach in the late 1990s transformed in 2004 into a line of business. Census contributed [to projects] through technical and programmatic assessment, data domain,

record validation, and linkage expertise, getting interested parties together, and understanding possibilities and constraints of the proposed applications.

Besides SNACC, LEHD, and the Food Assistance projects, Census was a partner on the Industry Coding project, described below.

Case 4. Incarceration and Education. The Profiles project examined a set of projects that used administrative records primarily for research on incarceration (e.g., earning a GED while in prison) and education. Members in the Incarceration and Education case study included researchers at the National Bureau of Economic Research (NBER), the University of Washington (UW), and the Education Research Center at the University of Texas at Dallas (UTD). In contrast to other case studies, researchers on this “project” were not partners with one another but instead had assorted projects and partnerships with various state agencies from which they obtained state-level administrative data. While case studies such as SNACC and LEHD involved survey data, the Incarceration and Education project, like much of the Food Assistance project, focused on linking different sources of administrative records.

The NBER project is evaluating outcomes from a housing study and involves administrative data from unemployment insurance agencies, schools, and state TANF agencies, as well as an examination of adult and juvenile arrests.

The overall goal of the administrative data component of the NBER project evaluation is to get administrative data for the project to compare treatment versus control groups for several outcomes: wages, arrests, participation in the Food Stamp Program or Temporary Assistance for Needy Families (TANF), and different aspects of schooling, including college enrollment . . . To get data, we are in touch with various types of agencies in different States. The data we obtain include employment data, TANF and food stamp receipt, arrest data, and national data on college enrollment from the National Student Clearinghouse.

The UW project uses administrative data from three agencies of the State of Washington: the Department of Corrections (DOC), which provided data on incarcerated persons; the State Board of Community and Technical Colleges (State Board), which provides data on persons who take the General Educational Development (GED) exam; and the Employment Security Department (ESD), which provides UI data.

The idea was to use administrative records to get a sample of men who spent time in prison and a matched comparison sample of men not in prison. There are data on women, but we used only the data for men. It was very research-focused. There were three substantive outcomes: employment, wages, and wage growth. We looked at men’s income before they went into prison and after they came out, in comparison with men who were not incarcerated. We think of it as a differences-in-differences approach. With a matched comparison sample, we made broader claims of an “incarceration effect” on labor market outcomes.

The Education Research Center at UTD has data transfers from the Texas Education Agency and the Texas Higher Education Coordinating Board.

From its data warehouse of administrative records, the Center creates research-ready datasets, documents them, and makes them available to researchers to study education outcomes and workforce issues . . . The proposed research projects cover topics such as student achievement and teacher quality; college readiness; school system effectiveness; high-performing, high-poverty Texas schools; and socioeconomic impacts of alternative education programs on Texas and regions in Texas . . . Data on over 10 million children are included in the dataset . . . The secondary school data are linked to higher education data (including private college attendance) using encrypted 10-digit identification numbers, and individual-level school data can be linked to work force data, such as unemployment insurance records.

Case 5. NASS/FSA. This project involves data-sharing activities between the National Agricultural Statistical Service (NASS) and the Farm Service Agency (FSA), both agencies of the U.S. Department of Agriculture (USDA). NASS provides state and county-level crop acreage, production, yield, and price data used by the FSA to implement, administer, and monitor their farm support programs. As requested by FSA, NASS also provides advice on the use of NASS data as well as statistical consultation. FSA provides NASS with three main types of data: (1) producer name and address information for list frame updating activities, (2) acreages of individual crops either at the producer level or various levels of aggregation, and (3) Common Land Unit (CLU) Geographic Information System (GIS) data for use in area frame data collection and as ground truth in the creation of the Cropland Data Layer. Members of the NASS/FSA case study stated:

The overall goal of the MOU [revised in 2004] was to document existing and potential data-sharing efforts and to improve access to existing data between the agencies, including NASS access to FSA micro-level data and FSA access to unpublished estimates by NASS.

NASS and FSA have a longstanding relationship within USDA as a principal statistical agency and a major program agency with overlapping stakeholders in agricultural production and programs. The 2004 MOU revised the 1992 MOU, which itself was a revision of a previous MOU. While the 2004 MOU included established practices, new arrangements were introduced such as direct access to portions of an agency's data.

Case 6. Industry Coding. The Industry Coding project is a cooperative activity by which the Bureau of Labor Statistics (BLS) and Census make use of administrative data to improve the Census' Business Register, which is the basis of the Economic Census and the sample frames constructed by Census for various economic surveys. Census ships a quarterly EIN (Employer Identification Number) extract file to BLS, which then matches the EINs against their Business Establishment List and returns matched records with selected BLS data variables appended. According to BLS and Census staff who participate in the Industry Coding project:

The goals of the MOU were to improve the quality of industry coding on the business register, lower costs, and lower respondent burden . . . There was an OMB directive for Census and BLS to initiate the Industry Coding project to lower costs for the economic census . . . The project can be considered an example of "good government."

Case 7. NTID/SSA/Cornell. This project is a partnership between a federal agency (Social Security Administration), a university (Cornell), and the National Technical Institute for the Deaf (NTID). NTID is one of eight colleges of Rochester Institute of Technology and serves 1,110 deaf and hard-of-hearing students in part through a Congressional appropriation through the U.S. Department of Education. Relative to the general population, fewer deaf and hard-of-hearing persons complete high school, complete college, or work and have lower family incomes.

The Social Security Administration (SSA) administers Supplemental Security Income (SSI) and Social Security Disability Insurance (SSDI). SSI and SSDI are the largest Federal programs that provide assistance to people with disabilities. SSI pays benefits based on financial need while SSDI pays benefits based on a person having “insured” status through working and paying Social Security taxes.

NTID partnered with SSA and a researcher at Cornell who had experience in economic analysis of disability and retirement programs. The project examined employment, earnings, and SSI/SSDI participation of nearly 12,000 deaf and hard-of-hearing persons who are former students or applicants to NTID (beginning with NTID data in the mid-1990s). As participants in the NTID/SSA/Cornell focus group recalled:

[NTID] launched the project. [The Cornell economist] was asked by . . . NTID on how best to demonstrate if NTID was using government funding wisely . . . NTID has a Congressional line item and has to justify their expenditures to Congress each year. Gallaudet University, a college for the deaf in Washington, DC, has to do the same . . . What is the impact of an NTID education? The impact can be asked different ways. For example, for those who come here versus those who don't, or for those who get a degree versus those who don't. It can also be asked for different types of degrees.

In contrast to some other case studies, the NTID/SSA/Cornell project involved linking administrative data to other administrative data (rather than to survey data). Data linkage was required between NTID records and SSA records in order to identify post-attendance employment and SSI/SSDI participation outcomes of NTID graduates, students and applicants. In addition, the project examines the receipt of SSI children's benefits by NTID applicants and graduates.

5. Emerging Themes: Four Elements of Success

To be successful, any project can use a dash of imagination and dose of initiative. But for some types of truly path-breaking projects, success hinges in part on whether the amounts of imagination and initiative are big enough. In contrast to new projects that closely resemble an agency's past accomplishments, projects that use administrative data for statistical purposes are stepping further beyond the boundaries of the familiar. These projects are fraught with even more risk than familiar projects, but they can result in commensurately higher rewards. The seven case studies examined by the Profiles project are examples of successes of data-sharing initiatives.

Participants in the Profile projects' conference calls offered insights on factors that contributed to the success of their projects. These factors are the conditions and behaviors that enabled the participants to address the organizational, legal, financial and technical challenges of their projects. The Profiles project distilled and summarized these factors into four "elements of success":

- vision and support by agency leadership;
- narrow but flexible goals;
- infrastructure; and
- mutual interest.

