Session 7

Benefits and Stewardship of Linked Survey and Administrative Data

Data Stewardship and Accountability at the U. S. Census Bureau[13]

Nancy A. Potok and Gerald W. Gates
U.S. Census Bureau

Introduction

Statistical agencies have long recognized the fundamental tension between their mandate to provide high-quality data that informs sound research and public policy development and their requirement to protect the privacy and confidentiality of their respondents. These dynamics often operate at odds with one another, as demands for richer data products face off against increasing public concerns about privacy, the increased availability of personal information on the internet, and newer, cheaper desktop data processing capability. However, a statistical agency's reputation for respecting privacy and confidentiality is critical to maintaining high response rates and, thus, the quality of its data.[14] The U.S. Census Bureau's mission to be the "preeminent collector and provider of data on people and the economy of the United States," requires that this tension be balanced successfully.

The Census Bureau's legal mandate, Title 13 of the United States Code, authorizes the collection of data, but it also establishes strict requirements for maintaining the confidentiality of data collected from its respondents. Indeed, the Census Bureau may not publish data about a particular establishment or individual that allows them to be identified. Even when the Census Bureau requires expert consultation from outside the agency, such experts are not permitted access to the data unless they are brought on as "Special Sworn Status" individuals[15] – effectively temporary staff – who are sworn to uphold the Census Bureau's confidentiality standards. Criminal penalties, specifically up to $250,000 in fines and 5 years imprisonment, further help to create an environment intolerant to such disclosures. Given the agency's strong legal mandate and ethical commitment to privacy and data confidentiality, how does it ensure that collected data result in useful, relevant and timely products?

[13] This paper has undergone a review more limited in scope than that given to official Census Bureau publications. It is released to inform interested parties about the Census Bureau's data stewardship approach to balancing confidentiality protections while providing quality data and to encourage discussion of these important issues.

[14] See Pat Doyle, Julia I. Lane, Jules J.M. Theeuwes, Laura V. Zayatz, Eds., Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies for a series of discussions on the tension between data access and confidentiality.

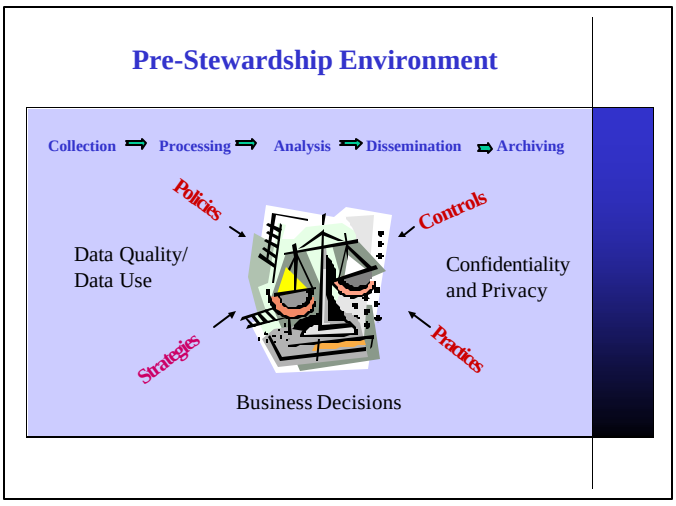
[15] Title 13 United States Code, Section 23(c) provides for the Census Bureau to "utilize temporary staff, including employees of Federal, State, or local agencies or instrumentalities, and employees of private organizations to assist the Bureau in performing the work authorized by this title," but only if such temporary staff is sworn to observe the limitations imposed by section 9 [which establishes confidentiality provisions].

A sound *data stewardship structure* within which such issues can be weighed provides a forum where the Census Bureau's can make balanced business decisions – data quality and access on one side of the scale and privacy and confidentiality on the other. The concept of "stewardship" is borrowed from environmentalists – the objective being to create a *sustainable balance* that supports one's needs over the long term.

Establishing a Basic Data Stewardship Structure

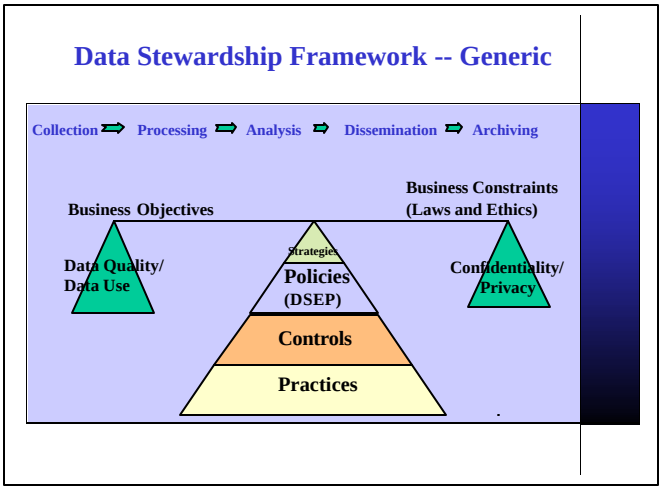
While data stewardship principles may exist, they are not always well coordinated or integrated, and/or they are applied in an ad hoc manner, depending on the particular circumstances involved. **Chart 1** demonstrates how business decisions that affect data-related operations -- collections, processing, analysis, dissemination, and archiving -- can become unbalanced and lose a corporate focus when there is no integration of strategies, policies, controls or practices, or they are not used systematically to make business decisions.

Chart 1 --



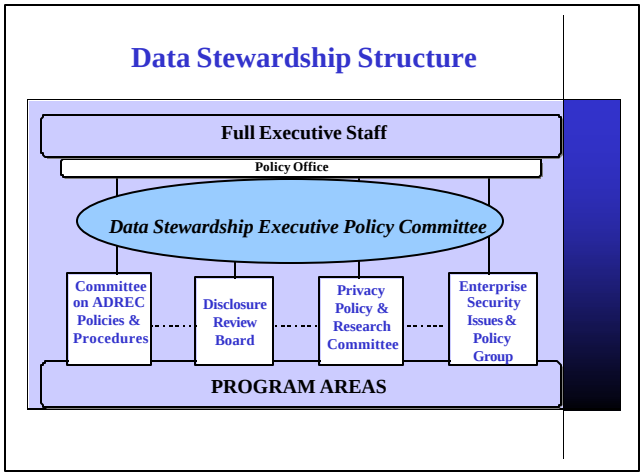
If strategies, policies, controls, and practices are fully integrated, the organization has a better chance of ensuring that business decisions will lead to the desired outcome. **Chart 2** illustrates how an otherwise *ad hoc* approach can be stabilized, achieving balance between business objectives and constraints. This better supports the data related operations.

Chart 2 --



The Census Bureau annually updates its 5-year strategic plan and communicates its strategic goals to employees and external stakeholders. In June 2001, the Census Bureau moved to address policy issues more consistently by establishing the *Data Stewardship Executive Policy (DSEP) Committee*. The DSEP Committee is composed of top bureau executives who are charged with identifying and developing policy issues related to data stewardship. This executive decision-making body is staffed by the Policy Office and supported by the analyses and recommendations of four DSEP staff committees: the Committee on Administrative Records Policy and Procedures (CARPP), the Disclosure Review Board (DRB), the Privacy Policy and Research Committee (PPRC), and the Enterprise Security Information and Policy (ESIP) Committee (see **Chart 3**).

