DEPARTMENTAL MANUAL	Number: 3430-001
SUBJECT: Division Data Element Standards Manual	**DATE:** June 14, 1995
	OPI: Office of Information Resources Management Division Information Management Division

FOREWORD

The Data Element Standards Manual is a dynamic document that the Office of Information Resources Management (OIRM) will update as the Departmental Data Administration Program is developed. This manual anticipates the future when the Department of Agriculture (USDA) will have a Data Repository populated by shareable agency data. By developing uniform standards and procedures for data now, USDA agencies will be positioned to build new or re- engineered software applications that ensure data sharing across USDA and provide better service to the public.

OIRM recognizing the strategic value of data and information, is developing a data management program under the Departmental IRM Strategic Plan. The Data Management Program is designed to move the Department into a corporate information environment by 1998.

The Data Element Standards Manual provides uniform standards and procedures that will aid USDA agencies in developing new or re-engineered software applications while ensuring data sharing across the Department. It will position the agencies to populate a USDA Data Repository as soon as that resource is acquired. In the interim, agencies can begin to redesign their data and build toward the corporate
information environment.

This manual is the result of a USDA-wide partnership. It is a technical
manual which outlines data element concepts; data element design and naming rules; and data element definition procedures. It is only one product of many planned under the USDA Data Management Program. It has been reviewed and approved by the Data Management Sub-Council of the USDA IRM Council.

This manual contains data element standards consistent with general principles and concepts successfully employed in government and industry. These standards differ from early data element standards in several ways. First, many data element standardization procedures rest on strict and elaborate naming conventions through which a data element's name becomes the key for search, retrieval and linkage to the

enterprise data model. In contrast, USDA's procedures use the full
search capabilities of data repository systems rather than relying on
the consistent building and use of "intelligent" data element names.
Thus, data about the data elements (i.e., metadata) are stored solely
in
the data repository database rather than being embedded in the data
element name. Similarly, data elements may be tied to data models
through the metadata contained in the data repository rather than
through the data element name.

Second, USDA can use previously defined data elements and use widely
accepted business names (e.g., social security number) when they exist
rather than having to construct specific data element names. If
neither
a previously defined data element nor an accepted business name is
available, a data element is named according to established guidelines
and recommended semantics, to minimize "writer's block" and encourage
precise naming techniques.

The use of business names, where acceptable business names (i.e.,
standard semantics) exist, will make data element names more
understandable to a wider audience.

Third, the USDA standards, while providing continual use of existing
stand-alone data dictionaries and repositories, require that they be
reconciled and remain consistent with the proposed concept and
development of an overall Agriculture Data Repository System (ADRS).
This data repository may eventually take the form of an "active" data
repository that will automatically enforce the use of standard data
elements, and thereby data integration at defined levels (e.g.,
application, organization, mission area, USDA).

Fourth, review and approval procedures for standard data elements are
being streamlined and will be provided via a separate manual. These
procedures are being designed specifically for USDA's new agency
structure and evolution to mission (i.e., functional) areas, as well as
for integration of data requirements within these areas and across
USDA.
These procedures will provide for the definition and review of data at
the highest level of USDA's interest. Thus, data of interest solely
within one mission area or agency are defined and reviewed at that
level. After the data element definition and review process, data
elements will be coordinated across all agencies and mission areas
prior
to approval as USDA standard data elements.

Finally, these data element standards are designed to be implemented
from an evolutionary perspective; that is, data elements will be
standardized as they are created. The threshold at which these
procedures must be used can be varied and defined by USDA policy (e.g.,
new development efforts, functional process improvement, reprogramming
of over 30% of an application). It should also be noted that these
procedures can be employed in situations other than those referenced
whenever it makes sense to do so. CHAPTER 1

GENERAL INFORMATION

1   PURPOSE

This manual promulgates standards for data element standardization
and management necessary to support the Department of Agriculture
(USDA) data administration policies. Key policy requirements
include:

a   Implementing data administration aggressively in ways that
     provide concise, unambiguous, and easily accessible data
     across USDA, and that minimize the cost and time required to
     transform, translate, or research differently named or
     described data that actually represent identical data
     requirements.

b   Standardizing and registering data elements to meet the
     requirements for data sharing and interoperability among
     information systems throughout USDA.

c   Promoting standardization of data elements in USDA consistent
     with requirements for sharing data among agencies, other
     government departments, state agencies, customers, suppliers,
     and other nations under treaty or international agreement.

d   Using applicable federal, national, and international
     standards and common industry practices before creating new
     USDA standards.

