

## **Census 2000 Special EEO Tabulation Files Technical Documentation for ASCII Files**

### **Appendix 9: Subject Matter Descriptions**

#### **INDUSTRY AND OCCUPATION**

The data on industry and occupation were derived from answers to long-form questionnaire Items 27 and 28 respectively, which were asked of a sample of the population 15 years old and over. Information on industry relates to the kind of business conducted by a person's employing organization and occupation describes the kind of work a person does on the job.

For employed people, the data refer to the person's job during the reference week. For those who worked at two or more jobs, the data refer to the job at which the person worked the greatest number of hours during the reference week. For unemployed people, the data refer to their last job. The industry and occupation statistics are derived from the detailed classification systems developed for Census 2000 as described below.

Respondents provided the data for the tabulations by writing on the questionnaires descriptions of their industry and occupation. These descriptions were data captured and sent to an automated coder (computer software), which assigned a portion of the written entries to categories in the classification system. The automated system assigned codes to 59 percent of the industry entries and 56 percent of the occupation entries. Those cases not coded by the computer were referred to clerical staff in the Census Bureau's National Processing Center in Jeffersonville, Indiana, for coding. The clerical staff converted the written questionnaire responses to codes by comparing these responses to entries in the *Alphabetical Index of Industries and Occupations*. For the industry code, these coders also referred to an Employer Name List. This list, prepared from the American Business Index (ABI), contained the names of business establishments and their North American Industrial Classification System (NAICS) codes converted to population census equivalents. This list facilitated coding and maintained industrial classification comparability.

#### **Occupation**

The occupational classification system used during Census 2000 consists of 509 specific occupational categories for employed people arranged into 23 major occupational groups. This classification was developed based on the *Standard Occupational Classification (SOC) Manual: 2000*, which includes a hierarchical structure showing 23 major occupational groups divided into 96 minor groups, 449 broad groups, and 821 detailed occupations. For Census 2000, tabulations with occupation as the primary characteristic present several levels of occupational detail. Some occupation groups are related closely to certain industries. Operators of transportation equipment, farm operators and workers, and healthcare providers account for major portions of their respective industries of transportation, agriculture, and health care. However, the industry categories include people in other occupations. For example, people employed in agriculture

include truck drivers and bookkeepers; people employed in the transportation industry include mechanics, freight handlers, and payroll clerks; and people employed in the health care industry include occupations such as security guard and secretary.

## **Industry**

The industry classification system used during Census 2000 was developed for the census and consists of 265 categories for employed people, classified into 14 major industry groups. From 1940 through 1990, the industrial classification has been based on the *Standard Industrial Classification (SIC) Manual*. The Census 2000 classification was developed from the 1997 North American Industry Classification System (NAICS) published by the Office of Management and Budget, Executive Office of the President. NAICS is an industry description system that groups establishments into industries based on the activities in which they are primarily engaged.

The NAICS differs from most industry classifications because it is a supply-based, or production-oriented economic concept. Census data, which were collected from households, differ in detail and nature from those obtained from establishment surveys. Therefore, the census classification system, while defined in NAICS terms, cannot reflect the full detail in all categories.

## **SEX**

The data on sex, which was asked of all people, were derived from answers to long-form questionnaire Item 3 and short-form questionnaire Item 5. Individuals were asked to mark either “male” or “female” to indicate their sex. For most cases in which sex was not reported, it was determined from the person’s given (i.e., first) name and household relationship. Otherwise, sex was imputed according to the relationship to the householder and the age of the person.

## **RACE**

The data on race, which was asked of all people, were derived from answers to long-form questionnaire Item 6 and short-form questionnaire Item 8. The concept of race, as used by the Census Bureau, reflects self-identification by people according to the race or races with which they most closely identify. These categories are socio-political constructs and should not be interpreted as being scientific or anthropological in nature. Furthermore, the race categories include both racial and national-origin groups.

