

Beneficiaries of Federal Housing Programs: A Data Reconnaissance

Anne B. Shlay
Temple University

Charles E. King
George Washington University

Abstract

This article reports on a reconnaissance of information systems containing data on the beneficiaries of direct and indirect federal housing expenditures. It covers data in the U.S. Departments of Agriculture, Housing and Urban Development, Veterans Affairs, and the Treasury; the Resolution Trust Corporation; financial regulatory agencies; and secondary mortgage market actors.

Data varied widely across agencies in availability, accessibility, and quality. Data are more systematically collected for low-income beneficiaries of housing programs than for the more affluent beneficiaries of indirect housing expenditures. The systems need improvements in data quality and coverage and database format, though they have improved recently. Many research topics can be explored with new and underused data systems: the characteristics of beneficiaries of rural housing programs, urban rental housing programs, low-income homeownership programs, and mortgage guarantee and insurance programs. But the lack of information on the systems themselves makes data difficult to locate and access.

Keywords: Housing; Discrimination; Data

Introduction

The go-go decade of the 1980s has given way to a sober reckoning in the 1990s. Spatial and racial inequality is increasing, and the number of people living in poverty has grown (Massey and Denton 1993). Cities grapple with problems of job loss and economic development while confronting population decline, urban decay, and widespread homelessness. Rural areas, more hidden from the public eye, mirror these trends. This geographically bound misery occurs in a context of fiscal retrenchment and federal cutbacks. For many, the future appears grim.

The economic scarcity facing communities makes it critical that existing public benefits be distributed fairly to those most in need. More than ever, it is crucial that economic fairness, equal opportunity, and real redistribution accompany the administration of public programs. Besides continuing the federal government's obligation to undo past discrimination, the Clinton administration has directed federal agencies to affirmatively further fair housing in federal housing programs (White House 1994).

To determine whether these standards are satisfied, we must ascertain who benefits from public programs. For this we need information on the characteristics of program beneficiaries, including data on the economic, demographic, disability, racial, and ethnic characteristics of families and individuals receiving assistance.¹

Learning the characteristics of beneficiaries is particularly important at the federal level because, even though states and localities have recently increased the aid they provide, the federal government remains the major revenue source and the chief administrator of public assistance programs. Critical areas for examination include federal assistance in housing, community and economic development, income maintenance and other forms of welfare, and health care.

The first step in conducting research on the beneficiaries of programs is to determine the types of information available for study. With funding from the Poverty and Race Research Action Council and Fannie Mae's Office of Housing Research, we conducted a reconnaissance of databases at the federal level. By "reconnaissance" we mean an investigation to gain information on data systems. The topic of investigation was federal housing programs. (The council has commissioned similar reconnaissance projects on databases related to federal health, education, and income maintenance programs.)

The goal of this reconnaissance was to determine the types of information available on the beneficiaries of federal housing programs. The focus was on data systems within agencies, not on research, monitoring, or evaluation studies that have also produced data on beneficiaries of federal housing programs. This article describes the results of the investigation.

¹ This research focuses on data on people who receive federal benefits. It does not cover people who are eligible for assistance but do not receive it or those who are on waiting lists for particular programs.

The data reconnaissance was directed at answering several questions:

1. What types of data are available, and what does each data system contain?
2. What is the form of these data?
3. Are the data accessible and publicly available?
4. What is the overall quality or reputation of the data?

Assessing the availability of information on the characteristics of federal housing program beneficiaries is a prerequisite to assessing the equity of federal housing program design and administration and the extent to which the duty to affirmatively further fair housing is being satisfied (White House 1994). To conduct a complete assessment of equity, however, would require substantial data in addition to that examined here, such as information on the characteristics of eligible nonbeneficiaries and on variations among beneficiaries in the types and levels of benefits received. Therefore, this reconnaissance is a preliminary effort to set in motion the research that would make possible a complete assessment of equity in federal housing program administration. It is intended as an overview for researchers who may not be familiar with basic federal housing data resources.

Methods

No prior research was available to guide the methods used in this investigation. Although the researchers had experience with some of the data systems and agencies studied, there was no readily available guidance about how to research a large number of diverse federal agencies. Therefore, an exploratory set of methods was developed out of the unique set of events that occurred after each agency was first contacted.

The initial work of this reconnaissance was to define what constituted a federal housing program. Our preliminary list of programs contained those characterized as direct housing expenditures: low-income housing subsidies, federal mortgage insurance and guarantee programs, housing rehabilitation programs, and other homeownership programs.

But many public expenditures on housing—in fact, the largest ones—are indirect and are therefore often overlooked as housing

programs. These include various tax programs—for example, the tax benefits of homeownership and the low-income housing tax credit (LIHTC). Therefore, this reconnaissance of data systems on federal housing programs was broadened to include both direct and indirect expenditures on housing.

The next step was to contact each of the major federal agencies involved in the provision of housing assistance: the Department of Agriculture, the Department of Housing and Urban Development (HUD), the Department of Veterans Affairs (VA), the Resolution Trust Corporation (RTC), and the Department of the Treasury. We then approached other types of housing institutional actors, including secondary mortgage market actors (Fannie Mae, Freddie Mac, and the Government National Mortgage Association [GNMA]) and regulators (the Federal Reserve Board and the Federal Financial Institutions Examination Council).² Since the goal of this research was to obtain information about internal data systems, our initial step upon contacting each agency was to determine the location within the organization of information on data. Although most agencies had public information staff, these people were not equipped to provide information on internal data systems. With few exceptions, agencies had no standard, established processes for providing such information.³ Therefore, finding the people who could talk about the data was a prerequisite to doing this reconnaissance.⁴

All agencies were initially contacted by telephone. Some then asked for a formal letter of inquiry. More often, the agency orally provided names of people to telephone. The formal letters that were sent to agencies asked questions about data, yet the replies usually provided the names and phone numbers of contact people, not information about the data systems themselves.

² We also examined the data systems of a small number of trade and nonprofit organizations to determine whether pursuing such organizations might be a viable avenue for acquiring data if agency data were not available. The organizations were the National Association of Realtors, the National Association of Housing and Redevelopment Officials, the Mortgage Bankers Association, the Housing Assistance Council, and the National Council of State Housing Agencies. Overall, this part of the reconnaissance indicated that such organizations are not the best sources of the data necessary for examining program beneficiaries.

³ A notable exception is the Department of the Treasury, which has staff and materials specifically targeted at providing information and documentation on publicly available data systems.

⁴ A list of people available to serve as references and contacts within each agency is available upon request.

Obtaining agency cooperation in this investigation was difficult, particularly in the early phases. Several agencies' initial response was to suggest filing a Freedom of Information Act (FOIA) request to obtain the information. To file the request, we had to know what to ask for. But if intelligence about the information were in hand, there would be no reason for the FOIA request in the first place because information about the data was the explicit goal of the research. In short, if we knew what to request, the request would be unnecessary. The "FOIA response" became the methodological catch-22 of the investigation.

In other cases, agency staff appeared suspicious and were hesitant to release any information. Getting to the right people and getting them to talk took a lot of time. Since there was no formal mechanism for requesting or obtaining information, obtaining it was often tied to building rapport with agency staff. Once trust was established, the method of learning about data systems was to "snowball" through an organization, gathering information while being passed from person to person. We believe that a detached, bureaucratic approach to obtaining information would have yielded little intelligence; this work required informal networking, not asserting the "right to know."

Ultimately, most people in most agencies were helpful, particularly those in charge of data collection, compilation, and analysis. For agencies for which limited information is available, the omission seems to be due not to the operations of staff themselves but to the overall posture of the agencies. In other words, it appears that people's behavior, in part, reflects the administrative stance of the organization, not the inclinations of individuals working for it. Of course, determining why some organizations have good, bad, accessible, unavailable, or no data remains a separate (and important) research topic.

