

IT Standard 16.0.0: SURVEY DESIGN AND STATISTICAL METHODOLOGY METADATA

Software and Standards Management Branch
Systems Support Division
Bureau of the Census

August 1998

Standard Number	Sequence: 0	Sub-standard: 0	Revision: 0
Category	[Completed by SSMB]		
Key Words	statistical metadata, survey design, terminology		

Abstract

This document describes the Survey Design and Statistical Methodology (SDSM) metadata standard. The SDSM is the content standard for a corporate metadata repository. The development of this standard is one of the key steps in designing a unified Bureau of the Census corporate metadata repository that will support an assortment of survey and census activities and products.

This standard will be used to develop a common corporate thesaurus and catalogue of reference materials, a metadata repository, and the tools to help analysts create, catalogue, populate, and retrieve metadata from the repository. As a reference library, it will also be used to aid in Internet data dissemination and automated integrated survey processing systems. The metadata items will also serve as one of the sources of conceptual entities and attributes for the modeling needed to develop the metadata repository. **This document does not describe views or presentations of the metadata. One of these, the Table of Contents view is presented in Appendix E.**

<p>This document is intended to be used as a specification for the development of systems and tools used to support population, access, and maintenance of the corporate metadata repositories of the Census Bureau.</p>
--

Table of Contents

Abstract	ii
1. Scope	1
1.1 Reason for Standard.....	1
1.2 Who or What is Affected.....	1
2 Assumptions.....	2
3 Survey Design and Statistical Methodology Metadata.....	2
3.1 Presentation Description.....	2
3.2 Universal Entity.....	3
3.3 Survey Design and Statistical Methodology Metadata Content Description.....	3
3.3.1 <u>Access and use constraint</u>	3
3.3.2 <u>Archive</u>	4
3.3.3 <u>Citation</u>	4
3.3.4 <u>Contact</u>	4
3.3.5 <u>Data</u>	6
3.3.6 <u>Data quality</u>	8
3.3.7 <u>Dataset</u>	14
3.3.8 <u>Dataset supplier</u>	15
3.3.9 <u>Descriptive statistics</u>	16
3.3.10 <u>Documentation</u>	19
3.3.11 <u>Emprise</u>	20
3.3.12 <u>Field</u>	20
3.3.13 <u>Format</u>	22
3.3.14 <u>Frame</u>	23
3.3.15 <u>Geography</u>	24
3.3.16 <u>Methodology</u>	25
3.3.17 <u>Planning and design</u>	25
3.3.18 <u>Pointer to object</u>	28
3.3.19 <u>Product</u>	28
3.3.20 <u>Program</u>	29
3.3.21 <u>Question</u>	29
3.3.22 <u>Questionnaire</u>	30
3.3.23 <u>Sample</u>	31
3.3.24 <u>Series information</u>	32
3.3.25 <u>Sponsor</u>	33
3.3.26 <u>Status</u>	33
3.3.27 <u>Survey</u>	33
3.3.28 <u>Survey instance</u>	34
3.3.29 <u>System</u>	36

3.3.30	<u>Technique</u>	36
3.3.31	<u>Theme</u>	36
3.3.32	<u>Time period information</u>	37
3.3.33	<u>Universe</u>	38
3.4	Definitions	38
4.	Training	43
5.	Reference Materials and Related Standards	43
5.1.	Reference Materials	43
5.2.	Related Standards	46
5.2.1.	Information Resource Dictionary System (IRDS)	46
5.2.2.	Data Element and Data Element Registry Standards.....	47
5.3	Other Standards	48
5.3.1.	Spatial Data Transfer Standard (SDTS)	48
5.3.2.	Content Standard for Digital Geospatial Metadata (CSDGM).....	49
5.3.3.	Standard for Cultural and Demographic Data Metadata (CDDM).....	49
6.	Conversion Requirements	49
7.	Compliance Methods and Timetable.....	49
7.1.	Compliance Methods	49
7.2.	Timetable	50
8.	Authority	50
8.1.	Policy	50
8.2.	Mandatory Standards	50
	Important Dates	50
	Contacts for Additional Information	51
Appendix A.	Metadata Repository.....	52
A1.	Introduction.....	52
A2.	Statistical Metadata Repository (MDR).....	52
A2.1	Purposes for Statistical Metadata Repository	53
A2.2	Architecture for Statistical Metadata Repository.....	54
A2.3	Metadata Management.....	54
A3	Who or What is Affected.....	56
Appendix B.	Business Data Model	57

Appendix C. Themes.....	58
C1 A Model for Cultural and Demographic Data.....	58
C2 The Theme Metadata Entity.....	59
Appendix D. Definitions of the Domains of Selected Metadata Entities.....	61
D1. Domain: Major Theme Description.....	61
D2. Domain: Minor Theme Description.....	66
D3. Domain: Geographic Entity Type.....	84
D4. Domain: Geographic Unit.....	85
Appendix E. Metadata Presentation:	
The SDSM Table of Content Presentation (TOC).....	92
1 Content.....	93
2 Planning.....	94
3 Design.....	95
4 Data Collection.....	98
5 Data Processing.....	101
6 Data Analysis.....	103
7 Data Dissemination.....	104
8 Data.....	105
Annexes to the Table of Contents View.....	106
Annex A Documentation.....	106
Annex B Survey Datasets.....	107
Annex C Product.....	108
Annex D Planning.....	109
Annex E Frame/Sample Concept.....	110
Annex F Observation models.....	111
Annex G Estimation model.....	112
Annex H Editing.....	113
Annex I Data Collection and Quality Control.....	114
Annex J Processing Details.....	116
Annex K Estimation.....	117
Annex L Descriptive Statistics.....	119
Annex M Data Dissemination.....	121
Annex N Data.....	122
Appendix F. Example application using CDDM.....	123
Appendix G. Use with Other Standards.....	129
Appendix H. Abbreviations and Acronyms.....	130

[BLANK]

1. Scope

1.1 Reason for Standard

This document defines the “Survey Design and Statistical Methodology” (SDSM) metadata standard. This standard defines the set of statistical metadata items (3.3, SDSM Content Description) needed for describing all aspects of survey design, processing, analysis, and datasets.

The terms and concepts defined in this standard will be used in the development of a unified Bureau of the Census (BOC) metadata repository (MDR) supporting survey and census activities and products. The MDR is a shared electronic library from which a wide variety of applications can extract or contribute metadata.

The SDSM standard is a specification for developing the tools and models needed to help BOC staff and others create, catalogue, manage, and retrieve metadata from the MDR, creating a catalogue of reference materials. The SDSM’s metadata items are the source of conceptual entities and attributes for the modeling needed to develop the MDR. The definitions of these metadata items may be thought of as a common thesaurus of survey design and statistical design terminology. As a reference library, the MDR will also be used to aid in Internet data dissemination and automated integrated survey processing systems. Appendix A provides more detail on the MDR.

The terms and concepts defined in the SDSM standard may also be used to enhance or extend systems implementing other standards such as the Government Information Locator Service (GILS) standard (FIPS 192) and the standards of the Federal Geographic Data Committee (i.e., the Spatial Data Transfer Standard (FIPS 173) and Content Standards for Digital Geospatial Metadata: Thematic Supplement for Geospatially Referenced Cultural and Demographic Data Metadata (CDDM)).

The SDSM is a *content standard*. This means that the standard specifies a set of metadata items and does not specify the physical format of the content, the services to be provided, or the syntax used. Each of these other properties is the subject of related standards mentioned elsewhere in this document.

1.2 Who or What is Affected

This standard is mandatory for all Bureau of the Census employees or contractors who document surveys.

Specifically, this standard **shall** be used as basis for the development of models for metadata repositories of the BOC and as part of specifications for any tools intended to populate, manage or make use of those repositories.

2 Assumptions

None.

3 Survey Design and Statistical Methodology Metadata

3.1 Presentation Description

Section 3.3 is an alphabetical dictionary of the *metadata items* in the SDSM content standard. An entry in the dictionary is organized as follows [example component]: A common use name, bold and underlined, for the metadata item or the *composite metadata item* [**Methodology**] follows the section number. This is followed by an indication, within parentheses, of a rule about inclusion, generally whether the item is repeatable, optional, or mandatory [(*optional, repeatable*)]. Next follows the business data model's definition of the composite [a structured approach to solve a problem], followed by the *metadata entity name* [**Methodology name**] and a definition of each component of the composite.

When a component is itself a composite, the details of which are defined elsewhere in the standard, it will be shown as a composite metadata item, i.e., bold and underlined [**Methodology description**], and its definition will be followed by a reference to the section of the standard providing those details [(See **Documentation** item for details)]. In this example, details of the composite metadata item, **methodology description**, will be found in section 3.3.11, **documentation**, and might include, in use, any or all of the items found there.

Example:

3.3.16 Methodology (*optional, repeatable*). A structured approach to solve a problem.

- a. **Methodology name** (*mandatory*). The designation assigned to a specific structured approach to solve a problem.
- b. **Methodology description** (*mandatory*). A specification or other formal or informal delineation of this particular structured approach to solve a problem. (See **Documentation** item for details).

3.2 Universal Entity

In addition to component-specific data elements, one data element may be included in any or all of the metadata items at any location(s). This metadata item is:

- **Tag** (*optional*). Tag provides the metadata producer a location for any additional information or identifier the metadata producer wishes to include for internal reference purposes. Information included in Tag need not be explained to nor necessarily understood by the general consumer of the metadata. Tag allows the metadata producer to interpret or translate the common set of metadata terminology and definitions to the metadata producer's unique terminology within the metadata itself. Metadata consumers may ignore all Tag metadata items and successfully obtain all information.

3.3 Survey Design and Statistical Methodology Metadata Content Description

3.3.1 Access and use constraint (*mandatory if applicable*). Information about any limitation or special rule to be applied in the process of accessing this object (data, dataset, product, or other component). This includes constraints due to security, confidentiality and privacy. It also includes data quality and related warning flags or indicators. This also includes use constraints, i.e., restrictions and legal prerequisites for using the item after access is granted. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the item.

- a. **Constraint designation** (*mandatory*). The specific constraint applicable to accessing or using this object, i.e., "Census Confidential," "Classified SECRET," "Privacy Act Restricted," etc.
- b. **Contact** (*mandatory, repeatable*). The individual or organization controlling access to or use of this object. (See **Contact** item for details).
- c. **Constraint authority** (*optional*). Specific authority under which access or use has been restricted or controlled, (q.v. "U.S.C. Title 13," or "Privacy Act of 1995"). As an alternative, or in addition this entry may contain a description, citation, or other reference to the applicable authorizing document or policy. (See **Documentation** item for details).

3.3.2 Archive (*mandatory if applicable*). Dataset stored for backup, contingency or statutory reasons. Typically, this dataset is stored in a way which requires additional processing before it can be used.

- a. **Abstract** (*mandatory*). A brief description or extract of the applicable documentation related to the content and/or creation of the data archived. (See **Documentation** item for details).
- b. **Archive supplemental information** (*optional, repeatable*). Information related to the archiving of this data, such as special access information.
- c. **Archive dataset** (*mandatory*). Information about archived dataset. (See **Dataset** item for details).

3.3.3 Citation (*mandatory if applicable*). The formal citation used to reference the material which describes the data elements or the formal citation for the survey or program dataset.

- a. **Version** (*mandatory*). The identification used to identify the variant of the original **citation**.
- b. **Originator** (*mandatory*). Information about the individual or organization responsible for the creation and/or promulgation of the cited material. (See **Contact** item for details).
- c. **Citation title** (*mandatory*). The title of the cited material.
- d. **Citation time period** (*mandatory if applicable*). The date or range of dates on which the cited material was created and/or promulgated. (See **Time period information** item for details).
- e. **Citation series information** (*mandatory if applicable*). The series information used to identify all material associated with the cited material. (See **Series information** item for details).

3.3.4 Contact (*mandatory if applicable*). Contact information for individual or organization that is knowledgeable about the subject.

- a. **Contact person** (*mandatory if no **Contact organization** is provided, else optional, repeatable*). A name of a person. This entity should provide the name of the person associated with the activity being documented. If the **Contact type** is "Person," this is this individual who is responsible for answering all queries. If the **Contact type** is "Organization," this is the individual who is most knowledgeable

about the subject, but may not be responsible for providing information in response to all queries.

- b. **Contact organization** (*mandatory if no **Contact person** is provided, else optional, repeatable*). The organization associated with the object or information of interest.
- c. **Contact address** (*optional*). The mailing address of the **Contact**.
- d. **Contact voice telephone** (*optional*). The telephone number to reach the **Contact**.
- e. **Contact electronic mail address** (*optional*). The address of the electronic mailbox of the **Contact**.
- f. **Contact facsimile telephone** (*optional*). The telephone number for a facsimile machine to reach the **Contact**.
- g. **Contact TDD TTY telephone** (*optional*). The telephone number by which hearing-impaired individuals can reach the **Contact**.
- h. **Hours of service** (*optional*). When **Contact person** or **Contact organization** is available.
- i. **Contact instructions** (*optional*). Supplemental instructions on how or when to contact the organization or individual. This may also include what specific information is required in order to get a response (for example, how to identify the particular product, to provide a self-addressed-stamped-envelope, or to include your electronic mailing address or postal mailing address).
- j. **Contact Domain** (*optional*). What in particular is this contact for or knowledgeable about. A given contact may know about the content of a dataset, for example, but may not know about access restrictions or about how the data was prepared.
- k. **Contact type** (*optional*). Specify "Person" or "Organization." Indicates whether providing information is the responsibility of a particular individual (indicated by "Person" in this field) or of the organization as a whole (indicated by "Organization" in this field).

3.3.5 Data (*mandatory if applicable*). **Data** are a record of factual information, typically responses to a survey or derivatives of such responses. An element of data is a logically primitive item of information.

- a. **Descriptive name** (*mandatory*). **Descriptive name** is the fully qualified designation (name) [see Annex A of ISO 11179, Part 5] assigned to a particular element of data. A specific data element which has been gathered in the process of carrying out a survey. The **descriptive name** is typically not abbreviated.
- 1) **Object class term** (*mandatory*). A component of the **Descriptive name** which represents the logical data grouping (in a logical data model) to which it belongs; e.g., “employee.”
 - 2) **Property term** (*mandatory*). A component of the **Descriptive name** which expresses the category to which the data element belongs.
 - 3) **Qualifier term** (*optional, repeatable*). A word or words which help define and differentiate the **Descriptive name** within a database or dataset.
 - 4) **Representation term** (*optional*). The form of the set of valid values for a data element, e.g., “amount,” “name.”
- b. **Variable name** (*optional, repeatable*). The name assigned to a particular element of data. A specific data element which has been gathered in the process of carrying out a survey. A **variable of interest** is distinguished from other **variable names** in that it is the name of one of the elements of **data** needed to accomplish the purpose of the **survey** through which it was gathered (or joined to). The **variable name** may also be called the “data identifier,” the “short name,” the “system name,” or the “program name.” The latter may be used because it is the name by which the **data** is referred in application programs developed to implement a **system** in which the **data** is used or meets a particular set of naming rules imposed by a particular development environment or language.
- 1) **Usage type** (*optional*). For each **variable name**, the **system** or application in which this **variable name** is used..
 - 2) **Variable of interest indicator** (*optional*). For each **variable name** and **system** or application in which this **variable name** is used this is an indicator of whether or not this **variable name** is a **variable of interest**.
- c. **Definition** (*mandatory*). The meaning or use of the **data**. (See **Documentation** item for details).
- d. **Theme** (*optional*). Describes the cultural and demographic content of the **data**. (See **Theme** item for details).

- e. **Concept** (*optional, repeatable*). The human perception of a property of an object class, described independently of any particular representation.
- 1) **Object class** (*optional, repeatable*). A set of ideas, abstractions, or things in the real world that can be identified with explicit meanings and boundaries and whose properties and behaviors all follow the same rules.
 - 2) **Object class version** (*optional*). The designation identifying this as a variant of this **object class**.
 - 3) **Property** (*optional*). A feature that humans naturally use to distinguish objects.
 - 4) **Property version** (*optional*). The designation identifying this as a variant of this **property**.
- f. **Descriptive Statistics** (*optional*). Information about the statistical characteristics of the data item. (See **Descriptive Statistics** item for details).
- g. **Representation** (*optional, repeatable*). The combination of a value domain, data type, and class.
- 1) **Representation class** (*optional, repeatable*). The attribute **class** indicates the general category of data. The representation **classes** are “compound,” “number,” “text,” “category,” “classification with units,” and “classification without units.”
 - 2) **Representation type** (*optional, repeatable*). The data **representation type** indicates the kind of data value to be provided. Data **representation types** are “compound,” “date,” “integer,” “enumerated,” “latitude and longitude,” “real,” “text,” “time,” or “system defined.”
 - 3) **Value domain** (*mandatory, repeatable*). Permissible value for the **data**. An enumerated **domain** is typically expressed as a list of all permissible values. A non- enumerated **domain** may be expressed as range of permissible values.
 - a) **Permissible value** (*mandatory*). A description (range) of a valid element of a value domain.
 - b) **Definition** (*optional*). What a particular value assigned to the **data** means or (rarely) the meaning of a range of **value domains**.

- h. **Derivation** (*mandatory, if applicable*). Algorithm, coding or rule applied to identified data (**data component(s)**) by which this **data** has been created.
- 1) **Data component** (*at least one, repeatable*). The **data** used in the creation of this **data**. (See **Data** item for details.)
 - 2) **Type of derivation** (*mandatory*). Identification of how this **data** was created. This includes concatenation, the application of a mathematical function or formula, or **technique** used.
 - 3) **Derivation formula** (*mandatory, if applicable*). The particular algorithm or rule used for combining the referenced **Variable component(s)**. (See **Technique** item for details.)

3.3.6 Data quality (*mandatory unless provided with an associated survey instance component, repeatable*). Information about sampling and non-sampling errors, as well as associated statistical reporting and adjustments intended to quantify and account for these errors is reported here. There are both direct and indirect measures of **data quality**. Direct measures deal with the survey itself, while indirect measures are the result of process evaluations or comparative studies. **Planning and design** provides documentation of processes put in place to avoid such errors. Many of these measures apply only to establishment surveys and many apply only longitudinal surveys.

a. **Sampling error estimation** (*mandatory if applicable*). Sampling error is that part of the difference between a population value and an estimate thereof, derived from a random sample, which is due to the fact that only a sample of values is observed. Obviously, sampling error does not apply when every entry in a frame, the complete population of interest, is measured (this is a “census”)¹. An estimate of sampling error can be derived from the particular sample selected for the survey (from [SPWP15] page 11).

The standard measure of the accuracy of an estimator is its mean-squared error. The mean-squared error is the expected value of the squared difference between an estimator and the value it is trying to estimate. The mean-squared error is composed of a sampling variance and a square of the bias component. The sampling variance, the uncertainty caused by the fact that data is collected from only a part of the universe, is often estimable from the sampling data itself. Estimation assumptions

1. Note that Yates(81), pg. 4, distinguishes censuses from surveys based on whether a count of occurrence is the purpose (in which case the term “census” is used) or the gathering of more elaborate data by trained investigators is involved (in which case the term “survey” is used). The terms “sample” and “complete” are used to distinguish whether or not sampling has been used.

can result in sources of bias (from [SPWP15] page 28). Provide a paper explaining the following were developed, if appropriate. (See **Documentation** item for details).

- 1) **Sampling variance estimation** (*At least one*). An approximation of the uncertainty caused by the fact that data is collected from only a part of the population of interest. The choice of sampling variance estimator is dependent on the sampling technique used. Supply a value using one of the following methods.
 - a) **Design based variance** (*optional*). A design based estimate of variance depends on how the sample was selected. (See Cochran, 77 or Wolter, 85 for formulas).
 - b) **Replication estimators of sampling variance** (*optional*). Random groups, balanced half-samples, and jackknife methods are all types of replication estimator.
 - c) **Generalized variance functions** (*optional*). A simple mathematical relationship connecting the variance or relative variance of a survey estimator to the expectation of the estimator. If the parameters of the model can be estimated from past data or from a small subset of the survey items, then variance estimates can be produced for all survey items simply by evaluating the model at the survey estimates, rather than by direct computations. This method is designed primarily for multistage sample surveys of households. It is also applicable primarily to dichotomous variables.

In general, GVFs are applicable to surveys in which the publication schedule is extraordinarily large, giving, for example, estimates for scores of characteristics, for each of several demographic subgroups of the total population, and possibly for a number of geographic areas. For surveys in which the number of published estimates is manageable, the preference is for a direct computation of variance for each survey statistic (from [Wolter-85] page 201ff).
 - d) **Taylor series variance** (*optional*). An estimator based on a Taylor series approximation of the area under the curve describing the sample results.
- 2) **Bias estimation** (*Mandatory, if available*). A study of the extent to which a systematic distortion of the results occurred.. (from [DST] page 26). Supply a value if available.

b. **Specification error** (*optional*). **Specification error** is error that occurs at the **planning and design** stage of a **survey** because **data** specification is inadequate or inconsistent with the objectives of the **survey**. It may result from the difference between the quantity to be measured, such as the price or volume of a good, or the number of square feet in a housing unit, and the data collector's ability to obtain this measure. **Specification error** can result from poorly worded questionnaires and survey instructions or may reflect difficulty of measuring abstract concepts (from [SPWP15] page 34).

- 1) **Questionnaire pretest** (*optional*). A **questionnaire pretest** is a study presenting the **questionnaire** to a small sub-sample of the survey universe and analyzing the results the same way the data will be collected and analyzed. Many problems involving unclear definitions, or the wording of the questions and instructions become clear during this process (from [SPWP15] page 40). (See **Survey** item for details).
- 2) **Cognitive study** (*optional*). This is also known as a "validation study." A **cognitive study** entails examining each stage of the data collection from beginning to end to detect error caused by improper operational definitions. This includes a review of data requirements, construction of the questionnaire and survey frame, data processing and editing procedures, nonresponse follow-up, and data aggregation and publication of results (from [SPWP15] page 39). (See **Documentation** item for details).
- 3) **Comparison to independent estimates** (*optional*). This is also called "comparisons of data series." The results of the survey are compared with similar, independent estimates. When the two estimates match up, both are usually presumed accurate. When the two estimates differ systematically, it is an indication that one of the estimates is biased. Sometimes the "true" value is considered to be bounded by the two estimates. If there is an indication of bias, one or more of the following procedures is instituted: (1) matching individual respondent records from the two data series; (2) contacting respondents; and (3) contacting the survey managers and data processing specialists to determine the source of the bias (from [SPWP15] page 40). (See **Documentation** item for details).
- 4) **Records check study** (*optional*). A **records check study** compares survey responses against administrative records (from [SPWP15] page 39). (See **Documentation** item for details).

c. **Coverage error** (*optional*). **Coverage error** is the error in an estimate that results

from failure to include specified units² in the conduct of a **survey** (undercoverage), and inclusion of some units erroneously either because of a defective frame or because of inclusion of unspecified units or inclusion of specified units more than once in the actual **survey** (overcoverage) (from [SPWP15] page 44).

- 1) **Birth rate** (*optional*). **Birth rate** is the frequency with which new units are added in the scope of the survey. This is compared with other periods covered by the survey as an indirect measure of coverage error. (See **Documentation** item for details).
- 2) **Death rate** (*optional*). **Death rate**³ is the frequency with which frame or sample units in the scope of the survey die or go out of business. This is compared with other periods covered by the survey instance as an indirect measure of coverage error. (See **Documentation** item for details).
- 3) **Out of scope rate** (*optional*). **Out of scope rate** is the frequency with which frame or sample units are found not to match the selection criteria defining the survey scope. (See **Documentation** item for details).
- 4) **Unclassified rate** (*optional*). Where a classification scheme is used (for example NAICS for establishment or employment surveys), the **unclassified rate** is the frequency with which a unit is found to have not been classified. (See **Documentation** item for details).
- 5) **Misclassification rate** (*optional*). Where a classification scheme is used (for example NAICS for establishment or employment surveys), the **misclassification rate** is the frequency with which a unit is found to have been classified incorrectly. (See **Documentation** item for details).
- 6) **Duplication rate** (*optional*). The **duplication rate** is the frequency with which the enumeration of units or data is found to have been repeated or duplicated. (See **Documentation** item for details).
- 7) **Sample attrition rate** (*optional*). The **sample attrition rate** is the frequency with which units in the sample stop reporting over time. (See **Documentation** item for details).

2. "Units" are the objects of interest of the survey, i.e. the individuals, households, enterprises, establishments, or other entities.