While these four "elements of success" can benefit many projects, including more familiar or routine ones, they have special nuances—and may be particularly critical—for the success of developing an interagency agreement for sharing administrative data.

Element 1. Vision and Support by Agency Leadership

Summary. Agency leadership is the element of success that assembles, organizes, prompts, motivates and facilitates all of the other elements of success. Cross-agency data-sharing projects can require significant involvement by agency leadership to provide or coalesce a vision for data-sharing activity. Data sharing can be an innovative activity that represents a departure from previous experience. Agency leadership can make these difficult projects a priority for the agency, assign appropriate staff, and ensure that policies and procedures are in place that are appropriate for data sharing, especially on data stewardship issues.

The time and expertise of staff are not the only resources that are limited in an agency. There are many demands placed on agency leadership, making their time and attention a scarce resource that they need to allocate with care. Projects on familiar topics may require little more than tacit approval by agency leadership. In contrast, cross-agency data-sharing projects can require significant support or attention to be successful. Their success may depend more critically on active involvement by leadership and management at the inception and conclusion of the MOU process, as well as continued involvement throughout the development of the MOU. One function of agency leadership is to provide or to coalesce a "vision" for the direction of the agency's work on a data-sharing activity. This vision is needed to address cultural barriers that may exist within an agency. Generally, such barriers include the intrinsic difficulties of innovation—of introducing anything new into an established organization. Furthermore, in a statistical agency, which may be a data-receiving partner, the use of administrative records can be a significant departure for a survey-oriented organization. Their use can even be seen as a challenge to the importance of traditional survey work. In a program agency, which may be a data-supplying partner, statistical uses of the agency's data may be considered of secondary importance to administering the program. Such uses may even be seen as a threat to protecting administrative data from unauthorized or inappropriate use or disclosure.

Agency leadership no doubt recognizes that its support is needed at two critical stages of a data-sharing project. At the earliest stage, agency leadership will need to approve a project "in principle" for efforts to continue towards developing an MOU between the two (or more) agencies. Later, when an MOU is on the desk and ready for signature, the data-sharing project can be legally finalized only with management support. But while such involvement at the start and the end are *necessary* for success, they may not be *sufficient*. The MOU is unlikely to be

drafted for signature—i.e., the project will be abandoned along the way—if agency leadership and management does not provide follow-up support after the initial broad approval.

Ongoing support for data-sharing projects can take various forms. Agency leadership can make these difficult projects a priority for the agency and its staff. If getting an MOU is always on the bottom of the to-do list, the project may never get done. Data sharing is too new and too complicated to happen on automatic pilot. Not only will staff need to be assigned to develop the project and draft an MOU, an agency will need to have data-sharing procedures in place. Setting priorities, allocating staff and developing procedures are inherently managerial activities and, therefore, support by agency leadership to have an appropriate structure in place is critical for success. Staff and procedures are two aspects of what the Profiles project calls “infrastructure”—a factor so important that the Profiles project considers infrastructure to be one of the four “elements of success” in its own right. One way of thinking of agency leadership and its role in data-sharing projects is that it is the element of success that assembles, organizes, prompts, motivates and facilitates not only infrastructure but all of the other elements of success.

For the SNACC project, Michael Davern of SHADAC provided initial leadership. Within federal government,

ASPE [the Office of the Assistant Secretary for Planning and Evaluation, HHS] was instrumental for the project. Michael O’Grady, Assistant Secretary at ASPE, had been at CRS as a staff person during the Clinton administration. The [Medicaid] undercount was a problem for him . . . When he became Assistant Secretary, he saw that the undercount issue was still being debated and he concluded that research was needed. He put his political capital behind the project as well as ASPE dollars, in addition to the funding provided by Robert Wood Johnson . . . Linda Bilheimer was important for the project. She had been at Robert Wood Johnson before going to NCHS. She was interested in the project, and then Christine Cox became involved. Linda helped bring SNACC participants together . . . Charles Louis Kincannon and Hermann Habermann, the Director and Deputy Director of the Census Bureau, were very supportive. Their initial support was important for starting the project. . . CMS, who came to SNACC somewhat later than others, had a nagging concern not only about the undercount but also the potential double-counts in administrative data.

Even as the participants in the SNACC conference call were recalling the contributions of agency leadership, the history they told points to important functions of “staff.” Indeed, the classifications of “agency leadership” “management,” and “staff” do not always have clear dividing lines. Nor are the boundaries obvious between “envisioning” and “developing” a project.

Two ASPE staff, Robert Stewart and George Greenberg, urged that a correct answer on the [Medicaid] undercount was important—that the issue was a pragmatic one for the policy community . . . Michael Davern, Sally Obenski and other SNACC participants hoped the reports from the project would solve most of the issues around the Medicaid undercount. Sally Obenski repeatedly pressed SNACC participants to identify what was wanted to avoid “scope creep.” . . . There was a champion for the project at each agency, which was key for

working through procedures at each agency. Dave Baugh was the champion at CMS and he would get an answer on following procedures appropriately. Christine Cox helped to negotiate difficult issues involving NCHS. There were large efforts at Census by Michael Berning, Shelly Martinez and Sally Obenski. The process took a long time to be able to deal with confidentiality rules, etc. At Census it was important to have buy-in and involvement from Charles Nelson and his supervisor, David Johnson; otherwise the project could not have moved forward.

Thus, as critical as agency leadership can be to launching a successful project, those who report to agency leadership may themselves help shape a vision and provide leadership for a project.

The Industry Coding and LEHD projects also provide examples of a project launch that depended on people at the Directorate or Division levels. For Industry Coding:

The MOU between BLS and Census was done primarily by the Economic Planning Coordination Division [at Census]. That Division worked with BLS to create the MOU and shepherd it through. The IRS needed to be involved with the MOU, which made the operation more complex . . . Project champions include Tom Mesenbourg, the Division Chief of EPCD [at the time]. At BLS it was Brian McDonnell who, in 1992, was Chief of the Division of Administrative Statistics and Labor Turnover. In addition, at BLS Tom Plewes, Assistant Commissioner for Employment and Unemployment Statistics, was involved. Staff at OMB were champions too. They wanted to do data sharing.

For LEHD:

In early 1997, Census executives held an off-site meeting and it was decided by these directorates that they wanted to move away from the “silo-approach,” with the Demographic Directorate focused on household data and the Economic Directorate building the business data . . . Following that meeting, Nancy Gordon asked Julia Lane to initiate an American Statistical Association proposal pilot project using Maryland UI data . . . There was strong support from two key directorates at Census, from Economic and Demographic, which had Frederick Knickerbocker (Assistant Director for Economic Programs) and Nancy Gordon (Associate Director for Demographic Programs).

The NTID/SSA/Cornell case study also shows support by “agency leadership,” only in this instance such leadership is located not only in a federal agency but also in academic institutions.

Gerry Walters, who has since retired [from NTID], was there from the beginning and launched the project . . . When Gerry Walters originally asked Richard Burkhauser [of Cornell] to work on the project, he [Richard] was paid as a private consultant. When Richard Burkhauser realized how interesting the data were and how they could trace individuals over time, he thought they could get major external funding to use for research and still provide the answers for the congressional demand. So, Richard went to the department chair and the Dean at Cornell and explained the opportunity. They were both very supportive of trying to get external funding for this project. It was submitted as one of the projects for National Institute on Disability and Rehabilitation/Department of Education funding. The project was also able to get the National Institute on Aging to provide funding,

and Bob Weathers [of SSA] was able to come to Cornell to work on the project due to this initial funding . . . Initially, SSA was simply responding to requests from NTID to link SSA data to NTID data . . . The exciting thing is that this evolved over time into something bigger . . . At SSA in 2004 the Deputy Commissioner of Policy, Paul Van De Water, played a strong role. Before him, the Associate Commissioner for ORES, Peter Wheeler, also played a role. Both were very supportive.