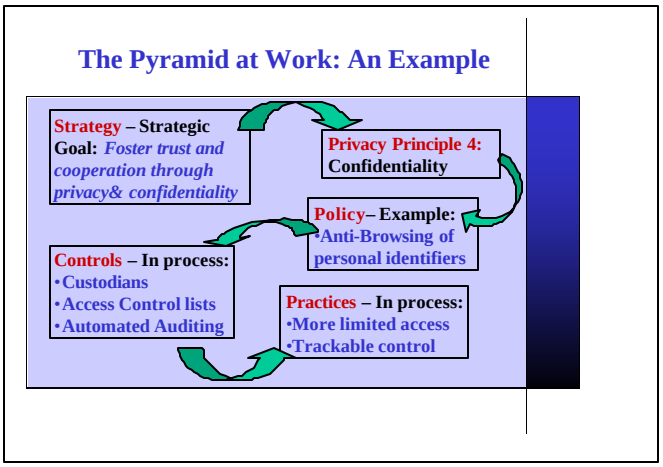
Chart 3 --



One goal of the DSEP Committee is to ensure that strategic goals, corporate ethics, policies, controls, and operational practices are integrated and consistent. This means that strategic goals are shaped by corporate ethics and drive policies. Policies in turn drive the creation of organizational controls, and these controls incorporate practices that ensure compliance. For

example, as shown in **Chart 4,** one of the Census Bureau's strategic goals is to *foster trust and cooperation through privacy and confidentiality.* In support of this goal, the Census Bureau developed a set of ethical standards called Privacy Principles, one of which is *Confidentiality.* This Privacy Principle resulted in the Census Bureau adopting a policy *prohibiting the browsing of records with personal identifiers by employees* and others who may have access to those records. The Census Bureau is currently working to establish *access control and auditing procedures,* such as identifying data custodians in each division responsible for monitoring access to personal identifiers. The result will be that fewer employees will have access to sensitive records, and those that do will have all their interactions with the data tracked and monitored by an automated audit system.

Chart 4 --



The DSEP structure has been successful in systematically establishing policies and procedures in several key areas. Accomplishments include the release of an *Administrative Records Handbook,* and documenting procedures for the negotiation, acquisition, access, and use of administrative record data. The DSEP Committee also has finalized a policy on appropriate data access and use for non-employees with Census Bureau Special Sworn Status. It is currently completing an analysis of how well existing policies support the Privacy Principles.

While the primary responsibility of the DSEP Committee is to serve as the policy-making body, it also gives considerable attention to controls and practices. However, translating policy decisions into day-to-day operational practices is a highly human resource-intensive activity. As a result, policy implementation is moving ahead more slowly than was originally anticipated. The Census Bureau has handled this challenge, in part, by establishing a new Policy Associates Program, which details competitively selected Census Bureau program staff for one year to the Policy Office to help implement new data stewardship policies.

Data Stewardship and the Use of Administrative Records

The benefits and stewardship of linked survey and administrative data, the subject of this panel, are of great interest to the Census Bureau's DSEP Committee, which uses its data stewardship framework to guide and support use of administrative records for statistical purposes. Using the

approach introduced in **Chart 4** above, the Census Bureau first looked to its strategic plan and whether administrative record data would support its goals. The Bureau's strategic goal of "*Fostering an environment that supports innovation, reduces respondent burden, and ensures individual privacy,"* supports use of data from administrative records. They minimize the cost of direct data collection, reduce the burden on respondents, improve and enhance census and survey collections, and enable the development of improved data products that inform public policy. This strategic goal drives the development of policies that balance the benefits of administrative record use against privacy and confidentiality concerns, particularly given that these benefits are primarily derived from linking administrative records to other datasets.

Policy issues surrounding use of administrative records are identified by the DSEP Committee, with subsequent policy analysis and recommendations developed by the CARPP (see **Chart 3** above). In addition to weighing the needs of the data user community and the public, the CARPP must give special consideration to the Census Bureau's data providers, including managing and safeguarding data in accordance with their legal authorities and policy requirements. The CARPP and the DSEP Committee have established a number of procedures for managing the use of administrative records at the Bureau.

Procedures for managing administrative records include consistent review criteria for all proposed projects; centralized custodial functions to control data access on a "need-to-know" basis; and centralized tracking of administrative record projects. In addition, personal identifiers on administrative records (e.g., Social Security Number and name) are maintained in a restricted environment by the custodian. Identifiers are stripped from the records before they are released to researchers. When necessary, the custodian replaces the personal identifiers with a "Protected Identification Key," or "PIK," to enable record linkage. Currently, the CARPP is developing a policy to guide the Bureau's record linkage activities, again seeking the balance between developing relevant, high-quality data products and providing appropriate privacy and confidentiality protections to respondents.

Enhancing the Basic Structure

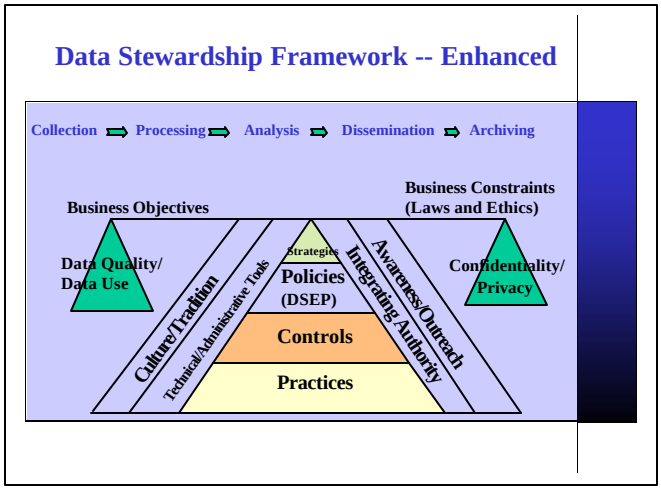
Although the basic data stewardship structure provides a mechanism for balancing data quality and access with privacy and confidentiality, that balance is still somewhat precarious. Looking back at the generic framework in Chart 2, it is useful, then, to consider ways to further stabilize this structure.

The Census Bureau has considered a number of sources for guidance in strengthening its data stewardship approach. First, it conducted a benchmarking exercise, making structured inquiry of six best practice-oriented private and government organizations about their policies, agency structures, and roles with regard to privacy. It also conducted a literature review consisting of recent privacy research both at the Census Bureau and elsewhere. The Census Bureau also drew on a General Accounting Office report issued in April 2001, *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information*, which provides a toolkit of

approaches to support data stewardship.[16] Lastly, the DSEP Committee commissioned an evaluation of the DSEP structure (executive body plus four staff committees). The evaluation targeted four areas for improvement -- the need to focus on employee awareness of the data stewardship structure; include stakeholders in policy discussions; be more systematic in assessing the operational impacts of policies; and restructure the role of the Security staff committee. The assessment activities also identified four key components that can help stabilize the data stewardship structure – culture and tradition, technical and administrative tools, awareness and outreach, and an integrating authority.

As shown in **Chart 5,** adding these steps to the data stewardship pyramid helps achieve a more stable balance between data access and use, on the one hand, and data protection, on the other.

Chart 5 --



- *Culture and tradition* form the basis for a statistical agency's approach to data stewardship. Al Zarate, Confidentiality Officer at the National Center for Health Statistics (NCHS) describes the Census Bureau as having a "culture of confidentiality."[17] Some organizations have cultures that focus predominantly on access to information. In an academic environment, for example, information sharing is the lifeblood of learning. The primary focus is on sharing research, not limiting access. Other organizations, like the National Security Agency, place a priority on keeping information highly controlled and access limitation is paramount. Survey organizations would not continue to do business without a focus on both confidentiality and access. The Census Bureau's culture and tradition fit this model well.

[16] U.S. General Accounting Office, *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information.* GAO-01-126SP, April 2001.

[17] Al Zarate, *Government Perspective on Data Stewardship for Statistical Data.* Paper presented for panel, "Statistical Data Stewardship in the 21st Century," Joint Statistical Meetings, New York, NY, August 11, 2002.

- *Technical and administrative tools* play an important role in a well-grounded data stewardship structure. Today, most organizations control disclosure by providing safe settings, where data can be used for legitimate statistical purposes, and by releasing safe data, where the data have been modified to hamper those who attempt to identify individual respondents. These tools allow organizations to more effectively accomplish the business objective of providing access to data while also ensuring confidentiality. They also play a role in restricting access and limiting uses within the organization. Need-to-know access and file-level auditing ensure that employees are not tempted to browse records or give others access, regardless of the motive. In deciding what tools to apply, the organization must be aware of external threats, assess the physical constraints on users, and take into consideration the impact on utility of the data for intended research.