3. In Establishment surveys, this is known as "out of business rate."

- 8) **Examination of OOS nonexistent classifications** (*optional*). This is a study of the causes of error which resulted in the failure to select units properly, that is the “out of scope” items for non-existent classifications. (See **Documentation** item for details).
 - 9) **Post enumeration survey** (*optional*). This is a second survey done of a sample of the same population, frequently using different data gathering instruments and procedures. The results of the **post enumeration survey** provides the second estimate for a “dual system estimate” or “capture, recapture” approach for determining **coverage errors**. (See **Survey** item for details).
 - 10) **Other studies** (*optional*). This includes any of the following and others: dual coverage independent estimates, match known population units against frame, check frame against alternatives, recheck unit listings, and study of frame components. (See **Documentation** item for details).
- d. **Response error** (*optional*). This is the difference between the values of the **data** collected and the correct values. **Response errors** may result from the failure of the respondent to report the correct value (respondent error), the failure of the interviewer to record the value reported correctly (interviewer error), or the failure of the instrument to measure the value correctly (instrument error) (from [SPWP15] page 57).
- 1) **Edit failure rate** (*optional*). The rate (percent of all items recorded) at which edit rules are violated. Edit rules may include tests for validity based on the contents of other data fields or a specified range or other permitted content within the data field. (See **Documentation** item for details).
 - 2) **Interviewer error rate** (*optional*). The rate (percent of all items recorded) at which the interviewer fails to record a reported value correctly. (See **Documentation** item for details).
 - 3) **Questionnaire pretest** (*optional*). A **questionnaire pretest** is a study presenting the **questionnaire** to a small sub-sample of the survey universe and analyzing the results the same way the data will be collected and analyzed. Many problems involving unclear definitions, or the wording of the questions and instructions become clear during this process (from [SPWP15] page 40). (See **Survey** item for details).
 - 4) **Cognitive study** (*optional*). This is also known as a “validation study.” A **cognitive study** entails examining each stage of the data collection from

- beginning to end to detect error caused by improper operational definitions. This includes a review of data requirements, construction of the questionnaire and survey frame, data processing and editing procedures, nonresponse follow-up, and data aggregation and publication of results (from [SPWP15] page 39). (See **Documentation** item for details).
- 5) **Records check study** (*optional*). A **records check study** compares survey responses against administrative records (from [SPWP15] page 39). (See **Documentation** item for details).
 - 6) **Detailed content reinterview study** (*optional*). The **detailed content reinterview study** involves canvassing respondents to determine information related to the information sought in the survey. This might include the respondent's understanding of the questions, the likelihood that the respondent reported the correct and complete information sought, the respondent's reaction to the questions asked and the information sought in the survey. (See **Documentation** item for details).
 - 7) **Interviewer variance study** (*optional*). A study of the variance of the information gathered by any given interviewer against all the information gathered for the entire survey. (See **Documentation** item for details).
 - 8) **Response variance study** (*optional*). The **Response variance study** determines the extent of response variation among classes of respondents and evaluates the causes for the variance. (See **Documentation** item for details).
- e. **Nonresponse error** (*optional*). **Nonresponse errors** result from a failure to collect complete information on all units in the selected sample. These are known as "unit nonresponse" and "item nonresponse." **Nonresponse errors** effect survey results in two ways. First, the decrease in sample size or in the amount of information collected in response to a particular question results in larger standard errors. Second, and perhaps more important, a bias is introduced to the extent that nonrespondents differ from respondents within a selected sample (from [SPWP15] page 68). **Nonresponse errors** are determined by collecting any or all of the following: unit response rate, weighted unit response rate, item response rate, item coverage rate, refusal rate, distribution of reason for nonresponse, comparison of data across contacts, link to administrative data for nonrespondents, estimate of nonresponse bias. (See **Documentation** item for details).
- f. **Processing error** (*optional*). This is the error in final survey results arising from the faulty implementation of correctly planned implementation methods. **Processing errors** include all post-collection operations, as well as the printing of

questionnaires. Most **processing errors** occur in **data** for individual units, although errors can also be introduced in the implementation of **systems** and estimates. (from [SPWP15] page 79).

- 1) **Keying error rate** (*optional*). The **keying error rate** is the frequency of keying verification failure. (See **Documentation** item for details).
- 2) **Coding error rate** (*optional*). Where data representing survey responses is coded, the **coding error rate** is the frequency of improper assignment of codes for the data provided. (See **Documentation** item for details).
- 3) **Edit failure rate** (*optional*). The **edit failure rate** is the frequency with which edit rules are violated. Edit rules may include tests for validity based on the contents of other data fields or a specified range or other permitted content within the data field. (See **Documentation** item for details).
- 4) **Imputation rate** (*optional*). The rate at which missing data must be provided through imputation, where applicable. (See **Documentation** item for details).
- 5) **Reclassification rate** (*optional*). This is the rate at which standard defined codes, such as NAICS, must be reassigned as the result of being erroneously coded. (See **Documentation** item for details).
- 6) **Following data values through processing study** (*optional*). The data for a sample of respondents is followed through each processing step to determine if the expected results are produced by each operation. (See **Documentation** item for details).

3.3.7 Dataset (*mandatory, repeatable*). Any permanently stored collection of information usually containing either case level data, aggregation of case level data, or statistical manipulations of either the case level or aggregated survey data, for multiple **Survey instances**.

- a. **Dataset name** (*mandatory*). The designation of the **dataset**.
- b. **Dataset abbreviation** (*mandatory*). The short name or acronym for the **dataset**. Typically, the **dataset abbreviation** is used in a system development environment or to satisfy programming language constraints.
- c. **Dataset creation time period** (*mandatory*). The time period the **dataset** was created. (See **Time period information** item for details).

- d. **Dataset version** (*mandatory*). The variant of the original **dataset**.
- e. **Dataset pointer** (*mandatory*). Pointer to the dataset. (See **Pointer to Object** item for details).
- f. **Dataset documentation** (*mandatory, repeatable*). The description and other documentation about the dataset. Also, **Dataset documentation** may be used to provide documentation related to sampling and non-sampling errors as well as other anomalies that may impact interpretation of the data provided. (See **Documentation** item for details).
- g. **Dataset theme** (*optional, repeatable*). Describes the cultural and demographic content of this dataset. (See **Theme** item for details).
- h. **Dataset format** (*optional, repeatable*). The format of the information (**data**) contained in the dataset. (See **Format** item for details).
- i. **Dataset supplier** (*mandatory if applicable, repeatable*). Source outside the Bureau of the Census. (See **Dataset supplier** item for details).
- j. **Dataset emprise** (*optional, repeatable*). An identifiable effort to generate new datasets NOT specific to a single **Survey instance**. (See **Emprise** item for details).
- k. **Dataset archive information** (*mandatory, if archived*). Information about the archived dataset. (See **Archive** item for details).
- l. **Access constraint** (*mandatory*). Level of restriction placed on access to the file. This is used by systems and appropriate personnel in determining whether the user can access the information. (See **Access constraint** item for details).
- m. **Dataset status** (*mandatory*). The state of and maintenance information for the dataset. Contains information about whom to contact concerning the status of this dataset. (See **Status** item for details)
- n. **Dataset survey instance** (*optional, repeatable*). The **Survey instance(s)** from which this dataset was generated, if applicable. (See **Survey instance** item for details).
- o. **Dataset frame** (*optional, repeatable*). The **Frame(s)** from which this dataset was generated, if applicable. (See **Frame** item for details).

- p. **Dataset sample** (*optional, repeatable*). The **Sample(s)** from which this dataset was generated, if applicable. (See **Sample** item for details).

3.3.8 Dataset supplier (*mandatory, repeatable*). A source outside the Bureau of the Census.

- a. **Supplier name** (*mandatory*). Name of the supplier.
- b. **Supplier contact** (*mandatory*). An individual knowledgeable about the supplier. (See **Contact** item for details).

3.3.9 Descriptive statistics (*optional*). Information about the statistical characteristics of the data. All statistics described here are available by executing SAS PROCs MEANS, SUMMARY, or UNIVARIATE. All notations and equations are consistent with [SAS pp.4-26].

- a. **Official statistics indicator** (*optional*). Indicates that these statistics may be referred to as "Official Census Statistics." Generally, this means that an officially approved verification process has been followed and that the Agency stands behind these results.
- b. **Number of observations** (*optional*). Number of all intended or planned observations whether data was obtained or not.
- c. **Number of non-missing observations** (*optional*). Number intended or planned observations for which data was obtained: n . X_i is the i^{th} non-missing observation. W_i is the i^{th} non-missing weight.
- d. **Number of missing observations** (*optional*). Number of observations for which data was not obtained.
- e. **Minimum value** (*optional*). The smallest value of the non-missing observations.
- f. **Maximum value** (*optional*). The largest value of the non-missing observations.
- g. **Range of values** (*optional*). **Maximum value - Minimum value.**
- h. **Sum** (*optional*). The total of values of non-missing observations.
- i. **Weighted sum** (*optional*). A total of non-missing observations that is given a measure of importance relative to other sums: $\sum W_i x_i$.

- 1) **Weighted sum technique** (*optional*). The particular technique used to develop this statistic. (See **Technique** item for details.)
- j. **Sum of weights** (*optional*). The total of weighted values of the non-missing observations.
- 1) **Sum of weights technique** (*optional*). The particular technique used to develop this statistic. (See **Technique** item for details.)
- k. **Arithmetic mean** (*optional*). A number obtained by adding several quantities together and dividing the sum by the number of quantities. The sample mean: $\bar{X} = \sum x_i / n$, an estimate of the population mean: $\mu = E(\chi)$.
- l. **Sum of squares** (*optional*). The total of the values of each observation multiplied by itself. The uncorrected **sum of squares**: $\sum W_i x_i^2$
- m. **Corrected sum of squares** (*optional*). The sum of squares corrected for the mean is the total sum of squares divided by the number of observations: $\sum W_i (x_i - \bar{x})^2$.
- n. **Variance** (*optional*). An estimation of the population variance, the expected value of the squared difference of the values from the population mean: $\sigma^2 = E(x - \mu)^2$, the sample **variance** measure of how the sum of individual observations deviated from the mean. It is the sum of the squares of differences of each observation from the mean divided by the number of observations in the population: $S^2 = \sum W_i (x_i - \bar{x})^2 / n$ or $\sum W_i (x_i - \bar{x})^2 / n - 1$ or $\sum W_i (x_i - \bar{x})^2 / \sum W_i$ or $\sum W_i (x_i - \bar{x})^2 / \sum W_i - 1$ depending on the handling of weighting.
- 1) **Variance technique** (*optional*). The particular technique used to develop this statistic. (See **Technique** item for details.)
- o. **Standard deviation** (*optional*). Square root of the variance of a set of numbers $S = \sqrt{\sum W_i (x_i - \bar{x})^2 / n}$.
- 1) **Standard deviation technique** (*optional*). The particular technique used to develop this statistic. (See **Technique** item for details.)
- p. **Standard error** (*optional*). The positive square root of the variance of the sampling distribution of a statistic. [See *A Dictionary of Statistical Terms (5th edition)* (DST) pg. 195]. The standard error of the mean: S / \sqrt{n} .

- 1) **Standard error technique** (*optional*). The particular technique used to develop this statistic. (See **Technique** item for details.)
- q. **Coefficient of variation** (*optional*). The standard deviation of a distribution divided by the arithmetic mean; sometimes multiplied by 100. It was proposed by K. Pearson (1895) for the purpose of comparing the variabilities of frequency distributions, but is sensitive to errors in the means and is of limited use. [See DST pg. 217]. **Coefficient of variation** is abbreviated as "CV." The percent **coefficient of variation**: $100 S / \bar{x}$.
- 1) **Coefficient of variation technique** (*optional*). The particular technique used to develop this statistic. (See **Technique** item for details.)
- r. **Skewness** (*optional*). A distribution which is not symmetrical. [See DST pg. 9]. The measure of sidedness: $\sum Z_i^3 xn / (n-1)(n-2)$.
- s. **Kurtosis** (*optional*). A term used to describe the extent to which an unimodal frequency curve is "peaked"; that is to say, the extent of the relative steepness of ascent in the neighborhood of the mode. [See DST pg. 111]. The measure of the heaviness of the tails or of flatness:

$$\sum Z_i^4 xn(n+1) / (n-1)(n-2)(n-3) - 3(n-1)^2 / (n-2)(n-3)$$
- t. **Students t** (*optional*). This distribution, originally due to "Student" (1908), is usually written in the form, as modified by R.A. Fisher (1925): The distribution is, among other things, that of the ratio of a sample mean, measured from the parent mean, to a sample variance, multiplied by a constant, in samples from a normal population. It is thus independent of the parent scale parameter and can be used to set confidence intervals to the mean independently of the parent variance. [See DST pg. 208]. Test of the null hypothesis, population mean = zero:

$$t = \bar{x} \sqrt{n} / S \text{ for } H_0$$
- u. **Median** (*optional*). The middle value when the values of the non-missing observations are arranged in order of value, and the **number of non-missing observations** is odd; the mean of the two middle values when the **number of non-missing observations** is even. The median is that value of the variate which divides the total frequency into two halves. As a partition value it may be defined for a continuous frequency distribution by equation. [See DST pg. 128].
- v. **Mode** (*optional*). The value at which the density of the population is at a maximum. For a sample, the most frequently occurring of the values

of the non-missing observations. The mode was originally conceived of as that value of the variate which is possessed by the greatest number of members of the population. Although the idea of the most frequently encountered or fashionable value of the variate is probably very old, it was not generally used in statistics until popularized by K. Pearson (1894). The concept is essentially used only for continuous distributions, although it can be extended to the discontinuous case. [See DST pg. 131]

3.3.10 Documentation (*optional, repeatable*). Any written or electronic material related to this element.

- a. **Documentation type** (*optional*). Memoranda, descriptions, reports, outlines, schedules, papers, summaries, manuals, specifications, procedures, program code, applets, formulas, algorithms, budgets, cost reports, progress reports, templates, presentations, diagrams, models, tables, URLs, database pointers, etc.
- b. **Documentation name** (*optional*). The designation of the document.
- c. **Documentation version** (*optional*). The designation identifying this variant of the document.
- d. **Documentation series** (*optional, repeatable*). If this document is part of a series, identifies the series in which it participates. (See **Series information** for content).
- e. **Documentation citation** (*optional*). The formal citation used to reference the document. (See **Citation** for content).
- f. **Documentation contact** (*optional*). Contact information for an individual or organization that is knowledgeable about the content of this document (See **Contact** item for details).
- g. **Documentation supplied-type** (*optional*). Indication whether supplier is submitting URL, citation, or no documentation.
- h. **Abstract** (*mandatory*). A brief description or extract of the applicable text of the **documentation**.
- i. **Documentation Theme** (*optional*). Describes the cultural and demographic content of the **documentation**. (See **Theme** item for details).
- j. **Documentation access and use constraint** (*mandatory*). Level of restriction placed on access to this documentation. This is used by systems and appropriate

personnel in determining whether or under what conditions the user can access the information. (See **Access and use constraint** item for details).

- k. **Documentation supplemental information** (*optional*). Other descriptive information about the item such as optional uses or applicable standards.
- l. **Documentation pointer** (*optional*). Pointer to the **documentation**. (See **Pointer to object** item for details).

3.3.11 Emprise (*mandatory, repeatable*). An identifiable effort to generate new datasets NOT specific to a single **Survey instance**.

- a. **Emprise name** (*mandatory*). The designation of the Emprise.
- b. **Emprise description** (*mandatory*). A summary or explanation of the Emprise. (See **Documentation** item for details).
- c. **Emprise planning and design** (*mandatory*). A record of the formulation and steps to be undertaken in accomplishing the Emprise. (See **Planning and design** item for details).
- d. **Emprise status** (*mandatory*). The state of and maintenance information for the Emprise. Contains information about whom to contact concerning the status of this Emprise. (See **Status** item for details).
- e. **Emprise dataset** (*optional, repeatable*). A dataset containing either case level data, aggregation of case level data, or statistical manipulations of either the case level or aggregated survey data, for multiple **Survey instances**. (See **Dataset** item for details).

3.3.12 Field (*mandatory, repeatable*). A description of the position and, possibly, the encoding of **data** on a physical medium or in a communication. A **format** contains one or more **fields**.

- a. **Field name** (*mandatory*). The designation of the **field**.
- b. **Field short name** (*optional*). An abbreviated designation of the **field** for use in a programming environment or to satisfy programming language constraints.
- c. **Variable** (*mandatory*). The **data** being stored retrieved or presented via this format. Note that **data** related attributes, such as **type** and **theme** are provided here. (See **Data** item for details).

- d. **Variable description** (*optional, repeatable*). The description of how the **data** is used in this **field**. Note that this is different from the description or definition of the content of the **data**, which is described under **variable**. (See **Documentation** item for details).
- e. **Subfield** (*optional, repeatable*). A physical area containing, or logical construct defining, a single element of data. (Note: **Subfields** do not contain **subfields**, i.e., *do not recurse*). (See **Field** item for details).
- f. **Field unit** (*mandatory if different from format unit*). The unit (i.e., measure) of length and position for this **field** entry in the **format**. This may include “character,” “bit,” “byte,” or “word.” If not provided, 8 bit characters are assumed. Applies to all succeeding entries for this **format** until another **field unit** entry is provided. Note that since a bit is 1 bit long and a byte is generally understood to be 8 bits long, **field unit explanation** is not needed if these units are chosen.
- g. **Field unit explanation** (*optional*). How **field unit** “character” or “word” is to be understood. Both character and word differ in meaning in different information system architectures. A word may consist of 8, 16, 32, 36, 60, or 64 bits (based on computer architectures used by the Census Bureau). A word may contain 1, 2, 4, 6, 8, or 10 characters. A character may contain 6, 7, 8, or 9 bits (again depending on the system architecture) or multiples of 8 bits (depending on **character encoding**). If not provided, **field unit explanation** is assumed to be 8 bits for both **field unit** “character” and “word.”
- h. **Character encoding** (*optional*). What character set applies to **record unit** “character.” If not provided, 8 bit ASCII is assumed.
- i. **Field type** (*mandatory unless provided in field representation*). The type of the field as stored. The types include “alphabetic,” “numeric,” “alphanumeric,” “bit,” “encoded.”
- j. **Field format** (*optional, repeatable*). Presentation or display conversion rule associated with the **field**.
- k. **Field length** (*mandatory, but see Note*). The length of the field in **record units** (or **field units** if applicable) as stored. Note: if **field starting position** and **field ending position** are provided or this is a variable length field, **field length** need not be provided.

- l. **Field starting position** (*mandatory, but see Note*). The beginning of the field in **record units** (or **field units** if applicable) as stored using this format. Note: if **field length** and **field ending position** are provided, **field starting position** need not be provided.
- m. **Field ending position** (*mandatory, but see Note*). The last position of the field in **record units** (or **field units** if applicable) as stored using this format. Note: if **field starting position** and **field length** are provided or this is a variable length field, **field ending position** need not be provided.
- n. **Field length variable** (*mandatory if applicable*). For a variable length field, what **variable** contains the length of the field. (See **Data** item for details).
- o. **Field termination** (*mandatory if applicable*). For a variable length field where the end of the field is indicated by an end of field mark or other indicator, the character or other maker that terminates the **field**.
- p. **Repetition count** (*mandatory if applicable*). The number of times the following set of **fields** is repeated.
- q. **Repetition variable** (*mandatory if applicable*). The **variable** containing the number of times the following set of **fields** is repeated. (See **Data** item for details).
- r. **Repetition termination** (*mandatory if applicable*). For a repeating set of **fields**, the character or other maker that terminates the set.
- s. **Repeating field** (*mandatory if applicable, repeatable*). Definition of the portion of the format that is repeated. (See **Field** item for details).

3.3.13 Format (*mandatory*). A description of the order and, possibly, the encoding of a set of **data** on a physical medium. A **format** contains one or more **fields**.

- a. **Format notes** (*optional*). Information about the format (source, mandate, requirements, etc.) that may be useful in understanding or maintaining it. (See **Documentation** item for details).
- b. **Format unit** (*optional*). The unit of length or positions for all fields in the **format**. This may include “character,” “bit,” “byte,” or “word.” If not provided, 8 bit characters are assumed. Applies to all entries for the **format** unless a **field unit** entry is provided. Note that since a “bit” is 1 bit long and a “byte” is generally understood to be 8 bits long, **format unit explanation** is not needed if these units are chosen.

- c. **Format unit explanation** (*optional*). How **format unit** “character” or “word” is to be understood. Both character and word differ in meaning in different system architectures. A word may consist of 8, 16, 32, 36, 60, or 64 bits (based on computer architectures used by the Census Bureau). A word may contain 1, 2, 4, 6, 8, or 10 characters. A character may contain 6, 7, 8, or 9 bits (again depending on the system architecture) or multiples of 8 bits (depending on **character encoding**). If not provided **format unit explanation** is assumed to be 8 bits for both **format unit** “character” and “word.”
- d. **Character encoding** (*optional*). What character set applies to **format unit** “character.” If not provided, 8 bit ASCII is assumed.
- e. **Field description** (*mandatory, repeatable*). A description of the position and, possibly, the encoding of a set of data on a physical medium. (See **Field** item for details).

3.3.14 Frame (*mandatory at the survey instance level, repeatable*). The set of all cases identified for a Survey instance based on a Survey’s Universe definition. Note that this is usually a dataset.

- a. **Frame name** (*optional*). The designation of this frame.
- b. **Frame documentation** (*optional, repeatable*). The description or other documentation related to the definition, creation, or use of this frame. (See **Documentation** item for details).
- c. **Frame creation time period** (*mandatory*). The **time period** over which the frame was created. (See **Time period information** item for details).
- d. **Frame access constraint** (*mandatory*). Level of restriction placed on this frame information. This is used by systems and appropriate personnel in determining whether the user can access the information. (See **Access constraint** item for details).
- e. **Frame data** (*mandatory, repeatable*). A particular data item used as a discriminator. (See **Data** item for details).
- f. **Frame dataset** (*mandatory, repeatable*). The frame dataset. (See **Data set** item for details).

- g. **Frame universe** (*optional*). The universe applicable to the frame. (See **Universe** item for details).

3.3.15 Geography (*mandatory if applicable*). Describes the locations covered by an item.

- a. **Geographic extent description** (*optional*). A brief description of the maximum geographic coverage of the item, generally a common name of the geographic region/entity/area. (See **Documentation** item for details).
- b. **Geographic entity** (*optional*). Describes the **geographic entity**(s) for which data is collected, analyzed, and/or presented.
- 1) **Geographic entity type** (*mandatory*). The type of geographic entity for which data is collected, analyzed, and/or presented. One of: "Hydrographic," "Legal, administrative," "Statistical, thematic." (See **Appendix D3**. for definitions.)
 - 2) **Geographic entity unit** (*mandatory*). The geographic unit(s) of the indicated geographic entity type for which data is collected, analyzed, and/or presented. (See **Appendix D4**. for definitions.) If **Geographic entity type** is "Hydrographic" select from "Coastal Drainage Area (CDA)," "Coastal, estuarine waters," "Drainage basin," "Estuarine Drainage Area (EDA)," "Fluvial Coastal Drainage Area (FCDA)," "Fluvial Drainage Area (FDA)," "National water quality assessment program study site," "Port," "Waterbody," "Wetlands," free text.
If **Geographic entity type** is "Legal, administrative," select from "United States," "American entity trust land," "American indian reservation," "American indian reservation with trust land," "Alaska native village," "Alaska native regional corporation," "Congressional district," "Consolidated city," "County," "Incorporated place," "Minor civil division," "Neighborhood," "Outlying area," "School district," "State," "Voting district," "ZIP code," free text.
If **Geographic entity type** is "Statistical, thematic" select from "Alaska native village statistical area," "Block," "Block group," "Census county division," "Census designated place," "Census tract or block numbering area," "Central city," "Central place," "Consolidated metropolitan statistical area," "Division," "Enumeration district," "Metropolitan statistical area," "National forest," "National monument," "National park," "Primary metropolitan statistical area," "Public domain," "Region: national," "The world," "Traffic analysis zone," "tribal designated statistical area," "Tribal jurisdiction statistical area," "Urbanized area," free text.

- c. **Geographic place name** (*optional, repeatable*). **Geographic place name** is the formal designation of the **geographic entity unit** identified above. Note that if the **geographic entity unit** identified is represented by a code, this item may not be applicable.
- d. **Geographic code** (*optional, repeatable*). **Geographic code** is the value of the **geographic entity unit** identified above.
- e. **Bounding coordinates** (*optional, repeatable*). **Bounding coordinates** establish the limits of coverage of a data set expressed by latitude and longitude values in the order western-most, eastern-most, northern-most, and southern-most. For data sets that include a complete band of latitude around the earth, the West Bounding Coordinate shall be assigned the value -180.0, and the East Bounding Coordinate shall be assigned the value 180.0.
 - 1) **West bounding coordinate** (*mandatory*). Western-most coordinate of the limit of coverage expressed in longitude.
 - 2) **East bounding coordinate** (*mandatory*). Eastern-most coordinate of the limit of coverage expressed in longitude.
 - 3) **North bounding coordinate** (*mandatory*). Northern-most coordinate of the limit of coverage expressed in latitude.
 - 4) **South bounding coordinate** (*mandatory*). Southern-most coordinate of the limit of coverage expressed in latitude.

3.3.16 Methodology (*mandatory if applicable*). A structured approach to solve a problem.

- a. **Methodology name** (*mandatory*). The designation assigned to a specific structured approach to solve a problem.
- b. **Methodology documentation** (*mandatory, repeatable*). The description of a structured approach to solve a problem. Other documentation, such as citations, algorithms, or code fragments, may be provided as appropriate. (See **Documentation** item for details).

3.3.17 Planning and design (*mandatory if applicable*). Information about the procedures, processes and methods used in the development of a particular enterprise, product or survey.

- a. **Planning and design overview** (*optional*). Brief description of the planning and design. (See **Documentation** item for details).
- b. **Variable of interest** (*mandatory, repeatable*). The name assigned to a particular datum about the population of interest. Specific data elements which have been gathered in the process of carrying out a survey. A **Variable of interest** is distinguished from other **data** in that it is one of the data items needed to accomplish the purpose of the survey through which it was gathered (or joined to). (See **Data** item for details).
- c. **Questionnaire design** (*mandatory*). The design (text, order, and conditions for skipping) of the questions used to obtain the data needed for the survey. (See **Documentation** item for details).
- d. **Geographic coverage** (*mandatory*). Describes the locations covered by the survey. These may be specified as location names, various FIPS codes for localities, Census map spots, and latitude and longitude ranges and should follow the recommendations contained in the CDDM standard for specifying geographic coverage. (See **Geography** item for details).
- e. **Implementation overview** (*mandatory*). Summaries of how the survey will be accomplished. (See **Documentation** item for details).
 - 1) **Edits** (*optional, repeatable*). Brief description of the process used to check for consistency and validity of data provided. (See **Documentation** item for details).
 - 2) **Quality control** (*optional, repeatable*). Description of the plan used to measure quality during a field listing procedure. (See **Documentation** item for details).
 - 3) **Data collection** (*optional, repeatable*). Obtaining information from respondents and the conversion of that data into a form, which can be processed. (See **Documentation** item for details).
 - 4) **NRFU** (*optional, repeatable*). Non-response follow-up. Follow-up procedures associated with handling full or partial refusal to respond by a respondent. (See **Documentation** item for details).
 - 5) **Training procedures** (*optional, repeatable*). Training of enumerators, clerks, their supervisors and other Census personnel. (See **Documentation** item for details).