2   INTRODUCTION

Information and data are principal USDA resources which, like
other
organizational resources, must be managed effectively. According
to the American National Standards Institute (ANSI), data is a
representation of facts, concepts, orinstructions in a formalized
manner suitable for communication, interpretation, or processing
by
human or automated means. Information is the meaning that is
currently assigned to data by means of the conventions applied to
that data.

The data required by an organization to perform its mission can be
graphically represented in data models. Data models contain data
entities, their attributes, and relationships between data
entities. Data entities are things or concepts important to
mission accomplishment for which information must be collected and
maintained (e.g., person, organizational unit). The data entity
attributes are items of data that describe the entity (e.g., name,
address, social security number of a person). The relationships
between data entities describe the business rules that govern the
association between data entities (e.g., a customer may have
multiple addresses). Definitions of relevant terms are contained
in Appendix A.

Data elements are the logical representation of data entity
attributes and provide the basis for physical implementation of

application data elements or database table columns in information
systems. We can discover and design data elements through data
modeling, which reflect functional policy and represent USDA's
data
requirements. Data elements can also be discovered through
analysis of functional information requirements. A data entity
attribute becomes a standard data element through application of
the procedures in this manual.

Use of standard data elements will enhance interoperability among
USDA information systems; facilitate data sharing; reduce data
handling costs; and lead to more accurate, consistent, and timely
data. The data element standards presented in Chapters 3 and 4 of
this manual provide the discipline necessary to design data
elements to maximize data sharing opportunities throughout USDA
and
to enforce data standards through the use of automated data
administration tools. The data element standards comprise three
major phases: preliminary analysis and design efforts (described
in Chapter 3), further efforts leading to formal definition of the
data element (discussed in Chapter 4), and a
review/coordination/approval process that will be presented in a
separate manual.

3   SCOPE

The scope of USDA data element standardization includes the data
requirements for all USDA Mission Areas. USDA Mission Areas are
currently defined as: 1) Farm and International Trade Services; 2)
Rural Economic and Community Development; 3) Food, Nutrition and
Consumer Service; 4) Natural Resources and Environment; 5)
Marketing and Inspection Services; and 6) Research and Economics;
and Administration.

4   OBJECTIVES

The primary objectives of data element standardization are
to:

a   Promote interoperability among separate agencies and
     mission areas;

b   Promote data sharing;

c   Control data redundancy;

d   Minimize data handling costs;

e   Improve data integrity; and

f   Reduce cost and time to develop, implement, and
     maintain information systems.

5   APPROACH

The basic approach to data element standardization outlined in this manual is to:

a   Use approved standard data elements whenever possible.

b   Have proposed standard data elements reviewed by Agency Data Administrators and/or Mission Area Data Managers. (This will ensure standard data elements are reused whenever possible.)

c   Name data elements using an agreed-upon structure.

d   Allow the proposed Agriculture Data Repository System to be used to automatically search for definitions of standard data elements rather than relying on rigorous naming conventions to serve as the search key.

6   STRATEGY

This manual includes the criteria and rules for standardization of data elements and associated semantics, format, and value sets throughout the USDA. It is the intent of USDA Data Administration, through this standard, to:

a   Develop standard data elements that satisfy USDA and agency mission needs and that support operational functional activities requiring the collection, storage, and exchange of data.

b   Develop standard data elements through data modeling efforts and the development and refinement of a department wide data model.

c   Develop an awareness of the value of managing USDA's data resources.

d   Provide guidance for the uniform description and representation of data.

e   Provide a single management standard that;

     (1) Implements USDA's data element standardization program.