- *Awareness and outreach* activities help ensure that business decisions are based on the valid concerns of external stakeholders, including respondents, privacy advocacy groups, and the data user community. Without adequate research and data on privacy attitudes and behaviors and data needs, it is easy to fall into an endless loop of supposition and speculation in the policy development process. The Census Bureau has conducted privacy attitude surveys for the past decade, to measure the public's awareness of confidentiality requirements and gauge concerns over the use of administrative records. Attitude surveys, focus groups, and cognitive interviews play an important role in understanding awareness of organization practices and identifying practices that may be misunderstood or not be acceptable. Messages that are conveyed to employees and to the public help reassure that data uses are important and that protections are appropriate. Message wording benefits from cognitive testing to ensure that what is intended is what is understood.

  An agency's marketing activities also support the agency's outreach efforts by emphasizing the organization's objectives and constraints and how its culture, tools and legal authority enforce its approach to data stewardship. It is critical, however, that messages accurately reflect practice (i.e., the "talk matches the walk") -- saying you do something when you don't can be worse than not saying anything at all.

- An *integrating authority* is critical to ensure integration of strategies, policies, controls and practices and to make most effective use of culture, tools and awareness. This typically entails a role for persons or groups to decide or advise on policies, controls and practices. The National Center for Health Statistics (NCHS) enlists its confidentiality officer for this purpose, who provides internal advice on data protection and access decisions. The Canadian government has established a Privacy Commissioner, who provides counsel and direction on matters affecting the privacy of Canadian citizens. Statistics Canada also has a privacy and confidentiality officer. In other instances, agencies are subject to Institutional Review Boards that review and approve survey research affecting human subjects. NCHS and the Census Bureau have also established Disclosure Review Boards to review and approve all publicly released data. Lastly, there is a trend among U.S. institutions to name a Chief Privacy Officer whose responsibility it is to implement privacy policies across the organization. Legislation recently enacted to establish a Department of Homeland Security requires affected federal agencies to establish a Chief Privacy Officer.

In short, there are several non-mutually exclusive options for establishing an integrating authority, all providing varying degrees of control. Some are purely internal, some external, and some provide a combination of the two orientations. The use of external decision makers is controversial and often resisted, but part of that resistance stems from a concern that such counsel generally lends itself to advocacy of privacy and confidentiality to the exclusion of balancing those concerns against the agency's need to provide quality data products. A redirection of the integrating authority's focus to a balanced data stewardship approach may alleviate this concern.

Conclusion

At this writing, the Census Bureau is deliberately working towards full implementation of the enhanced data stewardship framework illustrated in Chart 5. There are several data stewardship issues that will influence the way the Census Bureau – and the federal statistical community in general -- will function this decade. The impact of recent legislation like the USA Patriot Act and future implementation of new data sharing legislation (H.R. 2458), which passed through Congress in November 2002, need to be assessed and addressed. Additional challenges continue to arise.

As the Census Bureau explores the potential of using administrative records for statistical purposes, it needs a clear policy on record linkage methodology and standards for obtaining informed consent from respondents to conduct such matches. Also, administrative record procedures must include adequate controls on access and use of these data, which must be maintained in accordance with the requirements of the providing agencies. The Census Bureau is currently responding to new Office Management and Budget requirements for Privacy Impact Assessments, building on the Privacy Principles developed within the parameters of the data stewardship structure. A broad range of disclosure limitation approaches that permit safe release of data for public policy uses, must be developed, including contracting with experts to attempt unauthorized links of public data sets, and developing synthetic data sets to permit public users access to data while reducing the risk of identifying respondents.

Lastly, a key point bears repeating: developing and maintaining a viable data stewardship structure requires a significant commitment and investment of resources from an agency. Nevertheless, this more structured approach to data stewardship is integral to striking a balance between the tensions inherent in meeting data user needs and honoring the privacy and confidentiality of its respondents. In the end, privacy and confidentiality -- which are typically perceived as business constraints – can actually enable an agency's mission and business objectives by establishing the public's trust and cooperation as respondents.

Acknowledgments

The authors wish to thank Eloise Parker for her role in the preparation of this paper and Eloise Parker, Wendy Alvey, and Ta Shunna Marshall for their assistance with the presentation.

Recommended Resources on Data Access and Confidentiality

Doyle, Pat. Julia I. Lane, Jules J.M. Theeuwes and Laura V. Zayatz. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies.* North-Holland, published in conjunction with the Census Bureau (2001).

Duncan, George T. Thomas B. Jabine, Virginia A. de Wolf, Eds. *Private Lives and Public Policies.* Panel on Confidentiality and Data Access. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, and National Research Council. Washington, DC: National Academy Press (1993).

Federal Committee on Statistical Methodology (FCSM) (May 1994). *Report on Statistical Disclosure Limitation Methodology*. (Statistical Working Paper 22). Washington, DC: Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office.

Holz, V. Joseph, Robert Goerge, Julie Balzekas and Francis Margolin. *Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development.* A Report of the Advisory Panel on Research Uses of Administrative Data of the Northwestern University/University of Chicago Joint Center for Poverty Research (January 1998).

H.R. 2458. E-Government Act of 2002, containing the Confidential Information Protection and Statistical Efficiency Act of 2002. Legislation passed both Houses of Congress by November 15, 2002; pending presidential signature at this writing.

Privacy Protection Study Commission. *Personal Privacy in an Information Society* (July 1997). The principle of functional separation is addressed in Chapter 15.

U.S. Census Bureau. *Administrative Records Handbook.* May 2001. For inquiries about the handbook, please contact Eloise Parker, Administrative Records Coordinator, U.S. Census Bureau, FOB 3, Room 2430, Washington, DC 20233; (301) 763-2520; eloise.k.parker@census.gov.

U.S. General Accounting Office. *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information.* April, 2001. GAO-01-126SP.

Zarate, Alvan. *Government Perspective on Data Stewardship for Statistical Data.* Panel, "Statistical Data Stewardship in the 21st Century," Joint Statistical Meetings, New York, NY, August 11, 2002. (American Statistical Association CD-ROM Proceedings in process.)

Zarate, Alvan O. Jacob Bournazian and Virginia de Wolf. *Integrating Federal Statistical Information and Processes*. Federal Committee on Statistical Methodology (FCSM) Committee on Data Access and Confidentiality (November 8-9, 2000). <www.fcsm.gov/cdac/cdacpaper.pdf >.

SSA Policy Applications of Administrative Data Linked to SIPP[18]

Howard M. Iams
Social Security Administration

Abstract

The Social Security Administration (SSA) conducts policy analysis with the data from the Survey of Income and Program Participation matched to extracts from SSA's administrative records. SIPP represents the social characteristics of the U.S. population; SSA administrative records contain information necessary to administer the Old Age and Survivors and Disability Insurance Programs and the Supplemental Security Income program. SSA assesses the impact of policy changes to programs it administers on the distribution of income and poverty with these SIPP matched data. Using these matched SIPP records, SSA develops micro-simulation models to assist policy evaluation. These include models of eligibility and participation in the Supplemental Security Income and the Qualified Medicare Beneficiary programs as well as the retirement income and life histories of future retirees from the baby boom, World War II, and Depression birth cohorts. SSA also describes the beneficiaries served by its programs with these SIPP matched data. This paper discusses examples of these uses by SSA.

I. Introduction

The Social Security Administration's (SSA) Office of Policy relies extensively on the Census Bureau's Survey of Income and Program Participation (SIPP) matched to Social Security Administration records of benefits and lifetime earnings. A major focus is the impact of Social Security policy alternatives on the distribution of income to various sub-populations. A second is the development of statistical simulations of a projected population for policy evaluation. Linked data also describe who is being served by the programs administered by SSA. The programs include Title II benefits for Old Age Insurance, Survivor's Insurance, and Disability Insurance and Title XVI Supplemental Security Income benefits for disability and old age. The purpose of this paper is to briefly describe examples of these uses at SSA.