The racial classifications used by the Census Bureau adhere to the October 30, 1997, Federal Register Notice entitled, “*Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*,” issued by the Office of Management and Budget (OMB). These standards govern the categories used to collect and present federal data on race and ethnicity. The OMB requires five minimum categories (White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander) for race. The race categories are described below with a sixth category, “Some other race,” added with OMB approval. In addition to the five race groups, the OMB also states that respondents should be offered the option of selecting one or more races.

If an individual did not provide a race response, the race or races of the householder or other household members were assigned using specific rules of precedence of household relationship. For example, if race was missing for a natural-born child in the household, then either the race or races of the householder, another natural-born child, or the spouse of the householder were assigned. If race was not reported for anyone in the household, the race or races of a householder in a previously processed household were assigned.

**White.** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. It includes people who indicate their race as “White” or report entries such as Irish, German, Italian, Lebanese, Near Easterner, Arab, or Polish.

**Black or African American.** A person having origins in any of the Black racial groups of Africa. It includes people who indicate their race as “Black, African Am., or Negro,” or provide written entries such as African American, Afro-American, Kenyan, Nigerian, or Haitian.

**American Indian or Alaska Native.** A person having origins in any of the original peoples of North and South America (including Central America) and who maintain tribal affiliation or community attachment. It includes people who classified themselves as described below.

*American Indian.* This category includes people who indicated their race as “American Indian,” entered the name of an Indian tribe, or reported such entries as Canadian Indian, French American Indian, or Spanish American Indian.

*Alaska Native.* This category includes written responses of Eskimos, Aleuts, and Alaska Indians as well as entries such as Arctic Slope, Inupiat, Yupik, Alutiiq, Egegik, and Pribilovian. The Alaska tribes are the Alaskan Athabaskan, Tlingit, and Haida. The information for Census 2000 is based on the American Indian Tribal Classification List for the 1990 census, which was expanded to list the individual Alaska Native Villages when provided as a written response for race.

**Asian.** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. It includes “Asian Indian,” “Chinese,” “Filipino,” “Korean,” “Japanese,” “Vietnamese,” and “Other Asian.”

**Native Hawaiian or Other Pacific Islander.** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. It includes people who indicate their race as “Native Hawaiian,” “Guamanian or Chamorro,” “Samoan,” and “Other Pacific Islander.”

**Some other race.** This category includes all other responses not included in the “White,” “Black or African American,” “American Indian or Alaska Native,” “Asian,” and “Native Hawaiian or Other Pacific Islander” race categories described above. Respondents providing write-in entries such as

multiracial, mixed, interracial, or a Hispanic/Latino group (for example, Mexican, Puerto Rican, or Cuban) in the “Some other race” write-in space are included in this category.

**Two or more races.** People may have chosen to provide two or more races either by checking two or more race response check boxes, by providing multiple write-in responses, or by some combination of check boxes and write-in responses. The race response categories shown on the questionnaire are collapsed into the five minimum races identified by the OMB, and the Census Bureau “Some other race” category. For data product purposes, “Two or more races” refers to combinations of two or more of the following race categories:

1. White
2. Black or African American
3. American Indian and Alaska Native
4. Asian
5. Native Hawaiian and Other Pacific Islander
6. Some other race

There are 57 possible combinations involving the race categories shown above. Thus, according to this approach, a response of “White” and “Asian” was tallied as two or more races, while a response of “Japanese” and “Chinese” was not because “Japanese” and “Chinese” are both Asian responses. Tabulations of responses involving reporting of two or more races within the American Indian and Alaska Native, Asian, or Native Hawaiian and Other Pacific Islander categories are available in other data products.

Given the many possible ways of displaying data on two or more races, data products will provide varying levels of detail. The most common presentation shows a single line indicating “Two or more races.” Some data products provide totals of all 57 possible combinations of two or more races, as well as subtotals of people reporting a specific number of races, such as people reporting two races, people reporting three races, and so on.

In other presentations on race, data are shown for the total number of people who reported one of the six categories alone or in combination with one or more other race categories. For example, the category, “Asian alone or in combination with one or more other races” includes people who reported Asian alone and people who reported Asian in combination with White, Black or African American, Native Hawaiian and Other Pacific Islander, and Some other race. This number, therefore, represents the maximum number of people who reported as Asian in the question on race. When this data presentation is used, the individual race categories will add to more than the total population because people may be included in more than one category.

