Part I. Development of Policy on Collection of Race Information

A. Development. The legal/policy considerations underlying the collection of information about race by the Federal government are complicated by divergent principles, even though there is general consensus on ultimate goals of policy. On the one hand, policies are sensitive to the possibility of covert discriminatory abuses that can accompany racial identification of individuals on administrative records, for example, personnel records. At the same time, policies of non-discrimination and affirmative action require reasonably accurate information to measure the degree to which principles of non-discrimination are put into practice. The problem is to obtain good information about race of individuals, and at the same time neutralize that information to prevent its use in ways that could discriminate among those individuals.

The pragmatic solution—simple and sensible in concept, but complex in application—has been to keep records about individuals for administrative purposes (such as personnel records) separate from records kept for reporting statistical data about race for compliance purposes. This principle of separate record-keeping has dominated in the area of civil rights and equal employment compliance efforts, and has been the rule for the Equal Employment Opportunity Commission's (EEOC) reporting requirements. In a rudimentary way, this separate record keeping paralleled the practice of researchers and statisticians. The latter have a long tradition of protecting individual statistical data from administrative use. The Privacy Protection Commission (PPSC) in its report which grew out of two years of study, recommended legislation based on the formal concept of "functional separation" which was predicated on these earlier principles, and expressed in a cohesive set of rules (3).

It has happened that many of the statistical counts for civil rights reporting of the racial composition of particular groups have been rather primitive in form, and difficult or even impossible to verify, because identification of individual persons is avoided in those records. With the movement of compliance activities into the courts, opposing counsel have had to prove or disprove compliance with civil rights law on the part of particular organizations, or calculate suitable damages to individuals when those organizations have not complied. Considerable frustration with the quality of available data has been expressed by both groups of parties in such cases. In some Title VII (6) job discrimination cases, courts have criticized trial lawyers for their failure to bring good statistical evidence which can be verified in support of the allegations or denials. The lawyers have often employed as expert witnesses statisticians who have technical ability to perform sophisticated analyses to support their arguments. However, the raw data on race may be crude or in part missing, and the statistical results have often been expressed with levels of confidence which are unacceptable for purposes of evidence.

Much litigation has centered on class action cases which seek to show long-standing racial patterns of employment practice which are inconsistent with the racial composition of the particular job market from which employees are hired. Administrative records which provide the structure for the statistical proofs—for instance, records of job applications, testing, hiring, personnel evaluation, and promotion—ordinarily do not contain information on the race of individuals involved.2/ In search for a link between the administrative/personnel records and the statistical comparisons that would prove or disprove forbidden practices on a continuing basis, the Social Security Administration (SSA) has been approached as a potential source. The underlying records about individuals in SSA's records contain their names and ordinarily their social security numbers. The complainants may be employees representing groups of employees, or applicants representing groups of applicants, all of whom presumably have obtained a social security number (SSN) by making application to SSA. The application form for an SSN has regularly included a statistical question relating to the race of the applicant. Not surprisingly, therefore, SSA has been the focus of considerable attention and pressure on this issue—the question of basing evidence on individual records linked by the social security number to the information on race which SSA collects from individual applicants for its agency statistics.

The Federal government has been on both sides in civil rights litigation dealing with racial discrimination, and on both sides of the resulting question of appropriate use of race information for compliance. EEOC, as the agency primarily charged with Title VII enforcement, ordinarily has the task of proving discrimination on the part of private employers; and the Federal agencies have a general affirmative action duty to perform to prevent prohibited discrimination. On the other hand, the roles are sometimes reversed, and Federal agencies themselves, not excluding the EEOC, may be charged by employees or applicants with failure to uphold the non-discrimination rules in their own employment and related practices. The Justice Department, as the Federal government's principal law firm, may be asked to exercise responsibilities with respect to discrimination charges, or defenses, or efforts to establish guidelines, depending on the particular circumstances. It is clear that the Federal government as a whole has important concerns in balancing the competing public and private interests in the information about race which it collects, and that some of these concerns center on the Application for a Social Security Number, where statistical information on race is collected for storage in individually identifiable and retrievable form.