The SIPP matched data combine the SIPP survey information with SSA's administrative records. The content of the SIPP is well known (see the user's guide, U.S. Department of Commerce 2001), and the data are publicly available from the Census Bureau. Less well known are SSA's administrative records containing the material necessary to administer the Social Security Act (see Panis et. al. , 2000). [19] The matched SIPP permits analysts to use detailed SSA program

[18] The positions in the paper represent the author's professional judgement and do not represent the position of the Social Security Administration.

[19] The Numident includes basic information on birth and death dates. The Master Beneficiary Record contains monthly benefit status and payment amounts for Title II programs from 1951 to current month, while the Supplemental Security Record for the Title XVI program contains monthly benefit information from 1974 to the current month. The records include the SSA 831 Form for application for disability from 1974 to the current month. The Master Earnings File (MEF) contains detailed earnings information including Medicare taxable earnings and uncovered earnings. The Summary Earnings Record extract from the MEF contains Social Security taxable earnings and quarters of coverage for each year from 1951 to the current year minus 2 years. The Detailed Earnings Record extract from the MEF contains information from the W-2 tax form including total earnings, self-employment income, and nontaxable earnings for defined pension plan accounts.

information in combination with the socioeconomic and demographic information contained in SIPP. Through a joint agreement, SSA and the Census Bureau match individual respondent information provided in SIPP to the SSA records for administering the program for respondents providing Social Security numbers in the survey. They match about 90 percent of the adults in the 1990-1993 panels, about 85 percent of the adults in the 1996 panel, and about 74 percent of children in the 1996 panel.[20] SSA and the Bureau restrict access to these matchable administrative records to sworn census agents with approved research projects. The processing of the restricted data must take place at a secure Census Bureau or SSA site.

II. Policy Estimates

A primary use of SIPP matched data is the distributional impact from policy changes. This paper reviews three policy analyses conducted recently at SSA: cost neutral policies for increased widow benefits, childcare credits, and the removal of the retirement earnings test. The SIPP matched data were necessary for analysis of the distributional impact of policy change. The important function of the SSA administrative records is to provide SSA administrative details on benefits and lifetime earnings. Many survey respondents do not know these administrative details or would imperfectly recall a lifetime history. Examples would include the extent to which earned retired-worker benefits offset higher auxiliary benefits as a spouse or survivor and the lifetime history of annual earnings taxed for Social Security purposes (which changes across time). The important function of SIPP is to provide socioeconomic and demographic characteristics of a nationally representative sample including income, assets, marital history, fertility history, and pension coverage. In addition to these characteristics, the SIPP links husbands and wives in married couples. [21] Analysis of specific Social Security policy options requires both sets of information contained in the SIPP linked to SSA administrative data.

Widow Benefit Change

Older widows are much more likely to live in poverty than older married women. Because most aged widows receive Social Security benefits, one option for increasing widows' income would be to increase their Social Security benefits. The 1994-96 Advisory Council on Social Security (1996) proposed an increase in survivor benefits with some financing from reducing spouse benefits. This proposal would address both equity and adequacy issues connected with widow benefits.[22] Auxiliary benefits create inequities because wives/widows with their owned earned

[20] The incomplete matching of respondents to their own administrative records could influence results if the omission is selective. Several applications mentioned in this paper do not compensate for the 10-15 percent of adults without linkage other than reweighting population totals. SSA's microsimulation of the baby boomer's retirement (called Modeling in the Near Term or MINT) statistically generates an administrative linkage using a nearest neighbor or "hot-deck" linkage to a similar SIPP respondent. Analysts of beneficiary children use survey data when linkages are not available because of the lower match rate.

[21] SSA records only identify couples when a spouse/survivor is drawing benefits based on their current or former spouse's earnings record. In 2003, this includes about two-thirds of aged wives and most aged widows. No marital link is possible for those without benefits or those with only their own earned benefits.

[22] The increase in widow benefits would provide more adequate retirement income to qualifying widows, primarily survivors of couples with a working wife as well as a working husband. The spouse benefit reductions would affect couples with a non-working wife or a wife with much lower earnings than her

benefits often do not receive higher benefits than if they had not worked (Iams and Sandell 1998).

An analysis of the impact of such a change needs SIPP matched data to make the estimates. The analysis requires separate measures of each spouse's earned and auxiliary benefits that must be derived from SSA records. The estimate also requires the offset of the auxiliary benefit by the earned benefit for dually entitled beneficiaries (approximately one-third of beneficiary wives and two-thirds of widows). Most dually entitled beneficiaries would not know this information and could not report it in a survey. SIPP provides information on family income, poverty, and links husbands and wives, which is absent from SSA records.

Based on analysis of SIPP matched data for married couples, this policy shift would moderately decrease poverty rates among older women by reducing the poverty rate of widows slightly more than increases in the rate for couples (Iams and Sandell 1998, Table 2; Sandell and Iams 1997).

Childcare

Advocates have argued that periods of full-time child care reduce women's Social Security benefits, but perhaps more importantly, they argue that this has had a greater impact on minority and lower income women because they have more children. The legislative proposals in the 1980s would delete a few years with no earnings (called dropout years) because of full-time child care from the Social Security worker benefit computation, thereby increasing the lifetime average earnings and earned benefits of mothers.

The analysis to test the effect of providing additional dropout years for childcare could not be made without SIPP matched data. SIPP's fertility topical module identifies the birth years of children. But the policy test requires identifying the years with no earnings, which is not available from the SIPP. The SSA administrative data provide each year's earnings taxed for Social Security purposes. In addition, the policy analysis requires identification of women expected to receive only their own retired worker benefits, because changes in a woman's earned benefits have no impact on income if she receives higher benefits as a wife or widow. This requires SSA matched earnings records to estimate expected retired-worker and auxiliary benefits.

---

husband. Wives can receive half of their husband's benefits and widow's can receive their husband's full benefits, without paying Social Security taxes on any earnings of their own. (Divorced persons can receive these benefits if married for at least ten years.) Wives and widows can receive Social Security benefits based totally on their husband's earnings, based totally on their own earnings, or based on a combination (termed dual entitlement where earned benefits offset higher auxiliary benefits). About two-thirds of wives and the majority of widows receive their benefits based on their husband's benefit either as auxiliary or dual benefits. This creates an inequity between couples and survivors of couples with a working spouse and those without a working spouse. Those with a working spouse receive lower benefits than those without a working spouse given a similar level of total couple earnings over a lifetime. Changes that increase equity either reduce benefits of couples or survivors of couples with nonworking spouses or increase the benefits of couples or survivors of couples with working spouses. Lowering Social Security benefits may reduce the adequacy of retirement income. Thus, options to increase equity often reduce the adequacy of benefits.

Using SIPP matched data, Iams and Sandell (1994) estimated the impact of childcare dropout years on benefits expected for women born in the 1930s and 1940s. They found that childcare dropout years would increase the retirement benefits of some women, but the estimated benefit increases were small, were more likely for more privileged socioeconomic groups, and were lower among women born in the early baby boom than those born in the depression (Iams and Sandell 1994, Table 3 and Table 6). Iams and Sandell conclude that subsidizing child-care dropout years is not a well targeted policy, and the impact will decline over time as fewer women drop out of the labor force to care for young children.

Retirement Earnings Test

What is the impact of eliminating the retirement earnings test (RET) which reduces Social Security benefits of working beneficiaries with earnings above specified levels? SSA wanted to estimate the impact of legislation passed in 2000 that eliminated the earnings test for working beneficiaries aged 65-69. Although most agree the financial incentives of the RET affect earnings behavior, the size of the impact has been ambiguous for high and low earners.

The SIPP matched data provided the information needed for a study of the effects of the legislative change. SSA benefit records identify the benefits in each month of each year, and SSA earnings records contain annual Social Security taxable earnings. The SIPP data provide personal characteristics such as gender, educational attainment, health limits, per capita family income, and self-employment that would indicate differential effects on various groups of beneficiaries.

The study looked at changes in earning or not earning income, earnings levels, and applications for benefits. Removal of the earnings test in 2000 was not related significantly to changes in the presence of earnings (Song 2002). This suggests that it didn't change the decision to work or not work among beneficiaries aged 65-69. The earnings test removal significantly increased the earnings of high earners but not middle and low earners (Song 2002, Table 9). The removal also was associated with slightly increased applications for benefits among persons aged 65-69.