Thus, while the first element of success is phrased as “vision and support by agency leadership,” this report has used the term “agency” to refer to any type of institution or organization under study. The NTID/SSA/Cornell project drew vision and support from outside a government agency—in this case, from NTID and Cornell. Similarly, the Food Assistance project and the Incarceration and Education project benefited from vision and support of researchers and directors at many institutions, including Chapin Hall, the Jacob France Institute, the Ray Marshall Center, the University of Texas, and the University of Washington. People at these institutions were active partners with government agencies and often they were entrepreneurs who initiated data-sharing projects. Successful projects can benefit from “vision and support by agency leadership” that is found both inside and outside of government agencies.

For the NASS/FSA project the MOU was signed by the two agency’s Administrators. The MOU itself and the data-sharing activities it described were developed and conducted within the Office of Business and Program Integration (OBPI) at FSA with limited involvement by the Administrator. It was explained during the Profiles project’s conference call that FSA is an agency with a “large” staff and that “much is going out the door all the time.” Because of these conditions, “as long as privacy issues are taken care of” and a draft MOU had been seen by a key person with both administrative responsibilities and data-sharing expertise in OBPI, the MOU would not be questioned. Some activities simply “wouldn’t reach the level of the Administrator.” Thus, this first element of success may be rephrased for greater precision: it may be that “vision and support by agency leadership” is critical for success when “agency leadership” is interpreted as that level of authority within an agency at which resources are assigned.

The NTID/SSA/Cornell case study shows that as people holding various leadership positions change, it can be important to continue project support by those who remain and to garner support from those who newly fill the positions. Participants in the focus group noted that a Deputy Commissioner of Policy and an Associate Commissioner for ORES, both of whom had been “very supportive,” had both left SSA. A project that had been supported by one person may not be supported by his or her successor. However, in this case such a problem did not occur. “The current people in their [SSA] positions are still interested and support this project.” Similarly, once Gerry Walters retired from NTID, Sara Schley continued the project.

Innovative data-sharing projects are complex and take longer to develop than more familiar work does. As just noted, leadership can change in the interim. While “vision and support by agency leadership” is considered here to be a necessary element of success, perhaps a subtle but critical aspect of support is “empowerment” of key staff. As it was put in one focus group, “What is needed is, at the staff level, for people to feel that the project is theirs and then pursue it.” It turns out, as described in the next element of success, that the “goals” behind a project can

evolve as cross-agency discussions play out. Initial “goals” can turn out to be difficult if not legally impermissible to pursue. It will commonly be persons who report to “agency leadership” who are charged with the interagency relations by which new goals are developed and new means are identified by which to reach an initial broad vision of using administrative data for statistical purposes.

Element 2. Narrow But Flexible Goals

Summary. Narrow goals can be conducive to cross-agency discussions considering technical feasibility and data stewardship. Because databases are large, it can be important for the MOU to specify goals narrowly—sometimes down to the level of which particular fields in a database will be shared between agencies and how those fields will be used. Data stewardship issues of privacy, confidentiality and legal authority for sharing data can depend on what datasets are involved, which subset of variables from the datasets will be used, how they will be used, and the physical arrangements and procedural safeguards that are in place for data security. Goals must also be flexible—that is, to change as a result of cross-agency discussions. When some initial goal encounters a barrier or challenge that proved to be difficult or impossible to meet, the goal may shift to adapt. Flexibility also permits goals to be added over time as new uses of data come to be identified as a by-product of the two agencies working together. Goals changed and goals evolved in each of the seven projects. Projects make significant contributions even though the final form of the project may not have quite been the original goal or the original means, or had additional goals added with time.

A second “element of success” that helps develop data-sharing MOUs is for projects to have goals that are both narrow and flexible. In project management, the distinction between a broad steadfast *vision* and the narrow flexible *goals* adopted to pursue that vision mirrors the difference between *strategy* and *tactics* in a military campaign. While the motivating vision behind a project can be broad, identifying and articulating narrow goals and activities brings the specificity and clarity to a draft MOU that can help it to pass technical and legal review. At the same time, as more information on capabilities and constraints is shared during cross-agency dialogue, each agency’s goals may need to be flexible to reach an MOU. Narrow goals that are inflexible can result in dead-end projects when a barrier arises, as barriers inevitably do.

A qualification of the Profiles conclusions is needed. The discussion of the first two elements of success distinguish between a project’s vision and its goals. The former can be broad while the latter, it is argued below, should be narrow and flexible. In practice, the boundary line between “vision” and “goals” may not be as sharp as the discussion suggests. Moreover, there can be some iteration between how clearly a project’s vision is described and the goals that turn out to be feasible for implementing the vision.

Narrow Goals. Linking administrative records housed in a particular database with data from a particular survey may be considered a “broad” goal. For success, specifying goals more narrowly, sometimes down to the level of particular fields in a database and how they will be used, can be important for developing an MOU. Such specificity is needed because of technical feasibility and data stewardship.

The importance of specificity derives in part from the technical properties of administrative records. The NASS/FSA case study points to the benefits of narrow goals when administrative records are involved:

It must be understood that a large relational database is not an Excel spreadsheet. There are millions, billions of rows of data. There are dozens of variables in each table . . . There are constant changes. There are new code developments and preexisting code changes . . . It takes a long time before a database is released, which can involve going from 400 million records down to a first release of summarized totals of 60 million. FSA pulls a lot of data from different databases. There are accounting codes. There are transactions codes. It seems to be a simple request to provide data on “payments” to [agricultural] producers . . . It is not obvious what “payment” data is requested. There are different payments.

In other words, it would not be helpful for a data-receiving agency to ask its partner agency to “send us your data and we’ll decide what parts are helpful for us once we look around inside of it.” Instead, the particular uses of particular data need to be identified by the data-receiving agency—and agreed to by the data-supplying agency—so that an MOU can be written with specifics. Again, from the NASS/FSA case study:

If there is an interest in marrying databases for some general undefined purpose, it will be harder to develop an MOU . . . When developing an MOU, there is a lot of discussion within an agency, and across agencies, before talk about actually writing the MOU. That discussion is where 95 percent of the work is. An agency has to determine what it is trying to do, what questions it wants to answer, and identify different priorities and different needs.

For NTID:

The technical challenge of the project is managing a very large data set. There are lots of databases in our infrastructure. I am lobbying hard . . . to get all the databases talking to each other . . . Some people might feel overwhelmed that a project needs data going back to 1980, and the data need to be matched to SSA data. But the power of the outcomes data, on populations rather than a survey, is important.

The data stewardship issues of privacy, confidentiality and legal authority for sharing data can depend on what datasets are involved, which subset of variables from the datasets will be used, how they will be used, and the physical arrangements and procedural safeguards that are in place for data security. These critical matters receive attention at the stage of legal review. An MOU may not pass such a review without providing sufficient attention and detail—hence, the imperative of developing narrowly specified goals for the MOU. Grappling with the privacy and confidentiality issues, and the technical and procedural issues with which they intermingle, is important long before reaching the stage of formal legal review, even as early as the conception of a project.

Flexible Goals. Even though it is beneficial for goals to be narrow enough to discuss fruitfully in cross-agency discussions, goals must also be flexible to change as a result of those discussions. In some of the case studies, projects had goals that evolved. When some initial goal

encountered a barrier or challenge that proved to be difficult or impossible to meet, sometimes the goal shifted to one that would still serve the same broad vision. The projects might not have been successful without being able to adapt their goals.