This data reconnaissance occurred at two points in time—the summer of 1992 and the summer of 1993. Additional work was done in the summer and fall of 1994. Between the first two periods, an important political change took place: the start of the Clinton administration. This change made a huge difference in the availability of information and people's ability and willingness to discuss it. For some agencies, however, the "data chill" remains.

A central goal of this reconnaissance was to determine whether data are publicly available. Surprisingly, the answer to this question was often the most difficult to obtain. Despite repeated requests for answers, an informant would sometimes report that

the agency's position on public access to data was unknown—that no one had ever asked for data, so it was unclear how to respond. At other times, agencies that were in the process of compiling the data had not yet addressed public availability. Still other agencies seemed to want to avoid confronting the issue; they suggested filing an FOIA request. Thus, the data may be available to the public but only to those willing to expend a huge amount of time and effort.

Politics of data availability

An important and not surprising pattern is evident. The decade of the 1980s, including the administration of Ronald Reagan, is a period for which limited and lower quality data are available. Budget cuts, indifference to information, and disbelief in social science were all cited (off the record) as sources of the problem. Conversely, the Clinton administration has brought with it more attention to data collection, systematization, and accountability. In short, the current political climate is more conducive to providing data for research and advocacy related to government housing programs, although this trend varies by agency.

The change in the politics of data availability speaks to the importance of political mood, budget allocations, and political leadership. One source of new data on the characteristics of beneficiaries of federal housing programs has been the U.S. Congress, which has enacted legislation requiring agencies to report information to it. Therefore, an important method of persuading agencies to collect data as well as giving them the opportunity to do so is to pass laws requiring them to give this information to Congress.

Congressional mandate ensures that agencies will collect some data, but it does not guarantee that the data will be put together in a usable, accessible, and publicly available system. Our experience in this reconnaissance underscores the importance of the tone, posture, and priorities of the political administration in charge. Conducting this research under the Bush administration was a very different experience from conducting it under the Clinton administration. Changes in some agencies' attitudes toward data appear to be driven largely by the administration's renewed focus on fair housing and civil rights enforcement and monitoring, for which data are key.

Assessing data quality

Unlike tests of statistical significance or survey response rates, the field of research methods provides little guidance on how to measure data quality. Data collection methods are typically selected because they yield the highest quality data (at least, that is the intent). The utility of the data can be measured on the basis of validity and reliability. But no consensus exists on the parameters that define database quality. There is no platonic ideal that provides even the vaguest criteria for distinguishing good data from bad.

Yet data quality has important ramifications. The findings of any research effort can be criticized, accepted, or disqualified because of perceptions of the quality of the data on which they are based. Data quality can be the Achilles' heel of a research project, particularly one that addresses issues that are socially or politically contentious. Moreover, good data are essential for conducting good research. Therefore, a preliminary assessment of data quality was an important part of this research.

Data quality was assessed in two ways. The first method was to look at previous research experiences. If research had already been conducted with the same data and that research produced some information about the quality of the data, that information was used to assess the quality.

But for most of the data systems under investigation, this was not possible. Many databases were still being compiled or had never been made available to the research community or to the general public. Because experience with these data systems was limited, they had not been independently scrutinized for quality in terms of coverage, data irregularities, coding problems, or general errors.

Data systems that had not been independently evaluated were assessed by a "reputational method." People within the agency and familiar with the data were asked their opinions of the database overall and at various points in time. By asking about data quality over time, we could get some sense of whether data systems were being improved and upgraded or were staying the same.

These assessments represent knowledgeable actors' opinions of the overall quality of the data. In most cases, we could not verify these assessments. Further research and experience using the data will be necessary to confirm them.

Federal agencies

The federal government is involved with housing in a myriad of ways. Rental assistance and public housing are provided to some low-income people in both rural and metropolitan areas. Mortgage insurance and loan guarantees are available for veterans (VA loans) and for people purchasing lower priced housing (Federal Housing Administration [FHA] and Farmers Home Administration [FmHA]⁵ loans). The federal government also provides some home improvement loans and funds for new construction. It assists privately owned rental housing developments as well as providing capital grants and rental assistance for housing the elderly and disabled (Section 202), rental and cooperative housing subsidies and mortgage insurance to reduce mortgage interest rates on rental units for low-income families (Section 236), and mortgage insurance on multifamily housing for low- and moderate-income families (Section 221(d)(3)). Through the LIHTC, it subsidizes the development of affordable rental housing. The federal infrastructure on housing is vast and complex.

The most significant and influential federal housing program is not a direct benefit but an indirect subsidy to homeowners. Administered by the Department of the Treasury, this indirect “program” permits homeowners to deduct mortgage interest and other costs associated with obtaining housing finance from their income when paying income tax. In addition, homeowners may deduct local property taxes and defer payment of capital gains from selling their homes. The loss of revenue (i.e., the cost of this program) is a source of considerable controversy.

Department of Agriculture

The Department of Agriculture has several databases on programs administered through FmHA and is in the process of reorganizing its data systems. FmHA interest in data was sparked by the Housing and Community Development Act of 1987. Beginning in 1989, HUD and the Department of Agriculture were required to assess and report to Congress the extent of each agency’s compliance with the Fair Housing Act (U.S. Department of Agriculture 1992). Complying with this mandate requires data on the race, ethnicity, gender, age, family, and

⁵ In October 1994, the Farmers Home Administration was abolished and its functions transferred to other offices under the Department of Agriculture. Farm lending functions are now handled by the Consolidated Farm Service Agency, and the rural housing program is now administered by the Rural Housing and Community Development Service.

disabilities of program beneficiaries. To this end, the department is working toward having an on-line data system that will compile computerized program application and participation data from each county.

The characteristics of the two major Department of Agriculture databases are summarized in table 1, which includes the programs covered, unit of analysis, purpose of collecting the data, data elements, form of the data, whether and in what form they are available to the public, and the reputed quality of the data. The databases reviewed are the Multiple Housing Tenant File System (MTFS) and the New Farm and Housing Activity Report (NEWFAR).

Multiple Housing Tenant File System. MTFS is a computerized data system that provides information on all of FmHA's rental program beneficiaries. The information is compiled from a tenant certification form (Form 1944-8). This database provides information on race, ethnicity, gender, and household income. In addition, households are characterized as elderly or nonelderly and by whether individual members are disabled. Data are compiled annually for households and organized by county.

No racial data are provided for multifamily loan borrowers because they tend to be organized entities, not households.⁶ FmHA's rationale for excluding borrowers from the MTFS database is that "racial data on multiple family housing borrowers would not be useful to illustrate the racial characteristics of multiple family housing program beneficiaries" (U.S. Department of Agriculture 1992, 13).

Presumably, FmHA considers the *tenants* of these rental units, not the *borrowers* who develop the projects, to be the beneficiaries of these programs.

MTFS has only recently been used as a database. The data have been used in recent reports to Congress but were manually tabulated (U.S. Department of Agriculture 1992, 1993). The data are now being computerized, largely for reporting requirements and as part of the move to develop an on-line data system. FmHA staff consider the current data to be of high quality, but pre-1990 data are regarded as poorer in quality and unreliable.

⁶ A couple of people reported that MTFS contains information on a low-income single-family loan program, Section 502. This suggestion appears inconsistent with all else that is known about the database. There may be some internal confusion about these data.

Table 1. Department of Agriculture Program Beneficiary Data Systems

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Multiple Housing Tenant File System (MTFS)	All FmHA rental assistance programs	Tenants*	Congressional reports	Race, ethnicity, gender, income, elderly, disabilities	Computerized	In hard-copy form	Pre-1990: Poor Post-1990: Good
New Farm and Housing Activity Report (NEWFAR)	All FmHA single-family housing programs	Program applicants	Congressional reports	Race, gender, income	Computerized	No policy; data compilation in progress	Unknown

*Multiple Housing Program borrowers excluded.

The MTFFS data system is computerized, yet according to agency sources, the data in computerized form are not available to the public. People may request data items, which will then be provided in hard-copy (paper) form.