B. The SS-5: Nature of the Form. Race information is collected by SSA on its Form SS-5, Application for a Social Security Number. The SS-5 has long been an unusual form among Federal forms, particularly among application forms. Not the least of its unusual attributes is the fact...
that race information has been permitted to be collected on it at all. Indeed, so strong is the government's resistance to collecting race information on application forms, that for a short period in the history of the SS-5 when the Internal Revenue Service (IRS) had responsibility for its collection, that agency declined to include the question on race, and left a small but permanent gap in the statistical base.

Race information is linked to identity information collected from each applicant on the Form SS-5, but the race information is not related in purpose to that other information, nor used with it on an individual basis by SSA in managing its program operations. These circumstances also distinguish this form from other application forms. Because race data is collected for a different purpose from that of the other items of information on the form, and the stated reasons for collecting it generate a different set of expectations in the persons providing it voluntarily for statistical use, the disclosure rules for this item are also different from those that generally apply to release of agency data. Besides the usual confidentiality rules which apply to identification and other administrative data, the race information is subject to additional legal/policy constraints which narrowly restrict its use to statistical linkage or summary statistics.

C. The Enumeration Purpose of the Form SS-5. Apart from the race item, the SS-5 itself is distinctive. Although it is called an application form, yet the applicant expects and receives nothing in return for filling it out, other than a card indicating that SSA has assigned a unique number to a record which it creates for the express purpose of holding the identifying information supplied on the form. At the time of initial application, the applicant's primary uses of the account number may even be largely unrelated to SSA and its program—e.g., an application made on behalf of an infant by a parent who needs the number in order to open a bank account in the child's name, or by a non-income earner who needs the number to file an income tax return jointly with an employed spouse. Individuals are required by various laws to obtain (and disclose) a social security number in order to receive benefits or to discharge obligations quite unrelated to Social Security activities or programs.

Each number is intended to be unique to an individual—i.e., an individual can properly obtain only one number, and that number is properly assigned only to that individual. Moreover, the SSN is a unique number for all Federal agencies dealing with a particular individual. Although agencies are not required to use a numbering system to identify individuals in their records, most do, and they have been required by an Executive Order to adopt social security numbers for that function (1). As a consequence, the SSN is the standard individual identifier for Federal agency records about individuals. Thus, for a number of years, agencies have made social security number identification of the millions of individuals about whom they keep administrative records, e.g., for Veteran's Administration, Social Security, job programs, income taxes, and so on. For statistical purposes, in contrast, the Census Bureau has statistically not collected or kept social security numbers on its population records, except on a relatively small sample basis. Thus while Census has a comprehensive body of statistical information on the race of individuals, it does not routinely collect social security numbers in its Censuses and surveys. Its populations records may contain social security numbers acquired from other agencies' administrative records, but this has usually occurred on a relatively small sample basis. Thus while Census has a comprehensive body of statistical information on the race of individuals, it has not routinely maintained individual social security numbers on its Census population base as a link among other sets of statistical data about particular groups of individuals.

Outside the Federal government, state agencies, such as state motor vehicle administrations, have used the number as an identifier, for instance on the driver's license or on records related to it, or on welfare, housing, and criminal justice records. It is the numbering system likely to be used on hospital records in public and private institutions, whether or not the particular patient is covered by Medicare or other Federally-supported health programs. Although the Privacy Act of 1974 placed some limitations on future requirements to provide the social security number as a condition of initiating certain administrative actions in the Form SS-5, the Act provided that existing requirements could be continued when they were based on statutes or formal rules (4).

D. The Social Security Number as a Record Link. Following the initial SS-5 application, the individual applicant will ordinarily initiate a more substantial relationship with SSA by getting a job. At that time, payroll deductions are made and matched by the employer under the Federal Insurance Compensation Act (FICA) for the worker's Old Age, Survivors and Disability Insurance (OASDI) account, and the deductions are reported to IRS and SSA on the Form W-2 (formerly also on the quarterly Form 941). It is at this time that the individual's SS-5 information is linked readily with operational records in SSA and merged with files which have the character of "tax return information," thus introducing a new dimension to the information. Return information is subject to additional restrictions on disclosure imposed by the Internal Revenue Code (10).