This section reviews comments from the case studies in relatively great detail. In future data-sharing projects, the activity of developing and adapting goals is likely to be a core concern. The Profiles project can better inform those future projects by clearly explaining how successful projects exhibit flexible goals. Before proceeding, it is helpful to consider terminology.

Why conclude, as the Profiles projects has, that “goals” need to be flexible? Would it be just as appropriate—that is, would it fit the case studies’ results just as well—to conclude that “goals” can or should be steadfast while the “means” to achieve them need to be adaptable? The short answer is “yes, either way of *phrasing* the lessons is meaningful.” It is worthwhile to distinguish the *lessons* embodied in the elements of success from the *language* used to explain the lesson. The cross-agency team on the Profiles project has confidence in the project’s lessons—the four elements of success—but is not wedded to any particular language or terminology. The Profiles project could have used different sets of terms to discuss its findings. Instead of suggesting that two elements of success are *vision* by agency leadership and narrow flexible *goals*, it could just as well be proposed that agency leadership provides the broad *goals* for projects and that a project benefits from narrow flexible *means* or *approaches* to reaching those goals.

Even though there was a common “Profiles in Success” protocol across the case studies, and therefore common questions and terms, different participants *interpreted* the questions and terms a bit differently. The Profiles projects discovered that participants in the focus groups used the term “goals” sometimes more as a synonym for a strategy and sometimes more as a synonym for a tactic. The Profiles project had to discern what participants meant and then to adopt a standard terminology to convey the results and lessons of all the focus groups.

Something as fundamental to a project as the data it will use can change over time. The LEHD and SNACC projects each provide examples. For LEHD:

Originally, the project planned to combine employer and employee data by linking them through W-2 records. Because the original design of LEHD could not be met, due to the lack of access to the W-2 data, the project shifted to a different source of data—Unemployment Insurance (UI) records.

As data sources changed, so too did other aspects of “goals” including the final users of the project’s output:

The project’s original design was one primarily for internal use at Census for survey improvement as well as to develop a research infrastructure . . . In order to acquire these [UI] records, it was necessary to develop a partnership with the States.

The identity of the final users changed; in turn, so did the “goal” of the data product that would result from the project:

Discussions with States started with three things that the project thought may be helpful to States: identification of Predecessor/Successor relationships, having their wage records cleaned, and measures of labor market dynamics at detailed geographic level and age/sex categories. The only one of the three that States were interested in was the last one.

The LEHD project—as it is known today—reflects the evolving goals of source data, final users and data products.

In contrast to the change in LEHD, which shifted from using national data to State-level, the SNACC project shifted from State-based to a national approach:

The project did change its approach early. At the time, Michael Davern [of SHADAC] was in discussions for a State-based approach to the research. However, Census already had an MOU with CMS, which had MSIS [administrative data for Medicaid]. The SNACC participants decided to go broad first and do work at a national level, and then do work with certain participating States.

The flexibility needed to develop a new MOU continued to be exhibited by the SNACC project as cross-agency dialogue continued:

At the development stage of the project, it took most of our energy to get a clear definition of the objective and to define a project that everyone could live with. It was important to make sure everyone was working on the same project. At first, people wanted to do everything and so there was a scope issue. The team made tough decisions without which the project would have derailed. The team had to answer the question of what objectives are the important ones.

Participants of the SNACC project reached essentially the same conclusion as the Profiles project on this second element of success:

While it is good to get a project to be well-defined up front, it is also important to leave some flexibility which can serve the project. A project can deviate from what was initially planned and still meet project goals.

The NASS/FSA case study, like the SNACC and LEHD case studies, also reported changes in the course of developing the MOU:

At another point, goals changed. There was hesitation about what would go into the MOU. There was concern of doing something new . . . There was some negotiation, if you could call it that.

In contrast to NASS/FSA, LEHD, SNACC, for which it was clearly stated that the projects had “change,” the Food Assistance case study seemingly makes a different assessment:

With more time and more projects, what has happened is that the goals have not changed, but the ability to do the research changes . . . [emphasis added]

This focus group comment shows how, as noted before, different participants used the term “goals” differently from one another, as synonyms for either strategies or tactics. Other examples that echo the Food Assistance case study are from the Incarceration and Education case study. For its NBER component:

The basic research objective did not change over time. We did enlarge the types of data examined and increased the number of States involved in the overall project because participants now reside in over 35 States. We had to consider the question, “Which States should we include in the project?” We are focusing on the States with the most project participants, and doing a random sample of other States.

For its UTD component:

The general research goals—conducting research to improve PreK-16 education—have remained the same from the start of the project. As additional research partners got involved, the specific set of research projects evolved.

A close reading of each of these passages actually confirms that “goals” can change over time—at least in the meaning of “goals” as used by the Profiles project. The quote from the Food Assistance case study that “goals have not changed” can be interpreted, using the language of the Profiles project, as an indication that there was a steady unchanging *vision* by the researchers of linking survey and administrative data for government assistance programs to learn about program participation, eligibility and outcomes. Participants in the Food Assistance case study conducted several related projects implemented over time with various partners on behalf of various funders. These academic researchers discussed six distinct projects, each with its own project-specific goals and activities. At the *project* level—as opposed to the *research program* level—“goals” necessarily changed from project to project. Similarly, for the NBER project work the “goals” changed over time on the “types of data” and the States to be included. Thus, different participants in the case studies interpreted the notion of “goals” differently, sometimes in a broad sense for which the Profiles project substitutes the term “vision.”

Importantly, change does not necessarily mean compromising on an initial vision and settling for less. On the contrary, sometimes changes in goals augment the project. As projects move forward and develop, participants may learn of opportunities they did not know about before. The NTID/SSA/Cornell project provides an example:

Gerry Walters began the project way back in the 1990s. He wanted to work with IRS data. That’s how it started. So the goals are similar. It started as an interesting research question—getting information on earnings after students leave here. Gerry and others didn’t envision having SSA data available.

Thus, the possibility of using SSA data enhanced the project’s goals. With time, the involvement of SSA as a partner increased as well:

SSA became more interested in this project after some findings on the effect that graduation from NTID had on those who participated in the Supplemental Security Income (SSI) program. These findings showed that NTID students spent less time on the SSI program as adults . . . Now researchers at SSA, Cornell, and NTID all are involved.

An interesting feature of the NTID/SSA/Cornell project is that, with time and access to SSA data, the research design became an example of program evaluation where, in this instance, the “program” is NTID itself:

At the beginning . . . OMB did not have any interests as these have only come up in the last 2 or 3 years now that OMB is trying to exercise more control over funds. The initial intent was strictly to respond to Congress . . . So, the main project purpose remained the same—good use of taxpayer dollars—but the emphasis evolved from Congressional oversight concerns to OMB’s interest in program evaluation.

Thus, the project’s methodology and findings for NTID and, perhaps, its potential for serving as an example for other program evaluation initiations have increased interest in the project at the federal level.

The Industry Coding project provides another example in which the range of project activities grew with time as staff at the Census and BLS developed an ongoing data-sharing relationship.

Although the discussion is usually in terms of industry coding, BLS compares and returns to Census ownership code, multi-establishment employment indicators (MEEI), and physical address information. So the activity goes beyond industry coding. There can be a distinction between the goals as originally intended and what is occurring now. Without the support of BLS . . . these additional uses would not have been identified. So the goals changed over time.

Thus, the Industry Coding project’s goals evolved over time to broaden data sharing and the uses to which the available data are put. Success begets success.