Single-family housing data. Recent reports by FmHA on its single-family loan programs were based on annual tabulations of hard-copy forms (U.S. Department of Agriculture 1992). These data are on FmHA's guaranteed and insured housing loan programs (Section 502), home repair loan and grant programs (Section 504), and other rural single-family housing programs.

For the new on-line system, plans are to create a computerized database containing the location, race, ethnicity, gender, and income of program beneficiaries (borrowers or grantees) and program applicants. The database will include information on loan dispositions (whether an application is closed, withdrawn, denied, or accepted) and eligibility determination, but not on disabilities.

These new single-family loan data are very similar to data reported by financial institutions under the Home Mortgage Disclosure Act (HMDA). HMDA, discussed in detail later, requires that metropolitan lenders report the race, ethnicity, income, and gender of loan applicants. To be sure, FmHA-insured loans made by financial institutions are required to be reported under HMDA. But HMDA is required solely for metropolitan-based lenders that report on loans made within metropolitan statistical areas (MSAs). FmHA loans, by definition, are rural loans, so most are not included in the HMDA database. The creation of a computerized FmHA single-family lending database provides an unprecedented opportunity to look at rural lending and the beneficiaries of rural lending patterns.

These data will be collected as part of NEWFAR, which will replace all other single-family reporting systems. It appears that the planned database will yield high-quality computerized information. It could not be clearly determined from agency sources whether and in what form these data will be available to the public.

Department of Housing and Urban Development

Several HUD data systems were casualties of the Reagan-Bush years. A data system of tenant clients initiated during the Carter administration, the Multifamily Tenant Characteristics System

(MTCS), was virtually suspended. Data were neither routinely compiled nor computerized. This situation continued until 1989, when HUD, under Jack Kemp, began to rebuild its data systems.

Until 1990, HUD was required by the 1980 revisions to HMDA to compile the location of all FHA-insured loans by census tract. This work was discontinued in 1983 and, like MTCS, resurrected in 1989.

HUD is compiling a large number of data systems and has put renewed rigor and staff support into database maintenance. By necessity, HUD's multiple duties have generated multiple data sources. Data exist on rental housing subsidies, homeownership programs, housing rehabilitation programs, and block grant programs.⁷

Tenant and rental housing subsidy data may be found in three databases: MTCS, the Tenant Rental Assistance Certification System (TRACS), and the Housing Subsidy/American Housing Survey (AHS) Data Base (produced jointly with the U.S. Census Bureau). Data on single-family housing and homeownership programs may be found in five databases: the Computerized Homes Underwriting Management System (CHUMS), the FHA Single Family Lending Data Base, the Early Defaults and Foreclosure System, the Home Investment Partnership Act Management Information System (HOME MIS), and the Home Ownership for People Everywhere III Management Information System (HOPE III MIS). Data on the former Rental Rehabilitation Program may be found in the Rental Rehabilitation Program Management Information System (RRP MIS). Data on the Community Development Block Grant (CDBG) program may be found in two databases: the CDBG Performance Report Data Collection System and the CDBG Performance and Evaluation Report Data System.

In addition, a data set is currently being planned on the LIHTC. Although the LIHTC program is administered by the Department of the Treasury, HUD is taking the lead in developing a data system on it.

The characteristics of each of these databases are shown in table 2. Other information on federal housing programs may be gleaned from local Comprehensive Housing Authority Strategy

⁷ HUD is required by Congress to produce annual reports titled "The State of Fair Housing" and "Annual Report to Congress: Civil Rights Data on HUD Program Applicants and Beneficiaries." These reports provide useful introductions to HUD data sets and their limitations.

Table 2. HUD Program Beneficiary Data Systems

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Multifamily Tenant Characteristics System (MTCS)	Public and Indian housing, Section 8 vouchers and certificates	Tenant/household	Monitor Fair Housing Act compliance, congressional reports	Subsidy type, age, race, ethnicity, occupation, income, rent, family size	Partially computerized	Yes; database compilation in progress	Pre-1991: Poor Post-1991: Good
Tenant Rental Assistance Certification System (TRACS)	Section 8, Section 221(d)(3), Section 236, Rental Assistance Program, Section 202, Section 811	Housing unit	Maintain financial controls, monitor Fair Housing Act compliance, congressional reports	Subsidy type, age, race, ethnicity, occupation, income, rent, family size	Computerized	Yes; database compilation in progress	Unknown
Computerized Homes Underwriting Management System (CHUMS)	All HUD-administered single-family mortgage programs	Loan application	Management control system	Age, gender, income, credit history, processing time, down payment	Computerized	No	Good
FHA Single Family Lending Data Base	FHA-insured loans (MSAs only)	Pre-1990: Census tract Post-1990: Loan application	Fair Housing Act enforcement	Post-1990: Race, income, gender, loan amount, loan disposition, census tract, loan purchaser	Computerized	Yes	Pre-1983: Okay 1983-89: Bad Post-1989: Good

Table 2. HUD Program Beneficiary Data Systems (continued)

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Early Defaults and Foreclosure System	FHA-insured loans	Census tract, lender, HUD regional office, zip code	Congressional reports	90-day default, number of foreclosures	Computerized	Yes	Could not be determined
Housing Subsidy/American Housing Survey (AHS) Data Base	Public housing Section 8 vouchers and certificates, subsidized privately owned developments	Housing unit	Augment information on rental housing subsidy recipients	Housing costs/conditions, neighborhood conditions, household characteristics and composition	Hard copy and computerized	In hard-copy form; special arrangement to access computerized data	Good, although quality issues exist
National Tax Credit Data Base Project	Low-income housing tax credit	Projects placed in service	Develop sampling frame for in-depth research and annual updates	Address, date placed in service, number of units, credit amount, tax credit percentage, whether FmHA	Computerized and hard copy	When completed, yes	Unknown

Table 2. HUD Program Beneficiary Data Systems (continued)

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Community Development Block Grant (CDBG) Performance Report Data Collection System	Entitlement: CDBG	Entitled large cities, urban counties	Congressional reports	Budget and expenditures, activity type, national objectives and accomplishments, beneficiaries (race, low/moderate income, female headed)	Computerized	Yes	Not good, varies by year
CDBG Performance and Evaluation Report Data System	State CDBG	Localities	Congressional reports	Planned and actual expenditures, activity type, national objectives, accomplishments	Computerized	Yes	Not good, varies by year
Home Investment Partnership Act Management Information System (HOME MIS)	Home Investment Partnership Act (HOME)	Participating jurisdictions, community-based housing/home partnerships, units, payment	Congressional reports, disburse funds	Budget, project type, beneficiaries (unit type, income, rent subsidy, race/ethnicity, household size and type)	Computerized	Yes	Good

Table 2. HUD Program Beneficiary Data Systems (continued)

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Home Ownership for People Everywhere III Management Information System (HOPE III MIS)	Home Ownership for People Everywhere III (HOPE III)	Grant, property	Congressional reports, disburse funds	Budget, project type, beneficiary type (unit type, income, rent subsidy, race/ethnicity, household size and type)	Computerized	Yes	Good
Rental Rehabilitation Program Management Information System (RRP MIS)	Rental Rehabilitation Program	Project, grantees, units, and payment	Congressional reports, disburse funds	Budget, project type, number of units, beneficiaries (unit type, income, rent subsidy, race/ethnicity, household size and type)	Computerized	Yes	Good

(CHAS) statements and consolidated plans that accompany funding applications to HUD. In addition, an on-line data system is currently being explored with several organizations.

Multifamily Tenant Characteristics System. MTCS is a computerized database describing tenants residing in all housing programs operated by public housing authorities and Indian housing authorities. As of November 1994, recipients of Section 8 vouchers and certificates were being added to the database, which was still incomplete. Data items include information on the type of subsidy, age of household head, race, ethnicity, occupation, income, rent, and family size. MTCS is one of several data systems HUD uses to report to Congress on fair housing compliance (HUD 1990). Data for MTCS are collected from the 5058 Tenant Certification Form. This form is under revision, and the new version is expected to provide more information about HUD tenant beneficiaries.