Although SSA has resisted the concept of the SSN as a "universal identifier," the wide range of non-SSA uses for numbering a multiplicity of governmental and non-governmental records has important implications both in terms of the ubiquity of the number as a personal identifier and in terms of its value for record retrieval and linkage. From this discussion it can be observed that the SSN is a "universal identifier."
same individuals by their social security numbers, but which by law or custom do not contain individual information about race. In addition, survey information can be linked in with administrative information about individuals if SSN's are obtained in such surveys. In short, the application forms the basis for a common denominator in the SSN which links, actually or potentially, a large reservoir of government and non-government microdata uniquely associated with each individual who has at any time applied for a social security number. It thus permits measurement of the racial composition of the whole population enumerated, or of sub-populations based on attributes other than race. In the future, the question on ethnic background, which was added in the newly-revised Form SS-5, will produce a related body of attribute data which will have similar value for statistical analysis dealing with ethnicity.

SSA views race/ethnic information as being different from the other items of information on the Form SS-5 in two important complementary respects. The first is that race/ethnicity, unlike the other information on the form is not used operationally as a discriminant for SSA's own internal purposes of identifying a person in any way that affects determinations or decisions about that person individually, e.g., employment, benefit status, eligibility, or other operational relationships with SSA programs. The second is that SSA discloses race/ethnic information only in summary or unidentifiable form in its statistical publications, or in individual form only to other Federal statistical components which can assure that their statistical results will be presented exclusively in summary or other anonymous form.

Part II. Development of Legal Principles
A. Legal Principles in "Functional Separation."
"Functional Separation" is a term first used by the Privacy Protection Study Commission (PPSC), and defined by PPSC as "separating the use of information about an individual for a research or statistical purpose from its use in arriving at an administrative or other decision about that individual." (3) However, the concept of separate use is much older.

Since 1937 when the Social Security program began, the race item has been placed on the collection form at the request of SSA's statistical component (now the Office of Research and Statistics) to be used for the statistical and research needs of SSA. Other components of the Federal statistical establishment have shared in the use of the information under careful restrictions to limit their use to proper statistical activities and summary results which are consonant with SSA's principles for statistical use.

Within SSA, the distinction between operational use or statistical use by SSA was an internal administrative distinction, and not one set forth in Federal statute. Nevertheless, the rule was applied in a clear and definite way—information about an individual's race would be used by SSA only to prepare statistics in administering its programs, and those statistics would be disclosed only in summary or anonymous form to policy-makers in SSA or to others outside SSA. For release of race information outside SSA, this rule was applied formally in SSA's Regulation No. 1, which almost from the beginning of SSA's program implemented protections in the Social Security Act, and stated precise and detailed rules governing confidentiality of its information. (9)

Under Regulation No. 1, if SSA released its information on race, it did so only in summary form or to a Federal statistical user, mainly the Census Bureau, under conditions which could be enforced, that the use of the information to prepare summary or other anonymous statistical output, and to require SSA's explicit consent before any additional release of such statistical information could be made. This assured that redisclosure of SSA's data was not made in a form which constituted individual disclosure.

What made these restrictions and limitations clear and enforceable was the statutory provision contained in the Social Security Act (5), which precluded SSA from disclosing any information to anyone, except as its regulations might authorize.

For many years SSA's implementing Regulation No. 1 was strict in its limitation on statistical information, allowing routine disclosure to other Federal agencies in a form which would not allow identification of individuals, for their statistical and planning purposes only.

- Non-routine disclosures were also made in special situations, when authorized by the formal decision process of the Social Security Commissioner which was provided by Regulation No. 1.
- The 1967 Interagency Agreement between SSA and the Bureau of Census was the principal example of this type of arrangement. That agreement provided that linked data which resulted from release of data by SSA would be subject to the protection of Census provisions contained in Title 13 of the United States Code, as well as the Social Security Act. The statistical character of the data had to be preserved by the recipient, and carefully defined principles of use and disclosure were contained in the agreement, which was to be reviewed and modified on a continuing basis by an interagency group. Other than the release of data to Census under this agreement, other non-routine disclosures were generally limited to statistical activities within HEW, and continue to be protected from disclosure outside the Department.