- 6) **Imputation** (*optional, repeatable*). Where simulated data is substituted based on data from similar respondents adjusted for strata, and other factors. (See **Documentation** item for details).
 - 7) **Coding**. (*optional, repeatable*). The process of analyzing responses and selecting a code. (See **Documentation** item for details).
- f. **Estimation Models** (*mandatory*). Rules or methods for estimating the variance of the observed population from the universe of interest. (See **Documentation** item for details).
- 1) **Weighting** (*optional, repeatable*). The importance of an object in relation to a set of objects to which it belongs; a numerical coefficient attached to an observation, frequently by multiplication, in order that it will assume a desired degree of importance in a function of all the observations of the set. (See **Documentation** item for details).
 - 2) **Nonresponse adjustment** (*optional, repeatable*). Statistical adjustment applied to account for non-response (where respondent has been interviewed but refuses to respond). (See **Documentation** item for details).
 - 3) **Imputation** (*optional, repeatable*). Simulated data is substituted based on data from similar respondents adjusted for strata, and other factors. (See **Documentation** item for details).
 - 4) **Post-stratification** (*optional, repeatable*). Stratification of data based on data collected rather than as a part of sampling. (See **Documentation** item for details).
 - 5) **Seasonal adjustment** (*optional, repeatable*). Adjustment based on cyclic analysis. (See **Documentation** item for details).
 - 6) **Variance estimation** (*optional, repeatable*). Estimation of the variation to be expected in the responses. (See **Documentation** item for details).
 - 7) **Composite estimation** (*mandatory*). Estimation combining several techniques. (See **Documentation** item for details).
 - 8) **Estimation technique** (*mandatory*). Estimation combining several techniques. (See **Technique** item for details).

- g. **Observation Models** (*mandatory*). Models of the data acquisition process. (See **Documentation** item for details).
- 1) **Frame Coverage** (*optional*). A model of the impact of the frame design on the sampling coverage of the survey. (See **Documentation** item for details).
 - 2) **Sampling** (*mandatory*). The research strategy of collecting data from a part of a population with a view to drawing inferences about the whole. The “population” in this sense is often termed the “**universe.**” (See **Sample** item for details).
 - 3) **Measurement/observation** (*optional*). A model of the planned observations. (See **Documentation** item for details).
 - 4) **Non-response design** (*optional*). Statistical adjustment applied to account for non-interview (where a respondent is not interviewed) or partial non-response (where a respondent refuses to answer a particular question or provides an answer that is not relevant to the scope of the question). (See **Documentation** item for details).
- h. **Disclosure analysis** (*mandatory*). The process of protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, generation of simulated or synthetic data, etc. (See **Documentation** item for details).

3.3.18 Pointer to object (*mandatory if applicable*). An electronic pointer to the object of interest.

- a. **Pointer type** (*optional*). URL, database pointer, full file name, etc., of the documentation.
- b. **Pointer** (*mandatory*). The pointer to the object of interest.
- c. **File format type** (*optional, may not apply*). If the object of interest is a file, this indicates the file format (i.e. ASCII, JPEG, GIF, etc.).

3.3.19 Product (*mandatory, repeatable*). A finished deliverable of an Emprise or a Survey instance for outside the Bureau (external) use.

- a. **Product name** (*mandatory*). The name of the product.
- b. **Product release date** (*mandatory*). The date of release of the product.

- c. **Product version** (*mandatory*). The designation of this particular variant of the original **Product**.
- d. **Product series** (*optional*). Provides information about the series, if applicable. (See **Series information** item for details).
- e. **Product documentation** (*mandatory*). A description and other documentation of the product and its use. (See **Documentation** item for details).
- f. **Product emprise** (*optional*). An identifiable effort to generate new datasets NOT specific to a single **Survey instance**. (See **Emprise** item for details).
- g. **Product system** (*optional, repeatable*). The system which generated this product. (See **System** item for details).
- h. **Product survey instance** (*optional, repeatable*). The particular Survey instance that this product is related to. (See **Survey instance** item for details).
- i. **Product dataset** (*mandatory if applicable, repeatable*). Specific information about the dataset where a dataset is part of the product or input to the production of the product. (See **Dataset** item for details).
- j. **Product citation** (*mandatory*). The recommended reference to be used for the product. (See **Citation** for content).
- k. **Product status** (*mandatory*). The state of and maintenance information for the product. Contains information about how to find out about the current state of and maintenance information for this product. (See **Status** item for details).

3.3.20 Program (*mandatory*). A group of surveys related by a common theme. A Program can be made up of other Programs.

- a. **Program name** (*mandatory*). The name of the program.
- b. **Program documentation** (*mandatory, repeatable*). The description and other documentation about the program. (See **Documentation** item for details).
- c. **Program sponsor** (*optional, repeatable*). An organization which sponsors surveys or other activities conducted by the Census Bureau. (See **Sponsor** item for details).

3.3.21 Question (*mandatory if applicable*). A request for one or more related pieces of information from a respondent. A Question can contain other Questions. This is also referred to as a question block.

- a. **Question instruction** (*optional, repeatable*). Directions to the enumerator or respondent concerning answering this particular question.
- b. **Question note** (*optional, repeatable*). Documentation associated with this particular question. (See **Documentation** item for details).
- c. **Question label** (*mandatory, repeatable*). An identifier identifying this particular question or set of questions.
- d. **Question** (*optional, repeatable*). A request for information. (See **Question** item for details).
- e. **Question text** (*mandatory*). The question.
- f. **Response choice** (*mandatory if applicable, repeatable*). One of possibly several alternative answers. **Response choice** may be printed on the form, or provided as an alternative to be offered by the enumerator.
 - 1) **Response choice text** (*mandatory*). The response choice.
 - 2) **Response choice explanation** (*optional, repeatable*). A definition or other explanation associated with a particular **Response choice**. **Response choice explanation** may be printed on the form, or provided by the enumerator.
 - 3) **Response choice instruction** (*optional, repeatable*). Directions to the enumerator or respondent in the event a particular **response choice** is provided to this **question**. **Response choice instruction** may be printed on the form, or provided by the enumerator.
 - 4) **Skip pattern** (*optional, repeatable*). Rule to be applied in the event a particular **Response choice** is provided to this question. For example, this might be a branching instruction or an **NRFU** instruction.
 - 5) **Response choice key code** (*optional*). An aid to data preparation. This may be a field locator, a code to be entered for a particular **response choice**, reference or any other such aid to data preparation.
- g. **Key code** (*optional, repeatable*). An aid to data preparation. This may be a field locator, reference or any other such aid to data preparation.

- h. **Associated data** (*optional, repeatable*). The data associated with a response or response choice. (See **Data** item for details).
- i. **Questionnaire** (*mandatory, repeatable*). The **questionnaire** that this **question** is associated with. (See **Questionnaire** item for details).

3.3.22 Questionnaire (*mandatory if applicable*). An identifiable instrument containing questions for a particular survey instance.

- a. **Questionnaire name** (*mandatory*). The designation of the questionnaire.
- b. **Questionnaire documentation** (*mandatory, repeatable*). The description or other documentation associated with the questionnaire. (See **Documentation** item for details).
- c. **Questionnaire series** (*optional*). Identifies the series of which the questionnaire is or will be a part. (See **Series information** item for details).
- d. **Questionnaire time period** (*mandatory if applicable*). The date or range of dates on which the questionnaire was created and/or promulgated. (See **Time period information** item for details).
- e. **Questionnaire OMB clearance information**. (*mandatory*). The OMB (Office of Management Budget) clearance information authorizing the questionnaire.
 - 1) **Questionnaire OMB number**. (*mandatory*). The OMB clearance number authorizing the questionnaire.
 - 2) **Questionnaire OMB clearance time period** (*mandatory*). The range of dates on which the questionnaire is authorized to be used . (See **Time period information** item for details).
- f. **Question** (*mandatory, repeatable*). A request for information. (See **Question** item for details).
- g. **Questionnaire status** (*mandatory*). The state of and maintenance information for the questionnaire. (See **Status** item for details).
- h. **Questionnaire survey instance** (*optional, repeatable*). The particular **survey instance** that this **questionnaire** is related to. (See **Survey instance** item for details).

- i. **Questionnaire pointer** (*mandatory*). A pointer to the questionnaire, as a print image, file or CATI or CAPI system code. (See **Pointer to object** item for details).

3.3.23 Sample (*mandatory if applicable*). A dataset containing a subset of a Frame for a particular set of Survey Instances, selected with specific sampling technique. For a census, the sample incorporates the entire Frame.

- a. **Sample name** (*mandatory*). The designation of the sample.
- b. **Sampling technique** (*mandatory, repeatable*). The name or other identification of the specific process by which the entities of the sample have been selected. (See **Technique** item for details).
- c. **Sample creation date** (*mandatory*). The date that the sample was provided.
- d. **Sample version** (*mandatory*). A variant of the original **Sample**.
- e. **Sample documentation** (*mandatory, repeatable*). Description and other documentation associated with the sample. (See **Documentation** item for details).
- f. **Sample status** (*optional*). The state of and maintenance information for the sample design and development. (See **Status** item for details).
- g. **Sample source frame** (*mandatory*). The frame from which the sample is drawn. (See **Frame** item for details).
- h. **Sample selection criteria** (*mandatory*). Items and criteria used to make decisions about composition of a sample. (See **Documentation** item for details).
- i. **Sample size** (*mandatory*). The size of the sample.
- j. **Sample geographic coverage** (*mandatory*). Data items related to the geographic location of items in the sample. (See **Geography** item for details).
- k. **Sample dataset** (*mandatory, repeatable*). The **dataset** used in developing the **sample**. (See **Dataset** item for details).
- l. **Sample data** (*mandatory, repeatable*). The data used in the development of a sample. Specific information selected from all information available. (See **Data** item for details).

- m. **Sample error** (*optional, repeatable*). How sampling error is handled. (See **Documentation** item for details).
- n. **Subsample** (*mandatory*). A component of the sample. (See **Sample** item for details).

3.3.24 Series information (*mandatory if applicable*). Identifies the series of which the item of interest is or will be a part.

- a. **Series name** (*mandatory*). The name of the series of which the item is or will be a part.
- b. **Series issue** (*mandatory*). Information identifying the issue of the series of which the item is or will be a part.
- c. **Series date** (*mandatory*). The date of the issue of the series of which the item is or will be a part.

3.3.25 Sponsor (*mandatory if applicable*). An organization which sponsors surveys or other activities conducted by the Census Bureau.

- a. **Sponsor name** (*mandatory*). The name of the sponsoring entity.
- b. **Contact with sponsor** (*optional, repeatable*). Entity (individual or office) within the sponsoring organization who is knowledgeable about the sponsor's interest or other information related to the sponsor's relationship with the Census Bureau. (See **Contact** item for details).
- c. **Contact for sponsor** (*mandatory, repeatable*). Contact information for an individual or organization in the Census Bureau knowledgeable about the sponsor. (See **Contact** item for details).

3.3.26 Status (*mandatory if applicable*). The state of and maintenance information for the item of interest.

- a. **Progress** (*mandatory*). The state of the item of interest: update, finished, under revision, current, out of date, under review, etc.
- b. **Progress report date** (*mandatory*). The date of the progress information.
- c. **Maintenance and update frequency** (*optional*). The frequency with which changes and additions are made.

- d. **Date of last modification** (*mandatory if applicable*). The date of the last modification of the item of interest.

3.3.27 Survey (*mandatory if applicable*). An investigation about characteristics of a given universe. Note that for purposes of this standard, when the term “survey” is used, it is meant to include “census.”

- a. **Survey name** (*mandatory*). The formal designation or complete name of the survey.
- b. **Survey designation** (*mandatory if applicable*). The name by which the **survey** is known. This may be an abbreviation or acronym. For example, the “Current Population Survey” is known as the “CPS.” Therefore, its **Survey name** is “Current Population Survey” and its **Survey designation** is “CPS.”
- c. **Survey version** (*mandatory if applicable*). The identification used to identify the variant of the original **Survey**.
- d. **Survey documentation** (*mandatory, repeatable*). The description and other documentation of the survey. (See **Documentation** item for details).
- e. **Survey time period** (*mandatory if applicable*). The date or range of dates covered by the survey. (See **Time period information** item for details).
- f. **Survey program** (*mandatory if applicable*). The program under which the survey is accomplished. (See **Program** item for details).
- g. **Survey sponsor** (*mandatory if applicable, repeatable*). The organization sponsoring the survey. (See **Sponsor** item for details).
- h. **Survey planning and design** (*mandatory*). The planning and design of for this survey. (See **Planning and design** item for details).
- i. **Survey instance** (*mandatory, repeatable*). A particular instance of the survey. (See **Survey instance** item for details).
- j. **Survey universe** (*mandatory*). A **universe** used for the survey. (See **Universe** item for details).

3.3.28 Survey instance (*mandatory if applicable*). An identifiable activity which uses a **System**(s) to gather and process a set of **Data** from an identifiable set of cases, for a defined period of time, resulting in one or more specific deliverables.

- a. **Survey instance designation** (*optional*). The designation of the **Survey instance** (if not the same as the **Survey name**). This may be a generic identification (particularly for a **Survey instance component**) or a name by which this instance of the Survey is differentiated from other instance of the Survey.
- b. **Survey instance time period** (*mandatory if applicable*). The date or range of dates covered by the **survey instance**. (See **Time period information** item for details).
- c. **Survey instance status** (*mandatory*). The state of and maintenance information for the survey instance. (See **Status** item for details).
- d. **Survey instance system** (*mandatory, repeatable*). The system used in the development of this **Survey instance**. (See **System** item for details).
- e. **Survey instance survey** (*mandatory*). The survey of which this is an instance. (See **Survey** item for details).
- f. **Survey instance component** (*mandatory, if applicable, repeatable*). A **Survey instance component** of a **Survey instance** is a designated piece of the **Survey instance** that may have a unique sample and questionnaire but is considered to be integral to the **Survey instance**. A **Survey instance component** may be called a “wave,” “panel,” “component,” or other designation by developers or sponsors of the **Survey instance**. Note that since a “wave” has “panels,” a **Survey instance component** may itself have Survey instance components. (See **Survey instance** item for details).
- g. **Survey instance series** (*mandatory, if applicable*). The series information, if applicable, for this survey instance. (See **Series information** item for details).
- h. **Survey instance planning and design** (*optional*). The planning and design for this survey instance. (See **Planning and design** item for details).
- i. **Survey instance questionnaire** (*mandatory, repeatable*). The questionnaire developed for the survey instance. (See **Questionnaire** item for details).
- j. **Survey instance sample** (*optional*). The sample used for this survey instance. (See **Sample** item for details).

- k. **Survey instance sponsor** (*mandatory if different from Survey sponsor, repeatable*). The organization sponsoring this survey instance. (See **Sponsor** item for details).
- l. **Mode of data collection** (*mandatory unless provided with an associated Survey instance component, repeatable*). The actual technology used to collect the data (e.g., mail-back paper questionnaire, CATI, CAPI, fax, etc.).
- m. **Survey instance data quality** (*mandatory, repeatable*). Information about sampling and non-sampling errors, as well as associated statistical reporting and adjustments intended to quantify and account for these errors. (See **Data quality** item for details).
- n. **Survey instance dataset** (*mandatory, repeatable*). A dataset containing either case level data, aggregation of case level data, or statistical manipulations of either the case level or aggregated survey data, for a single survey instance. (See **Dataset** item for details).

3.3.29 System (*mandatory if applicable*). An identifiable process, either fully automated or computer assisted, which implements one or more **Techniques** to produce one or more deliverables. A **System** can be composed of other **Systems**.

- a. **System name** (*mandatory*). The name of the **system**.
- b. **System time period** (*mandatory if applicable*). The date or range of dates over which the system was provided. (See **Time period information** item for details).
- c. **System version** (*mandatory*). A variant of the original **system**.
- d. **System description** (*mandatory*). The description of the **system**. (See **Documentation** item for details).
- e. **Subsystem** (*mandatory if applicable, repeatable*). A **System** that is part of this **System**. (See **System** item for details).
- f. **System technique** (*mandatory, repeatable*). An identifiable algorithm which is used to implement all or part of a Methodology. (See **Technique** item for details).

3.3.30 Technique (*mandatory if applicable*). An identifiable algorithm which is used to implement all or part of a **Methodology**.

- a. **Technique name** (*mandatory*). The name of an identifiable algorithm which is used to implement all or part of a methodology.
- b. **Technique formula** (*optional, repeatable*). The formula (equation) implemented by this **technique**.
- c. **Technique documentation** (*mandatory, repeatable*). The description and other documentation related to an identifiable algorithm which is used to implement all or part of a methodology. (See **Documentation** item for details).
- d. **Technique methodology** (*mandatory, repeatable*). A structured approach to solve a problem. (See **Methodology** item for details).
- e. **Technique system** (*optional, repeatable*). **Systems** in which this **Technique** is used. (See **System** item for details).

3.3.31 Theme (*mandatory*). Describes the cultural and demographic content of the item.

- a. **Major theme** (*mandatory, repeatable*). A keyword or phrase identifying a major component or main-topic of the cultural and demographic content of the dataset. A minimum of one **Major theme** per dataset must be listed; multiple **Major themes** may additionally be listed. See **Appendix C. Themes** for a list of permitted values and **Appendix D. Definitions of the Domains** for definitions of those permitted values.
 - 1) **Definition classification** (*optional, repeatable*). Information defining terms and/or the data classification system(s). **Definition classification** consists of a description defining terms and/or the data classification system(s) or a reference to their location. The reference may be to either printed material or another dataset.
 - 2) **Minor theme** (*optional, repeatable*). A keyword or phrase describing a specific minor component or sub-topic of either the **Major theme** or previous-level **Minor theme**(s). Minor themes may be either grouped at one level (same-level entries) or nested (describe previous-level entry(s)). Nesting of **Minor themes** allows the data producer set to continue elaborating on previous-level entry(s) until he or she feels the component of the cultural and demographic content has been adequately described. See **Appendix C. Themes** for a list of permitted values and definitions of those values.
 - a) **Minor theme** (*optional, repeatable*). A keyword or phrase describing a specific minor component or sub-topic of either the **Major theme** or

previous-level **Minor theme(s)**. (See **Theme:Major theme:Minor theme** item for details).

3.3.32 Time period information (*mandatory if applicable*). Time period information specifies the date or range of dates of an event. Note: one of the following must be provided.

- a. **Single date** (*see note above*). This is the date of the event. This is not appropriate if the event takes place over a series (**multiple dates**) or a **range of dates**.
- b. **Multiple dates** (*see note above, more than one*). These are the dates on which the event occurred. This is not appropriate if the event takes place on a **single date** or **range of dates**.
- c. **Range of dates** (*see note above*). This is the period over which the event occurred. This is not appropriate if the event takes place on a **single date** or **multiple dates**. This element consists of two dates, a **beginning date** and an **ending date**.
 - 1) **Beginning date** (*mandatory if applicable*). This is the first date of a continuous set of dates over which the event occurred.
 - 2) **Ending date** (*mandatory if applicable*). This is the last date of a continuous set of dates over which the event occurred.

3.3.33 Universe (*mandatory*). The total defined set of interest to one or more surveys.

- a. **Universe name** (*mandatory*) The name of the universe.
- b. **Definition of Universe** (*mandatory*). Provides information about characteristics that distinguish this population from all other populations. (See **Documentation** item for details).
- c. **Population of interest** (*mandatory*). A description of the particular component of the universe being studied. (See **Data** item for details).
- d. **Universe documentation** (*mandatory, repeatable*). The description and other documentation of the universe. (See **Documentation** item for details).
- e. **Universe frame** (*mandatory, repeatable*). **Frame(s)** drawn from this **Universe**. (See **Frame** item for details).

3.4 Definitions

3.4.1 abstract: A brief description or extract of the applicable text of the document or object. [Team Definition]

3.4.2 citation: The formal citation used to reference the material which describes the data elements. We will be proposing a series of specific citation formats for this option. This is *to be determined*. [Team Definition]

3.4.3 URL: A pointer (*Uniform Resource Locator*) to a referenced document or object and its associated access service on the Internet or an intranet. The URL is the Client/Server access pointer as specified [RFC 1738]⁴ for use in the Hypertext Transfer Protocol (HTTP, as released by the HTTP Working Group of the Internet Engineering Task Force, IETF) [HTTP/3.0].

3.4.4 Statistical Metadata:⁵ Descriptive information or documentation about statistical data, i.e., microdata and macrodata. Statistical metadata facilitates sharing, querying, and understanding of statistical data over the lifetime of the data. The two types of statistical data (electronic or otherwise) are described as follows (Lenz, 1994):

- **Microdata** - data on the characteristics of units of a population, such as individuals, households, or establishments, collected by a census, survey, or experiment.
- **Macrodata** - data derived from microdata by statistics on groups or aggregates, such as counts, means, or frequencies.

The extensive nature of statistical metadata lends itself to categorization (Sumpter, 1994) into three components or levels:

- **Systems** - the information about the physical characteristics of the application's dataset(s), such as location, record layout, database schema, media, size, etc.;

4. Request for Comment 1738; this is an Internet Standard.

5. Defined with assistance from participants at Statistical Metadata Workshop, November 14-15, 1995, Bureau of Labor Statistics, Washington, DC.

- **Applications** - the information about the application's products and procedures, such as sample designs, questionnaires, software, variable definitions, edit specifications, etc.;
- **Administrative** - the management information, such as budgets, costs, schedules, etc.

The systems, applications, and administrative components help to differentiate the sources and uses of statistical metadata. Some authors (see, for example, Sundgren, 1991b, 1992, 1993) refer to the applications and administrative components of metadata as meta-information. We chose to use the term metadata because it seems to simplify the discussion.

3.4.5 Statistical Metadata Repository (MDR): A logically central statistical metadata repository that allows for the query, editing, and managing of metadata. Such a system provides a mechanism for looking up information about statistical products as well as their design, development, and analysis. Its design will be based on three main standards, the IRDS, SDSM, and the MMSD. The MDR is intended to be a source of metadata for all the BOC programs so comparisons of designs, processing, analysis, or data can be made across time and survey programs. The MDR will support both Internet data dissemination tools and automated integrated survey processing systems.⁶

3.4.6 Composite Metadata: The **Composite Metadata** may contain the full content of the document or other "object," and may contain another **Composite Metadata** (be recursively defined). The nature of the content of a document or other document may be limited by implementation constraints. What an object is context dependent. Generally this will be some form of text.

6. From: Gillman & Appel (1994), "Metadata Database Development at the Census Bureau."

3.4.7 Data Type:⁷ Data type indicates the kind of data value to be provided in a data item. Data types are compound, date, integer, enumerated, latitude and longitude, real, text, time, or system defined. A definition follows for each of these types. In the definitions below, the formats given in the examples are illustrative only and are not intended to provide a mandatory syntax. Possible value domain terms follow the data type definitions to which they apply.

3.4.7.1 Compound -The data element is actually described by a group of data elements either by or through intermediate compound data elements.

3.4.7.2 Date -A calendar date (Years, Months, and Days), any part or range of dates.

A.D. Era to December 31, 9999 A.D. - Express a year, month, or day.

YYYY for years.

Example: 1994 = 1994

YYYYMM for a month of a year (with month being expressed as an integer).

Example: October, 1994 = 199410

YYYYMMDD for a day of the year.

Example: October 17, 1994 = 19941017

B.C. Era to 9999 B.C. - Expresses a year, month, or day as above, preceded by the lower case letters "bc."

Example: 1010 B.C. = bc1010

B.C. Era before 9999 B.C. - Express a year only

YYYYYY . . . As many numeric characters as is needed to represent the number of the year B.C. preceded by the lower case letters "cc".

Example: 100000 B.C. = cc100000

A.D. Era after 9999 A.D. - Express a year only

As above, preceded by the lower case letters "cd" rather than "cc".

Example: 10255 A.D. = cd10255

Single Date consists of one date only.

Example: 1 date (1) October 17, 1990 = 19901017

Multiple Dates are more than one date. **Multiple Dates** are considered a single entry. The multiple dates are separated by commas.

⁷ These definitions are derived from Federal Geographic Data Committee - Subcommittee on Cultural and Demographic Data, "Cultural and Demographic Data Metadata", Final Draft, January 18, 1995.

Example: 3 dates (1) October 17, 1990, (2) October 17, 1992, and (3) October 17, 1994 = 19901917,19921916,19941917

Range of Dates is a single entry. **Range of Dates** consist of two dates. The beginning date always precedes the ending date. The two dates are separated by a comma.

Example: Between October 17, 1990 and October 17, 1992: 19901017,19921017

3.4.7.3 Integer - Integer numbers.

3.4.7.4 Enumerated - A selection of all possible values in a list.

3.4.7.5 Latitude and Longitude-Values for latitude and longitude are expressed as decimal of degrees. Whole degrees of latitude are represented by two-digit number from 0 through 90. Whole degrees of longitude are represented by a three-digit number ranging from 0 through 180. A decimal fraction of a degree is separated the whole number of degrees by a decimal point. Decimal fractions of a degree may be expressed to the precision desired.

Latitudes north of the equator are specified by a plus sign (+) or by the absence of minus sign (-) preceding the two digits designating degrees of latitude. Latitudes of the equator are specified by a minus sign (-) preceding the two digits degrees of latitude. A point on the Equator is assigned to the Northern A latitude of +90 (90) or -90 degrees specifies the position at the South Pole, respectively. The component for longitude may have any legal value.

Longitudes east of the prime meridian are specified by a plus sign (+) or by the minus sign (-) preceding the three digits designating degrees of longitude. West of the prime meridian is specified by a minus sign (-) preceding the digits designating degrees of longitude. A point on the prime meridian is assigned the Eastern Hemisphere. A point on the 180th meridian is assigned to the Western. One exception to this last convention is permitted. For the special describing a band of latitude around the earth, the East Bounding Coordinate element is assigned the value +180 (180) degrees.

3.4.7.6 Real - Real numbers.

3.4.7.7 Text - Characters. Text consisting of one or multiple blank spaces must be in quotations.

3.4.7.8 Time - The time of day (Hours, Minutes, and Seconds). One of three conventions may be used. One convention is selected for the metadata.