     (2) Fosters full participation of mission area and

f   Aggressively migrate to a managed data environment required by USDA regulation while preserving, to the extent possible, current investments in data.

g   Support the development of common data requirements and

formats to eliminate data definition redundancies and discrepancies.

h   Minimize the cost and time expended in transforming, translating, or researching the meaning of related data elements. These are not limited to, but may include, differently named but otherwise identical data elements (synonyms) or similarly named data elements with differences in definition or values (homonyms).

i   Improve the integrity and usage of data through data structuring rules and standards presented in this manual and by coordinating data element definitions among USDA agencies and Department-level organizations.

j   Document standard data, their definitions, and other attributes in a central data repository such as Agricultural Data Repository System (ADRS) as soon as that resource is acquired. The reference may be copied, but it will be centrally maintained using the ADRS. The ADRS will be expanded over time to support all components of USDA's Data Administration Program. Until that time, it is understood that it may be necessary for the separate agencies to maintain their own data dictionary/repository. Once the ADRS becomes operational, the dictionary/repository software systems will be coordinated to support metadata exchange. Agencies may continue to use dictionary facilities including those commonly provided with Computer-Aided Software Engineering (CASE) tools; but in all cases, once the ADRS is operational agencies will be responsible for ensuring the information in their data dictionaries is consistent with the information in the ADRS.

CHAPTER 2

DATA ELEMENT CONCEPTS

1   PURPOSE

The concepts discussed in this chapter are fundamental to the discovery, design, and definition of standard data elements.

As discussed in Chapter 1, the data required by an organization to perform its mission can be represented graphically in data models.

Data models contain data entities, their attributes, and relationships between data entities. Data elements are discovered and designed through data modeling and analysis of data entity attributes. Data modeling also ensures that data element definitions reflect functional policy and represent USDA's data

requirements. A data entity attribute becomes a standard data
element as a result of standardizing and documenting selected
characteristics that describe data semantics, data format, and
sets
of valid data values.

2   CONCEPTS

The following concepts should be well understood before attempting
to apply the rules for data element design, definition, and naming
described in the next chapters.

a   Data Standardization

(1)   Data semantics are the rules for meaning and usage of a
       particular string or value of data. Data can be
       effectively shared only to the extent that semantics are
       understood in a precise and unambiguous way by all who
       share the data.

(2)   Data standardization can be successful only with
       regulation and standardization of: 1) the information
       that describes the semantics of data; 2) the set of
valid
       values of data whenever the value set is to be purposely
       constrained; and 3) the representation of data in its
       most simple and useful form. These three sets of
       information about data, when specified more precisely,
       are commonly referred to as METADATA.

(3)   The approach to achieving data standardization begins
       with the premise that metadata are a very important set
       of corporate-level data that need to be defined,
created,
       reported, updated, and maintained in a non-redundant,
       efficient manner.

       This approach requires the existence of a highly
       structured database of metadata frequently called a DATA
       DICTIONARY or REPOSITORY.

b   Data Element

(1)   Data elements are named identifiers of each of the
       entities and their attributes that are represented in a
       database. They represent discrete instances of business
       facts. Examples of data elements include part number,
       social security number, unit identification code,
       individual last name, country code, and individual home
       address. By convention, data elements are the basic
       units of data that comprise a record or table of
       organized data suitable for processing. These basic
       units of data may, in fact, be compound data (e.g.,
       address), atomic data (e.g., state code or county code),
       primitive data (e.g., individual birth date) or derived

data (e.g., age or remaining balance). A data element must have a description, name, data type, structural composition, domain, and integrity constraints.

(2) Model the data required to support the mission. Data elements are derived from logical entities and their attributes identified in data models. Each data element represents an attribute of an entity in a data model.

(3) Data elements may also be derived through the analysis and re-engineering of information and data managed in existing files and databases.

c   Standard Data Element

(1) A standard data element must convey a single fact of importance to the mission, and the business rules (semantics) associated with the fact.

(2) A data element is standard when it has been designed, named, and its critical attributes documented and approved.

d   Metadata

(1) Data elements have definitive characteristics or
tributes
that collectively identify, quantify, and qualify facts about the data element itself. These facts include the description, authority, form, type, name(s), integrity rules, and value sets or domains. Metadata are facts (i.e., data) about data. It is the metadata for data elements that are standardized and maintained in the repository. Metadata for application data elements will be registered in the ADRS when acquired.