Changing trends in public attitudes about information collected by Federal agencies have shifted law and policy on the principles of access, use and disclosure of the Federal government's records. A major consequence of the movement of the 1970's toward greater openness in government through amendments to the Freedom of Information Act (FOIA), (7) including those made in Government in the Sunshine Act, has been a shift from the presumption of withholding to a presumption of disclosure. In the 1950's, FOIA judgements to release information were based principally on the requester's "need to know," with the burden on the requester to demonstrate his need. At the present time, such judgements are based in principle on the right to know, absent clear statutory requirements to withhold, or actual harm from disclosing.

Translating these general principles into SSA practice, the Social Security Act no longer blankets all information from disclosure. Rather, it
provides a foundation for exchanges within the Federal government and private organizations through which SSA's programs are carried out. Beyond that, it is essentially limited by FOIA principles. The shift from closed to open rules on disclosure has created much uncertainty for the researcher. For SSA, since most of its research and statistical records compile information about natural persons, the Privacy Act in 1974 established the basic structure for interagency disclosure.

The interface between Freedom of Information and Privacy Act principles which operate with respect to records about natural persons appears deceptively simple. The principles are clear, but their application is complex and subjective, and has been approached cautiously by the courts. General rules are difficult to state, and the courts have proceeded on a case by case basis with somewhat mixed results.

The Privacy Act itself deals with statistical information only in its final sanitized or summary form, except for releases which may be made to Census in individually identifiable form. The Tax Reform Act of 1976, which amended the Internal Revenue Code, superimposed its own confidentiality rules on information obtained through the taxing process, of which a large part consists of earnings reports filed by employers and self-employed persons under the Social Security program. The tax rules, if applied literally, limit statistical activity to IRS itself, and make no provision at all for redisclosure of earnings information for statistical purposes. Limited disclosures by IRS of specified tax information to named users in the Commerce and Treasury Departments are provided. They are mentioned in the context because an important attribute on the worker records is the race indicator, an item of importance to many statistical users, not the least of which are the epidemiologists studying environmental or occupational exposure to hazardous substances.

B. Privacy and Voluntariness of Response. The question of voluntariness of response to the SS-5 has two aspects. The application for an SSN is technically a voluntary form. SSA does not compel anyone to provide the agency with the requested information. The Social Security Act provides criminal sanctions for furnishing false information, but none for refusing to provide SS-5 information. In theory at least, the consequence of not answering all the questions on the form would at worst result in SSA not having sufficient information to warrant issuance of a number. However, this apparent voluntariness is largely illusory. Compliance with Federal duties such as tax paying, and the ability to receive numerous Federal benefits and services require the individual to have or apply for an SSN.

The race item on the form is different, in that its answer does not affect the issuance of an SSN in any way, for any of the administrative purposes the number is used for. There is a delicate balance in preserving the voluntary nature of the answer without endangering high response rate. This balance is between fairly representing the voluntary nature of the response, while at the same time not undermining the incentive to answer or interrupting the automaticity with which applicants answer all the questions on the form. To make the balance more precarious, the use of race statistics in civil rights litigation, etc., has introduced new pressures to allow identifiable data to be used in evidence, thus presenting a direct threat to the assurances of statistical confidentiality which are important for voluntary collection.

Nor is the balancing an academic exercise -- the Privacy Act requires the collecting agency to give reasonable notice to respondents. One of the self-help principles from which the Privacy Act is fabricated is the right to refuse to provide personal information that is not in one's interest to provide. When an agency collects personal information, it must tell the individual what the information is being collected for, what other uses can and will be made of it, whether the individual can be compelled to provide it, and the consequences of complying with or declining to comply with the request. On these principles, when response is mandatory, the informed respondent may presumably make his own trade-offs, for example in balancing the perceived risks of confiding illegal alien status to the Census Bureau, as against the perceived benefits in Congressional reapportionment or per capital Federal grants to state governments.