III. Micro-Simulations

SSA conducts policy evaluations with micro-simulation models created from SIPP matched to SSA administrative records. This paper discusses two models—Modeling Income in the Near Term (MINT) which projects life histories of the aged population 20 years from now in 2022 and a Financial Eligibility model for Supplemental Security Income and other means-tested programs.

Financial Eligibility Model

Policy analysis related to the Supplemental Security Income (SSI) program requires SIPP matched data. SSI pays benefits to the aged and nonaged disabled with limited income and limited assets which SIPP identifies.[23] The SSA administrative data are used to clarifying benefit

[23] SSI also requires the nonaged to have disabling health limitations which can be inferred from SIPP information.

eligibility status and actual benefits received from Social Security and the SSI program (Huynh et. al. 2001, Table 1 and Table 2).

SSA has developed a Financial Eligibility model that can be used to address a wide range of policy issues related to SSI, Social Security, Medicare, Medicaid and other programs. These include the following:

- What is the rate of participation in SSI and other means-tested programs? Is there a substantial pool of eligibles that do not participate in the program? Why? Davies et. al. (2002) find that about three-fifths of eligibles participate in SSI. Rupp and Sears (2000) and Sears (2002, Table 1) also find about three-fifths of eligibles participate in Qualified Medicare Beneficiary, Special Low-Income Medicare Beneficiary, and Qualified Individual buy-in programs which pays Part-B Medicare premiums with Medicaid funds.

- What are the costs and benefits of potential modifications of SSI program rules? The model can provide estimates on changes in program cost, number of eligibles, number of participants, average benefits, and distributional outcomes such as effects on the poverty rates and the poverty gap. The model is capable of estimating the potential effects of changes to the SSI program, such as the asset test, earned and unearned income disregards. For example, if SSI expenditures increased by 3 percent through changes in the Social Security benefit exclusion, then the poverty gap of aged women would decrease 1.1% (Rupp et. al. 2001, Table 3).

- If policy makers consider a range of alternative interventions, which one is the most effective? SSA has developed a methodology of cost-equivalent comparisons that can be used to assess which one of several policy alternatives are most effective in improving desired outcomes at given levels of funding availability. For example, Rupp et. al. (2001, Table 3) find that modifying the SSI asset limits is a relatively effective change in reducing poverty among elderly women.

- What are the interactions between SSI program changes and other programs? What is the effect of proposed changes in other programs, such as Social Security on SSI participation and cost? For example, how do proposals to introduce a minimum Social Security benefit affect SSI? What changes in SSI are necessary to facilitate desired distributional outcomes under a Social Security minimum benefit? What is the effect of changes in SSI eligibility rules on Medicaid participation and cost?

- What is the likely size of the SSI program in terms of costs and participation in the medium term? How do different demographic and socioeconomic factors, as well as potential policy changes affect this? For example, what is the likely effect of the increased proportion of successive cohorts with Social Security insured status and the aging of the baby boom generation on SSI participation and program cost?

SSA continues to develop and improve the Financial Eligibility Model to accomplish these objectives with the most recent SIPP data on income and assets matched to SSA records.

MINT

The Modeling Income in the Near Term (MINT) microsimulation model is designed to study the retirement of the baby boom birth cohort as well as the World War II and Depression birth cohorts. Policy makers have a strong interest in the differential effects of policy changes on the benefits, total income, and poverty level of the retiree population, as well as its subgroups. Of particular concern to policy makers is the economic well-being of future retirees in the baby boom cohort – those born between 1946 and 1964. Not only is this the largest birth cohort in U.S. history, but the earliest baby boomers will be eligible for retirement in 2008, and without program changes the Social Security (OASDI) Trust Fund is projected to be exhausted in 2041 (The Board of Trustees Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds, 2002). Aside from its sheer size, the baby boom cohort has distinguished itself from earlier cohorts in a number of ways that reflect the culture of the post-world War II period. The baby boom cohort experienced "unprecedented prosperity" and increased educational opportunities and attainment, as well as major changes in marital patterns and in the lifetime employment and earnings of women (Farley 1996; Levy 1998; O'Rand and Henretta 1999). Because of structural changes in mortality, marriage, lifetime earnings, and work patterns, we would expect the impact of policy changes to differ between current retirees and future retirees in the baby boom cohort.

When changes occur across time, policy analysis of the current beneficiary population may be misleading. Analysis of the future population targeted by legislation is preferable. This approach takes into account birth cohort differences and diversity and, consequently, is sensitive to shifts across cohorts in socio-economic relationship such as in women's lifetime earnings and work patterns. Accordingly, Modeling Income in the Near Term (MINT) projects the life histories of the baby boom cohort and the aged population to 2022.[24] SSA can estimate the impact of alternative Social Security policies on total income and poverty for subgroups defined by race, educational level, and marital status of the baby boom cohort in retirement.

The MINT projection of life histories relies heavily on the SIPP matched data. To enhance the data for analysis, MINT combines the SIPP panels of 1990, 1991, 1992, and 1993. The policy universe for most analyses is the surviving population born from 1931 through 1960 that is expected to reach retirement age and to receive Social Security retirement and survivor benefits in 2022. [25] The matched data provide important information that supplements the SIPP reported data. Statistical projections make use of these longitudinal SSA data to estimate life histories until death. SSA administrative records measure the annual earnings history, the monthly benefit history, and date of death through 1999. The MINT model makes independent statistical

[24] The U.S. Social Security Administration (SSA) created MINT with substantial input from the Brooking Institution, the RAND Corporation (Panis and Lillard 1999), and the Urban Institute (Toder et. al. 1999; Toder et. al. 2002). For a summary of the work completed by the Brookings Institution, RAND, and the Urban Institute for the initial MINT model see Butrica, et. al. 2001. Toder et. al. (2002) document the revision of MINT completed in 2002.

[25] Those born 1961-64 were dropped from the analysis because with fewer years of real data we are less confident in their projections of retirement income. The SIPP reported data for a person born in 1960 would be at age 30 in the 1990 panel, 31 in the 1991 panel, 32 in the 1992 panel, and 33 in the 1993 panel.

projections until death for each SIPP respondent's lifetime earnings, retirement income (Social Security benefits, pensions, assets, and earnings of working beneficiaries), and marital changes. The 1990-1993 panels of SIPP for middle aged persons born in 1931-1960 directly measures such choices as educational attainment, marriage and divorce history, current employment, pension plan participation, and savings.

MINT projects substantial changes in the characteristics of the baby boom retirees compared to earlier birth cohorts from World War II and the depression. Butrica and Iams (1999, Table 2) document with MINT the importance of both marital histories and earnings records to the projected Social Security benefits of married couples. MINT projects that spouse and widow benefits will be less important to the baby boom cohort than to earlier cohorts born in the depression and World War II (Butrica, Iams and Sandell 1999, Chart 2). MINT also projects that the proportion of women who divorce will be higher among the baby boom cohorts than earlier cohorts, but the proportion of these women eligible for benefits as a divorced spouse will decline (Butrica and Iams 2000, Table 3 and Chart 2). This occurs because MINT projects divorced women in the baby boom to be more likely to have their own earned retired-worker benefits, but they are less likely to have at least ten years of marriage necessary to be eligible for spouse/widow Social Security benefits.

Using the MINT data system, Toder et. al. (2002, Chapter 9) describe the characteristics of the aged population in 2020 and the retirement population at age 62 and age 67. These tables describe the projected change in socio-economic and demographic characteristics among the baby boom compared to earlier cohorts born in the 1930s and during World War II. MINT projects the baby boom cohort of beneficiaries at age 62 and age 67 to be more educated, to contain more minorities, and to contain fewer married couples than earlier cohorts. MINT projects retirement wealth among the baby boom to increase with shifts toward more income from pensions as well as non-financial wealth. MINT projects average levels of retirement income at age 67 to be higher in the early baby boom cohort than the depression cohort, but similar to the late baby boom cohort.