MTCS data are considered to be most reliable beginning in 1991 but unreliable before then. To maintain confidentiality, no household- or individual-level data (microlevel) are to be made available to the public. Project summaries of the data for the public and Indian housing programs will be available to the public in computerized form.⁸

Tenant Rental Assistance Certification System. TRACS is in the production stage. It is a database designed to track payments and maintain financial controls for all multifamily assisted, privately owned housing, including Section 8 housing programs and other forms of rental assistance. Section 8 programs include new construction, substantial rehabilitation, loan management set-asides, and property disposition set-asides. Other programs included in TRACS are Section 221(d)(3) Below Market Interest Rate, Rent Supplement, Section 236 Interest Reduction, Rental Assistance Program, Section 202, and Section 811 Supportive Housing for Persons with Disabilities. Programs operated by public housing authorities and Indian housing authorities are not included in the database (see MTCS).

TRACS is compiled from information collected on the 5059 Owners Certification Form, which contains information for each tenant on the type of subsidy received; household members' ages, races, and ethnicities; household income and net assets;

⁸To obtain the data (as well as tabulations, cross-tabulations, and custom-designed tapes), contact HUD USER, P.O. Box 6091, Rockville, MD 20851; (800) 245-2691.

occupation of household head; and rent. As currently designed, this data system is not useful for research because of issues of confidentiality. A database will be extracted from TRACS for research purposes.

Computerized Homes Underwriting Management System.

CHUMS is a data system designed to manage the underwriting of all HUD-administered single-family loan programs. It is the system that processes loan applications for each program. The unit of analysis is the loan application. The data include information on the applicant's age, gender, income, credit history, time to process the loan, and down payment.

The data are computerized and are regarded as good quality. The data are not publicly available because they contain information that would violate applicants' confidentiality.

FHA Single Family Lending Data Base. The FHA Single Family Lending Data Base is a database on FHA-insured single-family mortgages in MSAs. It includes no information on HUD-insured multifamily housing programs or on rural FHA loans. As required under 1989 revisions to HMDA, HUD collects lender reports on each FHA loan applicant that include the race/ethnicity, gender, and income of each applicant, the disposition of the loan, which institution (if any) purchased the loan on the secondary market, reasons for denying the loan, and assorted geographic information (including census tract). This information is currently available for 1990 to 1993.

The data are collected for two purposes. First, they are required to be collected under HMDA, and they are sent to the Federal Financial Institutions Examinations Council, which distributes them as part of the larger HMDA database. Second, HUD uses them for fair housing enforcement.

Before 1990, data on FHA loans were collected at the census tract level. No data were collected on individual loan applicants or the disposition of loans. Rather, the data represented the number and dollar volume of FHA loans at the census tract level. These data are available from HUD.

The quality of the pre-1989 FHA data is not regarded as good, partly because the database had to be reconstructed after compilation was suspended in 1983.⁹ The post-1989 data are

⁹ This assessment is based on the personal experience of Anne Shlay, the senior author of this article.

much better, although the database is still regarded as being in need of improvement.

Early Defaults and Foreclosure System. The Early Defaults and Foreclosure System contains data on 90-day defaults and foreclosures on FHA-insured loans. The data are available at various levels, including census tract, lender, HUD regional office, and zip code. The purpose of the data is to report to Congress. The data are computerized and are publicly available. The data quality could not be determined.

Housing Subsidy/American Housing Survey Data Base. The Housing Subsidy/AHS Data Base uses the AHS to determine the household characteristics, family composition, and housing conditions for all recipients of federal rental housing assistance and for income-eligible households that are not recipients. Programs included are public housing, Section 8 certificate and voucher programs, and private project-based assistance programs. This data system was compiled through a creative and constructive partnership between HUD's Department of Policy Development and Research and the U.S. Census Bureau.

The AHS is a biennial (formerly annual) national longitudinal survey of U.S. housing conducted by the Census Bureau. It contains a rich assortment of information on housing (including costs and conditions), household population characteristics, residential mobility, and neighborhood conditions. Population characteristics include race, ethnicity, age, income, education, and family composition.

The partnership between HUD and the Census Bureau was developed because of a serious limitation in the AHS. AHS respondent self-reports on whether respondents received a housing subsidy are often inaccurate. Therefore, the AHS has not been considered a good source of information on the residential situations of households receiving federal, state, or local housing subsidies or living in public housing because the data cannot reliably distinguish nonrecipients from recipients of housing assistance.

To deal with this limitation, the Census Bureau, in partnership with HUD, conducted a follow-on survey and obtained a more complete list of addresses of recipients of HUD rental housing assistance. To construct this database, street addresses from the follow-on survey were matched to addresses of AHS renter

respondents.¹⁰ In this way, the data could be used to conduct research that would more reliably distinguish recipients of federal housing assistance from other households.

Using the 1989 AHS and the new federal housing assistance indicators, a report and set of detailed data tables were compiled (Casey 1992).¹¹ A similar report is being prepared with 1991 AHS data, and reports using 1985 and 1987 AHS data are being considered.

Although these procedures have improved our capacity to assess the conditions and characteristics of households receiving rental housing assistance, two limitations may affect the quality of the data.¹²

First, the data do not represent the entire assisted rental housing stock. For a variety of reasons, HUD's address list included about 80 percent of the housing in either public housing or private developments. In addition, no matches were performed for FmHA, state, or local housing subsidy programs.

Second, mismatches between households' self-reported receipt of subsidies in the AHS and HUD's list raise uncertainties about the reliability of the Housing Subsidy/AHS Data Base. In some cases, households that said in the AHS database that they did not receive housing subsidies were reported as receiving subsidies on the HUD list. Yet in other cases, households that reported that they did receive housing subsidies were not on the HUD address list. These problems suggest that the data collection procedures used in constructing this system may need improvement (Newman and Schnare 1993).

Data aggregated to the national level are available in hard-copy form (Casey 1992). Data on household, housing, and neighborhood characteristics are provided for the following groups: public housing residents, Section 8 voucher and certificate recipients, households in subsidized privately owned developments, and income-eligible households that are not subsidized.

¹⁰ See Casey (1992) for a more specific discussion of the address-matching method.

¹¹ Copies of the report can be obtained from HUD USER, P.O. Box 6091, Rockville, MD 20851; (800) 245-2691.

¹² For a detailed discussion of issues associated with the coverage of the Housing Subsidy/AHS Data Base, see Newman and Schnare (1993, app. A).

The computerized Housing Subsidy/AHS Data Base has not been publicly released because of issues of confidentiality. The Census Bureau has made special arrangements allowing researchers to work with Census Bureau staff to customize computer tabulations of the data.

National Tax Credit Data Base Project. The LIHTC is the largest producer of affordable housing. By offering 10 years of tax credits (direct deductions from tax liability), the program provides indirect subsidies to individuals and corporations that invest in affordable housing developments (Wallace 1994). Projects developed with LIHTC assistance are required to allocate at least 20 percent of the units for households with incomes less than 50 percent of the MSA median or 40 percent of the units for households below 60 percent of median.¹³ Rents may not exceed 30 percent of the applicable income limit (either 50 or 60 percent of median).

The LIHTC is administered by the Department of the Treasury, but federal leadership on developing an LIHTC data system is coming from HUD. HUD has contracted with Abt Associates to assemble a rudimentary database that will eventually permit agencies to monitor this important program. This effort is the National Tax Credit Data Base Project.

The project will assemble data on all properties placed in service through the LIHTC since its inception in 1987. The data will include the address of the property, the number of residential units, the credit amount, the tax credit percentage, and whether the property is FmHA financed. The database is intended to be a sampling frame that will facilitate annual updates and future in-depth research on the LIHTC.