**Comparability.** The data on race in Census 2000 are not directly comparable to those collected in previous censuses. The October 1997 revised standards issued by the OMB led to changes in the question on race for Census 2000. The Census 2000 Dress Rehearsal data were the first to reflect these changes. First, respondents were allowed to select more than one category for race.

Second, the sequence of the questions on race and Hispanic origin changed. In 1990, the question on race (Item 4) preceded the question on Hispanic origin (Item 7) with two intervening questions. For Census 2000, the question on race immediately follows the question on Hispanic origin. Third, there were terminology changes to the response categories, such as spelling out “American” instead of “Amer.” for the American Indian or Alaska Native category; and adding “Native” to the Hawaiian response category. The 1990 category, “Other race,” was renamed “Some other race.” Other differences that may affect comparability involve the individual categories on the Census 2000 questionnaire. The 1990 category, “Asian and Pacific Islander,” was separated into two categories, “Asian” and “Native Hawaiian and Other Pacific Islander” for Census 2000.

Accordingly, on the Census 2000 questionnaire, there were seven Asian categories and four Native Hawaiian and Other Pacific Islander categories. The two residual categories, “Other Asian” and “Other Pacific Islander,” replaced the 1990 single category “Other API.” The 1990 categories, “American Indian,” “Eskimo,” and “Aleut,” were combined into “American Indian and Alaska Native.” American Indians and Alaska Natives can report one or more tribes.

As in 1980 and 1990, people who reported a Hispanic or Latino ethnicity in the question on race and did not mark a specific race category were classified in the “Some other race” category (“Other” in 1980 and “Other race” in 1990). They commonly provided a write-in entry such as Mexican, Puerto Rican, or Latino. In the 1970 census, most of these responses were included in the “White” category. In addition, some ethnic entries that in 1990 may have been coded as White or Black are now shown in the “Some other race” group.

## **HISPANIC OR LATINO**

The data on the Hispanic or Latino population, which was asked of all people, were derived from answers to long-form questionnaire Item 5, and short-form questionnaire Item 7. The terms “Spanish,” “Hispanic origin,” and “Latino” are used interchangeably. Some respondents identify with all three terms, while others may identify with only one of these three specific terms. Hispanics or Latinos who identify with the terms “Spanish,” “Hispanic,” or “Latino” are those who classify themselves in one of the specific Hispanic or Latino categories listed on the questionnaire — “Mexican,” “Puerto Rican,” or “Cuban” — as well as those who indicate that they are “other Spanish, Hispanic, or Latino.” People who do not identify with one of the specific origins listed on the questionnaire but indicate that they are “other Spanish, Hispanic, or Latino” are those whose origins are from Spain, the Spanish-speaking countries of Central or South America, the Dominican Republic, or people identifying themselves generally as Spanish, Spanish-American, Hispanic, Hispano, Latino, and so on. All write-in responses to the “other Spanish/Hispanic/Latino” category were coded.

Origin can be viewed as the heritage, nationality group, lineage, or country of birth of the person or the person’s parents or ancestors before their arrival in the United States. People who identify their origin as Spanish, Hispanic, or Latino may be of any race.

Some tabulations are shown by the origin of the householder. In all cases where the origin of households, families, or occupied housing units is classified as Spanish, Hispanic, or Latino, the

origin of the householder is used.

If an individual could not provide a Hispanic origin response, their origin was assigned using specific rules of precedence of household relationship. For example, if origin was missing for a natural-born daughter in the household, then either the origin of the householder, another natural-born child, or the spouse of the householder was assigned. If Hispanic origin was not reported for anyone in the household, the origin of a householder in a previously processed household with the same race was assigned. This procedure is similar to those used in 1990, except that for Census 2000, race and Spanish surnames were used to assist in assigning an origin.