(2) Reuse of standard metadata results in the implementation of standard data elements in operational information systems.

(3) A listing and description of ADRS standard metadata are provided in Appendix B.

e   Data Element Domain

A DOMAIN is the set of valid data values approved for a data element. Data element domains represent approved business policy and are controlled by data element stewards. A domain can be either specific or general.

(1) Specific Domain. A specific domain has a finite definition and an enumerable set of data representations as shown in the example below. A specific domain is defined by naming the acceptable values allowed in a prescribed set of data representations. In some cases, the set of acceptable values for a specific domain may

change over time, e.g., the set of office identifiers
for

              organizations within USDA.

              Example:

              Data element name: INDIVIDUAL EYE COLOR

              Data element domain values:
                                  Brown
                                  Gray
                                  Green
                                  Hazel
                                  Blue

      (2) General Domain.A general domain has a broad definition
          and a large (possibly infinite) set of acceptable
values.
          A general domain is described by establishing a range or
          set of possible values, but does not require listing all
          the possible values. Certain values or characters may
be
          restricted. An example of a general domain is shown
          below.

          Example:

          Data element Name: INDIVIDUAL PULSE RATE
                      Range: A general domain composed
                                of numeric values to one
                                decimal place within the
                                range 0.0 through 400.0.

CHAPTER 3

DATA ELEMENT DESIGN AND N   AMING RULES

1   PURPOSE

    The standards discussed in this chapter set guidance for designing
    and naming data elements that agencies and Department-level
    organizations within USDA will use. Implementation of these
    standards is covered in Chapter 4 and in a separate manual.

2   DATA ELEMENT DESIGN RULES

    The quality of the data element designs are key to sound data
    structures. Unless proper consideration is given to the design of
    individual data elements, the quality of the entire data structure
    will be affected. All data must be designed with sharing and
reuse
    in mind. Where feasible, data elements should be based upon the
    entities and attributes identified in approved enterprise data
    models, to ensure maximum sharing and compatibility of data.

Critical components of early design include identification of the purpose, reference, domain, description, and structure. The following rules are important to the design of a data element:

a   Where feasible, data elements should be designed to represent the attributes (characteristics) of data entities identified in data models. A model-driven approach to data element standardization provides a logical basis for, and lends integrity to, what is being standardized.

b   Data element attributes must be designed according to functional requirements and logical, not physical, characteristics. Physical characteristics include any reference to technology (hardware or software), physical location (databases, records, files, or tables), organization (section, branch or division), or application (systems, applications, or programs).

c   Data elements must be designed according to the characteristic
     of the object or concept they describe (i.e., what it is) rather than how, where and when the data element is used or who uses it.

d   Data elements must be designed so that they have singularity of meaning. A data element should reflect a single concept to
     promote sharing, reuse, and data independence from applications using the data element.

e   Data element descriptions should not contain conjunctions or phrases indicating multiple concepts, ambiguity of description, or process orientation. Descriptions should be void of technical jargon that may be unfamiliar to the reader.
     All acronyms and abbreviations must be defined within each description. The description must be more than just a reiteration of other characteristics of the data element (e.g., metadata such as data element name). It is good practice, when practical, to limit a data element description to one sentence. Data element descriptions, like definitions of words, should not be circular or contain references to the item being described.

f   Data element domain values must be homogeneous, mutually exclusive, and, for a specific domain, totally exhaustive and completely defined. As an example, values for the data element Employee Marital Status Code could be defined as married, single, divorced, widowed, legally separated, and annulled. These values are not mutually exclusive, however; any of the last four codes could apply to a person who is also
     single. The values actually relate to two distinct concepts (current marital status and marital history), and two separate
     data elements should really be defined. Domain values that

exceed a limited set do not need to be listed and defined
individually; the authoritative source of the values can be
referred to in the attributes (metadata) of the data element.
When referring to a source document, list examples of the
values along with the authoritative source.

g   Data element codes should avoid values that may be confused
with other values in the same domain (e.g., 0/O, 1/I, 2/Z,
5/S).

h   Data element values should avoid the use of embedded meaning
within all or part of the code. Mnemonic values may be used
only where the domain set is sufficiently limited and static
to allow such usage (e.g., the data element Sex Code with
values of M for male and F for female).

i   Data element purposes should not overlap or be redundant with
the purpose or use of other data elements (e.g., individual
birth date, current date, and current age). Current Age can
be computed from Individual Birth Date and Current Date.
Each

represents basic concepts applicable to many uses.