Local Time - A record of time in local time. Values follow the 24-hour timekeeping system for local time of day in the hours, minutes, seconds, and decimal fractions of a second (to the precision desired) without separators convention (general form HHMMSSSS.)

Example: 2:30 p.m. Eastern Daylight Time, with precision to the minute = 1430.

Local Time with Time Differential Factor - A record of time in local time and the Universal Time (Greenwich Mean Time). Values follow the 24-hour timekeeping system for local time of day in hours, minutes, and decimal fractions of a second (to the precision desired) without separators. This value is followed, without separators, by the time differential factor. Time differential factor expresses the difference in hours and minutes between local and Universal Time. It is represented by a four-digit number preceded by a plus (+) or a minus sign (-), indicating the hours and minutes the local time is ahead of or behind Universal Time, respectively. The general form is HHMMSSSSshhmm, where HHMMSSSS is the local time using 24-hour timekeeping (expressed to the precision desired), 's' is plus or minus sign for the time differential factor, and hhmm is the time factor.

Example: 2:30 p.m. Eastern Daylight Time, with precision to the minute = 1430-0400.

Universal Time (Greenwich Mean Time) - A record of time in Universal Time (Greenwich Time). Values follow the 24-hour timekeeping system for Universal Time of day in hours, seconds, and decimal fractions of a second (expressed to the precision desired) separators convention. The upper case letter "Z" directly follows the (or extreme right hand) time element of the 24-hour clock time expression. General form is HHMMSSSSZ, where HHMMSSSS is Universal Time using 24-hour and Z is the letter "Z".

Example: 2:30 p.m. Eastern Daylight Time, with precision to the minute = 1030Z

Single Time consists of one time only.

Example: 1 time (1) 10:00 a.m. Eastern Daylight Time = 1000

Multiple Times are more than one time. Multiple Times are considered a single entry. Multiple times are separated by commas.

Example: 2 times (1) 10:00 a.m. Eastern Daylight Time and (2) p.m. Eastern Daylight Time = 1000,2200

Range of Times is a single entry. Range of Times consist of two times. Beginning time always precedes the ending times. The two times are separated by a comma.

Example: Between 10:00 a.m. and 10:00 p.m. Eastern Standard Daylight Time = 1000,2200

3.4.7.9 System Defined - Any type unique to a particular application or implementation of this standard. Details are in documentation provided with the application or by the implementor.

4. Training

Training on this standard *should* be provided to all potential developers of tools for use with the BOC MDR. Training on the actual use of the BOC MDR and on the use of appropriate tools developed in compliance with this standard *shall* be provided via a Census Bureau-wide seminar coordinated by the IT Directorate.

5. Reference Materials and Related Standards

5.1. Reference Materials

Appel, M. V., Gillman, D. W., LaPlant, W. P. Jr., Creecy, R. H. (1996), "Toward Unified Metadata Systems and Practices," ISIS-96, Bratislava, Slovakia, May 21-24, 1996.

Bean, E. C. Jr. (1991), "ASCII BOX Files for Data Portability," Census Bureau internal document, Office of the Director - Demographic Programs; Washington, DC;1991.

Capps, C. (1995), "Overview of the Technical Architecture for FERRET," Census Bureau internal document, Demographic Surveys Division; Washington, DC;1995.

Census Bureau (1996a), "Statistical Design and Survey Methodology Metadata Content Standard," Draft, Census Bureau Internal Document, Statistical Research Division; Washington, DC; July 2, 1996.

Census Bureau (1996b), "Table of Contents for Statistical Design and Survey Methodology Metadata Content Standard," Draft, Census Bureau Internal Document, Statistical Research Division; Washington, DC; July 2, 1996.

Clinton, William (1994), "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," Executive Order 12906, Washington, DC; April 11, 1994.

Cochroan, William G., (1977). *Sampling Techniques* (3rd edition). New York, NY John Wiley & Sons, 1977.

[DST] Marriott, F.H.C., Wilson, Christopher, ed. *A Dictionary of Statistical Terms* (5th edition). International Statistical Institute; Essex, England, 1990.

FGDC (1994), Federal Geographic Data Committee, "Content Standards for Digital Geospatial Metadata," Federal Geographic Data Committee, Washington, DC; June 8, 1994.

FGDC (1995), Federal Geographic Data Committee - Subcommittee on Cultural and Demographic Data, "Cultural and Demographic Data Metadata," Final Draft, Federal Geographic Data Committee, Washington, DC; January 18, 1995.

[FIPS-123] National Institute of Standards and Technology. *Data Descriptive File for Information Interchange (DDF)*. Federal Information Processing Standard 123. U.S. Department of Commerce; Washington, DC; 1992. Adopts, with modifications, International Standard 8211-1985.

[FIPS 156] National Institute for Standards and Technology. *Information Resource Dictionary System (IRDS)*. Federal Information Processing Standard 156. U.S. Department of Commerce; Washington, DC; April 5, 1989.

[FIPS-173] National Institute of Standards and Technology. *Spatial Data Transfer Standard (SDTS)*. Federal Information Processing Standard 173. U.S. Department of Commerce; Washington, DC; 1992.

[FIPS-192] National Institute of Standards and Technology. *Application Profile for the Government Information Locator Service (GILS)*. Federal Information Processing Standard 192. U.S. Department of Commerce; Washington, DC; 1994.

Gillman, D. W. and Appel, M. V. (1994), "Metadata Database Development at the Census Bureau," Presented at the UN/ECE METIS Working Group Meeting, Geneva Switzerland, November 22-25, 1994.

Gillman, D. W., Appel, M. V., and LaPlant, W. P. Jr. (1996A), "Design Principles for a Unified Statistical Data/Metadata System," Proceedings of SSDBM-8, Stockholm, Sweden, June 18-20, 1996.

Gillman, D.W., Appel, M.V., LaPlant, W.P. Jr. (1996B), "Statistical Metadata Management: A Standards-based Approach," ASA Section on Survey Research Methods, Chicago, IL, August 4-8, 1996.

Graves, R. B. and Gillman, D. W. (1996), "Standards for Management of Statistical Metadata: A Framework for Collaboration," ISIS-96, Bratislava, Slovakia, May 21-24, 1996.

ISO/IEC JTC1 SC14 (1996), ISO/IEC 11179-1; *Specification and Standardization of Data Elements: Part 1 - Framework for the Specification and Standardization of Data Elements*. Committee Draft, September, 1996.

[ISO/IEC 11179-5]; *Specification and Standardization of Data Elements: Part 5 - Naming and Identification Principles for Data Elements*. International Standard 11179-5:1995, June 1995.

[ISO/IEC 10032]; *Reference Model for Data Management*, International Standard 10032:1995(E).

LaPlant, W. P. Jr., Lestina, G. J. Jr., Gillman, D. W., and Appel, M. V. (1996), "Proposal for a Statistical Metadata Standard," Census Annual Research Conference, Arlington, VA., March 18-21, 1996.

Lenz, H.-J. (1994), "The Conceptual Schema and External Schemata of Metadatabases," Proceedings of SSDBM-7, pp160-165, Charlottesville, VA, September 28-30, 1994.

Kendall, Maurice G. and William R. Buckland.(1960), *A Dictionary of Statistical Terms*, (2nd Ed.) Hafner Publishing Co.; 1960.

[OIRA-83] Office of Information and Regulatory Affairs. *Voluntary Standard: Procedures for Preparation of Abstracts for Public Use Statistical Machine-Readable Data Files*. Office of Management and Budget; Washington, DC; October 1983.

Prescott, Roland, Wilson, Christopher, ed. *The Dictionary of Demography*. Basil Blackwell, N.Y. 1985.

Reinvention Lab of the Census Bureau (1994), "Integrated Processing System," Systems Planning Document, Washington, DC; December 15, 1994.

[RFC1738] Berners-Lee, T., *et al.*, "Uniform Resource Locators (URL)," 12/20/1994; RFC 1738.

Rosen, B. and Sundgren, B. (1991), "Documentation for Reuse of Microdata from the Surveys Carried Out by Statistics Sweden," Research and Development Statistics Sweden, June 28, 1991.

[SAS].SAS Institute, *SAS Procedures Guide*, Version 6, 3rd Edition 1990

[SPWP15] Office of Information and Regulatory Affairs. *Statistical Policy Working Paper 15: Quality in Establishment Surveys*. Office of Management and Budget; Washington, DC; July 1988. .

StEPS (1996), "Standard Economic Processing System Document 1: Concepts and Overview," Draft 2, Internal Census Bureau Document, Washington, DC; April 16, 1996.

Sumpter, R. M. (1994), "Whitepaper on Data Management," Lawrence Livermore National Laboratory document, 1994.

Sundgren, B. (1991a), "Toward a Unified Data and Metadata System at the Australian Bureau of Statistics - Final Report, December 2, 1991.

Sundgren, B. (1991b), "Statistical Metainformation and Metainformation Systems," R&D Report Statistics Sweden, 1991:11.

Sundgren, B. (1992), "Organizing the Metainformation Systems of a Statistical Office," R&D Report Statistics Sweden, 1992:10.

Sundgren, B. (1993), "Guidelines on the Design and Implementation of Statistical Metainformation Systems," R&D Report Statistics Sweden, 1993:4.

Sundgren, B., Gillman, D. W., Appel, M. V., and LaPlant, W. P. (1996), "Toward a Unified Data and Metadata System at the Census Bureau," Census Annual Research Conference, Arlington, VA., March 18-21, 1996.

Szemraj, John A. and Billy R. Tolar. "Profile Development for the Spatial Data Transfer Standard"; paper delivered at *GIS/LIS*, Minneapolis, Minnesota; November 2-4, 1993.

Wolter, Kirk M. (1985), *Introduction to Variance Estimation*; Springer-Verlag, 1985, New York, NY.

Yates, Frank. *Sampling Methods for Censuses and Surveys*, (4th Ed.); Charles Griffin & Co. Ltd., 1981, London & High Wycombe, England.

Zeisset, P. T. (1993), "Meta-Information for Summary Statistics: The EXTRACT Experience," Proceedings of Statistical Metainformation Systems Workshop, Luxembourg, February 2-4, 1993.

5.2. Related Standards

5.2.1. Information Resource Dictionary System (IRDS)

IRDS is a standard which addresses the use, control, and documentation of the information resources of an enterprise (NIST, 1989). It is an application of another standard, **Reference Model for Data Management (RMDM)** (ISO/IEC, 1995b). RMDM specifies

a series of interlocking pairs of databases and schema, where (informally) the schema stored in the database at one level defines the database stored at the next lower level.

The IRDS uses three interlocking level-pairs of the RMDM specification. The levels are as follows:

- **application level** - which is the survey data itself;
- **dictionary level** - which describes the actual data in an application, i.e., contains the schema for the databases containing survey data;
- **dictionary definition level** - which describes the schema used at the dictionary level, i.e., contains a description of how the schema fit together (the MDR ISM model);
- **definition schema level** - which describes the schema used at the dictionary definition level, i.e., contains the modeling paradigm for the MDR ISM model (the meta-metamodel).

The level-pairs are each of the three pairs of levels going down the list above:

- application - dictionary
- dictionary - dictionary definition
- dictionary definition - definition schema.

These pairs can be viewed in a contextual frame (Graves and Gillman, 1996) by observing that the application - dictionary pair is the “operational” context, i.e., the operations of the BOC. The dictionary - dictionary definition pair is the “design and maintenance” context, i.e., the business activities of the BOC (c.f., the Business Data Model (see Appendix B) and the Data Element Registry (Appendix A, 2.3.3)). Finally, the dictionary definition - definition schema pair is the “policy and standards” context, i.e., standards, methods, and rules the BOC uses to determine the way it conducts business.

5.2.2. Data Element and Data Element Registry Standards

Data elements (or variables) are the fundamental units of data an organization collects, processes, or disseminates. A data element registry is a mechanism for managing data elements in a logical fashion. Data element registries organize information about data elements, provide access to the information, facilitate standardization, help identify duplicates, and facilitate data sharing.

Data element registries are like data dictionaries in that they contain definitions of data elements. But more than data dictionaries, they contain all the information about individual data elements that an organization requires. Data dictionaries are usually

associated with single datasets (files or databases), but a data element registry contains information about the data elements for an entire program or organization.

The information contained in a data element registry is part of an organization's metadata. Therefore, the registry itself will be part of MDR (see Appendix A, A2.3.3). When the information about data elements is organized into a registry or repository, it can be made available to a variety of people and processes.

The model for the data element registry will be based on the conceptual framework contained in the draft ANSI standard, **The Metamodel for the Management of Sharable Data (MMSD)**, ANSI X3.285. It, in turn, incorporates all the principles described in an emerging international standard, **Specification and Standardization of Data Elements**, ISO/IEC 11179 (ANSI X3L8, 1996). X3.285 provides a conceptual metamodel for building a data element registry. A complete data dictionary describing all the entities, attributes, and relationships of the conceptual metamodel is included in this document.

The MMSD metamodel provides a detailed description of the types of information which should belong to a data element registry. It provides a framework for how data elements are formed and the relationships among the parts. Implementing this scheme will provide the mechanism for delivering to users the information they need to understand an organization's data elements.

5.3 Other Standards

There are a number of other standards which will be of help in the design and use of the MDR and which may be thought of as extended by this standard. These include data interchange standards such as the **Spatial Data Transfer Standard (SDTS)**, FIPS-173, (FGDC, 1994); the US indexing system service standard, **ANSI/NISO Z39.50**; and its profiles for automated indexing on the Internet: the **Government Information Locator Service (GILS)**, FIPS-192; and proposed and existing **Wide Area Information Services (WAIS)**.

The three standards described below are all products of the Federal Geographic Data Committee (FGDC). The

5.3.1. Spatial Data Transfer Standard (SDTS)

The SDTS is designed to assist in moving the contents of Geographic Information Systems (GIS) databases between GIS servers and to exchange data between systems that have the capability of generating, analyzing, storing, or displaying geographic data. This is important because most of the data collected by the BOC has a geographic component to it.

The SDTS is designed to assist the movement of geographic information between two dissimilar GIS servers or to exchange data with systems that have at least some appropriate capabilities for generating, analyzing, storing, or displaying such data. The SDTS currently consists of four parts:

- Part 1 is a model of spatial phenomena, objects and features and a syntax and semantics for using the model. This part provides the rules for building any set of SDTS data transfer files. These files consist of a metadata part and a data part.
- Part 2 contains a list of attribute values and definitions. This is, in effect, the beginnings of a thesaurus for metadata descriptions of the meaning of the data being transferred.
- Part 3 provides a representation of an SDTS data file set, using the DDF. This is a profile, a selection of particular aspects of many ISO standards and related FIPS to be used in generating the data transfer file.
- Part 4 the *Topological Vector Profile (TVP)*, contains specifications for an SDTS profile for use with geographic vector data with planar graph topology.

5.3.2. Content Standard for Digital Geospatial Metadata (CSDGM)

The CSDGM provide a common set of terminology and definitions for the documentation of digital geospatial data.. The CSDGM is intended to provide the metadata components of the SDTS [FIPS-173]. The use of CSDGM supports Federal Agency compliance with Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure."

5.3.3. Standard for Cultural and Demographic Data Metadata (CDDM)

The CDDM standard includes a mapping to the metadata components of the CSDGM, and follows the syntax and semantic rules specified in the CSDGM. It is intended to be used as part of the metadata component of SDTS. Upon completion, this standard may be issued as a FIPS.

6. Conversion Requirements

The metadata repositories currently in use at the Census Bureau *shall* be compliant with provisions of this standard within one year of its issuance. Access tools and interfaces *shall* be

developed to allow the sharing of metadata compliant with provisions of this standard within 90 days of its issuance.

7. Compliance Methods and Timetable

7.1. Compliance Methods

In order to have access to the corporate metadata repository, tool developers *shall* comply with the provisions of this standard. Only tools and systems using, providing, or accessing metadata which are capable of being compliant with provisions of this standard *shall* be acquired or developed. Waivers to this standard will be allowed if fully justified by processing requirements or cost constraints. All applications for waivers to this standard *shall* be made in writing, following the procedures established through the IT Standards Program.

7.2. Timetable

This standard *shall* be effective at the date of issuance.

Existing metadata repositories *shall* be compliant within one year of the issue date.

Access tools and interfaces for existing metadata repositories *shall* be compliant within 90 days.

8. Authority

8.1. Policy (Draft)

It is the Census Bureau's policy to promote and facilitate the sharing of metadata between internal and external organizations. With the provision of appropriate tools, all BOC employees responsible for the design and development of Census products, publicly available datasets or datasets being exchanged with other agencies, and associated documentation *shall* provide appropriate metadata in accordance with this standard. With the provision of appropriate tools, all BOC employees responsible for the design and development of Census projects, systems, and associated datasets and documentation, which are intended to be used by more than one office or division within the BOC, *shall* provide appropriate metadata in accordance with this standard.

8.2. Mandatory Standards

None.

Note: You are invited to review the appendices referenced in this document and can get a hard copy from the Software and Standards Branch, SSD, Room 1385-3 or view the documents on the Intranet @ [to be provided later]

Important Dates	
Originally Issued	month day, year
Next Review Scheduled	month day, year
Compliance	See section 7.2

Contacts for Additional Information
<p>To find out:</p> <ul style="list-style-type: none"> ● where to send comments on specific standards ● where to send suggestions to improve standards ● what standards have been issued ● general information about IT Standards Program <p>Look at: IT Standards in Progress Bulletin Board IT Policies and Standards Bulletin Board</p> <p style="text-align: center;">or</p> <p>Contact: Software and Standards Management Branch in SSD 457-2142; Room 1385-3; 457-4606 (FAX) E-mail to <i>IT Standards Program</i> account</p>

Appendix A. Metadata Repository

A1. Introduction

Statistical metadata (see 3.4 Definitions) is the information and documentation needed to describe and use statistical datasets for the lifetime of the data. The main focus of the statistical metadata research at the Bureau of the Census (BOC) is the design and implementation of a prototype logically central statistical metadata repository (MDR) for use with Internet data dissemination systems and automated integrated survey processing systems.

Integrating the components of metadata management requires careful planning. Standards are the vehicles which will enable this work. International, national, U.S. Government, and BOC standards are all being brought to bear to solve these problems.

Too often, metadata is scattered, incomplete, or missing. Many times the only source for some information is from subject matter experts. The effective and efficient management of statistical metadata greatly increases the usefulness of statistical data. Since metadata is data, it can be stored and retrieved in a repository (see section 3.3) just as the data it describes is stored and retrieved in a database. Statistical metadata can also be used to facilitate survey design, processing, management, and analysis. It is the electronic storage and organization of statistical metadata which will allow the BOC to develop automated survey design, processing and dissemination systems in ways that were not possible before.

The eventual plan for the MDR is that it will be a catalogue of the metadata for survey designs, processing, analyses, and datasets for all surveys the BOC performs. Links to the data files, documentation, and images (such as questionnaire forms) will be made (Sundgren, *et al*, 1996; or Appel, *et al*, 1996).

A2. Statistical Metadata Repository (MDR)

The MDR is being designed to assist with two new types of tools which are under development at the BOC: Internet data dissemination (e.g., DADS, FERRET); and automated integrated survey processing systems (e.g., StEPS). These tools correspond to the end-user oriented purpose and production oriented purpose, respectively, of statistical systems.

- End-user oriented systems: supports users of statistical information, e.g., through Internet data dissemination systems; and
- Production oriented systems: supports the planning, design, operation, processing, and evaluation of statistical surveys, e.g., through automated integrated processing systems.

The current goal of the MDR project is to build a prototype for a standards-based, logically central metadata repository. There are three standards which form the basis for the design of the MDR:

- **Survey Design and Statistical Methodology (SDSM)** metadata standard (this document);
- **Information Resource Dictionary System (IRDS)** standard ANSI X3.138, (NIST, 1989); and
- **Metamodel for the Management of Sharable Data (MMSD)**, ANSI X3.285 (draft) see section 5.2.2.

The purpose of the SDSM and MMSD standard is to specify the basic metadata models necessary to build the MDR. IRDS specifies the framework in which the models will be organized.

The MDR will support metadata tools and metadata-driven systems. Metadata-driven systems at the BOC are primarily electronic data dissemination or automated survey processing systems that use metadata to guide the user and provide them with choices and information at each step. These systems can be either end-user oriented or production oriented.

A2.1 Purposes for Statistical Metadata Repository

There are many functions for which the MDR is being designed. Primarily, the MDR will be a standard tool for researchers and analysts to locate data and descriptions of surveys. Data dictionaries, record layouts, questionnaires, sample designs, and standard errors are examples of information that will be directly available. Links from subject types, e.g., income, race, age, and geography, to datasets will allow users to locate datasets by subject. Less obviously, users can compare designs of different surveys and find common information collected by different surveys.

Because the BOC manages data in a decentralized and non-uniform way, the MDR will bridge the gap between the data and the users who wish to find it. On one hand there is a need for the managers of each survey to create and manage their data in the most efficient way for their processing needs. On the other, there is a need for data users to be able to find and access data efficiently and effectively. The MDR will facilitate a solution for the data users while allowing the survey data managers to find a smooth transition to standard data management strategies.

The MDR will also help facilitate data administration at the BOC. Many surveys define data elements with the same name but with (slightly) different definitions. An aim of the MDR is to help people manage this problem. If definitions and other attributes of data

elements are standardized across surveys, through the use of a data element registry (a subset of MDR), then confusion generated by the differences in meaning will be reduced. Naming standards and conventions are also needed to reduce the confusion. The MDR will provide the information necessary for the user to understand the distinctions and similarities among data elements from multiple data sources. The design of the data element registry part of the MDR will be based on the MMSD.

A2.2 Architecture for Statistical Metadata Repository

There are several dimensions to the MDR model. Each dimension is a separate model in itself. They are listed here:

- **Business Data Model** - This model defines the metadata and the relationships necessary to describe the business (surveys and survey data) of the BOC. The SDSM standard defines the entities and attributes of this model.
- **Data Element Registry** - This model is based on the MMSD. A data element registry is a mechanism for managing the names, definitions, and other attributes of data elements. Implementation should provide easy access and links to the information about an organization's data elements. See 5.2.2, Data Element and Data Element Registry Standard for more details.
- **Metamodel** - The Metamodel is the repository's view of the business data model and the data element registry. This model describes the schema of information about datasets, such as specifications for record layouts or database schema. It also contains versioning, security, user type, and search category information.

The MDR architecture is the integration of these models. We will refer to this integrated model as the **integrated statistical metadata (ISM)** model. Repository access, query, and update tools will be built to let other tools, analysts, designers, and researchers use the repository. The repository models will help determine how these tools are designed.

A2.3 Metadata Management

The main aspects of managing metadata are content, storage, collection, delivery, and administration. This section will describe how the standards-based approach and the proposed design architecture address each of these aspects.

A2.3.1 Content and Storage

Content refers to the metadata which will be collected and stored in the MDR. These are the entities and attributes defined by the SDSM standard. Storage refers to how, i.e., the physical and logical mechanisms for storing the metadata. The prototype MDR being built uses Oracle RDBMS as its underlying storage mechanism and is based on the IRDS standard. The IRDS presents a framework for organizing the statistical metadata of the MDR. The ISM model is a logical view of how the statistical metadata will be organized in the MDR, and it fits into the framework of IRDS. The SDSM, MMSD, and other standards (5.3, Other Standards) define the content which is used to design the ISM.

A2.3.2 Collection and Delivery

The Collection and Delivery refer to creating, replacing, updating, deleting, and querying metadata in the MDR. Delivery (querying) will be part of the design of statistical information systems which work with the MDR. User interfaces for metadata-driven systems will let users query the metadata to locate data or other survey information. Query languages such as SQL (for RDBMS's) will allow the user to retrieve any metadata which is in the MDR. Other search mechanisms such as WAIS, key word, and hyper-text are available through the Internet.

Several metadata collection tools are in place at the BOC and other statistical agencies. SCBDOK (at Statistics Sweden), Document Management System (DMS - in use with FERRET at BOC), and the commercial document management system PCDOC (for 1997 Economic Censuses) are all designed or being designed under the framework outlined above. Metadata collection is recognized as a very difficult problem because of the fundamental changes that the survey design and analyst teams must go through to perform their work. At the BOC and other statistical agencies, metadata (mostly documents, often in the form of memos) is created either electronically or on paper for each survey, but it is just beginning to be stored in an organized repository, database, or document management system. Asking people to use a new system to capture this metadata and organize it represents a big change.

A2.3.3 Data Administration

The adoption of the SDSM standard is the start of data element standards by the BOC. The implementation of the MDR will require more than supplying information about data elements. Data administration is the active

management of the information about all the agency's variables. No function of this type, at the agency level, exists at the BOC.

The Data Element Registry standards described in section 5.2.2 detail the information that is required for accurate and complete data administration. They also describe the organization of a data element registry for that information. The registry, of course, will be part of the MDR.

There is a human side to data administration. Some of these functions are:

- designing and implementing naming conventions;
- designing rules for forming data element definitions;
- determining which data elements have the same meanings as others;
- ensuring all necessary information is properly supplied for each registered data element;
- working with data administrators of other agencies to facilitate the sharing of data.

Data administration will require a commitment from the BOC, but it will greatly enhance the usefulness of BOC data, make the MDR a better tool, and facilitate the sharing of data between groups within the BOC or with other agencies.

A3 Who or What is Affected

All BOC employees or sponsoring organizations who document BOC surveys are affected. All BOC disseminated data will be affected because it will now have a standard set of documentation which describes the data.

Appendix C. Themes

C1 A Model for Cultural and Demographic Data

Whereas for geospatial data there exist the fundamental units of points, lines, or areas, a corresponding data unit for cultural and demographic data is elusive at present and difficult to conceptualize. Cultural and demographic data are the data contained in or represented by the geographic framework units.

The identification of the basic cultural and demographic data unit seems to be centered on three questions:

QUESTION 1.

Does the data describe a (human) activity, for example an economic activity or a land use activity?