***Comparability.*** There are two important changes to the Hispanic origin question for Census 2000. First, the sequence of the race and Hispanic origin questions for Census 2000 differs from that in 1990; in 1990, the race question preceded the Hispanic origin question. Testing prior to Census 2000 indicated that response to the Hispanic origin question could be improved by placing it before the race question without affecting the response to the race question. Second, there is an instruction preceding the Hispanic origin question indicating that respondents should answer both the Hispanic origin and the race questions. This instruction was added to give emphasis to the distinct concepts of the Hispanic origin and race questions and to emphasize the need for both pieces of information.

Furthermore, there has been a change in the processing of the Hispanic origin and race responses. In 1990, the Hispanic origin question and the race question had separate edits; therefore, although information may have been present on the questionnaire, it was not fully utilized due to the discrete nature of the edits. However, for Census 2000, there was a joint race and Hispanic origin edit which for example, made use of race responses in the Hispanic origin question to impute a race if none was given.

## **EDUCATIONAL ATTAINMENT**

Data on educational attainment were derived from answers to long-form questionnaire Item 9, which was asked of a sample of the population. Data on attainment are tabulated for the population 25 years old and over. However, when educational attainment is cross-tabulated by other variables, the universe may change. (For example, when educational attainment is crossed by disability status, the data are tabulated for the civilian noninstitutionalized population 18 to 34 years old.) People are classified according to the highest degree or level of school completed. The order in which degrees were listed on the questionnaire suggested that doctorate degrees were “higher” than professional school degrees, which were “higher” than master’s degrees. The question included instructions for people currently enrolled in school to report the level of the previous grade attended or the highest degree received. Respondents who did not report educational attainment or enrollment level were assigned the attainment of a person of the same age, race, Hispanic or Latino origin, occupation and sex, where possible, who resided in the same or a nearby area. Respondents who filled more than one box were edited to the highest level or



degree reported.

The question included a response category that allowed respondents to report completing the 12th grade without receiving a high school diploma. It allowed people who received either a high school diploma or the equivalent, for example, passed the Test of General Educational Development (G.E.D.) and did not attend college, to be reported as “high school graduate(s).” The category “Associate degree” included people whose highest degree is an associate degree, which generally requires 2 years of college level work and is either in an occupational program that prepares them for a specific occupation, or an academic program primarily in the arts and sciences. The course work may or may not be transferable to a bachelor’s degree. Master’s degrees include the traditional MA and MS degrees and field-specific degrees, such as MSW, MEd, MBA, MLS, and MEng. Some examples of professional degrees include medicine, dentistry, chiropractic, optometry, osteopathic medicine, pharmacy, podiatry, veterinary medicine, law, and theology. Vocational and technical training, such as barber school training; business, trade, technical, and vocational schools; or other training for a specific trade, are specifically excluded.

## **AGE**

The data on age, which was asked of all people, were derived from answers to the long-form questionnaire Item 4 and short-form questionnaire Item 6. The age classification is based on the age of the person in complete years as of April 1, 2000. The age of the person usually was derived from their date of birth information. Their reported age was used only when date of birth information was unavailable.

Data on age are used to determine the applicability of some of the sample questions for a person and to classify other characteristics in census tabulations. Age data are needed to interpret most social and economic characteristics used to plan and examine many programs and policies. Therefore, age is tabulated by single years of age and by many different groupings, such as 5-year age groups.

## **INCOME IN 1999**

The data on income in 1999 were derived from answers to long-form questionnaire Items 31 and 32, which were asked of a sample of the population 15 years old and over. “Total income” is the sum of the amounts reported separately for wage or salary income; net self-employment income; interest, dividends, or net rental or royalty income or income from estates and trusts; social security or railroad retirement income; Supplemental Security Income (SSI); public assistance or welfare payments; retirement, survivor, or disability pensions; and all other income.

“Earnings” are defined as the sum of wage or salary income and net income from self-employment. “Earnings” represent the amount of income received regularly for people 16 years old and over before deductions for personal income taxes, social security, bond purchases, union dues, medicare deductions, etc.

For more detailed information on these topics and for information on other Census 2000 topics go to [www.census.gov/prod/cen2000/doc/sf3.pdf](http://www.census.gov/prod/cen2000/doc/sf3.pdf).