3   DATA ELEMENT NAMING RULES

The set of rules for naming data elements will make it easier to
determine if a data requirement is already being met within USDA
or
if it is a new requirement that needs to be fully defined and the
data collected and distributed as necessary. The development of a
standard data element is an iterative process as the data element
design and name are often modified during the process. Within
USDA,
the data element name will be known as the Business Name. Design
the data element unambiguously first and then apply the following
rules for naming:

a   The foundation for an entity, attribute, and data element is
the Business Name. Where there is wide Departmental
acceptance, and clear and unambiguous understanding and
agreement on the description and domain of a data element,
the
name by which it is currently and consistently referred may
be
designated the Business Name. It must be understood, however,
that this situation will prove to be the rare exception. If
department wide acceptance of the existing data
element/business name cannot be confirmed, the following data
element naming standard will be followed.

b   Business names should consist of the minimum number of words
that adequately identify the data element. In general, the
greater the number of words used in the name, the more narrow
or restrictive the data element becomes. Do not use the name
to record semantic information that more correctly belongs in
the attributes (metadata) of the data element.

c   Syntax Rules

When a business name is to be formatted, apply the following syntax:

(1)   Business Name will be a maximum of 50 characters, constructed of upper and lower case letters and being capitalized as in a title. This differentiates business names from technical programming names and orients them more to the end user (e.g., individual phone number).

(2)   Abbreviations and acronyms must be avoided inconstructing a business name. Exceptions to this rule include universally accepted abbreviations (e.g., FAX for Facsimile) and acronyms that are officially approved for usage. The ADRS will maintain lists of approved acronyms and abbreviations. Agencies are encouraged to develop and use consistent abbreviations that are managed within their own agencies. The objective is to comprise a consolidated list of these abbreviations for usage throughout the department at a future time.

(3)   Only alphabetic characters (A-Z, a-z), spaces ( ), and in rare cases, numbers (0-9) and hyphens (-) are permitted.

(4)   Each component of the business name is separated by a space. In exceptional cases, hyphens may be used to connect multiple words. (e.g., Loan Write- off Amount).

(5)   Plural words are not permitted (e.g., dependents birth dates).

(6)   Possessive forms of words are not permitted (e.g., individual's birth date).

(7)   Prepositions (e.g., at, by, for, from, in, of, to) are not permitted unless required for clarity (e.g., Power of Attorney).

(8)   Articles (e.g., a, an, the) are not permitted.

(9)   Conjunctions (e.g., and, or, but) are not permitted.

(10) Verbs are not permitted.

d   Semantic Rules

(1)   Business names must be clear, accurate, and self-explanatory.

(2)   Business names must not contain the names of organizations, computer or information systems,

directives, forms, rows or columns of screens, or
reports.

(3) Business names must not express multiple concepts,
either

implicitly or explicitly. For example, data elements
simply named 'Text' or 'Number' are inappropriate
because

they do not address a principal item.

   e   Structuring Conventions

(1) Business name consists of a required prime word,
modifiers and normally a class word. The prime word
will

be first unless a significant modifier is used. Each
business name will have one and only one description and
domain. The business name will be unique among all
other

data elements. If a business name appears similar to an
existing standard business name, but it has a different
domain or description, a separate business name must be
defined. The format for a business name is:

MODIFIER PRIME WORD   MODIFIER(S)   CLASS WORD
Optional Required   Optional   Required

(2) Business names are structured for consistent
identification (e.g., Birth Date is used in lieu of Date
of Birth). Rare exceptions to this standard may be
permitted for terms of wide acceptance with clear and
unambiguous understanding, and where the standard would
force an unnatural name (e.g., Power of Attorney Code
rather than Attorney Power Text).