There may be implicit sanctions for not reporting information which is legally regarded as voluntary. Refusal to provide information about one's financial status may result in loss of income, services, or benefits of various kinds under government programs which require proof of eligibility. All of these considerations may be relevant to the application information which uniquely identifies the SS-5 applicant.

Statistical information, however, falls in a different category. Not only is the reporting usually voluntary in the real sense, but there is seldom any direct advantage to the individual who provides it, nor any direct cost or disadvantage from not providing it. Given the possible risks of having one's personal information used or disclosed in ways the individual objects to, and the nebulous personal attachment to generalized social benefits which the statistics may generate, the respondent may prefer to leave voluntary questions unanswered. The race question on the SS-5 falls in this class of information.

The benefits to statistical users of having race information collected on the SS-5 are manifest. Collection is conducted in neutral circumstances which cause little incentive for bias or misreporting. Although there were substantial populations not enumerated in the past -- young children and women of all ages who never worked in covered employment -- this situation has been markedly changing. Most women now expect to be in the work force at some time in their lives, and Social Security coverage has been expanded to include the self-employed, domestic workers. At the same time, income tax filing reaches larger segments of the population, with obligatory use of the SSN for identification of individual and joint filers.

C. Recent Developments and Future Trends. From the standpoint of statistics and research, there is great general benefit to collect information about race (and add information about ethnicity) on the same form used to assign the number used throughout the government for record linkage. At
the same time, there is great pressure outside the statistical community, as noted, to make race information available for compliance processes and to remove the statistical constraints which have been placed on the disclosure of information. A particularly difficult situation arises in the use of race statistics to prove racial discrimination in the class action litigation discussed earlier, because of conflicting legal and ethical values.

SSA has traditionally produced statistical tabulations on request for organizations and agencies such as NAACP, EEOC and other litigants, providing summary information on race of applicants and employees involved in particular actions. The statistical summaries requested for use in evidence have been screened by SSA using its disclosure avoidance rules, and in most situations in the past, have met the parties' needs to establish the racial composition of the work force in question. In a few cases, however, the proof of discrimination in hiring, promotion, etc., has involved data for small departments of relatively small organizations, and has encountered problems of statistical disclosure in the small cells which occur with disconcerting frequency. Although the possibility of furnishing unidentified microdata has been explored on occasion, the conditions of use acceptable to SSA include the assurance that the authorized recipients be prohibited from matching the microdata to any other information in their possession or available to them, in such a way as to identify an individual whose racial characteristics have been provided by SSA. These conditions are not necessarily acceptable to the litigating parties, and there are important issues that remain to be resolved.

These issues were recently the focus of the Office of Management and Budget (OMB) "Interim Guidelines" for Federal agency collection of race/ethnic information on application forms. (2) In those guidelines, concurred in by EEOC and the Civil Rights Division of the Department of Justice, OMB stated several conditions acceptable to it for collecting race/ethnic information. On application forms for employment, race/ethnic information may be collected on a tear-off portion to be processed and maintained separately from the main form. However, the information could be cross-matched with individual application forms for compliance or enforcement purpose. The guidelines provide that on other application forms for a service or benefit, the race information may be collected directly on the form, and can be made mandatory only where the characteristic information is necessary to a determination of eligibility or amount of benefit. When not mandatory, the guidelines require the request to contain a standard statement indicating that the collection is solely for civil rights compliance purposes, and that response is voluntary.

The relationship of this guidance to Social Security's Form SS-5 is not wholly clear. The SS-5 is not an application for a job or a service/benefit. Its race/ethnic question is voluntary, but its collection is for statistical purposes. Use of it for compliance purposes, as described on the current SS-5 instructions approved by OMB, is in summaries or other forms of information which do not reveal names of individuals. It can be argued that the SS-5 is a type of application which is not subject to the guidelines. Or it can be said that it is generally covered, but has been given a partial exception from the required purpose statement. Under either interpretation, the underlying tension between statistical purposes and disclosure in administrative or judicial use remains to be resolved as each new situation occurs.