The data will be collected through the state housing finance agencies and will be available to the public in hard-copy or computerized form. Currently, there are no plans to collect data on tenant characteristics. Therefore, in the immediate future, the National Tax Credit Data Base cannot be used to monitor the racial, ethnic, or family characteristics of beneficiaries of the program—only incomes.

As designed, however, the National Tax Credit Data Base could be used to collect such information on LIHTC beneficiaries. Compiling the universe of properties placed in service through

¹³ Wallace (1994) reports that in practice 100 percent of the units are typically for households with incomes less than 60 percent of the local median.

the LIHTC is an essential first step toward collecting additional information.

CDBG Performance Report Data Collection System. The CDBG Performance Report Data Collection System is a computerized database on cities receiving CDBG funds under the large-city entitlement program. Data are currently available for 1979 to 1991. As of November 1994, the 1992 data were being automated. The data are available to the public.

This database contains a wide range of information about the CDBG program at the city and county levels that is taken from each city's grant performance report. The data include information on funding and expenditures, program activities, how the money was spent, and what the program accomplished.

Beneficiary data are also contained in this database but are only required and collected for certain programs. Program beneficiary data include household race and ethnicity, whether the household has a low or moderate income, and whether the household is female headed.¹⁴

No consensus exists on the quality of the data. Some people reported that the data were of good quality, but others said that the data quality was uneven—good in some years and bad in others.

CDBG Performance and Evaluation Report Data System. The CDBG Performance and Evaluation Report Data System is a computerized database for localities receiving funds under the state CDBG program. State governments receive CDBG funds and distribute them to localities. States report to HUD using a grant performance and evaluation report, which is used to compile the data system. The system contains information for 1982 to 1993 and is available to the public.

Data are reported by locality—specifically, the political entity that receives the CDBG funds—and include information on program expenditures, activities, national objectives, and accomplishments (e.g., number of units, number of jobs, persons

¹⁴ Under contract to HUD, a two-year evaluation of the CDBG entitlement program is currently being conducted by The Urban Institute. Using the CDBG Performance Report Data Collection System, the evaluation is examining local implementation of the CDBG program from 1982 to 1990. It is addressing local funding patterns, management strategies, and program beneficiaries. In addition, through site visits and other field research, it is addressing the local impact of the program.

served). Localities are not required to report beneficiary information, but some report it voluntarily. No beneficiary data are included in the automated database.

HUD management information systems. Three databases are set up as management information systems that are used for several purposes: the HOME MIS, the HOPE III MIS, and the RRP MIS. Each data system is used for reporting and research as well as for administering the program. They are updated regularly. The key to each system is that payments of project funds to participating jurisdictions are linked to the processing of information. Funds are not distributed until a complete set of forms has been received for each program and data have been entered into the system. Once the information is complete, the system triggers payments from the federal government to the participating jurisdiction.

This system advances several goals related to database management. First, it ensures that the data are timely because linking fund disbursement to data entry is a powerful motivator for providing current and complete information. Second, it should lead to better quality data because participating jurisdictions will use reporting practices that are least likely to delay payment. Hence, these systems appear to contain good-quality as well as current data.

The HOME MIS, HOPE III MIS, and the RRP MIS are computerized systems that contain information on their respective programs. Data include information on project expenditures and budget, project type, types of rental assistance, and program beneficiaries (race/ethnicity, household size, household type, and income). The RRP no longer exists, and the current management system is disbursing the remaining funds in the budget.

Comprehensive Housing Affordability Strategy and the consolidated plan. In fiscal year 1994, each local political jurisdiction was required to submit a CHAS to apply for funding from HUD. Each CHAS is intended to represent the local area's strategy for using the requested funding. The CHAS may be a useful source of information on housing programs at the local level. A wide range of programs must be addressed within each CHAS.¹⁵ To

¹⁵ These include HOME, HOPE I, HOPE II, HOPE III, HOPE for Youth, CDBG, Low Income Housing Preservation Section 601, Shelter Plus Care, Section 202, Section 811, Emergency Shelter Grants Program, Safe Havens for Homeless Individuals Demonstration Program, Supportive Housing Program,

facilitate access to these documents, the national HUD office asked each field office to submit local CHAS documents so that they could be maintained in a central depository. For fiscal year 1994, some CHAS documents are available at the HUD main library.

For fiscal year 1995, the rules governing jurisdictions' applications for funds at HUD are being revised to streamline the application process. These new rules will include the submission of the CHAS under a new "consolidated plan," which combines community development plans and CHAS in one document.

Local political jurisdictions are being provided with software designed to help them prepare these plans and submit them in electronic form. Although use of the software is voluntary, many of these plans could be available in computerized form. By November 1994, no decision had been made on whether these consolidated plans would be collected from the field offices by the national HUD office as the CHAS documents are.

Department of Veterans Affairs

The VA has a loan guarantee and insurance program. VA loans to veterans are made not directly by the VA but by private financial institutions. When asked about its data system, the VA suggested filing an FOIA request. Despite repeated attempts, nothing significant could be learned about data available in the VA.

Data on VA loans, though not available directly from the agency, are part of the HMDA database compiled by the Federal Reserve Board. All metropolitan lenders are required to disclose information on all loan applicants by loan type. Therefore, the HMDA database is a good database on VA loans. For each VA loan applicant, data are available on race, ethnicity, gender, and income. In addition, each record contains the census tract of the housing unit, the disposition of the loan, secondary mortgage market purchasers, and reasons for denying the loan. But VA loan data reported under HMDA are limited to metropolitan areas and do not include rural VA-guaranteed and VA-insured mortgage loans.

Moderate Rehabilitation SRO Program, Housing Opportunities for Persons with AIDS, Rural Homelessness Grant Program, and Revitalization of Severely Distressed Public Housing Program.

Resolution Trust Corporation

RTC is the government agency created under the Financial Institutions Reform, Recovery, and Enforcement Act of 1989 (FIRREA) to manage selling off assets and generally disposing of failed savings and loan institutions. The housing program component of the RTC mission is the Affordable Housing Program (Ely 1990).

The Affordable Housing Program is designed to promote low-income homeownership and to help provide more affordable rental housing (Johnson 1990). To this end, RTC sells housing in its inventory at below market value and with favorable financial terms. The guidelines indicate that single-family homes are to be sold to households with incomes not exceeding 115 percent of the area median. For multifamily dwellings, eligible purchasers include for-profit, public, and nonprofit agencies that agree to set aside at least 35 percent of the rental units for low-income and very low income residents under long-term deed restrictions.

This reconnaissance of RTC's data systems did not yield complete information. RTC maintains the Real Estate Owned Management System (REOMS) database (table 3), which contains information on the sold property location; the purchase date; the buyer name; the broker or asset manager; the institution from whose books the property was sold; and some combination of current book value, appraised value, and list price (Southern Finance Project 1992). RTC's monthly paper, the *Silver Lining*, indicates that data on households are available, including income, race, marital status, family size, first-time home-buyer status, and veteran status.

Our understanding is that these data are obtained through exit polling of auctions where assets are sold. Until recently, racial data reporting was voluntary, so such data were often missing. But now racial reporting on RTC buyers is mandatory.

No information could be obtained on whether RTC maintains data on the characteristics of tenants living in multifamily buildings that it sells. Apparently, REOMS contains information on multifamily purchasers but not necessarily on tenants. Repeated telephone calls to RTC were not returned.

Data may be obtained from RTC only through an FOIA request. Although the REOMS database is computerized, the data are alleged to be publicly available as hard copy only.

Table 3. RTC Program Beneficiary Data System

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Real Estate Owned Management System (REOMS)	Affordable Housing Program	Sold property	Monitor property inventory	Location, purchase date, buyer, broker, book institution, value, income, race, marital status, family size, first-time home-buyer status, veteran status	Computerized	File Freedom of Information Act request for hard-copy data	Poor/unknown

The data have a reputation for poor quality. One research report that used the data reported that the listings “clearly were marred by omissions and errors” (Southern Finance Project 1992, 8).