As the Census Bureau became more accustomed to the data, Census people started to look for more uses. That is where the use of physical location and employment indicators came from. Their use came as a by-product of the two agencies working together . . . Over time, there was a creep of goals and of benefits for the project, which was a good thing.

So data-sharing projects are not only about pursuing known possibilities. Sometimes they generate information about further data-sharing possibilities that were not previously known. The art of developing and managing these projects involves, in part, creating cross-agency relationships and environments of mutual trust and support. In such environments, undetected opportunities become detectable. An opportunity must be known before it can be seized.

Element 3. Infrastructure

Summary. One component of infrastructure is staffing. Cross-agency projects benefit from people who are results-oriented, experienced, and able to work cooperatively. Both the cumulative hours of staff time and the calendar time it took to organize a project were often

underestimated. A second component of infrastructure is policies and procedures appropriate for data-sharing activities. For laws and regulations that are established outside the agency, the agency may be responsible for providing or obtaining a legal interpretation of how the laws and regulations on privacy, confidentiality, and data stewardship apply to its activities. An agency may need to create, implement and manage the process by which the MOU-clearing process will be developed. Management may not know a priori the relevant laws and regulations, the technical requirements, or the data-sharing project's specific goals. Protecting data integrity and confidentiality are at stake. Ensuring that an agency's policies and procedures are conducive to data sharing is a management activity of paramount importance.

A key managerial function is ensuring that appropriate infrastructure is in place for conducting a project. Two components of project infrastructure discussed below are staff and the policies/procedures environment in which staff work. Physical infrastructure includes buildings, computers, and software by which data are actually coded and shared. Physical infrastructure did not seem to be a large barrier across the Profiles project's case studies, and is omitted from most of the discussion here; other barriers to project success were more salient. Managers recognize that they assign staff to projects. They do it all the time. There are nuances, though, whenever staff are assigned to a cross-agency project involving administrative records. In addition, the policies/procedures environment has issues specific to administrative records that need to be resolved by management to support developing an MOU.

Infrastructure Component 1. Staff. The main organizers and participants of a data-sharing project may be at various places and positions in an agency's organizational chart. In this discussion on staff, the Profiles project will refer to all of these participants as "staff" regardless of whether they are members of agency leadership or high-level management or staff. In any case, decisions about how a person's time and talent are allocated across an agency's workload are being made. Whether the decision-maker is a supervisor or the person himself or herself, there are important points about data-sharing projects and time requirements.

Amount of staff time. There are five major points about the amount of staff time required for data-sharing projects:

- 1) In terms of staff time and expertise, developing a data-sharing MOU is "costly" or "resource-intensive."
- 2) In the past, staff cost has been "high" compared to expectations—that is, historically the amount of staff time required has been systematically underestimated.
- 3) While staff cost may be "high" compared to expectations (at least historically), staff cost can be "low" compared to two other standards:
 - a) the project's benefits (i.e., the data-sharing project has benefits that exceed costs even though costs may be "high"), and
 - b) alternative means to obtaining data (e.g., sharing administrative records may be less costly than primary data collection through a new survey).
- 4) The amount of staff time required will almost certainly be lower in the future than in the past, at least for agencies that have already formed a data-sharing MOU.
- 5) The amount of staff time required for an agency new to data-sharing may be lower now that there are successful projects from which to learn.

It can be helpful for project planning to be aware of two aspects of “costly” staff time. First, the familiar aspect of the simple cumulative “hours” of staff hours (for an individual participant or for a project in total) can be high. Second, even for periods of the project in which hours of staff time are modest, the amount of *calendar* time can be long in which the cost is the cost of waiting. These two aspects provide a distinction by which a “few hours” of staff time that are stretched across several months or quarters is simultaneously a “little” time and a “long” time.

The case studies in the Profiles project exhibited both a “high” amount of staff time and a “long” amount of staff time:

It was hard to figure out what States wanted. There were a variety of possible measures that could be developed. The project iterated on what measures States would be interested in. The project made presentations to State Workforce Investment Boards and State Labor Market Information Agencies for almost every State. Three States took leadership on the project: Illinois, Florida, and California. [LEHD]

It took hundred of hours with lawyers [across the States] to go through the process. [LEHD]

Some staff from some agencies, such as Census and CMS, put in a lot of time that was not funded. [SNACC]

The process took a long time to be able to deal with confidentiality rules, etc. [SNACC]

The time issue is a big one. The time it takes to get a project in place to do the research is the biggest part of the project. Altogether, the MOU, funding, and actually getting the data in shape to be worked on takes a huge amount of time—more than expected. [Food Assistance]

One often underestimates both the time it takes to do this work and the timing of the work. [Food Assistance]

Even with a previous agreement in place, it can take two years even to renegotiate a data sharing agreement. The amount of time depends on who the contact person is and whether they want to see the agreement go through or see it as a hassle. Do they want it? Or do they not see the agreement as something they have the time to do? Is it a priority for them? We're lucky that a two-year period is not the usual case. It can take a year, though. For example, it takes time to simply get on the office calendar of the District Attorney. In another case, it took over a year to renegotiate some agreements with prior providers—and that's with an active contact person. We hope for six months, but some agreements drag on and on. In general, it is good to leave a long lead time to increase the odds of getting agreement. [Incarceration and Education]

In some cases, a “long” amount of time can be evidence of success. As noted in the second element of success, project goals that evolve over time can result in better, more ambitious projects that expand to use available opportunities. If this happens, it should be welcomed that a project lasts a “long” time—the project generated results that weren't anticipated at the

beginning. If expanded goals are added as additional “phases” of a project or, even more strongly, are classified under the “same” original project title, then long calendar times are a sign of success rather than problems.

After completing the case studies, the Profiles project considered why management tended not to allocate sufficient resources to data-sharing projects. While an under-staffing problem can easily result from unanticipated problems, focusing on “unanticipated problems” as an explanation seemed incomplete. After all, any new project has unanticipated problems. Wouldn’t the likelihood of unanticipated problems be taken into account for allocating enough resources to deal with unforeseen contingencies? What makes administrative records projects different from other projects? The explanation seems to be that these data-sharing projects are examples of innovation. The projects reviewed here were “first-generation” projects and simply so new, so outside of the usual routine and work plan, that the full extent of the necessary time was hard to anticipate. The technical and legal challenges grappled with by the case studies had not been considered before.

If a subsequent project were undertaken, it could speed much faster. This project was breaking new ground. [SNACC]

The projects discussed on this call were being conducted while Census was in the process of revamping its data stewardship policies and procedures. These projects were on the leading edge and unfortunately had to jump some hurdles, wait for decisions to be made, and processes to develop that new projects will no longer encounter. [Food Assistance]

For the first generation of data-sharing projects, the technical complexity, the cross-agency aspect, and their legal implications formed a perfect storm of challenges that regularly increased staff time above the expected level. Now, in 2008, there is more information for future projects on how to be successful than there was available for the projects in the case studies. Future projects—the ones that are under development now and those that are on the cusp of being imagined—will be a “second generation” of projects that can learn from the first generation. Moreover, infrastructures of policies and processes have been put in place, in Census and elsewhere, that represent innovations for their agencies and models for other agencies. While new projects will still constitute “innovations” for the agencies that begin them, hopefully innovation for second-generation projects will be easier than for first-generation projects.

The third of the five points listed above points out that even though costs have been higher than expected it does not necessarily follow that data-sharing projects have been or will be bad ideas. The relevant cost considerations for project development or evaluation is not whether costs are high compared to expectations but whether costs are high compared to project benefits or to other means of acquiring results (such as a new survey). Of course, whether or not any particular project has a favorable benefit-cost analysis is up to the people closest to the project to decide.