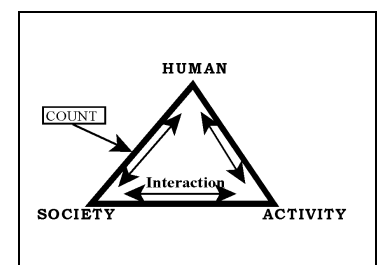
QUESTION 2.

Does the data describe an aspect of humans, for example their health or age?

QUESTION 3.

Does the data describe an aspect of (human) society, for example its political, social, or historical aspects?

It appears all cultural and demographic data fits into at least one and possibly more of these categories similar to the way in which typical geographic data may be categorized as points, lines, and areas. Loosely applying the "geometric" model, one can consider the human as the point, society as the line (linking humans and bounding activities), and activities as taking place over or in relation to an area. And just as there appears to be a common reference to all typical geographic features by position, there appears to be a common reference to all cultural and demographic data by count, such as its number or amount. Figure C-1 depicts this model graphically.



The standard uses this conceptual data model in identifying the geospatially referenced cultural and demographic data metadata core and defining geospatially referenced cultural and demographic data user-defined metadata elements. The model additionally requires the introduction of new production rules for use with the geospatially referenced cultural and demographic data user-defined metadata elements as fundamental to describing the cultural and demographic data subject matter content (themes). Theme prioritizing, i.e., hierarchical or nested relationships, is the most difficult task for the metadata producer given the complexity of ideas which can make up a dataset.

C2 The Theme Metadata Entity

The following is an expanded version of the **Theme** entity, containing a list of the allowed values for **Major theme** and **Minor theme** below:

3.3.28 Theme (*mandatory*). Describes the cultural and demographic content of the item.

- a. **Major theme** (*mandatory, repeatable*). A keyword or phrase identifying a major component or main-topic of the cultural and demographic content of the dataset. A minimum of one **Major theme** per dataset must be listed; multiple **Major themes** may additionally be listed. Here is a list of permitted values:

“Agriculture” “Archeology” “Architecture, art” “Archives, museum, museum collections” “Attitudes, outlook” “Banking, finance, insurance” “Business enterprise” “Communications” “Construction” “Customs” “Diet, traditional foods” “Domestic trade, service” “Economics” “Education” “Energy” “Environment” “Family life” “Finances, employment (federal)” “Finances, employment (state, local)” “finances, employment (corporate)” “Fisheries” “Foreign trade, commerce” “Forests, forest products” “Geography” “Government, politics” “Health, nutrition” “Heritage” “History” “Housing” “Income, expenditures, wealth” “Industrialization” “Labor force, employment, earnings” “Language” “Law enforcement, courts, prisons” “Manufacturers” “Migration, naturalization” “Mining, mining products” “National defense, veterans affairs” “Nationality” “Navigation” “Parks, recreation” “Population” “Prices” “Science, technology” “Settlement” “Skills” “Social insurance, Human services” “Transportation” “Vital statistics” “War”

- 1) **Definition classification** (*optional, repeatable*). Information defining terms and/or the data classification system(s). **Definition classification** consists of a description defining terms and/or the data classification system(s) or a reference to their location. The reference may be to either printed material or another dataset.
- 2) **Minor theme** (*optional, repeatable*). A keyword or phrase describing a specific minor component or sub-topic of either the **Major theme** or previous-level **Minor theme(s)**. Minor themes may be either grouped at one level (same-level entries) or nested (describe previous-level entry(s).) Nesting of **Minor themes** allows the data producer set to continue elaborating on previous-level entry(s) until he or she feels the component of the cultural and demographic content has been adequately described. Here is a list of permitted values:

“Abortions” “Accidents, safety” “Acquisitions” “Acreage” “Activities” “Age composition” “Aid” “Air, land, water” “Aircraft” “Amount” “Alien registration”

“Armed forces” “Arrests, sentences, appeals” “Assets, liabilities” “Attendance”
“Benefits” “Biota” “Births, deaths” “Boundary” “Building” “Businesses”
“Campaigns” “Capacity” “Capital punishment” “Cargo” “Care” “Catch” “Cause”
“Cemetery” “Climate” “Commercial” “Commuters, drivers” “Composition”
“Conditions” “Consumption” “Contaminants” “Courts” “Crime” “Criminal justice
system” “Criminals” “Crops” “Currency” “Debt” “Disability” “Disposal”
“Distribution” “Educational status” “Educators” “Emigration” “Employees”
“Employers” “Employment” “Endangered species” “Enrollment” “Equipment”
“Ethnic stock” “Expenditure” “Facilities” “Family relationships” “Farms” “Fees,
royalties” “Fertility” “Fertilizers” “Finances” “Fish kills” “Flow of funds
accounts” “Foreign trade” “Forest system, national” “Franchising” “Freight”
“Funding” “Government” “Grants, credits” “Gross national product” “Health”
“Highways” “Holdings” “Housing unit” “Illiteracy” “Immigration” “Income”
“Index” “Industry” “Institutions” “Insurance” “Investments” “Investors”
“Irrigation” “Jurisdiction” “Labor force” “Landing strip” “Law enforcement”
“Licenses” “Life expectancy” “Livestock” “Local expert” “Location” “Machinery”
“Marital status” “Migration” “Monument” “Morbidity” “Mortgages” “Mutual
funds” “National defense” “Nativity” “Naturalization” “New, failure” “Nutrition”
“Occupation” “Offenses” “Officers” “Officials” “Organizations” “Organizations,
labor” “Ownership” “Parks and forests” “Parole” “Passengers” “Patents,
trademarks” “Payroll” “Personal wealth” “Pesticides” “Pipelines” “Place of work”
“Postal service” “Poverty status” “Precipitation” “Price” “Prisons” “Processing”
“Producers” “Production” “Products” “Programs” “Protected areas” “Proximity”
“Quality” “Quantity” “Race” “Rail” “Recipients” “Recovery” “Registrations”
“Regulations” “Remote access” “Research and development” “Reserves”
“Residence” “Resources” “Revenue” “Roads” “Rural” “Sales” “Savings”
“Securities” “Sediments” “Sex composition” “Shellfish” “Shipbuilding”
“Shipments” “Site” “Size” “Space program” “Students” “Subscriptions”
“Subsistence” “Sunshine” “Supply, demand” “Tax” “Temperature” “Tourism”
“Trade origin, destination” “Traffic hubs” “Transactions” “Travel” “Travelers”
“Type” “Unemployment” “Urban” “Usage, utilization” “Value” “Value added”
“Vehicles” “Vessels” “Veterans” “Veterans affairs” “Victims” “Visitation”
“Volume” “Vote cast” “Voters” “Wages, earnings” “Waste” “Wildfires” “Wind
speed” “Years completed”

- a) **Minor theme** (*optional, repeatable*). A keyword or phrase describing a specific minor component or sub-topic of either the **Major theme** or previous-level **Minor theme(s)**.
(See **Theme:Major theme:Minor theme** item for details).

Any other term may be used in **Major theme** or **Minor theme**, however if that term does not appear in the above list, it must be defined using **Definition classification**. See **Appendix D** for definitions of these values.

Appendix D. Definitions of the Domains of Selected Metadata Entities**D1. Domain: Major Theme Description**

The following descriptions elaborate on each item in the domain, suggesting but in no way limiting the types of data within each **Major Theme** Description.

Agriculture -

Farms and farm population. Includes data on land use and irrigation; farm cooperatives; farm income, expenditures and debt; farm output, productivity and marketing; foreign trade in agricultural products; specific crops; and livestock, poultry, and their products.

Archeology -

Information from or pertaining to studies of past human life as revealed by relics left by ancient people.

Architecture, Art -

The humanities. In particular, includes data relating to the planning and building of structures (including methods or styles of buildings) and to objects produced through the use of skill and imagination as well as to those people responsible for creating the buildings and objects. Includes those areas the humanities which are generally considered the "fine arts."

Archives, Museum Collections -

Information from or pertaining to records or objects which generally have a common theme and have been gathered together from numerous sources.

Attitudes, Outlook -

Relating to the thoughts, mental positions, opinions and ways in which individuals, groups of persons, or societies view past, present, and future events. This information is often obtained from surveys and opinion polls.

Banking, Finance, Insurance -

A nation's finances, various types of financial institutions, money and credit, securities, and insurance.

Business Enterprise -

The place and behavior of business firms and business initiatives. Includes data on the number, type, and size of businesses; the financial data of corporations (domestic and international); business investments, expenditures, and profits; business sales and inventories; consumer cooperatives; and business failures.

Communications -

The usage, finances, and operations of postal services and of the various communications media, including telephone, telegraph, radio, television, newspapers and periodicals, and books. Includes data on advertising expenditures.

Construction -

The construction industry and the various indicators of its activity and costs.

Customs -

Traditional courses of action generally historically associated with particular societies or cultures.

Diet, Traditional Foods -

Data on the current or past (traditional) eating habits of individuals, groups of individuals, societies, or cultures.

Domestic Trade, Service -

Distributive trades and services industries (of which franchised businesses are a part). Includes data on cover sales or receipts, establishments, employees, and payrolls.

Economics -

Economic data on the general economic situation and outlooks. Includes resource economics and resource valuation methodologies.

Education -

Formal education, at various levels, for public and private schools. Includes data on school-age population and school enrollment, educational attainment, education personnel, the financial aspects of education, libraries, adult education, and vocational training.

Energy -

Fuel resources, electric energy, hydropower, nuclear power, and alternate energy sources. Includes data on the exploration and reserves of energy fuels; production, supply, and distribution of energy; use and consumption of energy; and financial information.

Environment -

The environmental aspects of geographic areas, including data on environmental trends, quality, hazards, pollution, threatened and endangered species/wildlife, and pollution and hazards abatement and control expenditures, and contingency planning.

Family Life -

Information or data related to activities engaged in by "families." Families may be either a "nuclear" family (persons related by birth, marriage, or adoption) or an "extended" family (persons often not related by blood but considered part of the family).

Finances, Employment (Federal) -

The financial structure and the civilian employment of the Federal government.

Finances, Employment (State and Local) -

The revenues, expenditures, debt, gross assessed value and employment of State and Local governments.

Finances, Employment (Corporate) -

The finances and employment of corporate or private entities.

Fisheries -

Recreational or commercial fishing and the fish processing industry.

Foreign Trade, Commerce -

The flow of goods, services, and capital between countries. Includes data on changes in the official reserve assets of a country, its international investments, foreign assistance programs, and import duties.

Forests, Forest Products -

The area, ownership, and timber resources of commercial timberland. Includes forestry statistics covering the National Forests and the U.S. Forest Service cooperative programs; product data for lumber, pulpwood, woodpulp, paper and paperboard, and all other forestry products.

Geography -

The physical environment of geographic areas. Includes data on land use and ownership, renewable and renewable resources (and their consumption).

Government, Politics -

All elections, including presidential, congressional and gubernatorial elections. Congressional legislation, population of voting age, voter participation, and campaign finances.

Health and Nutrition -

Health expenditures and insurance coverage; medical personnel; hospitals and other care facilities; and the incidence of acute and chronic conditions. The nutritional intake of the population, per capita food consumption, and government food programs.

Heritage -

Data concerning legacies, or something that has come from ancestors or predecessors of the past.

History -

Records and explanations of past events.

Housing -

Housing units and their characteristics, occupants, and financing; public housing; and non residential buildings.

Income, Expenditures, Wealth -

The gross national product, national and personal income, saving and investment, money income, poverty, and national and personal wealth.

Industrialization -

Information on or pertaining to branches of economic activity, including businesses or manufacturers, which produce goods or services.

Labor Force, Employment, Earnings -

Distribution of the labor force by occupation and industry affiliation, and the supply of, demands for, and conditions of labor.

Language -

Data associated with human speech and the study of human speech, including the units nature, structure, and development of languages or a language.

Law Enforcement, Courts, Prisons -

The criminal justice system. Includes data on law enforcement, crimes, judicial, correctional, and prevention programs.

Manufacturers -

Manufacturing as a whole and the more detailed data for major industry groups and their products.

Migration, Naturalization -

Immigration, naturalization, alien registration, and emigration.

Mining, Mining Products -

Mineral industries and their products, general summary measures of production and employment, and more detailed data on production, prices, imports and exports, consumption, and distribution for specific industries and products.

National Defense, Veterans Affairs -

The national defense and its human and financial costs; active and reserve military personnel; ships, equipment, and aircraft; and programs and benefits for veterans.

Nationality -

Related to the "national" character of either a particular nation or an ethnic group with a larger unit (as a nation) as well as to the legal relationships involving allegiance of an individual and his protection by the state.

Navigation -

Information pertaining to the movement of crafts or vessels from one place to another. Includes the waterways used by all forms of marine traffic, commercial shipping lanes, canals, locks, aids to navigation, and traffic densities.

Parks, Recreation -

Parks, forests and other recreation areas, recreational activities, the arts and humanities, and domestic and foreign travel.

Population -

The growth, distribution, and characteristics of a population.

Prices -

Producer and consumer price indexes, actual prices for selected commodities, and budgets for consumers.

Religion -

Data on personal or institutionalized systems of religious attitudes, beliefs, and practices.

Science, Technology -

Scientific, engineering, and technologic resources. Includes data on research and development, funding, and personnel and education, and space program outlays and accomplishments.

Settlement -

Data on the movement and location of populations. Includes information on places and regions and the ways in which people chose to live as individuals or groups.

Skills -

Data on occupations, crafts, or trades requiring training and proficiency or technical expertness and the people having that expertise.

Social Insurance, Human Services -

Expenditures on social welfare, programs for old-age, survivors, disability, and health insurance; retirement programs; government unemployment and temporary disability insurance; Federal supplemental security income payments and aid to the needy; and child and other welfare service. Also data on workers' compensation, vocational rehabilitation, the American Red Cross, United Way Campaign, and philanthropic trusts and foundations.

Transportation -

Air, water, and land transportation. Includes passenger and cargo data, costs and revenues, safety/accidents, commuting, and the characteristics of all forms of public transportation.

Vital Statistics -

Births, deaths, abortions, fertility, life expectancy, marriages, and divorces.

War -

Information on preparation for, the act of, and the consequences resulting from hostilities between groups of persons. Includes the historical, social, and military aspects of conflicts and data on the persons engaged in the conflicts and the science of warfare, including the weapons used.

D2. Domain: Minor Theme Description

A brief definition of each item (the demographic application of an item was selected when several definitions were available). These domain items are offered only as suggestions; the list of items is not meant to be either exclusive or comprehensive. If the provided items do not meet the metadata producer's needs, alternative items may be found in the common set of terminology and definitions for spatial features provided in FIPS PUB 173, Spatial Data Transfer Standard (SDTS), or any subject matter thesaurus.

Abortions -

Termination of pregnancy before the fetus is capable of survival as an individual; the termination may be spontaneous or induced.

Accidents, Safety -

Accidents are chance happenings which result in injury or death. Safety is the preventive measure taken for accident avoidance.

Acquisitions -

Acquired objects, often used in reference to land.

Acreage -

Measurement of land area in acres or hectares.

Activities -

Participation in an action, often organized. Organized activities include the labor force, armed forces, and higher education.

Age Composition -

The composition of a population according to the number of persons within single year ages or age categories. Age is often reported in terms of Age-Sex Composition.

Aid -

The transfer of money or its equivalent from wealthy to poorer. Aid is often used as a summary term in the context of international relations.

Air, Land, Water -

Existing in or part of the atmosphere, the Earth's dry surface, or a body of water (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Aircraft -

A device that is used or intended to be used for flight in the air; aircraft may be heavier or lighter than air, powered or unpowered, and have fixed or rotary wings.

Amount -

A total or aggregate.

Alien Registration -

The legal requirement for a foreign-born resident who has not been naturalized and is still a subject or citizen of a foreign country to register with the country in which he or she is currently living.

Armed Forces -

The military forces of a country. The armed forces of the United States include persons on active duty with the U.S. Army, Air Force, Navy, Marine Corps, or Coast Guard.

Arrests, Sentences, Appeals -

Within the context of the Criminal Justice System: Arrests are the seizing and holding of persons under the authority of the law. Sentences are the judicial opinion of what punishment is to be inflicted on a convicted person or the penalty meted out/Appeals are the requests for new hearings.

Assets, Liabilities -

Assets are anything owned that has commercial or exchange value. Liabilities are financial obligations or claims against assets.

Attendance -

Persons or number of persons attending.

Benefits -

Gains or advantages received by an individual; benefits generally fulfill a need.

Biota -

The animal and plant life of a region (a broad, homogeneous geographical area).

Births, Deaths [Mortality] -

Births mark the beginning of life. Deaths mark the termination of life. [Mortality is the process by which deaths occur in a population.]

Boundary -

A nonphysical line indicating the limit or extent of an area or territory (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Building -

A permanent walled and roofed construction (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Business [Corporations]-

Private (nongovernment) firms involved in the buying and selling of goods or services.

Campaigns -

Operations undertaken to attain some political, social, or commercial goal.

Capacity -

A means of measuring a volume of space or the amount which can be handled.

Capital Punishment -

Use of the death penalty as a means of criminal justice.

Cargo -

Any freight, that is, merchandise transported by a carrier; the term cargo is usually but not always confined to freight carried by ships or other vessels.

Care -

Attentiveness. The term care is often used in conjunction with health and refers to the attention paid toward maintaining a person's physical well-being.

Catch -

The total quantity caught at one time or over time; the term catch is often used as a measure in the fisheries industry.

Cause -

Something that produces an effect; cause can be applied to any event, circumstance, or condition or any combination of these that brings about or helps bring about a result.

Cemetery -

A place for burying the dead (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Climate -

Weather typical of a region as indicated by the mean temperature, rainfall, wind velocity, and other measurements.

Commercial -

Used or exploited for financial gain (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Commuters, Drivers -

Commuters are persons who travel regularly from one place to another, usually to and from their place of work by one of several modes of transportation (private or public). Drivers are persons who operate their own vehicles either for profit or for personal use.

Composition -

The specified mixture or combination of one or more elements or ingredients (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes). Demographers perceive composition to be the distribution of members of a population according to its characteristics.

Conditions -

Existing circumstances.

Consumers -

Persons who use or consume goods and services with the use being personal rather than for manufacturing, processing, or resale.

Consumption -

The purchase and use of goods and services to satisfy needs and desires, either by an individual or an institution.

Contaminants -

Something that contaminates or has the potential to make impure through contact or mixture.

Courts -

One element of the Criminal Justice System; a judge or group of judges with the authority to hear civil and criminal cases and to render judgements.

Crime -

Violations of the criminal law for which an individual can be punished through the use of formal sanctions applied by governmental authorities.

Criminal Justice System -

The administration of justice including all stages of the process (arrests, court appearances, trials, correctional institutions and community-based treatment) and the agencies involved (police, courts, prisons and community-based programs for released offenders).

Criminals -

Persons committing acts, which have been determined by a governmental authority to be crimes.

Crops -

Cultivated plants or agricultural produce.

Currency -

The money issued by the monetary authorities that forms part of an economy's money supply.

Debt -

Money, services, or materials owed to one person, company, government, etc. as a result of a previous agreement.

Disability -

A handicap which is generally limiting (as in a work disability where the handicap restricts a person's choice of jobs). When used in reference to insurance, disability is a benefit provided for the loss of time in the event of sickness or an accident. The benefit may come from several sources such as workers' compensation, companies or unions, or Federal, State, or local government.

Disposal -

The means of getting rid of or throwing out.

Distribution -

Either the division of something among several persons or entities or the separation of merchandise into different categories and group levels.

Educational Status -

The legal status of an individual in relation to education; that is, a person's current enrollment as a student.

Educators -

Persons trained in either teaching or the theory and practice of education.

Emigration -

The movement out of a sovereign state.

Employees [Personnel] -

A general term for all those who work for a wage or salary and perform services for an employer.

Employers -

A person, organization, or corporation that has workers in its employ.

Employment -

Occupational activity usually, but not necessarily, for pay.

Endangered Species -

A species which is threatened or faced with the danger of extinction.

Enrollment -

The number entered on a register. When applied to education, enrollment specifically is the total number of students registered in and attending a given school at a given time (generally the fall of the year).

Equipment -

Goods that are used to produce other goods or services but are not themselves used up in the process. The term equipment is mainly for such capital assets as machinery, tools, furnaces, appliances, and vehicles of various kinds.

Ethnic Stock -

Social groups distinguished by various traits, including language, national or geographic origin, customs, religion, and race.

Expenditure -

An actual payment or the creation of an obligation to make a future payment for some benefit, item, or service rendered.

Facilities -

All buildings, equipment, structures, and other stationary items located or brought to a single site or upon contiguous sites, adjacent sites, or sites linked by common ownership or a common environmental concern.

Family Relationships -

The basic network of family ties between either the members of a "nuclear" family (persons related by birth, marriage, or adoption) or the relationships existing between more "extended" family members (persons often not related by blood but considered part of the family).

Farms -

A tract of crops or grazing land, as well as the group of buildings with and often surrounding a farmhouse, including barns, sheds, and other outbuildings, used for agricultural production (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes). The Bureau of the Census specifies a farm's agricultural operations must be under the day-to-day control of an individual management and that \$1,000 or more in agricultural products must be, or could be, sold in a year.

Fees, Royalties -

Fees are either a fixed charge or a charge for a professional service. Royalties are shares of a profit or product reserved by the grantor of a license or lease or shares of proceeds of sale paid to an author, composer, inventor, or proprietor for the right to use or sell their invention or service.

Fertility -

The physiological capacity to produce live, born children. Fertility is used by demographers to denote the production of live births by individuals, couples, or populations.

Fertilizers -

Natural or artificial materials added to the soil to increase the fertility.

Finances -

The monetary status of a person, company, organization, etc., specifically their assets and liabilities.

Fish Kills -

Acts or events resulting in occurrences of a significant number of dead fish.

Flow of Funds Accounts -

[To be defined by the Federal Geographic Data Committee.]

Foreign Trade -

The exchange of goods and services between countries.

Forest System, National -

The forests which are protected by the Government and may be harvested only under controlled conditions.

Franchising -

The assignment by one company to another or others of the right to supply its products. A franchise is a contractual arrangement entered into for a specified period of time.

Freight -

All merchandise, goods, products, or commodities shipped by rail, air, road, or water other than baggage, express mail, or regular mail.

Funding -

The gathering together of cash, securities, or other assets for a particular purpose or program.

Government -

Governing is the authoritative exercise of power and government is the formal system of offices endowed with the authority to make binding decisions for a society or institution. More specifically, government is the persons who fill the positions of authority in a society or institution.

Grants, Credit -

Grants are revenues, generally from a Government agency, for a specific research project or program having a specified purpose. Credit is a sales or purchase accompanied by a promise of later payment.

Gross National Product (GNP) -

Total flow in money terms of all goods and services that an economy produces during a specified period.

Health -

A state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity.

Highways -

A class of road based on design, weatherability, governmental designation and the Department of Transportation functional classification system. Highways are hard-surfaced roads with controlled access which can be either Interstate and U.S. numbered or primary State routes.

Holdings -

Land held or property (for example, stocks or bonds) owned.

Housing Unit -

A dwelling occupied as a living quarter; it may be a house, an apartment, a mobile home or trailer, a group of rooms or a single room occupied as a separate living quarter. Housing units may be vacant or occupied.

Illiteracy -

The quality or condition of being unable to read and write.

Immigration -

Movement into a sovereign state.

Income -

An amount of money or its equivalent accruing over a specified period to an individual, family or household, firm, or a nation.

Index -

A statistical yardstick expressed in terms of percentages of a base year or years.

Industry -

A generic term designating any branch of economic activity. Industries are usually grouped according to the kinds of goods or services produced.

Institutions -

Organized social groups or organizations.

Insurance -

A method whereby those concerned about some form of a hazard contributes to a common fund out of which losses to the contributors are paid.

Investments -

The use of money for the purpose of making more money, to gain income or to increase capital, or both.

Investors -

Individuals or organizations whose principal concerns in the purchase of a security are regular dividend income, safety of the original investment, and, if possible, capital appreciation.

Irrigation -

Used for the supplying of water by artificial means to land for agriculture purposes (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Jail -

A confinement facility usually administered by city, county, or other local government that received individuals pending arraignment, trial, conviction, and sentencing and houses sentenced persons usually for terms of less than one year. Jails readmit probation, parole, and bailbond violators and absconders; temporarily detain juveniles pending transfer to juvenile authorities; hold mentally ill persons pending their movement to appropriate health facilities;

hold individuals for the military, for protective custody, for contempt, and the courts as witnesses; release convicted inmates to the community upon completion of their sentences; and transfer inmates to Federal, State, or other local correctional authorities.

Jurisdiction -

Jurisdiction refers to the authority of a sovereign power to govern or legislate or to the actual limits or territory within which this authority may be exercised. Within the context of the Criminal Justice System, jurisdiction may also refer to the power, right, or authority to interpret and apply the law or to the actual limits or territory within which this authority may be exercised.

Labor Force -

An economically active population. The term Labor Force may be restricted to those who participate, or try to participate in the production of a nation's economic goods and services.

Landing Strip -

An unimproved area of land or water that is used or intended to be used for the landing and takeoff of aircraft. The term landing strip includes the area's buildings and facilities, if any.

Law Enforcement -

Part of the Criminal Justice System. Law enforcement refers to both the agencies and procedures of enforcing the law.

Licenses -

Grants by a government to an authority of the right to supply a good or provide a service or by one company to another to manufacture its products or use its technology or facilities. Licenses may also refer to the documents which show proof of legal ownership or entitlement or compliance with a requirement for payment thereon or that is issued as a means of ensuring premises or persons are fit for the purpose in which they are engaged.

Life Expectancy -

The average number of years of life remaining at any specified age.

Livestock -

Domestic animals, raised for home use or for profit.

Local Expert -

A person knowledgeable in the relevant field of study or study area.

Location -

The place, site or space occupied by a specified feature (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Machinery -

Machines in general or as a functioning unit.

Marital Status -

The legal status of an individual in relation to marriage; that is, the place of an individual in the succession single-married-widowed (or divorced).