Some of the possible ways of resolving these tensions would be substantial -- even radical -- changes in approach to the statistical basis for collection of racial information. At one extreme, the collection of race/ethnic information could be made mandatory for administrative and compliance use. With this approach, the statistical purpose would be subordinated, and the anonymous statistical character of the information would be altered. On the other hand, all of the race information which is presently maintained as statistical data, such as the SS-5 information could be transferred, together with identifiers, to a protected statistical environment, for example the Census Bureau, and made subject to disclosure rules like those of Title 13. At the present time, it is not entirely clear how the various competing public interest claims to the data will be balanced out between these extremes.

(This paper is based on the author's work in the Office of Research and Statistics, Social Security Administration, but the opinions expressed are her own, and not necessarily the formal views of the agency.)

FOOTNOTES
1. See for example the Instructions for filling out form EEO-1 and similar reporting forms.
2. In its instructions for completing Form EEO-1, for instance, EEOC informs reporting employers that they may keep post-employment records of race of their employees (as distinguished from applicants) unless prohibited by state law, but EEOC recommends keeping the information separate from personnel or decision records.
3. "Microdata" is used in this paper to mean data sets containing records with information about the individual persons in a defined study population.

REFERENCES
(1) Executive Order 9397; Numbering System for Federal Accounts Relating to Individual Persons, November 22, 1943.
(2) Interim Guidelines for the Collection of Race, Ethnic Background, Age, and Sex Information on Applications Made by Individuals for Benefits From Federal Programs. 44 FR 238, Monday, December 10, 1979.
(4) Privacy Act of 1974, Section 7.
(5) Social Security Act, Section 1106: 42 USC 1306.
(7) 5 USC 552.
Annotated Bibliography on Legal/Policy Issues in the Collection of Race/Ethnic Data


This directive provides standard classifications which have been developed as a result of the needs of both the executive branch and the Congress to provide for the collection and use of race/ethnic data by Federal agencies. These standard classifications pertain to recordkeeping, collection, and presentation of data on race and ethnicity within Federal program administrative reporting and statistical activities.

Interim Guidelines for the Collection of Race, Ethnic Background, Age, and Sex Information on Applications Made by Individuals for Benefits from Federal Programs. Federal Register/vol. 44 No. 238/Monday December 10, 1979/Notices.

OMB issued these interim guidelines regarding the collection of information on the race, ethnic background, age, and sex of individuals applying for a benefit under Federal programs so that Federal Departments and agencies collecting this type of information may be assured of their compliance with various civil rights provisions and other statutes.


This notice was intended to publicize the Department of Justice Memorandum, dated November 23, 1979 and sent to the heads of all Executive Departments and Agencies, which dealt with collecting and reviewing data on the race, ethnic background, age, and sex of persons applying for benefits or services under a federally assisted program. Collection of this data was cited as being significant in determining compliance status of various Federally Assisted Programs.


This memorandum included an attached draft which proposed revisions of the interim guidelines issued by OMB on December 10, 1979. The revisions resulted from several comments that suggested an expansion of the guidelines to include the collection of information on handicap status.


Quashing of a subpoenas served upon SSA to provide the racial identification of approximately 100 persons holding Social Security numbers was ordered on March 6, 1980. The government's position in this matter was set forth in an attached memorandum which stated: 1) that under 5USC552a(b)(7), 42USC1306, 20CFR601 the Privacy Act is not a mandatory disclosure provision and 2) under 5USC552(b)(6) disclosure would be an unwarranted invasion of personal privacy.


On February 28, 1980 the Custodian of Records was served with a subpoena commanding him on behalf of the plaintiff EEOC to produce at deposition the SSA records which would individually identify the race and sex of 114 persons listed by name and Social Security number, all of whom are past or present employees of the defendant. The Court concluded that the information sought by EEOC is not subject to mandatory disclosure under the Freedom of Information Act because it is included within Exemption 6 of that Act.


Federal government use of a single, unduplicated numerical identification system of accounts is acknowledged as an extremely desirable practice. E.O. 9397 therefore orders that any Federal department, establishment or agency establishing a new system of permanent account numbers pertaining to individual persons utilize exclusively, Social Security Account numbers.

Bibliography prepared by Mark Wechsler and Robin Allen.