The fourth point above conveys that renewing an MOU should be easier than forming an MOU from scratch. Several case studies pointed to the benefits of previous relationships.

If someone wants an MOU with FSA to get access to a detailed data file, it is important to know if they want access to all the data. There are start-up costs to get someone familiar with the data. A more limited request is easier . . . The bottom line lesson is that NASS had familiarity with FSA data so the data-sharing wasn't a huge resource drain on FSA . . . Another lesson learned is the importance of having a working relationship in place. [NASS/FSA]

The preexisting relationships helped tremendously. They helped establish credibility. There were fairly well-established procedures. I didn't spend a lot of time explaining who I was. [Incarceration and Education]

There are differences in how easy it is to work with an agency [that houses criminal justice data] depending on whether the relationship is ongoing . . . or whether the agency is new to the project. For older agencies, when we contact them, we can say, "You are an agency that has worked with the project before." These agencies understand who needs to sign off. [Incarceration and Education]

The project was in place at SSA in the early 1990's, and the MOU was renewed in 2004. In 2004, the MOU renewal process was fairly straight-forward. [NTID]

Fortunately, CMS already had a relationship with Census. The project was not starting at ground zero. [SNACC]

Chapin Hall, RMC, and JFI each have worked with agencies in their States on projects besides the ones under discussion. RMC has worked with Texas agencies since 1980s, as has Chapin Hall in Illinois. [Food Assistance]

When planning for renewing an MOU, it would be good to recognize that the expectation of "renewing is easier" is not the same expectation as "renewing is cost-free." Renewing is precisely the time at which MOUs get reevaluated, especially by the legal team and especially if there are new activities involved or new legislation has been passed. Even if not renewing an existing MOU, agencies that have already completed data-sharing MOUs should have an easier time completing the next one compared to an agency developing its first MOU.

The fifth and final point listed above provides some relief to agencies considering developing their first data-sharing MOU: "The amount of staff time required for an agency new to data-sharing may be lower now that there are successful projects from which to learn." That is, costs should be lower than for the past projects. Several data sharing projects have been completed. These premise of the Profiles project is that agencies can learn from successes. There is better information now at hand than at the outset of the first-generation data-sharing projects.

One aspect of "better information" that may be helpful for designing a new project is considering long-term planning for staff through incorporating expectations and into responsibilities and job descriptions In SNACC:

Agencies rarely designate one person to promote data sharing projects who is versed in the development of the complex interagency agreements that must accompany these developmental projects. What is needed is to put this activity into a job description. That way, even if that person leaves, the functions will continue with the next person in the position.

Similarly, for NTID:

When SSA agreed to do the project, they agreed to devote resources to it in the MOU. So, this project is actually part of official job duties . . . The challenge [at NTID] of maintaining the project is simply the lack of time, juggling too many things at once. Twenty other things are calling. In terms of priorities, this project is clearly specified as a responsibility of mine. Supervisors are very excited by it. It is a crucial project for them.

Expertise of staff. Besides the amount of staff hours to devote to developing MOUs, a supervisor can consider the expertise and personal qualities of the assigned staff. As noted, data-sharing MOUs are complicated, and will involve many persons before they are done:

Everyone brought different expertise to the project . . . It was essential to bring others at certain point for substantive expertise to help the process. [SNACC]

Besides the quantity and variety of staff skills, there is personal quality—persistence—that was mentioned in some of the case studies:

A lesson learned is that persistence really counts. We have been doing this over ten years. Just in the last couple of years has the project reached a point of presentation and publication. It takes a long time to get there. But the project has really been worth it. Persistence is key. It is not a quick and dirty project, not even one that takes one or two years. [NTID]

One lesson learned is it takes considerable time and persistence. It is a challenge to identify the right agency contact to handle a data-sharing agreement. [Incarceration and Education]

People on these projects—even when staffing is closer to sufficient than in the past—will be working hard for there to be success. They will need to be goal-directed and focused. In an ideal world, they will be someone who already has experience in the field:

For these projects, everyone is gaining new competencies and new ways of working together. There is a steep learning curve the first time a researcher works with administrative data, but then the researcher has a better set of expectations and the work gets easier. It is possible to estimate costs better, which questions to ask, and which things various parties agree to and which things they don't. For the later [of the six] projects, it was clearer what questions to ask. [Food Assistance]

A second hard-to-identify aspect of staff is their organizational culture, which describes an agency's staff collectively rather than individually. Organizational culture can be hard to

recognize and harder to change, but it is important because it can affect the success of the project. Because administrative records projects are so new, more so in some agencies than others, one aspect of organizational culture is how receptive staff are to change “in general.” More specifically, staff in a survey-oriented culture may not fully appreciate the value of an enhanced role of administrative data and staff in a program agency may not fully appreciate the value of sharing administrative records with a survey agency. In particular, the perception by survey staff that the use of administrative records indicates that a survey is “deficient” can make staff defensive. The Profiles project recognizes that administrative and survey data each have strengths and limitations, but that view might not be shared or shared widely in the organizational cultures of the data-sharing agencies.

The specialized expertise that lawyers bring to the process of reviewing MOUs has been appraised differently by different participants in the case studies. One view is that lawyers make the process longer and worse, and the other is that they make the process longer and better. Of course, if there were no legal review, the process would be shorter almost by definition. So the question is whether the extended length generates benefits that are worth the delay. From the point of view of data stewardship, legal review can be critical to identifying needed safeguards that sustain public confidence and continuation of these data-sharing agreements. This second perspective arose in the Food Assistance case study:

It is a waste of time for researchers to do the MOUs. Researchers are not good at it, and then they wonder why it does not happen. Researchers want to handle data. They won't make the right decisions. They are too expeditious. With MOUs, everyone needs to be represented properly. Lawyers need to run the process.

In the end, some attention simply must be given to legal details. It is noted that lawyers within States may necessarily be involved when State-level administrative records are to be shared: pertinent statutes vary across States.

In summary, cross-agency projects benefit from people who are results-oriented, supportive of the project's goals, experienced with the data, and able to work cooperatively with people in their agency and the partner agency.

Infrastructure Component 2. Policies and Procedures. It may seem unusual for the Profiles project to consider policies and procedures to be an aspect of “infrastructure.” But policies and procedures resemble physical infrastructure in several ways. They are applied or used across different projects instead of developed project-by-project—giving policies and procedures longevity, like physical infrastructure, instead of being re-promulgated anew each Monday morning. They are part of the background or environment in which staff conduct projects. Like physical infrastructure, staff may take for granted the existence or the particular forms of policies and procedures.

Agency leadership, unlike staff, have responsibility for policies and procedures—as opposed to laws and regulations which, for discussion's sake, we take to be determined outside the agency. Even for laws and regulations that are established outside the agency, the agency may be responsible for providing a legal interpretation of how the laws and regulations apply to its

activities, or at least be responsible for contacting the appropriate legal authority for that interpretation.

A subtle threat to the success of a data-sharing agreement may be posed by an agency's policies and procedures. Agency leadership or upper management may make the mistake of thinking of policies and procedures in the same way that staff do: long-lived, applicable across projects, something taken for granted. Yet, completing a data-sharing MOU may call for non-routine decisions by non-routine people. It is a managerial responsibility to consider whether the policies and procedures an agency has in place are conducive to developing a data-sharing MOU, which is a different animal from an agency's routine MOUs. Established procedures may not cover all the steps in the process by which a data-sharing MOU will be cleared. If not, then management will need to create from scratch, or at least close the holes in, its MOU-clearing process. Actually, management may need to create, implement and manage the process by which the MOU-clearing process will be developed. While it may seem excessively bureaucratic and finely detailed to worry about having the right "process to create a process," management may not know *a priori* the relevant laws and regulations, the technical requirements, or the data-sharing project's (narrow) goals—in short, just who needs to be involved in either clearing a data-sharing MOU or in creating a clearance process in the first place. Protecting data integrity and confidentiality are at stake. Ensuring that an agency's policies and procedures are conducive to data sharing is a management activity of paramount importance.