Migration [Mobility] -

Relatively permanent movement of persons over a significant distance. A distinction is usually made between movements within the boundaries of one country and those between two countries.

Monument -

A structure erected as a memorial (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Morbidity -

The opposite of health; sickness in the broadest sense.

Mortgages -

Written conveyances of title to a property, but not possession, to obtain the payment of a debt or the performance of some obligation under the condition that the conveyance is to be void upon final payment.

Mutual Funds -

Investment companies that issue shares on a continuing basis and stand ready to purchase shares from shareholders on demand.

National Defense -

The means whereby a nation protects itself against its enemies.

Nativity -

Birth, especially the place, conditions, or circumstances of being born.

Naturalization -

To grant full citizenship to a person of foreign birth.

New, Failure -

New is being recently obtained, acquired, or established. Failure is becoming bankrupt or insolvent.

Number -

A count; the numeric measure of units or individuals.

Nutrition -

The composite of all the processes by which a living being absorbs food.

Occupation -

The work a person does; usually grouped by some system of classification.

Offenses -

Crimes or transgressions of the law.

Officers -

Individuals holding an office of authority or trust in an organization. Officers are also individuals performing police duties (within the context of the Criminal Justice System).

Officials -

Persons who hold an office or position, particularly those who act in a subordinate capacity for an institution such as a corporation or governmental agency.

Organizations -

A structured system of roles and functional relationships designed to carry out policies, or the programs inspired by the policies.

Organizations, Labor -

As defined by Section 2(5) of the National Labor Relations Act "means any organization of any kind, or any agency or employee representation committee or plan, in which employees participate and which exists for the purpose, in whole or part, of dealing with employees concerning grievances, labor disputes, wages, rates of pay, hours of employment, or conditions of work."

Ownership -

Possession of a legal title with the rights to enjoy the benefits derived from any assets accompanying or accruing from such title.

Parks, Forests -

Places or areas developed for public use or recreation. A park is further described and a place or area set aside for recreation or preservation of a cultural or natural resource (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Parole -

The release of a prisoner before his or her term has expired on condition of continued good behavior or the duration of such continued release.

Passengers -

Persons who travel on a train, aircraft, ship, bus, or other conveyance without participating in its operation; such persons often pay for the right to travel.

Patents, Trademarks -

Patents are rights assigned for seventeen years to the inventor of a device or process to exclude others from making or selling the invention without permission. Trademarks are distinctive identifications of a manufactured product or of a service in the form of a name, logo, motto, and so on.

Payroll [Salary] -

The wages earned by a firm's employees for a certain period of time. Various deductions from payrolls are identified on the employee's records.

Personal Wealth -

The totality of a person's savings.

Pesticides -

Substances that are applied to control animal or plant pests. This term refers to all use classes including herbicides, insecticides, fungicides, nematocides, algicides, piscicides, wood preservatives, and fumigants.

Pipelines -

Closed conduits with pumps, valves and control devices for conveying fluids, gases, or finely divided solids.

Place of Work -

The geographic location at which workers carry out their occupational activities during a specified time period.

Postal Service -

The public department responsible for the transportation and delivery of the mails; also, referred to as the Post Office.

Poverty Status -

A measure of poverty based on poverty levels originally defined by the Social Security Administration in 1964 and prescribed by the Office of Management and Budget in the Poverty Level Directive 14 as the standard to be used by Federal agencies. The poverty status indicates whether the income of a family falls above or below the poverty level.

Precipitation -

Water droplets or ice particles condensed from atmospheric water vapor and sufficiently massive to fall to the earth's surface, such as rain or snow.

Price -

The amount of money a seller receives for goods and services.

Prisons -

A State or Federally administered correctional facility (including privately operated under State or Federal authority) housing adults with sentences usually of one year or more. Prisons include prison farms; reception, diagnostic, and classification centers; road camps; forestry and conservation camps; youthful offender facilities; prison hospitals; and drug and alcohol treatment facilities.

Processing -

A kind of manufacturing in which one material is converted into another.

Producers -

Individuals who manufacturer goods and services.

Production -

Any form of activity that adds value to goods and services including creation, transportation, and warehousing until used.

Products [Services] -

Goods and services made available to consumers.

Programs -

An organized list of procedures, often set out as a means of meeting a policy or solving a problem.

Protected Areas -

Areas for which there are restrictions on use and/or development to protect/preserve natural resources.

Proximity -

The nearness or closeness of a person or object to another.

Quality -

A measure of the degree or grade of excellence.

Quantity -

An exact number or amount.

Race -

Race, in its pure, biological sense, is a subpopulation that differs significantly from others in the frequency of one or more genes. Race has come to culturally mean color, language, religion, or nationality and is rarely identified scientifically and is usually identified by physical characteristics and presumed genealogy. The Bureau of the Census collects race information reflecting self-identification.

Rail -

Set of parallel rails on which a train generally used for the transportation of freight or the extended transport of passengers runs.

Rate -

A relation between actual and potential of the cost of a unit.

Recipients -

Persons, institutions, or governments who receive.

Recovery -

The period of a business cycle that follows a depression.

Registrations -

Continuous recordings of events for legal or statistical purposes or the documents issued upon those recordings.

Regulations -

Principles, rules, or laws designed to control or govern behavior.

Remote Access -

Means of entering remotely located areas.

Research, Development -

Any scientific investigation leading to discovery of new techniques and products and their commercial application, together with refinement and improvement of existing technologies and products.

Reserves -

Any additional claims of company shareholders that reflect increases in the value of company assets in the balance sheet.

Residence -

The place in which a person resides or the official home or location of a corporation.

Resources -

Anything a nation uses to produce goods and services. Resources are usually categorized as natural resources, labor, capital, or enterprise.

Revenue -

The grand total of all resources received from the sale of a firm's product or service during a stated period.

Roads -

Open ways for the passage of vehicular traffic which are passable in most seasonal conditions and trafficked by mainstream motor vehicles (all types). Several recognized road classification systems exist.

Rural -

Rural is a complement of urban. Rural connotes a predominance of agriculture and a low population density with a social life in small communities.

Sales -

The transfer of title to an item or items or the agreement to perform a service in return for cash or the expectation of a cash payment. Sales may also refer to the revenue received from the sale of goods.

Savings -

The amount of existing income that is not spent on consumption.

Securities -

Documents that identify legal ownership of a physical commodity or legal claims to another's wealth.

Sediments -

Material suspended in water or in air or the material deposited onto the surface underlying this water or air.

Sex Composition -

The composition of a population according to the number or proportion of males to females within single year ages or age categories. Sex is often reported in terms of Age-Sex Composition.

Shellfish -

An aquatic animal having a shell or shell-like exoskeleton.

Shipbuilding -

The business of constructing ships.

Shipments -

The act of shipping goods or the goods and cargo shipped.

Site -

A contiguous property unit, land, plot of ground, or area.

Size -

The physical dimensions, proportions, magnitude, or extent of something (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Space Program -

The name given a country's efforts of space exploration. The United States Space Program refers to the expenditures, agencies, efforts, and goals of space exploration.

Students -

Persons attending a school, college, or university.

Subscriptions -

A purchase made by signed order, as for a periodical for a specified period of time or for a series of performances.

Subsistence -

The level of living just sufficient to maintain a productive population. Subsistence is implicitly based on nutritional standards, but may be expanded to include the notion of "not-for-profit."

Sunshine -

The radiant energy, especially heat and visible light, emitted by the sun. Sunshine is often expressed in terms of number of days or hours.

Supply, Demand -

The amount of a commodity potential purchasers would like to buy.

Tax -

A charge levied by a government against the income of an individual, corporation, or other profit-making center of activity.

Temperature -

A specific degree of hotness or coldness as indicated on or referred to a standard scale (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Tourism -

The practice of traveling for pleasure and the business of providing tours and services for tourists.

Trade Origin, Destination -

Trade origin is the location from which a product was shipped. Trade Destination is the final location to which the product is shipped.

Traffic Hubs -

Centers of travel. Companies often maintain these central locations through which the majority of their routes pass as a means of cost containment.

Transactions -

Agreements between two or more parties that establish a legal obligation.

Travel -

To go from one place to another; to journey.

Travelers -

Persons who travel.

Type -

To classify according to a shared trait or characteristic which distinguishes an identifiable group, kind, or category.

Unemployment -

A condition of not being employed. According to the U.S. Government, unemployment is the condition of a member of the labor force who seeks work but does not find it.

Urban -

Associated with large aggregates of people or with their typical institutions and way of life. In official statistics, urban delimits places classified principally by population size and/or civil status.

Usage, Utilization -

Usage is the action or amount of using. Utilization is the state of being used.

Value -

The worth of property, goods, services, and so on. Value is generally considered to be the purchasing power, though may refer to a nonmarket value which, although hard to measure, is an important component of resource valuation.

Value Added -

The part of the value of produced goods that is developed in a company. Value added is determined by subtracting from sales the costs of materials and supplies, energy, contract work, and so on.

Vehicles -

A device for carrying passengers, goods, or equipment. A common conveyance or vehicle is the automobile.

Vessels -

A craft designed to navigate on the water, also an airship.

Veterans -

Persons who have been members of the armed forces of their country whom are not currently serving on active duty. In the United States, veterans include former U.S. Army, Air Force, Navy, Marine Corps, and Coast Guard personnel, or persons who served as Merchant Marine seamen during WWII. Veterans also include persons serving in the National Guard or Military Reserves only if they were ever called or ordered to active duty.

Veterans Affairs -

All procedures, actions, and occasions pertaining to veterans.

Victims -

Persons whom crimes are committed against or who are injured or die under any of various conditions.

Visitation -

Official visits, often for the purposes of inspection, or the legal right to visit.

Volume -

Space occupied or cubic capacity as measured in cubic units (FIPS PUB 173, Spatial Data Transfer Standard, Part 2, Annexes).

Vote Cast -

The number of votes which are cast.

Voters -

Persons making a formal expression of preference for a candidate for office or for a proposed resolution of an issue; those persons who actively participate in voting and cast a vote.

Wages, Earnings -

Compensation of employees who receive a stated sum per piece, hour, day, or any other unit or period; generally refers to all compensation paid including salaries.

Waste -

The damaged, defective, or superfluous materials produced by a manufacturing process, or refuse from places of human or animal habitation.

Wildfires -

A fire that travels and spreads rapidly.

Wind Speed -

A horizontal measure of the rate at which air is in motion, most commonly measured in knots (nautical miles per hour) using an anemometer.

Years Completed -

The number of years invested by a person in an activity; when applied to educational attainment refers to the highest grade of regular school attended and completed.

D3. Domain: Geographic Entity Type

A brief definition of each item .

Hydrographic -

Geographic areas whose boundaries are based on the properties, distribution, and circulation of water or are delineated by a Federal agency for hydrographic data collection, analysis, or presentation. These areas can be either points (specific sites such as water quality collection sites), lines (navigation routes), or areas.

Legal, Administrative -

Geographic areas whose boundaries result from charters, laws, treaties, or other administrative or governmental action. These geographic areas are either governed by elected officials or are without elected officials but created to administer elections and other governmental functions. These areas can be either points (boundary monuments), lines (trails), or areas.

Statistical, Thematic -

Geographic areas whose boundaries are not legally defined and which have no governmental standing. These geographic areas are delineated by Federal agencies (most notably the Office of Management and Budget and the Census Bureau) for either data tabulation purposes or to delineate thematically similar units. These areas can be either points (specific sites such as weather collection sites), lines (trails), or areas.

D4. Domain: Geographic Unit

A brief definition of each item.

If Geographic Entity Type is "Hydrographic:"

Coastal Drainage Area (CDA) -

That component of an entire watershed that meets the following criteria: 1) it is not part of any Estuarine Drainage Area (EDA) or Fluvial Drainage Area (FDA); 2) it drains directly into an ocean, estuary, or the Great Lakes; and 3) it is composed of only the downstream-most hydrologic cataloging unit (HUC) in which the head-of-tide is found.

Coastal, Estuarine Waters -

A unique geographic component of wetlands, this area consists of deepwater tidal habitats, tidal wetlands, and drainage channels adjacent to the sea in which the tide ebbs and flows.

Drainage Basin-

A geographic unit (valley) whose area contributes water to and is drained by a drainage system (one stream and its tributaries). Drainage basins are separated by divides. A drainage basin may also be referred to as a watershed; if so, technically the drainage rim is considered to be part of the watershed.

Estuarine Drainage Area (EDA) -

That component of an estuary's entire watershed that empties directly into the estuary and is affected by tides. EDAs may be composed of a portion of a single hydrologic cataloging unit (HUC), an entire HUC, more than one HUC, or several HUCs and portions of several adjacent HUCs.

Fluvial Coastal Drainage Areas (FCDAs) -

Those components of coastal watersheds that lie "upstream" Coastal Drainage Area (CDA) boundaries.

Fluvial Drainage Area (FDA) -

The component's of an estuary's entire watershed "upstream" of the Estuarine Drainage Area (EDA).

National Water Quality Assessment (NAWQA) Program Study Site -

Geographic units identified by the U.S. Geological Survey (USGS) for the purpose of monitoring water quality. The USGS divides the United States into 60 sites delineated on the basis of river basins, ground water, or political boundaries.

Port -

An area where ships may ride secure. The area often has facilities for the taking on and discharging of cargo from ships; the facilities may include a place where foreign goods are cleared through a customhouse and aliens are permitted to enter a country.

Waterbody -

A geographic area delineated by the extent of its surface water; no particular size criteria exists.

Wetlands -

A geographic unit containing lands transitional between terrestrial and aquatic systems where the saturation with water is the dominate factor in determining the nature of soil development and the types of plant and animal communities living in the soil and in its surface. Wetlands include a wide variety of areas; several classification schemes are used by Federal agencies/organizations.

If Geographic Entity Type is "Legal, Administrative:"

Nation -

An area under the jurisdiction of a sovereign government, excluding the United States.

United States -

The 50 states and the District of Columbia.

The following "**Legal, Administrative**" geographic entities or areas are recognized within the United States. A majority of the **Geographic Unit** definitions are adapted from definitions included by the Census Bureau in its standard decennial census products.

American Entity trust land(s) -

American Indian Reservation (AIR) associated on- or off-reservation lands held in trust for either the tribe (tribal trust lands) or an individual tribe member (individual trust lands) by the Federal government.

American Indian reservation (AIR) -

A legal entity over which American Indians have governmental jurisdiction. The boundaries of AIRs were established by treaty, statute, and/or executive or court order and are identified by either the Bureau of Indian Affairs (BIA) for Federal reservations or the individual States for State reservations.

American Indian reservation with trust land(s) -

An AIR and its associated on- or off-reservation lands held in trust for either the tribe (tribal trust lands) or an individual tribe member (individual trust lands) by the Federal Government.

Alaska Native village (ANV) -

A local Alaskan governmental unit consisting of a tribe, band, clan, village, community, and association recognized pursuant to the Alaska Native Claims Settlement Act (ANCSA) with boundaries identified by an appropriate authority.

Alaska Native Regional Corporation (ANRC) -

A corporate entity established by the ANCSA. Alaska is divided into 12 ANRCs that cover the entire State, except for the Annette Islands Reserve. Boundaries are established by the Department of the Interior in cooperation with Alaska Natives. (A 13th ANRC without associated land area was established for Alaska Natives who are not permanent residents of Alaska.)

Congressional District (CD) -

An area established by State officials or the courts for the purpose of electing a person to the U.S. House of Representatives. Within each State, these areas must contain, as nearly as possible, an equal number of inhabitants.

Consolidated City -

An incorporated place that has combined its governmental functions with a county or subcounty entity but contains one or more other incorporated places that continue to function as local governments within the consolidated government.

County -

Primary legal subdivision of every State except Alaska (referred to as parishes in Louisiana.) The legal geographic unit county is expanded in this documentation to include the following statistically equivalent areas to provide "complete" national coverage: boroughs and census areas in Alaska which has no counties, cities that are independent (independent cities) of any county organization located in Maryland, Missouri, Nevada, and Virginia, that part of Yellowstone National Park in Montana, and the District of Columbia.

Incorporated Place -

A governmental unit, incorporated under State law as a city, town (except in New England, New York, and Wisconsin), borough (except in Alaska and New York), or village, having legally prescribed limits, powers, and functions.

Minor Civil Division (MCD) -

Primary legal subdivision of 28 states, created to either govern or administer an area rather than a specific population. The several types of MCDs are identified as town, township, district, and include both functioning and non-functioning governmental units. The legal geographic unit MCD is expanded in this documentation to include the following statistically equivalent areas to provide "complete" national coverage: census subareas in Alaska, census county divisions (CCD) delineated by the Census Bureau for statistical purposes in 21 states, and the District of Columbia.

Neighborhood -

A subarea within a local governmental unit, usually an incorporated place or county, having legally defined boundaries.

Outlying Area -

Areas, other than States and the District of Columbia, under the jurisdiction of the United States. For the 1990 Census, these areas included American Samoa, Guam, the Northern Mariana Islands, Palau, Puerto Rico, the Virgin Islands of the United States, and several small islands in the Caribbean Sea and the Pacific Ocean. (Outlying areas are considered statistically equivalent to States.)

School District -

The territory administered by the elected or appointed authorities of a State, county, or other local governmental unit to provide educational services to a resident population.

State -

Primary legal subdivision of the United States, including the District of Columbia as a statistically equivalent area. (Outlying Areas are also statistically equivalent, however, they are noted as a unique geographic entity/area in this documentation.)

Trail -

A cleared path, beaten track, or improved surface, as through woods or wilderness, not usually trafficked by vehicles because of width, conditions, or access restrictions.

Voting District (VTD) -

Any of a variety of areas, such as election districts, precincts, legislative districts, or wards, established by States and local governments for voting purposes. Also referred to as election precincts (EP) in the 1980 census.

ZIP Code -

An area encompassing a section of a street, a collection of streets, an establishment, a structure, a group of post office boxes, and so forth which is assigned the same 5-, 7- or 9-digit code by the U.S. Postal Service.

If Geographic Entity Type is "Statistical, Thematic: "

Alaska Native Village Statistical Area (ANVSA) -

A 1990 census entity that represents the densely settled extent of an ANV as delineated for the Census Bureau by officials of the ANRC in which the ANVSA is located, or by other appropriate officials.

Block -

The smallest entity for which the Census Bureau collects and tabulates decennial census information.

Block Group (BG) -

A combination of census blocks that is a subdivision of a census tract (CT) or block numbering area (BNA). A BG consists of all blocks whose numbers begin with the same digit in a given CT or BNA.

Census County Division (CCD) -

A statistical subdivision of a county, established cooperatively by the Census Bureau and State and local government authorities, for the presentation of decennial census data in 21 states which do not have legally established MCDs or whose MCDs do not serve a legal or administrative governmental purpose, are not well known, have poorly defined boundaries, and/or have frequent boundary changes.

Census Designated Place (CDP) -

A statistical entity, defined for each decennial census according to Census Bureau guidelines, comprising a densely settled concentration of population that is not within an incorporated place, but is locally identified by a name. Known as unincorporated places for the 1940 through 1970 censuses.

Census Tract or Block Numbering Area (CT/BNA) -

A CT is a small, relatively permanent statistical subdivision of a county in a metropolitan area (MA) delineated by a local committee of census data users. CTs are designed to be relatively homogeneous units with respect to population characteristics, and usually contain between 2,500 and 8,000 inhabitants. A BNA is an area delineated by State officials or by the Census Bureau for the purpose of grouping and numbering decennial census blocks in counties or statistically equivalent entities in which census tracts have not been established. CTs and BNAs are equivalent in the Census Bureau's geographical hierarchy.

Central City -

The core incorporated city(ies) or MCDs of a MA. From the 1950 through 1980 censuses, the central city may also refer to an urbanized area (UA), usually consisting of the most populous incorporated city(ies).

Central Place -

The core incorporated place(s) or CDP(s) of a UA, usually consisting of the most populous place(s) in the UA. First used in the 1990 census to recognize a CDP as the most populous place in a UA.

Consolidated Metropolitan Statistical Area (CMSA) -

An area designated as a metropolitan statistical area (MSA) and having a population of 1,000,000 or more may become a CMSA if component parts are recognized as primary metropolitan statistical areas (PMSAs) and if local opinion favors the designation. Prior to 1990, CMSAs were called standard consolidated statistical areas (SCSAs) and standard consolidated areas (SCAs).

Division -

Any subdivision of international and/or national region(s) established for the collection/presentation of data. Divisions are intended to represent relatively homogeneous areas that are the subdivision of regions.

Enumeration District (ED) -

A small geographic area established and used by the Census Bureau as the basic unit for data collection and tabulation in decennial and special censuses prior to 1990. An ED usually consisted of the area to be assigned to an individual enumerator for canvassing. The ED was replaced by the address register area (ARA) for data collection purposes in the 1990 decennial census.

Metropolitan Statistical Area (MSA) -

A geographic entity, defined by the Federal Office of Management and Budget (OMB) as a statistical standard for Federal agencies, consisting of one or more contiguous counties (cities and towns in New England) that contain a core area (city with a population of 50,000 or more or a UA) within a large population nucleus and adjacent communities having a high degree of economic and social integration with the core. If the MSA is defined on the basis of a UA core, the total MSA population must be at least 100,000 (75,000 in New England). MSAs were referred to as Standard Metropolitan Areas (SMAs) during the 1950's and Standard Metropolitan Statistical Areas (SMSAs) from 1960 through 1980.

National Forest -

A large expanse of forest that is protected by the Federal government and that may be harvested only under controlled conditions.

National Monument -

A natural landmark or a structure or site of historic interest set aside by the Federal government and maintained for enjoyment or study by the public.

National Park -

A tract of land declared public property by the Federal government with a view to its preservation and development for purposes of recreation and culture.

Primary Metropolitan Statistical Area (PMSA) -

An area meeting the requirements of an MSA and having a population of one million or more. Two or more PMSAs may be defined within this area if statistical criteria are met and local opinion favors the designation.

Public Domain -

Lands owned directly by the Federal government.

Region: International -

Any statistical or thematic region(s) outside the United States established for the collection/analysis/presentation of data.

Region: National -

Any statistical or thematic region(s) within the United States established for the collection/analysis/presentation of data.

The World -

Any statistical or thematic region(s) consisting of both international and national coverage established for the collection/analysis/presentation of data.

Traffic Analysis Zone (TAZ) -

A special-purpose geographic entity delineated by a metropolitan planning organization (MPO) for tabulating transportation statistics from the decennial census.

Trail -

A cleared path, beaten track, or improved surface, as through woods or wilderness, not usually trafficked by vehicles because of width, conditions, or access restrictions.

Tribal Designated Statistical Area (TDSA) -

An area defined by an American Indian tribal government that does not have a land area certified by the BIA or a State government. A TDSA cannot exist in Oklahoma.

Tribal Jurisdiction Statistical Area (TJSA) -

An area defined by American Indian tribal governments to approximate their former reservation boundaries. TJSA's exist only in Oklahoma.

Urbanized Area (UA) -

An area consisting of a central place(s) and adjacent fringe that together have a minimum residential population of at least 50,000 people and generally an overall population density of at least 1,000 people per square mile of land area.

Appendix E. Metadata Presentation
The SDSM Table of Content Presentation (TOC)

SURVEY DESIGN AND STATISTICAL METADATA STANDARD
TABLE OF CONTENTS

- CHAPTER 1** **Content** (*mandatory*).
- CHAPTER 2** **Planning** (*optional*).
- CHAPTER 3** **Design** (*optional*).
- CHAPTER 4** **Data Collection** (*optional*).
- CHAPTER 5** **Data Processing** (*optional*).
- CHAPTER 6** **Data Analysis** (*mandatory*).
- CHAPTER 7** **Data Dissemination** (*optional*).
- CHAPTER 8** **Data** (*mandatory*).

1 Content (*mandatory*). The **Content** refers to the nature of the information that is the subject of the survey, i.e., what the universe is, a description of the data collected, and a description of the resulting products. May contain definitions, and data standardization and coding information.

1.1 **Universe** (*mandatory*). The total defined set of interest to one or more Surveys.

1.1.1 **Universe Name** (*mandatory*). The name of the universe.

1.1.2 **Definition of Universe** (*mandatory*). Provides information about characteristics that distinguish this population from all other populations.

1.1.3 **Universe Detail Description** (*mandatory*). The description of the universe.

1.1.3.1 **Documentation** (*mandatory*). See Annex A for documentation components.

1.2 **Population of Interest** (*mandatory*). A description of the particular component of the universe being studied.

1.3 **Variables of Interest** (*mandatory, repeatable*). The name assigned to a particular population of interest. Specific elements of data which have been gathered in the process of carrying out a survey. A Variable of Interest is distinguished from other data in that it is one of the elements of data needed to accomplish the purpose of the survey through which it was gathered (or joined to).

1.4 **Datasets and Products** (*mandatory, repeatable*). A general description of the sets of data and outputs for a particular universe.

1.4.1 **Survey Datasets** (*optional*). A dataset containing either case level data, aggregation of case level data, or statistical manipulations of either the case level or aggregated survey data, for a single Survey Instance.

1.4.1.1 **Dataset Description** (*mandatory*). See Annex B for detail.

1.4.2 **Products** (*mandatory*). Finished deliverables of a Project or a Survey Instance for outside the Bureau (external) use.

1.4.2.1 **Product Description** (*mandatory*). See Annex C for detail.

2Planning (*optional*). Documentation related to the **planning** and management of the design; the conduct of the survey and the analysis, dissemination and disposition of the data. This includes documentation related budgeting, manning, and training.

2.1 **Abstract** (*mandatory*). A brief description or extract of the applicable text of the product.

2.2 **Contact** (*mandatory*). Contact information for an individual or organization that is knowledgeable about the product.

2.3 **Variables of Interest** (*repeatable*). The name assigned to a particular population of interest. Specific elements of data which have been gathered in the process of carrying out a survey. A Variable of Interest is distinguished from other data in that it is one of the elements of data needed to accomplish the purpose of the survey through which it was gathered (or joined to).