Department of the Treasury

The Treasury Department is the federal agency that contains the Internal Revenue Service (IRS), which is charged with administering federal tax policy. The IRS receives all filed federal income tax returns and collects income tax revenues. Therefore, it is an important source of information on the economic situations of households.

The Treasury Department also administers the LIHTC, but it maintains no data on that program. As mentioned earlier, Abt Associates is developing a database on the LIHTC under contract to HUD.

IRS Individual Income Tax Return Public Use File. The IRS, through collecting federal income taxes, is charged with administering one of the federal government’s largest indirect housing subsidies—the homeowner’s deduction of mortgage interest from taxable income—as well as other indirect benefits, including the deduction of local property taxes and the deferral of capital gains. These deductions result in significant tax savings for many homeowners and can be important economic incentives for homeownership, but they represent a huge cost to the government in lost revenues. Therefore, determining who receives these indirect benefits is important.

The IRS maintains a publicly available data system, the Individual Income Tax Return Public Use File, which contains individual tax returns. It consists of two subsystems: the Super File and the Regular File (table 4).

The data system is “disclosure proofed”: Social Security numbers and other identifying information are not included in the files.¹⁶ A state identification number is provided for all individuals and households with adjusted gross incomes less than \$200,000. To protect the anonymity of high-income people, no identification number is provided for those with incomes greater than \$200,000.

¹⁶ Data files with Social Security numbers are available for sale to state governments.

Table 4. Treasury Department Program Beneficiary Data Systems

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Individual Income Tax Return Public Use File: Super File	Mortgage interest deductions, property tax deduction, deferred capital gains	Individual tax return, 1985 and 1988	Research and policy analysis	All items on tax returns reported to IRS excluding identification information*	Computerized	Yes, for \$10,000 per year	Good
Individual Income Tax Return Public Use File: Regular File	Mortgage interest deductions, property tax deduction, deferred capital gains	Individual tax returns, 1966 to present	Research and policy analysis	All items on tax returns reported to IRS excluding identification information*	Computerized	Yes, for \$2,150 per year	Good

*State identification number reported only for individuals with adjusted gross incomes less than \$200,000.

Both the Super File and the Regular File contain samples of individual tax files. The samples are randomly selected from different income strata. The major differences between the files are the sample size and the time frame covered. The Super File is available for two years: 1985 and 1988. It contains 364,562 records of households and individuals with adjusted gross incomes less than \$200,000. Data on higher income households are provided only in summary form. The cost of the Super File for each year is \$10,000.

The Regular File is available for years from 1966 on. It contains varying numbers of records per year, typically between 80,000 and 100,000. Data for 1966 to 1978 are available from the National Public Archive for \$90 per year. Data from 1979 on can be obtained from the IRS for \$2,150 per year.

Each record contains the complete set of information from the federal income tax forms filed with the IRS by individuals and households. This includes all information on income, deductions, capital gains, tax paid, and refunds.

No data are provided other than that on each tax return. Because there is no information on the race, ethnicity, gender, or disabilities of individuals, the data cannot be used to directly measure these characteristics for beneficiaries of the mortgage interest deduction and other indirect subsidies to homeowners administered by the Treasury Department.

Data from the Regular File can also be found in hard-copy form in a report produced by the IRS (1991). The report provides information on returns filed and sources of income, exemptions and itemized deductions, tax computations, and high-income returns. The IRS also issues a quarterly bulletin that provides additional analyses and information (IRS 1993).

Financial regulatory agencies

Among the most popular federal housing benefits are federal loan guarantee and insurance programs, which protect private lenders against loss associated with mortgage default. Although many state governments are implementing programs, most loan guarantees originate with the federal government within three separate agencies: HUD (FHA loans), the VA (VA loans), and the Department of Agriculture (FmHA loans). The major source of information on these programs is the HMDA database.

HMDA data

Lender reports on the census tract location of their residential lending have been an important source of information on the investment activities of private financial institutions since the late 1970s, when HMDA was passed (Shlay and Goldstein 1993). Before 1990, these data contained only information on lending in a census tract and not to individual households. Therefore, these data could not be used to investigate the characteristics of individuals and households receiving or not receiving credit.

To rectify this limitation, as part of amendments to HMDA in FIRREA, lenders were required to report information on all mortgage loan applicants beginning in 1990. The characteristics of the post-FIRREA database are summarized in table 5.

Institutions required to report under HMDA include federally regulated depository institutions with assets of \$10 million or more and mortgage bankers and subsidiaries. Data include the lender regulator; the race/ethnicity, income, and gender of the loan applicants; the disposition of the loan; the loan amount; the loan purpose (single family, home improvement, refinancing, multifamily); the loan type (conventional, FHA, VA, FmHA); the loan purchaser (if purchased); and the census tract, county, and MSA of the property for which the application was made. In addition, the lender may voluntarily report up to three reasons for denying a loan application.

To compile this database, each lender reports the application information on a loan application registrar (LAR), which is sent to the appropriate regulatory agency (e.g., HUD, the Federal Reserve Board, the Office of the Comptroller of the Currency) for review. Each regulatory agency sends the data to the Federal Reserve Board, which compiles and computerizes the database for the entire United States.

The Federal Financial Institutions Examination Council is responsible for distributing HMDA data. Data for one year for the entire United States can be purchased for \$500. Data can be purchased for single MSAs; the cost varies with the size of the data set.¹⁷ Three sets of data are available: (1) the LAR, (2) a reporter directory, and (3) census tract housing and population data. The LAR data do not contain reporting lenders' names. To

¹⁷ The cost of the data depends on the number of diskettes required for storage. For 1 to 4 diskettes the cost is \$25; for 5 to 9 diskettes, \$75; for 10 or more, \$150 (Center for Community Change 1993).

Table 5. Financial Regulatory Systems Program Beneficiary Data System

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Home Mortgage Disclosure Act (HMMDA)	FHA loans, VA loans, FmHA loans (MSAs only)	Loan applicant	Regulate Community Reinvestment Act and Fair Housing Act compliance	Race, income, gender, loan type, loan purpose, loan amount, loan disposition, census tract, loan purchaser	Computerized	Yes	Varies, improving

Note: Reporting lenders under the following regulatory agencies: Office of the Comptroller of the Currency, Federal Reserve Board, Federal Deposit Insurance Corporation, Office of Thrift Supervision, National Credit Union Association, and HUD.

determine the name, a unique identification number in the LAR file can be linked to the reporter directory, which contains the names and addresses of each reporter. To determine the social and economic geography of the location and disposition of the loans, LAR data can be linked to U.S. census tract housing and population data.

HMDA data accessibility. HMDA data are publicly available. Indeed, the HMDA legislation originated with the goal of making financial institutions more responsive to local communities. Nonetheless, controversy over whether HMDA data are accessible abounds. When HMDA was implemented in the late 1970s, the information was difficult to obtain because it had to be requested from each financial institution. Amendments to HMDA continue to work on making data more accessible, but making them more accessible has not been a straightforward and logical process. Amendments to HMDA in 1980 led to the creation of central depositories in each MSA responsible for maintaining hard-copy HMDA reports and to the construction of computerized HMDA databases. Yet some depositories were not informed about the nature of the data and therefore were not particularly effective.¹⁸

Computerization of HMDA data was intended to improve access to the data as well. But computerization made the information more useful only to those who had access to a mainframe computer, because during the early 1980s personal computers were neither plentiful nor powerful. Since the augmentation of HMDA under the 1989 FIRREA amendments, the national HMDA database has been huge, consisting of several computer tapes for one year alone, thereby further complicating accessibility.

Maintaining public access to these data in a reasonable format has proved an enormous challenge for regulators. Activities being pursued to improve HMDA's accessibility include distributing the data better, developing user-friendly software to access and analyze the data, and putting the data on line and even on the Internet.