A partial list of the many details involved with policies and procedures arose in the Incarceration and Education project:

Part of the process for conducting the project is to meet the different requirements of different states. We go back to States: "Does this meet your requirement?" Our own system is a non-network machine. State agencies are used to storing on mainframe. There is also an issue of how to archive data. In addition, in some states, but not others, each person signs a form. There are issues about transmission. We prefer encrypted files delivered by overnight carrier. Different people have different views about the best way to transmit. Some data are transmitted via secure File Transfer Protocol.

Sometimes, this infrastructure component of policies and procedures must be built with considerable time and effort rather than simply adapting close precedents that may not exist. For example, for the data-sharing agreements between Census and States in LEHD:

There were many questions for the first two or three States. The MOUs with Florida, Illinois and California each look a little different from one another. Then there was convergence to a standard. Ronald Prevost developed an MOU template that could work with all States.

Also for LEHD:

[At the time] Census was developing a whole new data stewardship structure, including a bunch of new policies. Trying to navigate a maelstrom of policies being developed was challenging. The project made it work by taking each specific policy as it was developed, and tried to run LEHD scenarios to see whether it worked. . . . Within Census, review was a

challenge because the level of infrastructure was not there. Census was accustomed to bringing in one or two different [data] sources at a time. With twenty different sources, Census needed to set up a process

The Incarceration and Education project provides an example by which infrastructure was developed by legislation.

TEA's interpretation of the original agreement—that the data were to be used by John Kain for his research and that upon his death the data need to be destroyed—presented a major barrier to the continuation of the project . . . the underlying issue of whether the data had to be destroyed was resolved by legislation. Texas passed legislation establishing Education Research Centers (ERCs) within the State (of which the UTD Center is one) that would conduct research for purposes that would benefit the State.

The Incarceration and Education project also showed how infrastructure, in the form of a standardized agreement, can be designed to facilitate data access:

The State of Washington has an interagency agreement or data-sharing process. I am a State employee at the University of Washington. So the interagency agreement process was a pre-existing arrangement I used. I gave the grant proposal to the agencies. At two agencies . . . they were very research-oriented and familiar with the interagency agreement process.

Thus, it can be a managerial responsibility within an agency and a political responsibility for a State legislator to consider how the policies and procedures either promote or inhibit data sharing.

Element 4. Mutual Interest

Summary. To reach successful conclusion, data-sharing arrangements benefit the various partners to the MOU. Each agency has its own set of statutory and policy requirements to protect the confidentiality of its own data. An agency that hopes to receive data is advised to recognize that its data-supplying partner is in a different environment and that the details of that environment are known better by the partner than the data-receiving agency. Developing an awareness and sensitivity to each other's policies and rules is achieved through cross-agency negotiation and discussion. Together, the partners can discern if an MOU is possible that can accomplish something that serves the mutual interest of both agencies.

The notion of the importance of “mutual interest” is simple yet profound. When various parties are considering entering an agreement—be it a commercial contract or an interagency MOU—there will be no finalized signed agreement unless it is in the mutual interest of both parties. While that point might seem obvious, it contains a subtle aspect that bears emphasizing:

Principle of Mutual Interest: When two parties sign an agreement, the act of signing shows that the agreement is considered to be in the mutual interest of each party—that is, the benefits of the entering and executing the agreement outweigh the costs—from the *separate distinct perspective* of each.

The Principle of Mutual Interest explains that each party evaluates a proposed agreement from *its own* position or perspective, with all constraints and incentives involved in that agency's environment and position. A way of appreciating the principle is to consider a potential MOU between an agency that would like to receive data and an agency that collects the data. What prompts the data-collecting agency to enter the MOU and to share the data with the (potential) data-receiving agency is *not*, in the final analysis, whether the data would benefit the data-receiving agency. Of course data access benefits the data-receiving agency—that is why the data-receiving agency is issuing a request for the data in the first place. That story is only half—one side—of the Principle of Mutual Interest. The second half is the subtle one. What the data-receiving agency could fail to realize is that the data-collecting agency will enter the MOU only if the benefits of the MOU outweigh the costs from the *data-collecting agency's* point of view. The actual point of view of the data-collecting agency is *not* necessarily the same as what the data-receiving agency thinks its partner's point of view "should be." Cross-agency negotiations can break down from not seeing the world through the eyes of your partner.

An agency's constraints were recognized by Cox, Berning and Martinez (2006) in a study of the SNACC project: "The fundamental basis for the policy and legal issues in linking source files from different organizations is that each organization has its own set of statutory and policy requirements to protect the confidentiality of its own data." Thus, an element of success is for Agency A to recognize that Agency B has a *different* environment that can be shaped by statutes and policy requirements that are different from Agency A's. The specific differences may not be known at the start of cross-agency work, in part because neither agency may have had much reason to learn about the other's environment. It is these differences that the agencies consider in the course of developing an MOU for a project, and for an MOU to be finalized the differences need to be recognized, respected and accommodated.

The lesson of the Principle of Mutual Interest is especially pertinent for cross-agency projects that involve administrative data because these are innovative activities. As a result, some legal issues, such as who "owns" or is "responsible" or "steward" for integrated data, need to be ironed out for the first time. As noted in Cox et al: "As we worked to define and document the process to transfer confidential data collected in the NHIS to the Census Bureau for linking to CMS data, we noted that each agency had well-defined data stewardship policies designed to provide the highest level of protection for the data collected by that agency. However, these processes did not necessarily lend themselves to the provision or receipt of data between our agencies and the transfer or rights and responsibilities that accompany data integration between to such Federal agencies." Resolving such complexities depends on mutual respect and working towards mutual interest. Despite such challenges, SNACC's process proved successful:

The project had a high degree of sensitivity to each other's policies and rules. Staff from each agency knew their own agency's requirements. The team knew that each agency's requirements differed, and the team wanted to satisfy each agency involved. this was a priority throughout the project . . . It is always an issue of how to bring a group together as a team to be on the same page. It took time to build the team. It took many iterations of the project plan . . . In the end, people recognized that there was a need to balance everyone's

interest. People conceded that elements of the project might not be the top of one's own priority list, but people were still willing to work on it. The project ended up balanced.

A comment provided from the Education and Incarceration projects is consistent with the Principle of Mutual Interest:

[Stage agencies] are interested in how their participation [in a data-sharing project] benefits them.

LEHD too was successful because careful attention was given to products that would benefit the States that were providing the data. As noted above:

Discussions with States started with three things that the project thought may be helpful to States: identification of Predecessor/Successor relationships, having their wage records cleaned, and measures of labor market dynamics at detailed geographic level and age/sex categories. The only one of the three that States were interested in was the last one.

One last example of the importance of mutual interest is that the principle applies not only at an agency level but also a personal level. When contacting people, it can be important to recognize the position and perspective of that individual:

It is helpful to distinguish between the concerns of the agency from the personal concerns. Who is the contact for the agency? At the personal level, is the project too much work? Or is the person facing deadlines right now, and if we call back in a couple months they would be more receptive? [Education and Incarceration]

6. Summary and Conclusions

This report reviewed reasons for sharing administrative data between agencies for statistical purposes. It explained the particular motivations and topics behind seven case studies that involved cross-agency MOUs for sharing administrative data. The case studies were conducted using focus groups composed of a project's original participants convened by conference calls. The challenges they described for the projects, and the means by which these challenges were met, became the basis for several themes that emerged across the case studies. These emerging themes took the form of four "elements of success" that can facilitate the development of a data-sharing MOU.