(3) The prime word identifies the object to which the
business name refers. For example, an organization may
need to maintain information about offices, so an entity
of Office would exist. The prime word for this entity
would also be called Office (e.g., Office entity may
contain the business names Office Identification, Office
Name, and Office Telephone Number, and Office Telephone
Use Code). The prime word is frequently the noun given
to

an entity in a data model. Prime words are to be placed
at the beginning of the name while maintaining
readability. Prime words in some business names may be
used as modifiers in other business names.

(4) Modifiers further refine the prime word or classword.

(5) Modifiers may be used with class words and prime words
to

define business names more accurately; however, the use
of modifiers should be restricted to the minimum
essential for clear meaning. A common misuse of naming
conventions is to overload the business name with

semantic information that should be captured as metadata of the data element. The use of multiple modifiers in a name for purposes of uniquely identifying the data element may be an indication that multiple concepts are represented. The result may be limited potential for reuse and sharing.

(6)   A class word is used to designate the general category of

data described by a business name. A class word is a noun that is, in effect, a shorthand notation for the general domain of the data element. Examples of class words are Amount, Code, Name, and Date. Appendix C contains a list of authorized class words and their standard abbreviations.

(7)   A class word is the last word of a business name and follows the prime word and its modifiers. Class words are mandatory except when their usage would cause awkward

redundancy (e.g., Forestland Acreage instead of Forestland Acreage Quantity, and Crop Yield instead of Crop Yield Quantity). This exception pertains only to those data elements where the class is extremely obvious to discern by simply looking at the business name; if the

implied class is not obvious--use a class word in the business name.

(8)   Class words are reserved words and should not be used as modifiers except in extenuating circumstances. If a class word is used as a modifier it must be fully justified.

f   Standard Business Name

When a proposed data element has been defined according to the

standard in Chapter 4, named according to the procedures in this chapter, and reviewed, coordinated, and approved according to the procedures that will be provided in a separate manual, it becomes a USDA standard data element, and its name becomes a standard business name. Standard business names will normally be limited to 50 characters in length. In

special cases where 50 characters are insufficient a waiver may be requested to exceed the 50 character limitation. Procedures for requesting a waiver will be provided via a separate manual.

CHAPTER 4

DATA ELEMENT DEFINITION PROCEDURE

1   PURPOSE

     The procedures presented in this chapter have been established to
     support the efficient development of USDA standard data elements.

     After undergoing these procedures, data elements will be ready to
     enter the USDA data element standardization approval process.

2   INTRODUCTION

     As described earlier, a data element is standard when it has been
     designed and named according to the rules specified in Chapter 3,
     its critical attributes have been defined and documented according
     to the standard outlined below, and it has been approved according
     to the procedures that will be provided in a separate manual.

3   APPROACH

     There are four basic approaches for discovering information
     requirements, translating information requirements to data
     requirements, and specifying those data requirements as standard
     data elements:

     a   Top-down enterprise wide data modeling efforts based on
          information engineering methodologies

     b   Bottom-up data re-engineering activities associated with
          systems re-engineering/modernization projects

     c   New data requirements resulting from legislative action or
          policy change

     d   Some combination of the above

     Regardless of the approach, data modeling techniques should be
used
     to accomplish the following processes and resolve semantic
     ambiguities:

     a   Data analysis to discover information requirements

     b   Data design to translate information requirements to data
          requirements

     c   Data definition to specify standard data elements

     The data element standards in this chapter focus on the last step,
     data definition, to specify standard data elements. When data
     elements are defined within the general context of a data model,
     there is greater certainty of avoiding redundancy and duplication.

     Also, the integration of all data requirements into a USDA data
     model provides a comprehensive, department wide view that
     facilitates data sharing.

4   PROCEDURE

Regardless of who discovers a specific information requirement or
translates it into a general data requirement, the data
requirement
must be documented and approved using the following procedure (to
be specified as a standard data element) and made available for
reuse.