IV. Descriptions of Beneficiaries

SSA also produces several reports of the socioeconomic and demographic background characteristics of its current beneficiaries using SIPP matched records. These reports describe the characteristics of beneficiaries served by SSA and the importance of SSA administered benefits as an income source. The SSA record match identifies the SSA program beneficiaries and benefit amounts actually paid to beneficiaries (Huynh et. al., 2001). The SIPP based characteristics are unavailable from SSA records used in administering its programs. Tabulations include SIPP based demographic characteristics, sources of income, family income, poverty level, family and household size, household type, home ownership and receipt of assistance for energy, for housing, for Food Stamps, for health insurance. [26] The Performance and Accountability Report

[26] For example, the SSI Annual Statistical Report (2001d) reports characteristics of Supplemental Security Income Title XVI recipients, and the Annual Statistical Report on the Social Security Disability Insurance Program (2001a) reports characteristics of Disabled Insurance Title II

(Social Security Administration 2001c) contains measures of adequacy of income of beneficiaries including the reduction in the poverty gap due to SSI benefits, SSI as a percent of total income, and the percent participating in an employer sponsored pension plan.

Conclusion

SIPP data linked with SSA administrative data benefit from the strengths of surveys and administrative data. The linked data have become a critical source of information for policy analysis, evaluation of legislation, and statistics to inform policymakers.

Bibliography

The Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Trust Funds. 2002. THE 2002 ANNUAL REPORT OF THE BOARD OF TRUSTEES OF THE FEDERAL OLD-AGE AND SURVIVORS INSURANCE AND DISABILITY TRUST FUNDS. Washington, D.C.: U.S. Government Printing Office.

Butrica, Barbara and Howard M. Iams. 1999. "Projecting Retirement Income of Future Retirees with Panel Data: Results from the Modeling Income in the Near Term (MINT) Project. Social Security Bulletin. Vol. 62, No. 4, pp. 3-8.

Butrica, Barbara and Howard M. Iams. 2000. "Divorced Women at Retirement: Projections of Economic Well-Being in the Near Future". Social Security Bulletin. Vol. 63, No. 3, pp. 10-24.

1994-1996 Advisory Council on Social Security. 1996. Reports of the Technical Panel on Assumptions and Methods Technical Panel on Trends and Issues in Retirement Savings and Presentations to the Council. Vol. II. Washington, D.C.: Social Security Administration.

Butrica, Barbara, Howard M. Iams, and Steven H. Sandell. 1999. "Using Data for Couples to Project the Distributional Effects of Changes in Social Security Policy". Social Security Bulletin. Vol. 62, No. 3, pp. 20-27.

Butrica, Barbara, Howard M. Iams, James H. Moore, and Mikki D. Waid. 2001. Methods in Modeling Income in the Near Term (MINT I). ORES Working Paper Series No. 91. Washington, D.C.: Office of Research, Evaluation and Statitistics.

Davies, Paul, Minh Huynh, Chad Newcomb, Paul O'Leary, Kalman Rupp, and Jim Sears. 2002. "Modeling SSI financial eligibility and simulating the effect of policy options". Social Security Bulletin. Forthcoming.

Farley, Reynolds.1996 THE NEW AMERICAN REALITY: WHO WE ARE, HOW WE GOT THERE, WHERE WE ARE GOING. New York: Russell Sage Foundation.

Huynh, Minh, Kalman Rupp, and James Sears. 2001. The assessment of Survey of Income and Program Participation (SIPP) benefit data using longitudinal administrative records. Paper presented at the Federal Statistical Committee on Methodology meeting. SIPP Working Paper Series No. 238. Washington, D.C.: Census Bureau.

---

beneficiaries. SSA also includes SIPP information in the Annual Report to the President and Congress on the Supplemental Security Income Program (Social Security Administration 2001b)

Iams, Howard M. and Steven H. Sandell. 1994. "Changing Social Security Benefits to Reflect Child-Care Years: A Policy Proposal Whose Time Has Passed". Social Security Bulletin. Vol. 57, No. 4 (Winter), pp. 10-24.

Iams, Howard M. and Steven H. Sandell. 1998. "Cost-Neutral Policies to Increase Social Security Benefits for Widows: A Simulation for 1992". Social Security Bulletin. Vol. 61. No. 1, pp. 34-43.

Levy, Frank. 1996. The New Dollars and Dreams: American Incomes and Economic Change. New York: Russell Sage Foundation.

National Research Council. 1997. Assessing Policies for Retirement Income: Needs for Data, Research, and Models. Edited by Constance F. Citro and Eric A. Hanushek. Washington, D.C: National Academy Press.

O'Rand, Angela M. and John C. Henretta. 1999. *Age and Inequality: Diverse Pathways through Later Life.* Boulder, Co: Westview Press.

Panis, Constantijn, Roald Euller, Cynthia Grant, Melissa Bradley, Christine Peterson, Randall Hischer, and Paul Steinberg. 2000. SSA Program Data User's Manual. June. Santa Monica, California: RAND.

Panis, Constantijn and Lee Lillard. 1999. Near Term Model Development Part II. Final Report. Santa Monica, Ca: RAND.

President's Commission to Strengthen Social Security. 2002. Report of the President's Commission to Strengthen Social Security. Washington, D.C.

Rupp, Kalman and James Sears. 2000. "Eligibility for the Medicare Buy-in Programs, Based on a Survey of Income and Program Participation Simulation". Social Security Bulletin. Vol. 63, No.3. pp. 13-25.

Rupp, Kalman, Alexi Strand, and Paul S. Davies. 2001. "The Potential of the Supplemental Security Income Program to Fight Poverty Among Elderly Women." Proceedings of the Institute for Women's Policy Research 2001 Conference, The Status of Women: Facing the Facts, Forging the Future.

Sandell, Steven H. and Howard M. Iams. 1997. "Reducing Women's Poverty by Shifting Social Security Benefits from Retired Couples to Widows". Journal of Policy Analysis and Management. Vol 16, No. 2 (Spring), pp. 279-297.

Sears, James. 2002. " 1996 QMB/SLMB". Unpublished paper. Washington, D.C.: Social Security, Office of Policy.

Social Security Administration. 2001a. Annual Statistical Report on the Social Security Disability Insurance Program. Washington, D.C.: Social Security Administration, Office of Policy, Office of Research Evaluation and Statistics.

Social Security Administration. 2001b. Annual Report to the President and Congress on the Supplemental Security Income Program. Baltimore, Md: Social Security Administration, Office of the Chief Actuary.

Social Security Administration. 2001c. Performance and Accountability Report Program. Baltimore, Md: Social Security Administration, Office of Finance, Assessment, and Management.

Social Security Administration. 2001d. SSI Annual Statistical Report. Washington, D.C.: Social Security Administration, Office of Policy, Office of Research Evaluation and Statistics.

Toder, Eric, Cori Ucello, John O'Hare, Melissa Favreault, Caroline Ratcliffe, Karen Smith, Gary Burtless, and Barry Bosworth. 1999. Modeling Income in the Near Term- Projections of

Retirement Income Through 2020 for the 1931-1960 Birth Cohorts. Final Report. Washington, D.C.: The Urban Institute.

Toder, Eric, Lawrence Thompson, Melissa Favrealt, Richard Johnson, Kevin Perese, Caroline Ratcliffe, Karen Smith, Cori Uccello, Timothy Waidmann, Jillian Berk, Romina Woldmariam, Gary Burtless, Claudia Sahm, and Douglas Wolf. 2002. Modeling Income in the Near Term: Revised Projections of Retirement Income Through 2020 for the 1931-1960 Birth Cohorts. Final Report. Washington, D.C.: The Urban Institute.

U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. 2001. SIPP Survey of Income and Program Participation User's Guide. Washington, D.C.

Discussing Potok and White's Papers Presented in Session 7: Stewardship of Linked Survey and Administrative Data

Olivia Blum
Israel Central Bureau of Statistics

Potok and White scan policies and restrictions that statistical agencies put upon them. The deriving question is why. Why do we go through Acts of - privacy, confidentiality, asking for consent, access to information etc.? Why a secondary use of administrative data accelerates the need to refer to these deeds and acts?