The elements of success described in the report are: vision and support by agency leadership; narrow but flexible goals; appropriate infrastructure; and mutual interest.

- **Vision and support by agency leadership** is the element that assembles, organizes, prompts, motivates and facilitates all of the other elements of success. Cross-agency data-sharing projects are too new and complicated to happen automatically. An MOU may be unlikely to be drafted for signature—that is, the project will be abandoned along the way at the development stage—if agency leadership and management do not provide follow-up support after the initial broad approval for project development. Agency leadership can make these difficult projects a priority for the agency,

assign appropriate staff, and ensure that policies and procedures are in place that are appropriate for data sharing, especially on data stewardship issues.

- **Narrow but flexible goals** are the second Element of Success. It can be important for the MOU to specify goals narrowly—sometimes down to the level of which particular fields in a database will be shared between agencies and how those fields will be used. Narrow goals can be conducive to cross-agency discussions that take into account both technical feasibility and data stewardship issues. Goals can also be flexible, changing as a result of interagency discussions. In the successful data-sharing projects we studied, when some initial goal encountered a major barrier, the goal shifted to adapt. Flexibility also permits goals to be added over time as new uses of data come to be identified as a by-product of the two agencies working together. The art of developing and managing these projects involves, in part, creating cross-agency relationships and environments of mutual trust and support. In such environments, undetected opportunities become detectable. Goals that are flexible can result in expanded activities as initial success begets more success.
- **Infrastructure**, the third element, has two components: staffing, and policies and procedures. Cross-agency projects benefit from people who are results-oriented, experienced, and able to work cooperatively. While the projects we studied had capable people, the projects were consistently understaffed perhaps because they were innovative “first-generation” projects and simply so new that the full extent of the necessary time was hard to anticipate. The second component of infrastructure is appropriate policies and procedures to support data sharing. For laws and regulations that are established outside the agency, the agency may be responsible for providing or obtaining a legal interpretation of how the laws and regulations on privacy, confidentiality, and data stewardship apply to its activities. An agency may need to create, implement and manage the process by which the MOU review will occur. Management may not know *a priori* the relevant laws and regulations, the technical requirements, or the data-sharing project’s specific goals.
- **Mutual Interest** is a fourth Element of Success. Each partner agency has its own set of statutory and policy requirements to protect the confidentiality of its own data. Developing an awareness and sensitivity to each other’s policies and rules is achieved through cross-agency negotiation and discussion. Together, the partners can discern if an MOU is possible that can accomplish something that serves the mutual interest of both agencies.

The report’s distillation of the experiences of the “first generation” of data-sharing projects is meant to make future data-sharing projects easier to develop. While these types of projects are not routine, especially for agencies that have not done any data-sharing previously, the benefits of data-sharing can merit the cross-agency efforts such projects require. Data-sharing can be time-consuming and resource-intensive. At the same time, though, it can be a *relatively* low-cost route to certain goals. For example, the costs of collecting certain administrative data and survey data may have already been incurred by two agencies. The additional cost of merging the two datasets can be less than the cost of new primary data collection or the cost of trying to capture the rich array of survey data in an administrative database.

Any innovation can be a challenge. By striving to meet such challenges, the agencies that develop a data-sharing partnership can be serving their respective agency missions as fully as possible.

References and Selected Research Products of Case Study Projects

Congressional Research Service. *Medicaid/SCHIP Enrollees: Comparison of Counts from Administrative Data and Survey Estimates*. Memorandum. Peterson, Chris L., and April Grady. Updated March 30, 2005. Available at http://www.aei.org/docLib/20050407_CRSMemoonMedicaidUndercount.pdf

Congressional Research Service. *SCHIP Original Allotments: Funding Formula Issues and Options*. Peterson, Chris L. April, 2006.

Cornman, Stephen, Howard Bradsher-Fredrick, Carol Comisarow, Christine Cox, Mark Denbaly, Rochelle Wilkie Martinez, William Sabol, and Michelle Vile. *Model Data Use Agreement for Sharing Administrative Records: Recommended Components Annotated With Examples*. 2009. Forthcoming at <http://www.fcsn.gov>

National Research Council. *Principle and Practices for a Federal Statistical Agency, Third Edition*. Committee on National Statistics. M.E. Martin, M.L. Straf, and C.F. Citro, eds. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press. 2005.

SNACC Case Study

Cox, Christine S., Michael A. Berning, and Rochelle Wilkie Martinez. "Data Policy and Legal Issues in Creating and Managing Integrated Data Sets." Proceedings of the 2006 Federal Committee on Statistical Methodology Conference. (2006)

Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, and the U.S. Census Bureau. "*Phase I Research Results: Overview of national Medicare and Medicaid Files.*" *Final Report*. February 2007. Available at: <http://www.census.gov/did/www/shadac/shadac.html>

Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, The National Center for Health Statistics, and the U.S. Census Bureau. "*Phase II Research Results: Examining Discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC).*" *Final Report*. March 2008. Available at: <http://www.census.gov/did/www/shadac/shadac.html>

Research Project to Understand the Medicaid Undercount: The University of Minnesota's State Health Access Data Assistance Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, The National Center for Health Statistics, and the U.S. Census Bureau. "*Phase III Research Results:*

Refinement in the Analysis of Examining Discrepancies between the National Medicaid Statistical Information System (MSIS) and the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC)” Final Report. July 2008. Available at:
<http://www.census.gov/did/www/shadac/shadac.html>

Research Project to Understand the Medicaid Undercount: The University of Minnesota’s State Health Access Data Assistance Centers for Medicare and Medicaid Services, the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, The National Center for Health Statistics, and the U.S. Census Bureau. “*Phase IV Research Results: Estimating the Medicaid Undercount in the National Health Interview Survey (NHIS) and Comparing False-Negative Medicaid Reporting in the NHIS to the Current Population Survey (CPS). Final Report.* April 2009. Available at:
<http://www.census.gov/did/www/shadac/shadac.html>

LEHD Case Study

On the Map and other LEHD data products are available at:
<http://lehd.did.census.gov/led/index.html>

Food Assistance Case Study

Lynch, Victoria, Dean M. Resnick, Jane Staveley, and Cynthia M. Tauber. *Differences in Estimates of Public Assistance Reciprocity Between Surveys and Administrative Records.* Washington DC/Baltimore MD: Bureau of the Census, Jacob France Institute. January 2008. Available at:
<http://www.ubalt.edu/jfi/jfi/reports/TANFJan2008.pdf>

Tauber, Cynthia, Dean Resnick, Susan P. Love, Jane Staveley, Parke Wilde, and Richard Larson. *Differences in Estimates of Food Stamp Program Participation Between Surveys and Administrative Records.* Washington DC/Baltimore MD: Bureau of the Census, Jacob France Institute, U.S. Department of Agriculture, and the Maryland Department of Human Resources. June 2004. Available at:
<http://www.census.gov/acs/www/Downloads/ACS/FoodStampFinalReport.pdf>

NTID/SSA/Cornell Case Study

Weathers II, Robert R., Gerard Walter, Sara Schley, John Hennessey, Jeffrey Hemmeter, and Richard V. Burkhauser. “How Postsecondary Education Improves Adult Outcomes for Supplemental Security Income Children with Severe Hearing Impairments.” *Social Security Bulletin* 67(2): 101-131, 2007. Available at
<http://www.ssa.gov/policy/docs/ssb/v67n2/v67n2p101.pdf>