2.4 **Survey** (*mandatory*). An investigation about characteristics of a given Universe.

2.4.1 **Survey Description** (*mandatory*). See Annex D for detail.

2.5 **System** (*mandatory, repeatable*). An identifiable process, either fully automated or computer assisted, which implements one or more Techniques to produce one or more deliverables. A System can be composed of other System.

2.5.1 **System Description** (*mandatory*). See Annex D for detail.

2.6 **Project** (*mandatory*). An identifiable effort to generate deliverables NOT specific to a single Survey instance.

2.6.1 **Project Description** (*mandatory*). See Annex D for detail.

3Design (*optional*). Information about the procedures, processes and methods used in the development of a particular emprise, product or survey.

3.1 **Frame** (*optional*). The particular set of objects from which a sample is to be drawn.

3.1.1 **Frame Concepts** (*optional*). See Annex E for detail.

3.1.2 **Frame Update Specification** (*mandatory*). The specification for determining the particular set of objects from which a sample is to be drawn. There may be multiple frames, which themselves may be selected using particular criteria.

3.2 **Sample** (*optional*). A dataset containing a subset of a Frame for a particular set of Survey Instances, selected with specific sampling Technique. For a census, the Sample incorporates the entire Frame.

3.2.1 **Sample Concepts** (*optional*). See Annex E for detail.

3.2.1 **Sample Specification** (*mandatory*). Specification for the process of selecting particular respondent from a set of potential respondents (or “frame”).

3.2.2.1 **Stages of sampling** (*mandatory, if applicable*). Steps in the development of a sample.

3.2.2.2 **Stratification used** (*mandatory, if applicable*). The particular respondent or other characteristics (stratification rule) used in determining eligibility for sampling.

3.2.2.3 **Methods of selection** (*mandatory, if applicable*). Method used in selecting the particular respondent, frame, or other unit sampled (systematic, simple random, etc.).

3.2.2.4 **Sample sizes** (*mandatory, if applicable*). The size of the sample.

3.3 **Listing Operations Specifications** (*optional*). Specifications for the process (usually accomplished on the site) of verifying addresses and the existence of a particular respondent or other object interest.

3.4 **Coverage** (*mandatory*). An examination of the procedure for determining whether more or less units were actually surveyed than originally planned.

3.4.1 **Overcoverage/Interruptions** (*mandatory*). More objects are surveyed than planned.

3.4.2 **Under coverage** (*mandatory*). Less objects are surveyed than planned.

- 3.5 **Observation models** (*mandatory*). Models of the data acquisition process. A model is a formalized expression of a theory or the casual situation which is regarded as having generated observed data. In statistical analysis, the model is generally expressed in symbols, that is to say in a mathematical form, but diagrammatic models are also found. [See DST pg. 132]. See Annex F for some of the Observation models.
- 3.6 **Estimation models** (*optional*). Rules or methods for estimating the variance of the observed population from the universe of interest. See Annex G for some of the estimation models.
- 3.7 **Information Sources and Contact Specification** (*optional*). Entities contacted and the procedures used to contact them.
- 3.7.1 **Data Collection Specification** (*optional*). The procedure used to obtain the data for the survey.
- 3.7.2 **Measurement Instrument** (*optional, repeatable*). The actual questions, forms, scripts, etc. used in collecting the survey data.
- 3.7.2.1 **Questionnaire Design** (*optional, repeatable*). The design (text, order, and conditions for skipping) of the questions used to obtain the data needed for this survey.
- 3.7.2.1.1 **Type of Questionnaire** (*optional*). (paper vs. electronic: see item 4 Data Collection Mode). Mode of delivery of the **Measurement Instrument**.
- 3.7.2.2 **Questionnaire Development and Updates** (*optional*). Related to the management of the development of the questionnaire.
- 3.7.2.2.1 **Update Specification** (*optional*). Documentation dealing with modifications made to the questions asked in the survey.
- 3.7.2.3 **Data Collection Mode** (*optional*). The actual technology used to collect the data. (e.g. mail-back paper questionnaire, CATI, CAPI, fax, etc.)
- 3.7.2.3.1 **Case Selection Specification** (*optional*). The processes used to select cases.
- 3.7.2.3.2 **Case Preparation Specification** (*optional*). The procedure(s) used in preparing cases.

3.7.2.3.3 **Quality Control Specification** (*optional*). Description of the mechanisms used to assess error in data collection. (See items 3.10 Quality Control Specification, 4.3.6 Data Collection Quality Control Procedures).

3.8 **Planned Observation Register Documentation** (*optional*). A record of the observation process. For example if there is a mail-out questionnaire, this may be a record of the mail preparation process, the mailing itself, etc.

3.9 **Data Preparation and Editing Specifications** (*optional*). Specifications for any data quality improvement tool by which erroneous or highly suspect data are found, and if necessary, corrected. Specifications for the process of obtaining and capturing data from respondents and preparing that data for further processing.

3.9.1 **Editing Specifications** (*optional*). Specifications for any data quality improvement tool by which erroneous or highly suspect data are found, and if necessary, corrected. Refer to Annex A for detailed summary. Refer to Annex H for details.

3.9.2 **Coding Specifications** (*optional*). Specifications for analyzing responses and selecting a code.

3.10 **Quality Control Specifications** (*optional*). Description or a compilation of the specifications. Refer to Annex I for detailed summary.

4 Data Collection (*optional*). Obtaining information from respondents and the conversion of that data into a form which can be processed.

4.1 **Interview Collection Period** (*optional*). The period of time during which data is obtained from respondents.

4.2 **Measurement Instrument** (*optional, repeatable*). Questionnaire in paper, electronic, or other forms.

4.3 **Procedures** (*mandatory*). Record of instructions, results, correspondence, budgets, and schedules pertaining to survey data collection and the management of surveys.

4.3.1 **Interviewer Training Procedures** (*optional*). The material associated with training survey interviewers. "Training" is the process of educating the person performing the process in the procedures needed, what to do in unusual "anomalous" situations, and the importance of applying these procedures consistently. An "Interviewer" is an individual who interviews respondents for a survey regardless of the method used (face-to-face or telephone). Interviewers are also known as "Enumerators."

4.3.2 **Outreach and Promotion Procedures** (*optional*). Publicity and other methods intended to increase respondent cooperation for the purpose of decreasing the impact of non-response.

4.3.3 **Listing Operations Procedures** (*optional*). The process(es) associated with verifying addresses and the existence of a particular respondent or other object interest.

4.3.4 **Check-in/Check-out Procedures** (*optional*). The process of recording of the number of forms (**Measurement Instruments**) sent out versus the number received by batch, frame, or other unit.

4.3.5 **Non-Response Follow-Up Procedures** (*optional*). Instructions for conducting the data collection for cases or units that did not provide substantial information in the first interviewing attempt.

4.3.6 **Data Collection Quality Control Procedures** (*optional*). Procedure used to collect data for the survey. Description of the inspection and other procedures used to assess error during data collection. Refer to Annex I for detailed summary.

4.3.7 **Survey Preparation Procedures** (*optional*). Instructions or memos describing how a survey will be implemented or organized. These instructions include how surveys are managed and general instructions on how data is collected.

- 4.3.8 **Field Observation Procedures** (*optional*). Instructions for reviewing the survey interview process, processing of questionnaires, and the entire process of collecting data.
- 4.3.9 **Data Capture Modes Description** (*optional*). Method of converting the data as acquired by the measurement instrument into a form or format that can be further analyzed, stored, and processed. For example, data keying, CATI, CAPI, OCR are data capture modes.
- 4.3.10 **Coverage Procedures** (*mandatory*). Instructions for handling Under coverage or Overcoverage situations. That is, instructions for identifying and accounting for total observations that are more than or less than expected.
- 4.4 **Reports And Records** (*mandatory*). Reports from all phases of data collection.
- 4.4.1 **Interviewer Notes** (*optional*). Method used to contact the source of the survey data. This may include rules and procedures associated with conducting the survey.
- 4.4.2 **Data Collection and Quality Control Reports** (*optional*). Obtaining information from respondents and the conversion of that data into a form which can be processed. The results of testing interviewers and data keyers for error rates. Refer to Annex I for detailed summary.
- 4.4.3 **Editing Reports** (*optional*). Documentation reflecting the outcome of any data quality improvement tool by which erroneous or highly suspect data are found, and if necessary, corrected. Refer to Annex H for detailed summary.
- 4.4.4 **Coding Reports** (*optional*). The results of analyzing responses and selecting a code.
- 4.4.5 **Planned Observation Register** (*optional*). A record of the observation process. For example if there is a mail-out questionnaire, this may be a record of the mail preparation process, the mailing itself, etc.
- 4.5 **Production of the Final Anomalous Observation Register** (*optional*). Creation of a record of the determination and handling of all anomalous observations (i.e., undercount, overcount, partial response, erroneous response, non-response).
- 4.5.1 **Treatment of Overcoverage/Interruptions Objects** (*optional*). Where there are more or different respondents than planned for a given sample or frame.
- 4.5.2 **Treatment of Non-response** (*optional*). Total imputation or weighted adjustments of data when the potential respondent has refused to answer the entire questionnaire.

- 4.5.3 **Treatment of Partial Non-response** (*optional*). Use of imputation of data when the potential respondent has refused to answer the selected portions of the questionnaire.
- 4.5.4 **Counting of Anomalous Observations** (*optional*). How all or certain anomalous observations were counted.
- 4.5.5 **Derived Variables** (*optional*). A statistic obtained by an arithmetical operation or aggregation from the primary observations. In this sense, almost every statistic is "derived." The term is mainly used to denote descriptive statistical quantities obtained from data which are primary in the sense of being mere summaries of observations.

5 Data Processing (Computer Systems) Documentation of all computer processes needed to support of the survey. The stage of a project, following collection and receipt of the original material and preceding report-writing, during which the information is entered onto a machine-readable medium (or directly into a computer system) and eventually used to produce tabulations and statistical analyzes. The following states may be involved: editing, coding, data cleaning, and rearrangement of information to form a summary set of variables amenable to analysis. In modern practice, the latter stages are based exclusively on computer-related devices and the punch card methods of the past are rapidly disappearing. A wide range of general purpose software is available but complex surveys or special analyzes may still require specific programs. [See Prescott, Roland, Wilson, Christopher, ed. *The Dictionary of Demography* (TDD) pg. 51].

- 5.1 **Interviewer Administrative Processing** (*optional*). Electronic or manual calculations used to assess interviewer hours worked and net pay. Also the dissemination of this kind of data to requesting offices. See Annex J for detail.
- 5.2 **Data Preparation and Editing Processing** (*optional*). Electronic or manual manipulations of any data quality improvement statistics by which erroneous or highly suspect data are found. The process of obtaining and capturing data from respondents and preparing that data for further processing.
 - 5.2.1 **Editing Processing** (*optional*). Electronic or manual manipulations of any data quality improvement statistics by which erroneous or highly suspect data are found. Refer to Annex H for detailed summary.
 - 5.2.2 **Coding Processing** (*optional*). Electronic or manual analysis of responses and selecting a code.
 - 5.2.3 **Survey Preparation Processing** (*optional*). Electronic or manual programs that produce listings or materials for field workers. These materials include management listings and reports. See Annex J for detail.
- 5.3 **Data Capture Processing** (*optional*). The electronic procedures used to store data in an electronic form so that it can be centrally located, analyzed, and disseminated.
- 5.4 **Coverage Processing** (*mandatory*). Electronic or manual manipulations for handling Under coverage or Over coverage situations. That is, instructions for identifying and accounting for total observations that are more than or less than expected.
- 5.5 **Non-Response Follow-Up Processing** (*optional*). Electronic or manual calculations that store information on the nonresponse follow-up operation. This includes production of statistics and listings regarding the operation.

- 5.6 **Outreach and Promotion Processing** (*optional*). Electronic or manual manipulations for the preparation of publicity materials and administration of publicity events in order to improve respondent coverage.
- 5.7 **Data Collection and Quality Control Processing** (*optional*). Electronic and manual calculations for the components of the data collection process. This includes creating interviewer lists, processing evaluations, and producing statistics on the status of operations. Refer to Annex I for detailed summary.

6 Data Analysis (*optional*). Documentation related to all statistical processes used to analyze the survey results or those used for displaying or presenting the resultant information.

6.1 **Other inferences and analyses** (*optional*). Use of sample data to build predictive models which can be applied to other population. One example would be regression models.

6.1.1 **Disclosure Analysis** (*optional*). The process of protecting the confidentiality of data. It involves limiting the amount of detailed information disseminated and/or masking data via noise addition, data swapping, etc.

6.1.2 **Microdata Disclosure Avoidance** (*optional*). Method used for preventing legally protected data from being accessed in any way to unauthorized users. These include Masking, Data Switching, Confidentiality Edits.

6.1.3 **Macrodata Disclosure Avoidance** (*optional*). Processes and procedures used to ensure that **Macrodata** cannot be decomposed in such a way that the data associated with particular respondent or set of respondents.

6.2 **Estimation** (*mandatory*). A detailed description with formulas of all steps in estimation, including (where pertinent). See Annex K for detail.

6.3 **Descriptive Statistics** (*optional, repeatable*). Information about the statistical characteristics of the data. Refer to Annex L for details.

7 Data Dissemination (*optional*). The processes used to distribute non-confidential data to as many users and as cheaply as possible.

7.1 **System** (*optional*). An identifiable process, either fully automated or computer assisted, which implements one or more Techniques to produce one or more deliverables. A System can be composed of other Systems. See Annex M for detail.

7.2 **Product** (*optional*). A finished deliverable of a Project or a Survey Instance for outside the Bureau (external) use. See Annex C for detail.

7.3 **Archiving** (*optional*). Storing data for backup, contingency or statutory reasons. Typically, this data is stored in a way which requires additional processing before it can be used. See Annex M for detail.

7.4 **Data Transmission** (*optional*). The process or procedure by which collected data is moved from the point of collection to where it is being accumulated for further processing or storage.

7.4.1 **System** (*optional*). An identifiable process, either fully automated or computer assisted, which implements one or more Techniques to produce one or more deliverables. A System can be composed of other Systems.

7.4.2 **Mode** (*optional*). If more than one mechanism for accomplishing data transmission is available, which mode is used.

7.4.3 **Problems** (*optional*). Record of difficulties encountered in accomplishing **Data Transmission**.

8Data (*mandatory*). Any information gathered as the result of a survey or added to a survey from

- 8.1 **Released Data** (*optional, repeatable*). Data released for general, unrestricted use by the public. See Annex N for detail.
- 8.2 **Active Data** (*optional, repeatable*). Any dataset currently being used for some official purpose. Examples of such use include frame development, survey quality evaluation, research, longitudinal studies and time series analysis. This may include: **respondent data**, and can be further divided into microdata (actual data items collected from respondents or imputed from other sources) and macrodata (summary data aggregated into meaningful groupings from individual microdata records or from other macrodata). All data associated with a particular survey or census; can also be further divided into microdata and macrodata. See Annex N for detail.
- 8.3 **Archived Data** (*optional, repeatable*). Data stored for backup, contingency or statutory reasons. Typically this data is stored in a way which will require additional processing before it can be used. See Annex N for detail.

Annexes to the Table of Contents View

Annex A Documentation. Any written or electronic material related to this element.

Document Name (*mandatory*). The familiar or common name of the documentation.

Abstract (*mandatory*). A brief description or extract of the applicable text of the project dataset.

Theme (*mandatory*). Describes the cultural and demographic content of the documentation.

Citation (*mandatory*). The formal reference used to identify the material which describes the data elements.

Version (*mandatory*). A variant of the original citation.

Originator (*mandatory*). The name of a person who initiates the document.

Origination date (*mandatory*). The date the document was initiated.

Origination title (*mandatory*). The title of the document when it was initiated.

Origination time (*mandatory*). The time the document was initiated.

Series no. (*mandatory*). Any numerical identification that uniquely identifies the document.

Pointer to Object (*mandatory*). Uniform Resource Locator. A unique address on a network or machine where a document is located. To access the address, a user must use an Internet protocol such as http, gopher, or ftp.

Pointer type (*optional*). URL, database pointer, full file name, etc., of the documentation.

Pointer (*optional*). The pointer to the object of interest.

Distribution Information (*mandatory*). The name of the distributor or person releasing the document. If the document is released in large quantities, information on how many copies and where they were sent is requested.

Annex B Survey Datasets Any permanently stored collection of information usually containing either case level data, aggregation of case level data, or statistical manipulations of either the case level or aggregated survey data, for multiple **Survey instances**.

Survey Dataset Name (*mandatory*). The name of the survey dataset.

Creation Date (*mandatory*). The date the survey dataset was created.

Version (*mandatory*). A variant of the original survey dataset.

Dataset Description (*mandatory*). The description of the survey dataset. Any documents specifically concerning the description of the dataset. See Annex A for documentation components.

Date of Last Modification (*mandatory*). The date of the last modification of the survey dataset.

Access Constraint (*mandatory*). Type of restriction to the amount and type of data a user can have access to.

Contact (*mandatory*). Contact information for an individual or organization that is knowledgeable about the dataset.

Contact person (*mandatory*). A person associated with the component. This entry should provide either the Contact Person or the Contact Organization, dependent upon whether the person or organization has the more significant association to this element.

Contact phone (*mandatory*). The telephone number by which individuals can speak to the contact.

Contact email (*mandatory*). The address of the electronic mailbox of the contact.

Time period of contact (*mandatory*). When **Contact person** or **Contact organization** or is available.

Contact organization (*mandatory*). The organization associated with the dataset. The dataset producer should provide either the Contact Person or the Contact Organization, dependent upon whether the person or organization has the more significant association to the survey.

Contact type (*mandatory*). Specify "Person" or "Organization". Indicates whether providing information is the responsibility of a particular individual (indicated by "**Person**" in this field) or of the organization as a whole (indicated by "**Organization**" in this field).

Annex C Product. A finished deliverable of an Emprise or a Survey instance for outside the Bureau (external) use.

Product Name (*mandatory*). The common or familiar name of the product.

Release Date (*mandatory*). The date the product was released.

Version (*mandatory*). A variant of the original product.

Product Description (*mandatory*). Any documents specifically concerning the description of the dataset. See Annex A for documentation components.

Contact (*mandatory*). Contact information for an individual or organization that is knowledgeable about the product. See Annex B for Contact detail.

Annex D Planning Information about the procedures, processes and methods used in the development of a particular emprise, product or survey. This Annex gives the detail for items 2.4 Survey, 2.5 System and 2.6 Project, Appendix E, Chapter 2.

Survey/System/Project Name (*mandatory*). The name of the survey/system/project.

Documentation (*optional*). Any documents specifically concerning the description of the survey/system/project. See Annex A for documentation components.

Personnel (*optional*). Projected personnel requirements to design and implement this survey, system, or project as proposed.

Scheduling (*optional*). Particular dates associated with the start and completion of each element of the subject procedure or survey, system, or project.

Training (*optional*). Projected training requirements for personnel designing and implementing this survey, system, or project as proposed.

Budget (*optional*). Projected funding requirements to design and implement this survey, system, or project as proposed.

Annex E Frame/Sample Concept The research strategy of collecting data from a part of a population with a view to drawing inferences about the whole. The “population” in this sense is often termed the “universe.” This section shows the detail for items 3.1.1 Frame Concepts and Section 3.2.1 Sample Concepts, Appendix E, Chapter 3.

Definition of Frame/Sample (*optional*). The definition of the frame/sample. The frame is the list, map or other specification of the units which constitute the available information relating to the universe of interest which is being sampled. A sample is a dataset containing a subset of a Frame for a particular set of Survey Instances, selected with specific sampling Technique. For a census, the Sample incorporates the entire Frame.

Variables of Interest (*optional*). The name assigned to a particular datum about the population of interest from which the content of this **Data Element** is derived. Specific elements of data which has been gathered in the process of carrying out a survey. A **Variable of Interest** is distinguished from other data in that it is one of the elements of data needed to accomplish the purpose of the survey through which it was gathered (or joined to).

File Description (*optional*). A description of the file that constitutes the frame. This may be a data layout, a data definition, a FIPS 123 (DDF) definition and information necessary to access the data. This may also be a URL to the location of the file description for the file that constitutes the frame.

Object Type (*optional*). The format a requested file or object is in. For example, an object may have an .html, text, .gif, .jpg, .wpd, .rtf, or unencoded format.

Pointer to Object (*optional*). The URL or software link used to access an object.

Eligibility Requirements (*optional*). A description of the independent criteria to be applied in selecting the source or sources for the frame.

Annex F Observation models. Models of the data acquisition process. This section shows the detail for item 3.5 Observation models, Appendix E, Chapter 3.

Sampling (*mandatory*). The science that guides quantitative studies of content, behavior, performance, materials, and causes of differences.

Non-response (*optional*). Refers to many sources of failure to obtain observations on some elements selected and designated for the sample.

Measurement/Observation (*optional*). A model of the planned observations.

Frame coverage (*optional*). A model of the impact of the frame design on the sampling coverage of the survey.

Total model (*optional*). A comprehensive, inclusive model of the impact of the sample design on the survey results.

Annex G Estimation model. Rules or methods for estimating the variance of the observed population from the universe of interest. This section shows the detail for item 3.6 Estimation models, Appendix E, Chapter 3.

Completed estimations (*optional*). Error estimations based on analysis of the completed survey.

Point estimations (*optional*). The methods or estimation equations used to develop the estimated values of a parameter where such estimations have been required. [See Kendall, Maurice G. And William R. Buckland (1960) *A Dictionary of Statistical Terms*, pg. 220].

Variations and Other Quality Measures (*optional*). Description of measure of quality on the data, with the sample variance estimate being foremost; the variance, a measure of the reliability/precision of a sample estimate, and its computation formula depend on the design of the survey sample. Other measures may be item nonresponse rates, proxy rates, etc.

Annex H Editing. The processes of verifying that data or responses are accurate and consistent by any of several criteria. This section shows the detail for editing specifications, reports, procedures, and processing.

Pre-Edits (*optional*). Specifications, reports, procedures, and processing for the automated process devised to detect blatant errors on the questionnaire. Pre-edits are performed before program edits are applied to the data.

Program Edits (*optional*). **Program Edits** consist of Range checks and consistency edits.

Clerical Edits (*optional*). Specifications, reports, procedures, and processing for the automated process devised to detect blatant errors on the questionnaire. Pre-edits are performed before the program edits are applied to the data.

Field Edit (*optional*). Specifications, reports, procedures, and processing applied during data collection to prepare the data for further processing.

Tabulation Edits Processing (*optional*). Electronic or manual detection of individual errors by: 1) checks on aggregated data, or 2) checks applied to the whole body of records. Also referred to as Macro-edits.

Annex I Data Collection and Quality Control The specifications, reports, procedures, and processing involved in the collection or compilation of data.

Description (*optional*). Narrative text explaining the specification, report, procedure or process.

Reinterview (*optional*). Specifications, reports, procedures, and processing for the reporting on analysis of data regarding quality control and/or response variance data.

Production Standard (*optional*). Specifications, reports, procedures, and processing indicating which cases have failed or were rejected by the quality control standards acceptable levels.

Field Observation (*optional*). Specifications, reports, procedures, and processing for reports containing supervisory field representatives' assessments of subordinate field representatives.

Evaluation (*optional*). Specifications, reports, procedures, and processing for reports filed following the development and evaluation of new technology.

Address Coverage (*optional*). Specifications, reports, procedures, and processing for reports of the whole structure within structure misses.

Check-In/Check-Out (*optional*). The statement of work associated with the process of recording of the number of forms (**Measurement Instruments**) sent out versus the number received by batch, frame, or other unit.

Listing Operations (*optional*). Operations (usually accomplished on the site) for verifying addresses and the existence of a particular respondent or other object interest.

Interviewer Training (*optional*). The material associated with training survey interviewers. "Training" is the process of educating the person performing the process in the procedures needed, what to do in unusual "anomalous" situations, and the importance of applying these procedures consistently. An "Interviewer" is an individual who interviews respondents for a survey regardless of the method used (face-to-face or telephone). Interviewers are also known as "Enumerators."

Data Preparation (*optional*). The process of obtaining and capturing data from respondents and preparing that data for further processing.

MIS Processing (*optional*). Specifications, procedures, and reports reflecting the overall progress of an operation. This includes budgeting statistics, nonresponse statistics, system status, and overall workload, for example.

Non-response (*optional*). Documentation showing total and percent refusal by the potential respondent to answer all or part of the questionnaire or the inability to locate the potential respondent.

Causes (*optional*). The reason the information was not obtained.

Actions (*optional*). Procedure used in the case of non-response

Rate (*optional*). The percent of total units in a survey that classified as non-response units.

Other Documentation (*optional*). Data collection or quality control reports that are not classified as reinterview, non-response, production standard, field observation, address coverage, check-in/check-out or listing operations.

Annex J Processing Details. How processing is accomplished.

Cost and Progress Processing (*optional*). Electronic or manual calculation and approval of the hours worked and the measure of accomplishment of each field worker.

Payroll Processing (*optional*). Electronic or manual calculation and approval of net pay for each field worker.

Documentation (*optional*). Narrative text explaining the procedures, processes, or results of the object of interest. See Annex A.

Flow diagram (*optional*). Chart of the relationships among the data, processes and resources.

Component descriptions (*optional*). System modules, or objects and methods (depending on the program design paradigm used).

System (*optional*). An identifiable process, either fully automated or computer assisted, which implements one or more Techniques to produce one or more deliverables. A System can be composed of other Systems.

Annex K Estimation. Rules or methods for estimating the variance of the observed population from the universe of interest.

Procedures (*optional*). The specific steps or systems used in estimation.

Weighting (*mandatory, if applicable*). The measurement of the relative importance of an item in a statistical population.

Calculated Adjustment Factors (*optional*). The coefficient attached to an observation as its weight in a procedure involving weighting. [See DST pg. 219].

Independent Control Counts (*optional*). Counts of data records obtained independently of any process dependent on those counts.

Historical Weights (*optional*). Weights used for previous time series, seasonal, or other adjustments.