As of November 1994, the Federal Financial Institutions Examination Council was introducing a new system for accessing HMDA data. The Home Mortgage Data Analysis Report System, or HOMER, is intended to be a tool to aid in the analysis of HMDA data.

¹⁸ Some depositories, apparently unaware of their responsibilities, actually threw away their local HMDA data.

The Center for Community Change has nearly completed the design of the HMDA Query and Reporting System, personal computer software for producing standard reports and customized analyses of the HMDA data in a menu-driven environment. The center will provide HMDA data with the software.

A number of nonprofit organizations, several foundations, and HUD are currently developing a user-friendly, on-line data system that would enhance the availability of HMDA and AHS data and possibly other databases as well.¹⁹ Using such data, at present, often requires access to tape drives, typically a main-frame computer, a substantial amount of time to set up the database, and a fair amount of technical sophistication. An on-line system would permit users to dial up the system through a modem to access and manipulate the data.

As of November 1994, Essential Information, a nonprofit organization, was preparing to place HMDA data on the Internet. This effort would make the data available to the growing number of people who are able to connect to the Internet.

The initial plans are to post a series of prepared data tables. One set of tables will contain information by lender. Another will contain aggregate information by census tract. Some summary data will be provided as well. These data will be posted for each MSA. Essential Information will have available computer mapping software (Archview 2) for nonprofit fair housing groups to examine the spatial patterns associated with the data.

Essential Information is also preparing to develop a procedure to use the HMDA data interactively from the Internet. This would permit people to do more of their own analyses of the data. Plans are also under way to expand the HMDA Internet node to include call report and bank branching data.

HMDA data quality. There is no consensus on the quality of HMDA data. Each regulatory agency is required to monitor the quality of the LARs received from their reporting institutions. Mistakes and errors are expected to be resolved before the data

¹⁹ These groups include ACORN (the Association of Community Organizations for Reform Now), OMB Watch, Unison Institute, the Center for Community Change, the Community Reinvestment Coalition of North Carolina, the National Council of La Raza, the Pittsburgh Community Reinvestment Coalition, the National Urban League, the National Fair Housing Alliance, and the National Community Reinvestment Coalition. The major foundations participating have been the Surdna Foundation and the Bauman Foundation. HMDA researchers have also participated.

are submitted to the Federal Reserve Board. But it has been reported that errors are not always rectified, often because (according to regulators) lenders fail to cooperate.

The Federal Reserve Board reports that it cleans the data before release. Yet numerous users have reported extensive problems with “dirty” data, including incorrect census tract numbers and impossibly high or low values.²⁰ There have been reports of poor coverage, particularly of FHA lending. Errors or high values may be attributable to sloppy reporting by lenders or lack of follow-up by regulators.

Nonetheless, the quality of the HMDA database remains an open question. There is evidence of some effort to improve HMDA coverage and quality, although, clearly, greater attempts must be made. Also unclear is the scope of the problem relative to the size of the database. The pattern of errors is important as well: Are they randomly distributed through the database, or do they relate to some underlying dimension within the database?

HMDA and federal loan guarantee programs. HMDA data include measures of the type of loan for each loan applicant. Loans are classified as either conventional or government insured (FHA, VA, or FmHA). Therefore, the data may be used to determine the racial, economic, and gender characteristics of people receiving each type of government-insured loan within metropolitan areas. In addition, when combined with census tract data, the data allow us to examine the spatial patterns associated with insured loans.

HMDA and the secondary mortgage market. In addition to reporting the disposition of each loan applicant, HMDA reporting provides loan purchase data. For each loan, the type of loan purchaser is reported: Fannie Mae, Freddie Mac, GNMA, FmHA, commercial bank, savings bank or savings association, life insurance company, affiliate institution, or other. It is also noted if a loan is *not* sold in a given calendar year.

HMDA data also include the race, ethnicity, income, and gender of each loan applicant. The availability of data on the characteristics of loan applicants and loan purchasers means that it is possible to identify the racial, economic, and gender dimensions of the secondary mortgage market. Moreover, since HMDA data also include the census tract of the property, a

²⁰ The existence of dirty HMDA data was discovered through personal communication with HMDA users as well as Anne Shlay’s lengthy experiences with cleaning HMDA data.

spatial analysis of secondary mortgage market activity can be performed.

HMDA data do not capture the entire pool of secondary market loan purchases, in part because HMDA reporting is restricted to metropolitan regions. HMDA reporting has captured much less than the purchase activity by Fannie Mae (Wilson 1994). There is no reason to expect that the coverage for Freddie Mac would be much different. Whether these missing data differ substantially from those included in the HMDA database is currently unknown.

Secondary mortgage market actors

A review of potential data sources from secondary mortgage market actors is included here largely because these actors play a prominent role in enhancing (or limiting) access to homeownership and because their activities can influence public programs, particularly loan guarantee programs. Moreover, a major focus of public policy is homeownership, which is indirectly and heavily subsidized. Therefore, this reconnaissance examined data sources maintained by secondary mortgage market actors that may provide information on who benefits from secondary mortgage market activity.

Three institutions were contacted: the main purchaser of government-insured mortgages (GNMA) and the two main private mortgage purchasers (Freddie Mac and Fannie Mae). GNMA reported that it maintains no databases that can be used to study beneficiaries of housing programs. Apparently, the current best source of information on the racial, gender, and economic characteristics associated with GNMA-purchased loans is HMDA data.

Fannie Mae and Freddie Mac have worked together to develop internal databases that provide information on an array of useful characteristics associated with loans that they purchase. The collection of these data is required under the Federal Enterprise Safety and Soundness Act of 1992. Under this legislation, Fannie Mae and Freddie Mac must report to HUD, beginning in 1994, the racial, economic, gender, and first-time home-buyer characteristics associated with all loans that they purchased after January 1, 1993.²¹

²¹ Reporting data collected on loans purchased before January 1, 1993, is optional; the agencies are required to use their "best effort" to collect data on such loans.

Fannie Mae and Freddie Mac have worked together to design similar systems and have collected identical data. The characteristics of the agencies' databases are summarized in table 6. The data are collected for all purchased loans, including rural loans, so they cover a larger geographic universe than HMDA data.

Metropolitan data are geographically coded (geocoded) by census tract. Rural data are geocoded by block number. Data items are the number of borrowers, gender of borrowers, race of borrowers, age of borrowers, monthly income, first-time home-buyer status, date of mortgage note, year house built, number of bedrooms, appraisal amount, purchase price, monthly housing expense, monthly debt expense, and eligible rents (for investment properties).²² In addition, data are collected on whether the down payment, closing costs, or secondary financing is provided as a subsidy from the federal government, from the state or local government, or from another source.

The data are computerized. As of November 1994, the 1993 data have been submitted to HUD. HUD may release summary tables of the data but may not release any of the individual loan records. It has signed a nondisclosure agreement to withhold the data from the public to preserve confidentiality, although a limited amount of data will be made available through HUD. Both Fannie Mae and Freddie Mac report that they are seeking 100 percent compliance in reporting.

Conclusion

This reconnaissance of data systems on the beneficiaries of federal housing programs points to the emergence of new opportunities for research on fair housing and equal opportunity. New data systems are being developed. Some promise to become state-of-the-art systems; even on-line systems are being explored. The availability of data and their increased accessibility to the public mean that now is the time to learn more about the business of politics—who gets what and why.

It is widely recognized that federal data systems need dramatic improvements in the quality of reporting, in coverage, and in database format. As the revolution in communications technology continues, it is a telling illustration of budget priorities that massive information systems at the highest level of government

²² Age of borrowers and year house was built are reported at the lender's option.