First, these are all implementation symbols of social norms and values. This is the social glue that we use and create day by day. Theses acts come to corroborate the social contract within the national group. The second function is ensuring the survival of the statistics organization through legal agreements with the direct users and with the public.

The statistics bureau is responsible for providing the users with quality data under changing circumstances. No bureau can allow itself stagnation with regard to attitude/policy, ways of action and tools. People, social structure and processes, technology, they all change and therefore social values and behavior. Adaptation to theses changes serves the quality data objective in the long run, meaning that it is required to maintain the functionality and therefore the mere existence of the organization.

As for the public, the agreement with the public has a give and take pattern. The data collector asks for private information and gives in return analyzed information that allows the policy makers to act efficiently and effectively and allows the individual to choose, based on empirical findings, what to eat, where to live etc. Private data and the consent to link individual administrative files are given in return to meta information and aggregated results and in return to explicit way of handling these data: Storage in secured sites, no accessibility to people and uses not specified, etc. Breaking this contract means no data, no quality data, no consent to link records, no support in a changing reality and changing environment, and therefore, no justification to the existence of the organization.

The acts and the careful processes of handling linked data, as described in both papers, bring about additional costs since the pure professional considerations are not the only guiding lights when coming to link administrative records. The resultant questions aim toward the quality data market in a broader perspective: Who are the participants in this market? What are the mechanisms to make it stable? What does the statistics agencies have an influence on?

There are three core participants: The direct data supplier, whom the data describe, the data user and the statistics agency who demands the data. The challenge derives from the statistics agency's role to serve and protect both, the supplier and the user.
As for the operating forces and mechanisms, I would like to make an analogy to the economic commodities market, which has two basic parameters, quantity and price. In the quality data market the detailed data represents the quantity while the quality represents the price. In this market the data supplier prefers to be less exposed while the data demander is seeking more

detailed data (see supply and demand curves in diagram1). The statistics agency and the user consider rich data, obtained by linking records, as quality data. However, the first is obliged to protect the privacy of the data suppliers whilst the last does not. The equilibrium point in this market is not stable; everyone wants to get out of it.

There are several mechanisms to be engaged in stabilizing the equilibrium point:

1. Acts, controls, policies, practices, as described in Potok and White's papers. These restrictions move up the whole supply curve, i.e., the public is willing to allow the statistics agency to link records and to have more detailed information for the same price in quality terms (see diagram2).

2. Pushing the suppliers up along the supply curve, which means reducing antagonism by overt presentation of the benefits drawn from rich linked data and by encouraging and enabling the public to use statistics on a daily basis.

3. Partnerships or business relations with the suppliers of the administrative records, in the public and government sectors, in the private and business sectors. This is a mechanism that comes to ensure the obtaining of the administrative data. It is a prerequisite to the existence and stability of this market.

4. Pulling the demand curve of the statistics agencies, vertically, toward a less invasion of privacy with no quality loss. Meaning, reducing the correlation between quality and quantity, which can be done by developing methodologies that enable the statistical estimates to rely on less detailed information (see diagram3).

5. Reducing antagonism of the public by visible fairness of the redistribution center. The state administration serves as a redistribution center of the national resources. Although it seems to be irrelevant to the statistics world and statistics agencies have no control over it, the conduct of public administration with regard to benefits, subsidiaries, infrastructure investments etc. has a direct influence on the cost/benefit analysis of the individual when asked to supply data or to give consent to use linked data.

Potok and White focus on the first mechanism as activated in their statistics agencies (US Census Bureau and Statistics Canada, respectively). It is the one that statistics agencies have a more influence on. This mechanism stipulates the approval of record linkage not only vis a vis the public, but also within the statistics bureau and vis a vis the relevant government oversee functions. However it is not clear how far the supply curve can be pushed and when additional restriction costs more than its contribution.

The second mechanism, in which the empirical findings are either published or made accessible to the public, is presented as an integral part of the record linkage program in White's paper. Data suppliers, whether they are individuals or administrative data holders (third mechanism), should have an ongoing interest to supply the data and to allow its use.

The forth Mechanism is an ongoing challenge for today's statisticians while the feasibility of the implementation of the fifth one is unclear.