Weighting Controls (*optional*). Checking mechanisms for the efficacy of weighting procedures.

Base Weights (*optional*). The weights of a weighting system for an index number computed according to the information relating to the base period instead, for example, of the current period. [See DST pg. 15].

Special Weighting Factors (*optional*). The coefficients of a linear function of the values of the sample units used to estimate population, stratum, or higher stage unit totals are called raising, multiplying, weighting or inflation factors of the corresponding sample units. If the raising factors of all the sample units are equal, the common raising factor is called the raising factor of the sample, and the sample itself is called self-weighting. It should be noted that the raising factors depend not only on the sampling plan but also on the method of estimation. [See DST pg 167].

Ratio Estimators (*optional*). An estimator which involves the ratio of two variates, e.e., a ratio whose numerator and denominator are both subject to sampling errors. The term occurs particularly in sample survey theory. If the members of a population each bear the values of two characteristics, x and y , and the total of x , say X , is known for the population, the corresponding total of y , say Y , can be estimated by multiplying X by a sample ratio consisting of the sample total of y divided by the sample total of x . [See DST pg. 170]

Non-Interview Adjustment (*optional*). Statistical adjustment applied to account for non-interview (where respondent is not interviewed).

Longitudinal Weights (*optional*). Weights applied across series.

Other Weights (*optional*). Types of weighting not previously listed.

Nonresponse adjustment (*mandatory, if applicable*). Statistical adjustment applied to account for non-interview (where respondent is not interviewed).

Imputation (*mandatory, if applicable*). Where simulated data is substituted based on data from similar respondents adjusted for strata, and other factors.

Type (*optional*). Type of imputation used. Imputation type: Hot Deck, Cold Deck, Ratio, etc.

Post-stratification (*mandatory, if applicable*). Stratification of data based on data collected rather than as a part of sampling.

Composite estimation (*mandatory, if applicable*). Estimation combining several techniques.

Seasonal adjustment (*mandatory, if applicable*). Adjustment based on cyclic analysis.

Variance estimation methods (*mandatory, if applicable*).

Estimations and analyses (*optional*). Analysis of the data, considering anomalies of data collection.

Overview (*optional*). A high-level description of the program, project, system, data, metadata, or object of interest.

System flow (*optional*). Chart of the relationships among the data, processes and resources.

Component descriptions (*optional*). System modules, or objects and methods (depending on the program design paradigm used).

System (*optional*). An identifiable process, either fully automated or computer assisted, which implements one or more Techniques to produce one or more deliverables. A System can be composed of other Systems.

Annex L Descriptive Statistics. Information about the statistical characteristics of the data. The detail sections for 6.3 Descriptive Statistics, Appendix E, Chapter 6..

Number of Observations (*optional*). Number of all intended or planned observations whether data was obtained or not.

Number of Non-missing Observations. Number intended or planned observations for which data was *not* obtained.

Number of Missing Observations (*optional*). Number of observations for which data was obtained. 6.3.3

Minimum Value (*optional*). The smallest value of the non-missing observations.

Maximum Value (*optional*). The largest value of the non-missing observations.

Range of Values (*optional*). **Maximum Value - Minimum Value.**

Sum (*optional*). The total of values of non-missing observations.

Weighted Sum (*optional*). A total of non-missing observations that is given a measure of importance relative to other sums.

Sum of Weights (*optional*). The total of weighted values of the non-missing observations.

Arithmetic Mean (*optional*). A number obtained by adding several quantities together and dividing the sum by the number of quantities.

Sum of Squares (*optional*). The total of the values of each observation multiplied by itself.

Corrected Sum of Squares (*optional*). The corrected total is the total sum of squares divided by the number of observations.

Variance (*optional*). A measure of how the sum of individual observations deviated from the mean. It is the sum of the squares of differences of each observation from the mean divided by the number of observations in the population.

Standard Deviation (*optional*). Square root of the variance of a set of numbers.

Standard Error (*optional*). The positive square root of the variance of the sampling distribution of a statistic. [See DST pg. 195]. The standard error of the mean.

Coefficient of Variation (*optional*). The standard deviation of a distribution divided by the arithmetic mean; sometimes multiplied by 100. It was proposed by K. Pearson (1895) for the purpose of comparing the variabilities of frequency distributions, but is sensitive to errors in the means and is of limited use. [See DST pg. 217]. The percent **Coefficient of Variation**.

Skewness (*optional*). A distribution which is not symmetrical. [See DST pg. 9 and 189]. The measure of sidedness.

Kurtosis (*optional*). A term used to describe the extent to which an unimodal frequency curve is "peaked"; that is to say, the extent of the relative steepness of ascent in the neighborhood of the mode. [See DST pg. 111]. The measure of the heaviness of the tails or of flatness.

Students t (*optional*). This distribution, originally due to "Student" (1908), is usually written in the form, as modified by R.A. Fisher (1925): The distribution is, among other things, that of the ratio of a sample mean, measured from the parent mean, to a sample variance, multiplied by a constant, in samples from a Normal population. It is thus independent of the parent scale parameter and can be used to set confidence intervals to the mean independently of the parent variance. [See DST pg. 208]. Test of the null hypothesis, population mean = zero.

Median (*optional*). The middle value when the values of the non-missing observations are arranged in order of value, and the **Number of Non-missing Observations** is odd; the mean of the two middle values when the **Number of Non-missing Observations** is even. The median is that value of the variate which divides the total frequency into two halves. As a partition value it may be defined for a continuous frequency distribution by equation. [See DST pg.128].

Mode (*optional*). The value at which the density of the population is at a maximum. For a sample, the most frequently occurring of the values of the non-missing observations. The mode was originally conceived of as that value of the variate which is possessed by the greatest number of members of the population. Although the idea of the most frequently encountered or fashionable value of the variate is probably very old, it was not generally used in statistics until popularized by K. Pearson (1894). The concept is essentially of use only for continuous distributions, although it can be extended to the discontinuous case. [See DST pg. 131].

Annex M Data Dissemination. The process of distributing data.

Overview (*optional*). A high-level description of the program, project, system, data, metadata, or object of interest.

Flow diagram (*optional*). Chart of the relationships among the data, processes and resources.

Annex N Data. The collection of facts, ideas, or instructions stored in a formalized manner suitable for communication, interpretation, or processing.

Data Set (*optional, repeatable*). A collection of facts, ideas, or instructions stored permanently in a formalized manner suitable for communication, interpretation, or processing. See Annex B.

Record Layout (*optional*). Describes the physical layout of the data (preferably in Data Definition File format).

Appendix F. Example application using CDDM

1 Identification Information:

1.1 Citation:

1.1.1 Citation Information:

1.1.1.1 **Originator:** Population Estimates and Population Distribution Branches. U.S. Department of Commerce, Bureau of the Census

1.1.1.2 **Publication Date:** 19961008

1.1.1.3 **Title:** Estimates of Housing Units, Households, and Persons per Household for the United States, Regions, and States: 1990 to 1995

1.2 Description:

1.2.1 **Abstract:** This dataset contains estimates of housing units, households, and persons per household for the United States, its Census regions and divisions, and each state for the years 1990 through 1995. The estimates were produced by modeling base data from the 1990 decennial census. The data is presented in 14 blocks of information.

1.2.2 **Purpose:** Through the use of datasets produced by the modeling and estimation of population, housing, and households, the Bureau of the Census provides data users with timely information on an annual basis. This information could not otherwise be provided as the size and cost of a census operation does not lend itself easily to the production of statistics that meet consumer demands of continuously developed current information.

1.3 Time Period of Content:

1.3.1 Time Period Information:

1.3.1.1 Range of Dates/Times:

1.3.1.1.1 **Beginning Date:** 1990

1.3.1.1.2 **Ending Date:** 1990

1.3.1.2 **Currentness Reference:** 1990 (decennial census)

1.4 Status:

1.4.1 **Progress:** Complete

1.4.2 **Maintenance and Update Frequency:** None planned

1.5 Spatial Domain:

1.5.1 Bounding Coordinates:

1.5.1.1 **West Bounding Coordinate:** -187.652154

1.5.1.2 **East Bounding Coordinate:** -66.775965

1.5.1.3 **North Bounding Coordinate:** 71.441470

1.5.1.4 **South Bounding Coordinate:** 18.865460

1.5.2 Geographic Extent:

1.5.2.1 **Geographic Extent Description:** The United States (48 contiguous states plus Alaska and Hawaii)

1.5.2.1.1 Geographic Entity:

1.5.2.1.1.1 **Geographic Entity Type:** Legal, administrative

1.5.2.1.1.2 **Geographic Entity Unit:** United States

1.5.2.1.2 Geographic Entity:**1.5.2.1.2.1 Geographic Entity Type:** Statistical, thematic**1.5.2.1.2.2 Geographic Entity Unit:** Census regions and divisions (Northeast Region, New England division, Middle Atlantic division, Midwest Region, East North Central division, West North Central division, South Region, South Atlantic region, East South Central division, West South Central division, West Region, Mountain division, Pacific division)**1.5.2.1.3 Geographic Entity:****1.5.2.1.3.1 Geographic Entity Type:** Legal, administrative**1.5.2.1.3.2 Geographic Entity Unit:** State**1.5.2.2 Tag:** None of the three Geographic Entities provided have any "missing geography" so the mandatory if applicable compound element "Missing Geography" and its data elements "Missing Geographic Entity Unit Identifier" and "Missing Geographic Entity Unit" are not included in this metadata example.**1.6 Keywords:****1.6.1 Theme:****1.6.1.1 Theme Keyword:** housing**1.6.1.1.1 Main Theme:** housing**1.6.1.1.1.1 Definition/Classification:** "Methodology for State Housing Unit/House Estimates," Extract from Bureau of the Census Report Series P25-1123**1.6.1.1.1.1.1 Minor Theme:** persons**1.6.1.1.1.1.1.1 Scalar:** persons in households**1.6.1.1.1.1.1.2 Scalar:** change in persons in households**1.6.1.1.1.1.1.3 Scalar:** percent change in persons in households**1.6.1.1.1.1.2 Minor Theme:** estimated persons**1.6.1.1.1.1.2.1 Scalar:** persons per households estimate**1.6.1.1.1.1.2.2 Scalar:** change in persons per households estimate**1.6.1.1.1.1.2.3 Scalar:** percent change per persons in household estimate**1.6.1.1.1.1.3 Minor Theme:** households estimate**1.6.1.1.1.1.3.1 Scalar:** total households estimate**1.6.1.1.1.1.3.2 Scalar:** change in total households**1.6.1.1.1.1.3.3 Scalar:** percent change total households**1.6.1.1.1.1.4 Minor Theme:** householder**1.6.1.1.1.1.4.1 Minor Theme:** Age**1.6.1.1.1.1.4.2 Minor Theme:** 15 - 24 years old**1.6.1.1.1.1.4.3 Minor Theme:** 25 - 29 years old**1.6.1.1.1.1.4.4 Minor Theme:** 30 - 34 years old**1.6.1.1.1.1.4.5 Minor Theme:** 35 - 39 years old**1.6.1.1.1.1.4.6 Minor Theme:** 40 - 44 years old**1.6.1.1.1.1.4.7 Minor Theme:** 45 - 49 years old**1.6.1.1.1.1.4.8 Minor Theme:** 50 - 54 years old**1.6.1.1.1.1.4.9 Minor Theme:** 55 - 59 years old

- 1.6.1.1.1.4.10 **Minor Theme:** 60 - 64 years old
- 1.6.1.1.1.4.11 **Minor Theme:** 65+ years old
- 1.6.1.1.1.4.12 **Scalar:** 4/1/90 census (corrected) total households with householder
- 1.6.1.1.1.4.13 **Scalar:** estimate of total households with householder
- 1.6.1.1.1.4.14 **Scalar:** change in total households with householder
- 1.6.1.1.1.4.15 **Scalar:** percent change in total households with householder
- 1.6.1.1.1.5 **Minor Theme:** housing unit
- 1.6.1.1.1.5.1 **Scalar:** 4/1/90 census (corrected) housing unit
- 1.6.1.1.1.5.2 **Scalar:** housing unit estimates
- 1.6.1.1.1.5.3 **Scalar:** amount of change in housing units
- 1.6.1.1.1.5.4 **Scalar:** percent change in housing units
- 1.6.2 **Place:**
- 1.6.2.1 **Place Keyword Thesaurus:** FIPS Publication 5-2, "Codes for the Identification of the States, the District of Columbia and the Outlying Areas of the United States, and Associated Areas," 19870528
- 1.6.2.2 **Place Keyword:** United States
- 1.6.2.3 **Place Keyword:** State: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming
- 1.6.2.4 **Place Keyword Thesaurus:** "1990 Census of Population and Housing Data," (1990 CPH), Appendix A, Area Classifications
- 1.6.2.5 **Place Keyword:** Census regions and Census divisions: Northeast Region, New England division, Middle Atlantic division, Midwest Region, East North Central division, West North Central division, South Region, South Atlantic division, East South Central division, West South Central division, West Region, Mountain division, Pacific division
- 1.7 **Access Constraints:** Primarily available online; Internet access required for retrieval
- 1.8 **Use Constraints:** None (acknowledgment of the Bureau of the Census appreciated for products derived from this file.)
- 2 **Data Quality Information:**
- 2.1 **Attribute Accuracy:**
- 2.1.1 **Attribute Accuracy Report:** Estimates were modeled from 1990 census data. To test the accuracy of the estimates, the April 1, 1990 state estimates of population, housing units, and households were compared to the 1990 census tabulations based on special modified age, race editing procedures. Mean absolute percent errors of these estimates for states, by type of estimate were obtained. Even though the estimation models used were fairly simplistic, the April 1, 1990 estimates of housing units and households performed at a surprising degree of accuracy. The U.S. April 1, 1990

housing estimates when measured against the U.S. April 1, 1990 decennial census counts exhibited an error of 240,000 units, or +0.26 percent of the total housing stock.

The household estimation model works well for total households. However, its dependence on the accuracy of state population estimates by age is a shortcoming which has not been measured.

Note: It is believed statistics (and estimates) decline in reliability as geographic detail increases and sample size decreases. Therefore, subnational estimates are less reliable than national estimates.

2.1.2 Quantitative Attribute Accuracy Assessment:

2.1.2.1 Attribute Accuracy Value:

	U.S. percent error	State mean absolute percent error
Population		
Total	0.63	1.55
Ages 15-24	-2.64	3.33
Ages 25-34	1.47	6.85
Ages 35-44	0.30	2.48
Ages 45-54	1.28	2.87
Ages 55-64	1.78	5.29
Ages 65-74	2.02	2.50
Ages 75+	0.01	0.79
Housing Units		
Total	0.26	1.75
Households		
Total	1.25	1.54
Ages 15-24	-1.75	8.46
Ages 25-34	2.17	5.18
Ages 35-44	1.37	2.30
Ages 45-54	1.78	3.00
Ages 55-64	2.81	5.61
Ages 65-74	1.69	2.25
Ages 75+	-3.10	3.06

Overall: The U.S. April 1, 1990 housing estimate error was +0.26.

2.1.2.2 Attribute Accuracy Explanation: Comparison of April 1, 1990 state estimates of population, housing units, and households to 1990 census tabulations based on special modified age, race editing procedures. Mean absolute errors of estimates per state obtained and are reported by type of estimate.

- 2.2 Logical Consistency:** 100% consistent. The dataset is an ASCII file. Printouts of the ASCII file have been visually checked by technicians to ascertain the correct information is placed in the 14 data blocks in accordance with the file layout described at the end of the file.
- 2.3 Completeness:** 100% complete. The dataset contains all estimates for 3 levels of geography: (1) the national level, (2) the Census district and region level, and (3) the state level. The states are limited to the contiguous 48 states plus Alaska and Hawaii (trust territories are NOT included.).
- 2.4 Tag:** As the spatial component of the dataset is indirect rather than direct (coordinates such as latitude/longitude or x/y values are not a part of the dataset), the mandatory if applicable Positional

Accuracy compound element and its nested data elements are not applicable and have not been provided.

2.5 Lineage:

2.5.1 Source Information:

2.5.1.1 Source Citation:

2.5.1.1.1 Citation Information:

2.5.1.1.1.1 **Originator:** U.S. Department of Commerce, Bureau of the Census

2.5.1.1.1.2 **Publication Date:** 1990

2.5.1.1.1.3 **Title:** Summary Population and Housing Characteristics 1990, CPH-1 and Population and Housing Unit Counts 1990, CPH-2

2.5.1.2 **Source Scale Denominator:** None

2.5.1.3 **Source Time Period of Content:**

2.5.1.3.1 Time Period Information

2.5.1.3.1.1 Single Date/Time:

2.5.1.3.1.1.1 **Calendar Date:** April 1, 1990

2.5.1.3.2 **Source Currentness Reference:** Census day, 1990

2.5.1.4 **Source Citation Abbreviation:** CPH

2.5.1.5 **Source Contribution:** Base data for the year 1990 (reported in the dataset). Additionally, used to model estimates for housing unit/households the years 1991 through 1995 (reported in the dataset). Data also served as the "quality control" check for the estimate model through the comparison of the actual 1990 data to modeled 1990 data.

2.5.2 Source Information:

2.5.2.1 Source Citation:

2.5.2.1.1 Citation Information:

2.5.2.1.1.1 **Originator:** U.S. Department of Labor, Bureau of Labor Statistics and U.S. Department of Commerce, Bureau of the Census

2.5.2.1.1.2 **Publication Date:** Quarterly, 1990 through 1995

2.5.2.1.1.3 **Title:** Current Population Survey

2.5.2.1.2 **Source Scale Denominator:** None

2.5.2.2 **Source Time Period of Content:**

2.5.2.2.1 Time Period Information

2.5.2.2.1.1 Range of Date/Time:

2.5.2.2.1.1.1 **Beginning Date:** 1990

2.5.2.2.1.1.2 **Ending Date:** 1995

2.5.2.2.2 **Source Currentness Reference:** 6 weeks prior to publication date

2.5.2.3 **Source Citation Abbreviation:** CPS

2.5.2.4 **Source Contribution:** Data from a monthly survey conducted by the Bureau of the Census for the Bureau of Labor Statistics used to model estimates for housing unit/households the years 1991 through 1995 (reported in the dataset).

2.5.3 Process Step:

2.5.3.1 Process Description: 1990 decennial census data "modeled" to determine housing unit/households estimates for the years 1991 through 1995. The models were developed by statisticians at the Bureau of the Census. An example model for determining the state household estimate for 1993 is: $HH93j = ((HHR93j/HHR90j) * CHHR90j) * PA93j$, where j = age groups (15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+) HH93 = Estimate of households: July 1, 1993 HHR93 = U.S. householder rate from the March 1993 CPS HHR90 = U.S. householder rate from the March 1990 CPS PA93 = Total population estimate: July 1, 1993

2.5.3.2 Source Used Citation Abbreviation: CPH and CPS

2.5.3.3 Process Date: Unknown

3 Distribution Information:

3.1 Distributor:

3.1.1 Contact Information:

3.1.1.1 Contact Organization Primary: Population Estimates Branch U.S. Department of Commerce, Bureau of the Census Contact Address:

3.1.1.2 Address Type: mailing and physical address

3.1.1.2.1 Address: Federal Building 3, Room 2011

3.1.1.2.2 City: Washington

3.1.1.2.3 State: District of Columbia (DC)

3.1.1.2.4 Postal Code: 20233

3.1.1.2.5 Country: U.S.A.

3.1.1.3 Contact Voice Telephone: (301) 457-2380

3.1.2 Tag: As the label by which the dataset is known to its distributor is its title, this mandatory if applicable data element is not applicable and not included as part of the metadata.

3.1.3 Distribution Liability: No warranty, expressed or implied is made and no liability is assumed by the Bureau of the Census for the use of the data.

3.1.4 Standard Order Form:

3.1.4.1 Non Digital Form: Available as paper copy upon request.

3.1.4.2 Digital Form:

3.1.4.2.1 Digital Transfer Information:

3.1.4.2.1.1 Format Name: ASCII

3.1.4.2.1.2 Format Version Number: None

3.1.4.2.1.3 Format Version Date: None

3.1.4.2.1.4 Format Information Content: File layout consists of 14 blocks of information, field location, length, and descriptions included at end of file.

3.1.4.2.1.5 Transfer Size: 112K

3.1.4.2.2 Digital Transfer Options:

3.1.4.2.2.1 Online Option:

3.1.4.2.2.1.1 Computer Contact Information:

3.1.4.2.2.1.1.1 Network Address:

3.1.4.2.2.1.1.2 Network Resource Name:

<http://www.census.gov/population/estimates/housing/hsehld95.dat>

3.1.4.2.3 **Fees:** None

3.1.4.2.4 **Ordering Instructions:** To request a paper copy of this file, contact the Distributor

3.1.4.2.5 **Turnaround:** 1 - 2 weeks when requesting a paper copy

3.1.4.2.6 **Custom Order Process:** None available

3.1.4.2.7 **Technical Prerequisites:** Internet access required for to use Digital Transfer Option

3.1.4.2.8 **Available Time Period:**

3.1.4.2.8.1 **Time Period Information:**

3.1.4.2.8.1.1 **Range of Dates**

3.1.4.2.8.1.1.1 **Beginning Date:** 199606

3.1.4.2.8.1.1.2 **Ending Date:** 200001

4 **Metadata Reference Information:**

4.1 **Metadata Date:** 199612

4.2 **Metadata Standard Name:** Content Standards for Digital Geospatial Metadata: Thematic Supplement for Geospatially Referenced Cultural and Demographic Data Metadata

4.3 **Metadata Standard Version:** 11.15.96

4.4 **Metadata Time Convention:** local time

Appendix G. Use with Other Standards.

Section 7 of this document addressed the relationship of the SDSM metadata standard to a number of other standards. This appendix addresses how the SDSM is used with other standards.

The following standards are related in various ways to the SDSM:

Standard	How Used with the SDSM	Notes
Spatial Data Transfer Standard	Dataset Transfer Standard with SDSM Metadata	FIPS ⁸ -173
FGDC ⁹ Digital Geospatial Metadata Standard	Basis Several sections are incorporated	Future FIPS. Will be cross-walked via tools under development
FGDC Digital Geospatial Metadata Standard: Cultural and Demographic Supplement	Basis Themes concept and definitions are incorporated in the SDSM	Future FIPS. Will be cross-walked via tools under development
Application Profile for the Government Information Locator Service (GILS)	SDSM will either extend or will be registered under GILS	FIPS-192. Will be cross-walked via tools under development
Data Definition Format	Being considered as the future basis for statistical data interchange standard and is used as the source for the SDSM specification for dataset formats	FIPS-123 also ANSI/ISO ¹⁰ 8211-1985
Information Retrieval (Z39.50): Application Service Definition and Protocol Specification	SDSM may be registered under as a profile of Z39.50	ANSI/NISO ¹¹ Z39.50 GILS is a profile of Z39.50
Specification and Standardization of Data Elements	Data defined in the SDSM will have been registered in accordance with this standard	IS 11179

8. US Federal Information Processing Standard.

9. US Federal Geographic Data Committee.

10. International Standard issued by the International Standards Organization.

11. A US National Standard issued by the American National Standards Institute and the National Information Systems Organization.

Appendix H. Abbreviations and Acronyms

This section contains a list of all abbreviation and acronyms used in this document.

Abbreviation or Acronym	Meaning
ANSI	American National Standard Institute; the U.S. member of ISO and IEC.
ASCII	American Standard Code for Information Interchange; the character set used on most computers in the US.
bit	binary digit (base 2 number system)
BOC	Bureau of the Census
CAPI	Computer Aided Person Interview
CATI	Computer Aided Telephone Interview
CDDM	Standard for Cultural and Demographic Data Metadata
CPS	Current Population Survey
CV	Coefficient of Variation
DADS	Data Access and Dissemination System
DMS	Document Management System
FERRET	Federal Electronic Research and Review Extraction Tool
FGDC	Federal Geospatial Data Committee
FIPS	Federal Information Processing Standard
GIF	Graphics Interchange Format
GILS	Government Information Locator Service; GILS is a profile of Z39.50
GIS	Geographic Information Systems
HTTP	Hyper-Text Transfer Protocol
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force, the group responsible for proposing standards for

Abbreviation or Acronym	Meaning
	use on the Internet
IRDS	Information Resource Dictionary System, American National Standard ANSI X3.138.
ISM	Integrated Statistical Metadata
ISO	International Organization for Standardization
IT	Information Technology
JPEG	Joint Photographic Experts Group: a standardized image compression technique
JTC1	Joint Technical Committee 1 (Information Technology Standards) of ISO/IEC
MDR	Statistical Metadata Repository
MMSD	Metamodel for the Management of Sharable Data, American National Standard ANSI X3.285
NAICS	North American Industrial Code Standard
NCITS	National Committee for Information Technology Standards, (manages the development in the US of computer related standards, designated X3).
NISO	National Information Systems Organization (manages the development in the US of library related standards, designated Z39).
NIST	National Institute for Standards and Technology
NRFU	Non-response follow-up
OCR	Optical Character Recognition
OMB	Office of Management and Budget
OOS	Out of Scope
PROC	a SAS PROC edure
RDBMS	Relational Data Base Management System
RFC	Request For Comment (Internet standard proposed by the IETF)

Abbreviation or Acronym	Meaning
SAS	SAS Institute, Incorporated
SCBDOK	Statistics Sweden metadata collection
SDSM	Survey Design and Statistical Methodology
SDTS	Spatial Data Transfer Standard
SPWP	Statistical Policy Working Paper
SQL	Structured Query Language
SRD	Statistical Research Division
SSD	Systems Support Division
SSMB	Software and Standards Management Branch
StEPS	Standard Economic Processing System
TDD	Telecommunications Devices for the Deaf
TOC	Table of Contents
TTY	Teletype
U.S.C.	U. S. Code (i.e., the formal designation for Federal Laws)
URL	Universal Resource Locator
WAIS	Wide Area Information Servers, a standard Internet search service; WAIS is a profile of Z39.50
Z39.50	Information Retrieval (ANSI/NISO Z39.50): Application Service Definition and Protocol Specification