Table 6. Secondary Mortgage Market Program Beneficiary Data Systems

Data System	Program Coverage	Unit of Analysis	Purpose of System	Data Elements	Form	Public Availability	Quality
Fannie Mae Purchased Loan Data Base	All loans purchased by Fannie Mae	Purchased loan	Report to HUD	Gender, race, age, income, first-time home-buyer status, year house built, sale price, appraisal amount, census tract or block number, number of borrowers, date of mortgage note, eligible rents, housing expense, debt expense, number of bedrooms, presence of subsidy	Computerized	A limited amount	Unknown; data compilation in progress
Freddie Mac Purchased Loan Data Base	All loans purchased by Freddie Mac	Purchased loan	Report to HUD	Gender, race, age, income, first-time home-buyer status, year house built, sale price, appraisal amount, census tract or block number, number of borrowers, date of mortgage note, eligible rents, housing expense, debt expense, number of bedrooms, presence of subsidy	Computerized	No	Unknown; data compilation in progress

remain in hard-copy form, requiring manual tabulation. Luckily, this situation appears to be changing.

But simply having data in a format so that they *can* be used efficiently does not mean that they *will* be used. Information must be readily available on the existence of data, their form and contents, and where and from whom they may be obtained. Determining the availability of data on public expenditures should not require a full-fledged research project or rigorous searches of complex bureaucracies. It should not be necessary to seek out the right people to query and hope that they will cooperate. Finding data should not be a time-consuming nightmare of following paper trails.

Major improvements are being made to data systems, but it is important to recognize that the commitment to maintaining these systems may be fragile and unable to withstand changes in the political administration. Data systems maintained within government agencies may fall victim to the vicissitudes of politics and budget cuts. Yet the ability to account for beneficiaries of programs as important as housing should not depend on politics. In a democratic society, good-quality, accessible information systems about public expenditures are not frills; they are fundamental components of a monitoring system needed to ensure equity and protect civil rights.²³

If agencies neither can nor will provide the information themselves, a central public information office with a strong public presence and mission could be charged with this responsibility. A central database repository could be created as well.

Beneficiaries' privacy rights must be protected in the process of disseminating information. This research does not suggest that data should be distributed so that individuals' confidentiality is violated. Yet technical issues related to privacy should not hinder access to information about public spending.

Many research topics can be explored with these newly discovered, uncovered, reconfigured, and underused data systems, but for the moment, several broad areas are suggested:

²³ Data that have resulted from HMDA in 1975 are one of the few data systems to withstand changes in several political administrations. There are two reasons for their staying power. First, HMDA data are required to be collected by federal law, and the legislative mandate is to collect and disseminate data, not to report to Congress. Second and most important, HMDA data have a large, articulate, and well-organized community constituency.

1. Characteristics of beneficiaries of rural rental housing programs using the newly compiled MTFS
2. Characteristics of beneficiaries of FmHA rural single-family lending programs using NEWFAR
3. Characteristics of beneficiaries of urban rental housing programs using the improved MTCS
4. Characteristics of home buyers in the Affordable Housing Program administered by RTC using REOMS
5. The racial, ethnic, and spatial dimensions of the secondary mortgage market using data reported under HMDA
6. Characteristics of beneficiaries of metropolitan federal loan guarantee and insurance programs using data reported under HMDA

In addition, data quality must be seriously investigated. This reconnaissance suggests that data quality remains a big issue for many agencies. The void in experience with many of these systems among people in the research community means that the strengths and weaknesses of these data are unknown. Many databases are developing a reputation for quality; whether they live up to their reputation when used for research applications is an important question.

The partnership between HUD and the U.S. Census Bureau has led to the production of a much-needed database on participants and income-eligible nonparticipants in federal rental housing programs. Such endeavors should be promoted and expanded.

Are data systems being put into place that will permit research on program beneficiaries for the largest and most important programs? This reconnaissance suggests a tentative answer to this question, depending on the criteria for defining important programs. Certainly, data systems are being configured that will permit more research on the beneficiaries of low-income housing programs. For many low-income people, these programs are key to maintaining housing stability and preventing homelessness.

Yet direct expenditures on low-income housing programs are meager compared with many of the indirect housing programs, including those that subsidize homeownership through preferential tax treatment of mortgage interest and capital gains and the

indirect subsidies that go to developers of low-income housing through the LIHTC. No data system is currently in place or being put into place that will permit any assessment of the racial characteristics of beneficiaries of programs administered by the Treasury Department through the IRS. Therefore, these programs cannot be monitored to ensure compliance with the Fair Housing Act. The gulf in our information systems on the side of indirect expenditures is enormous.

Except for information on guaranteed and insured mortgages from HMDA, data being compiled appear to be better for monitoring poor beneficiaries than more affluent beneficiaries. Household characteristics are being collected much more systematically for programs targeted to low-income households than for those targeted to higher income households.

This bias in federal data systems directs our attention toward poor beneficiaries, who are often the targets of public criticism, and away from wealthier beneficiaries, who may go unnoticed as receiving largesse from the public trough. This focus on the poor appears to have political and ideological ramifications.

Expanded data systems will permit more complete reporting by federal agencies to Congress, which should lead to more efficient management of programs and fairer distribution of benefits. Data must be made available to researchers outside of government to foster a renewed dialogue on ensuring that benefits flow fairly to those most in need.

Authors

Anne B. Shlay is Associate Director of the Institute for Public Policy Studies and Associate Professor of Geography and Urban Studies at Temple University. Charles E. King is a Ph.D. candidate in Political Science at George Washington University.

This research was funded by the Poverty and Race Research Action Council and the Office of Housing Research of Fannie Mae. The authors are grateful to the many informants who took time to share their information; to Margery Turner, John Goering, Chester Hartman, Steven Hornburg, Florence Roisman, and an anonymous reviewer for helpful comments; to Barbara Urso for computer assistance; and to Emma Rose Giloth for providing perspective along the way.

References

- Casey, Connie H. 1992. *Characteristics of HUD-Assisted Renters and Their Units in 1989*. Washington, DC: U.S. Department of Housing and Urban Development.
- Center for Community Change. 1993. *CRA Reporter* 4(2 and 3).
- Ely, Bert. 1990. The Resolution Trust Corporation in Historical Perspective. *Housing Policy Debate* 1(1):53–78.
- Internal Revenue Service. 1991. *Individual Income Tax Returns, 1988*. Washington, DC: Statistics of Income Division.
- Internal Revenue Service. 1993. *SOI Bulletin: A Quarterly Statistics of Income Report*. Washington, DC: Statistics of Income Division.
- Johnson, Sara E. 1990. RTC's Affordable Housing Program: Reconciling Competing Goals. *Housing Policy Debate* 1(1):87–130.
- Massey, Douglas S., and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.
- Newman, Sandra J., and Ann B. Schnare. 1993. Last in Line: Housing Assistance for Households with Children. Paper read at the Fannie Mae Annual Housing Conference on Housing Policies for Distressed Urban Neighborhoods, June, Washington, DC.
- Shlay, Anne B., and Ira J. Goldstein. 1993. Proving Disinvestment: The CRA Research Experience. Institute for Public Policy Studies Working Paper No. 1. Philadelphia: Temple University.
- Southern Finance Project. 1992. Fortunate Sons: Three Years of Deal Making at the Resolution Trust Corporation. Briefing paper. Charlotte, NC.
- U.S. Department of Agriculture. 1992. *The Beneficiaries of USDA Housing Programs in Rural America: Racial and Ethnic Characteristics of Farmers Home Administration Housing Program Applicants, Borrowers, and Tenants*. Washington, DC.
- U.S. Department of Agriculture. 1993. *Initial Results of FY 1992 MFH Occupancy and Fair Housing Survey*. Farmers Home Administration mimeo. Washington, DC.
- U.S. Department of Housing and Urban Development. 1990. *Who Benefits? Civil Rights Data on HUD Program Applicants and Beneficiaries*. Washington, DC.
- Wallace, James E. 1994. Financing Affordable Housing in the United States. Paper read at the Tri-Country Housing Conference on Social Housing in the Year 2000, June, Kenilworth, England.
- White House. 1994. *Executive Order on Fair Housing*. WL 11403, January 17.
- Wilson, John. 1994. Interview by Anne B. Shlay. June 10, Washington, DC.