Diagram 1

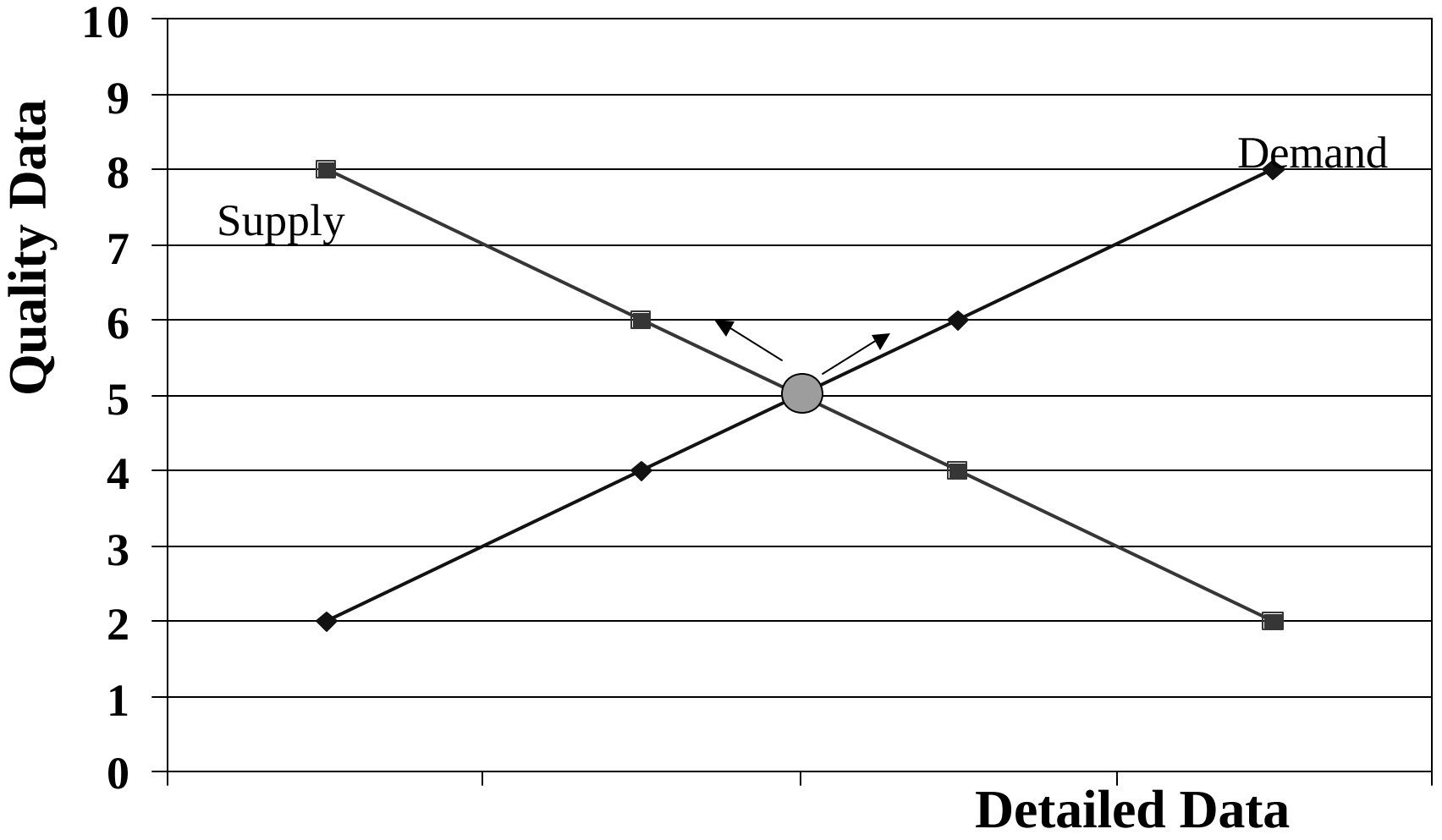


Diagram 2

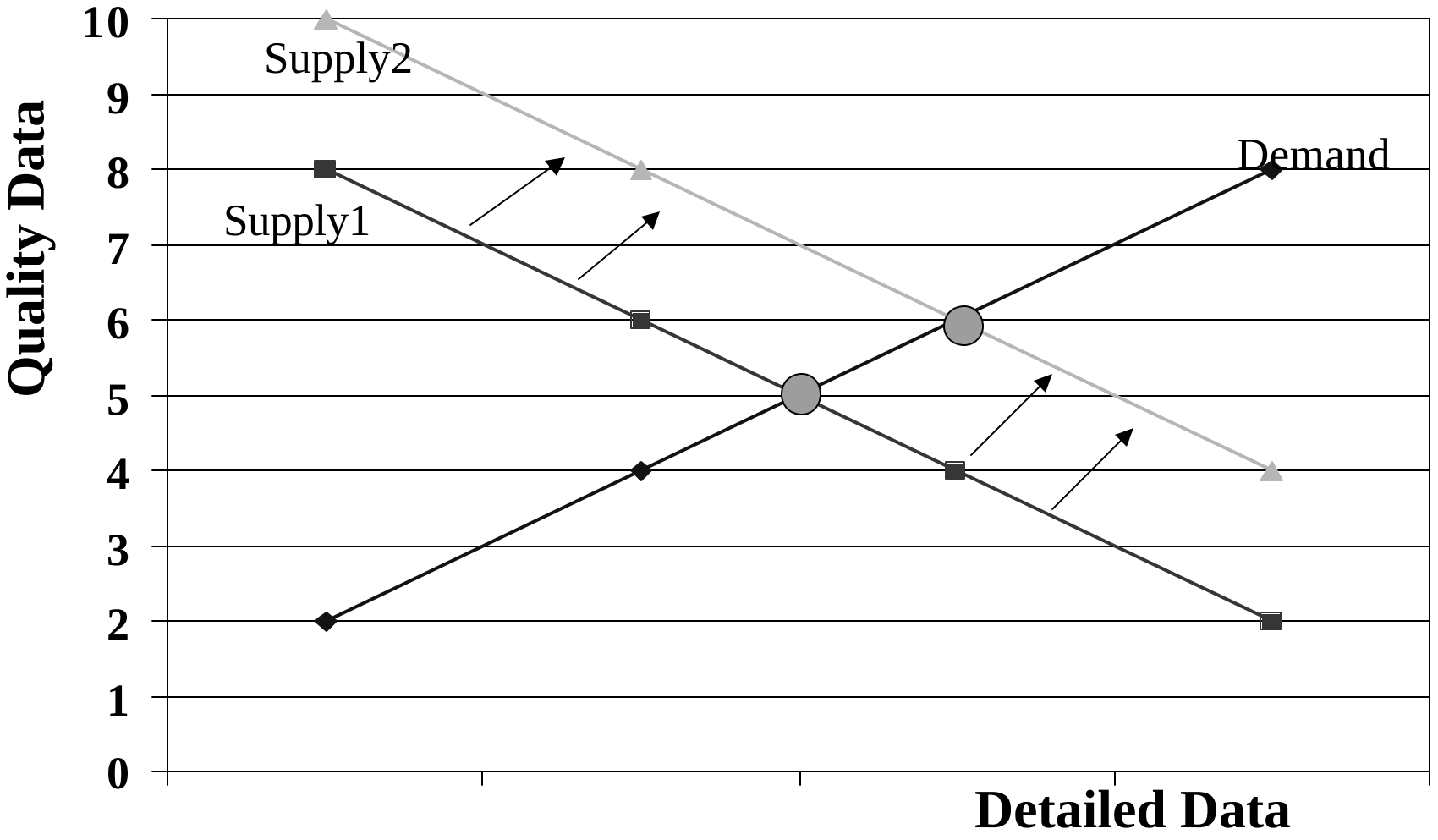
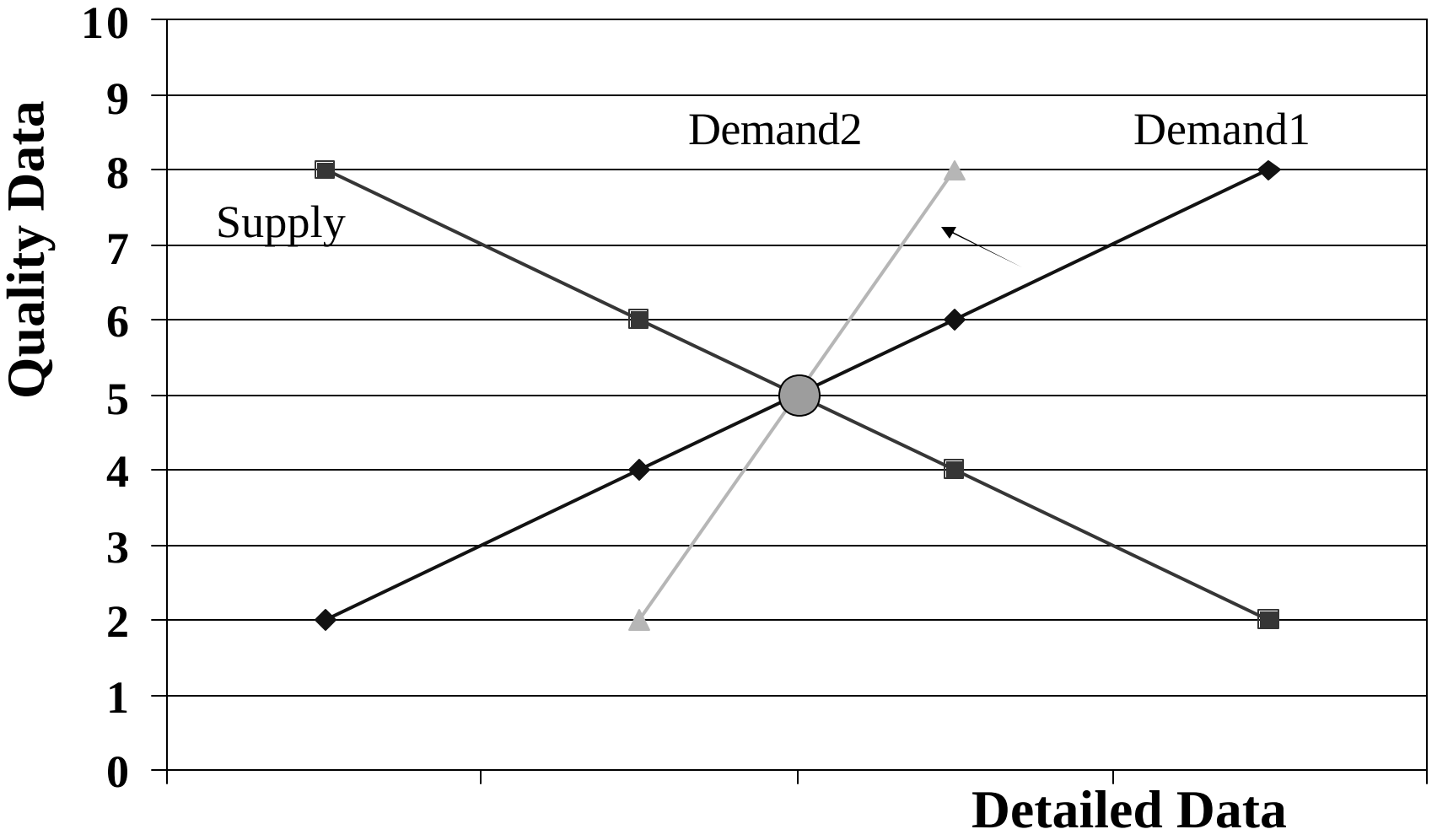

Quality Data
10
9
8
7
6
5
4
3
2
1
0
Supply2
Supply1
Demand
Detailed Data

Diagram 3


Quality Data
10
9
8
7
6
5
4
3
2
1
0
Demand2
Demand1
Supply
Detailed Data