As discussed earlier, data elements have definitive
characteristics
or attributes that collectively identify, quantify, and qualify
facts about the data element itself. The standard facts or
METADATA that must be documented to specify a data requirement as
a
standard data element are the following:

            Data element description
            Data element name
            Data type
            Data element domain
            Data element precision
            Data element length
            Data element alias(es)
            Data element integrity rules
            Data element unit of measure reference
            Data element sensitivity
            Data steward
            Data element authority
            Data element authoritative source
            Data element model reference
            Data element system reference
            Data element Mission Area reference
            Data element status

These metadata are explained in Sections b and g below, and in
Appendix B.

a   Gather Necessary Documentation. Collect any available
     documentation that may provide information for or assist in
     completing the USDA standard definition of the data
element(s)
     proposed for standardization. USDA data models and the ADRS
     will be primary sources of information for developing a USDA
     standard data element. Additional references and resources
     include the following:

            Functional information requirements
            Mission Area and agency data models and process
            models
            Agency data dictionaries
            Federal Information Processing Standards (FIPS)
            Dictionary of Business Terms

USDA directives
                   System documentation

      b   Develop Preliminary Definition. Document preliminary data
            element attributes. Following the rules in Chapter 3,
develop
            the data element design and name. The preliminary definition
            attributes are:

                   Data element description
                   Data element name
                   Data type
                   Data element domain
                   Data element precision
                   Data element length
                   Data element alias(es)

            (1)   Complete Data Element Description. Describe the data.

            (2)   Complete Data Element Name. When there is wide
                  acceptance across the USDA and clear and unambiguous
                  understanding of a business name, the data element
should
                  assume the same name. Lacking a clear and unambiguous
                  business name, develop a data element name for the data
                  element by applying the naming rules in Chapter 3.

            (3)   Complete Data Type. Potential data types include
                  numeric, alphabetic, alphanumeric, string, date,
                  currency, time, graphic, sound, or video. Numeric data
                  may be integer, decimal, or floating point. String data
                  may be fixed or variable in character length. Note that
                  some metadata attributes may not apply to some data
                  types; for example, precision and length do not apply to
                  graphic, sound, or video data types.

            (4)   Complete Data Element Domain. Develop text to describe
                  the overall meaning or general characteristics of the
                  data element domain. For a data element with a specific
                  domain and the class word CODE or NAME, enter each value
                  (data element domain value identifier) and a description
                  for each value (data element domain value description
                  text). For the data element Individual Eye Color in
                  Section 2e (1) of Chapter 2, the list of values would be
                  brown, gray, green, hazel, and blue. In this case, the
                  values are self-explanatory; if they were not, a
                  description of each would be needed. If the specific
                  domain of a data element is excessively large, a source
                  document should be referenced, and a few examples should
                  be identified, instead of listing the entire domain.

                  For quantitative general domains, enter the allowable
                  range of the domain values (i.e., data element low-range
                  identifier and data element high-range identifier). For
                  the data element Individual Pulse Rate in Section 2e (2)
                  of Chapter 2, the low range identifier would be 0.0 and
                  the high range identifier would be 400.0.

(5)   Complete Data Element Precision. For numerical data
         elements, indicate the number of decimal places
required.

(6)   Complete Data Element Length. Define data element
length
         as the maximum number of characters or digits required
to
         accommodate the longest instance of the data element.

(7)   Complete Data Element Aliases. List all known synonyms
         for the data element that are in current usage as
         aliases. Attributes within commercial off-the-shelf
         (COTS) software will be linked to the USDA attribute
name
         through aliases. This is an interim step until USDA
         creates its own repository. At that time, linkages
         between USDA standard data elements and other elements
         such as those in COTS packages will be described
         explicitly in the repository.

c   Research Existing Standard Data Elements. Examine USDA data
     models and existing data elements and their attributes to
     determine whether the data requirement is already satisfied
by
     an existing standard data element. Also examine standard
data
     elements beyond USDA (for example, in FIPS and ANSI
     publications). Compare the descriptions, names, types,
     domains, precisions, and lengths.

d   Identify Potential Matches. Identify data elements whose
     required attributes match or approximate the intended
     attribute values metadata of the data element under
     development.

e   Resolve Potential Matches. From those identified as
potential
     matches that cannot be resolved, analyze each data element
     having a domain that either matches, includes all of the
     values of (superset), or approximates the intended domain of
     the data element under development. If more than one such
     element is identified, determine which best represents the
     data element under development.

f   If Matched, Use Existing Standard Data Element. Select the
     data element from the previous step having mandatory
attribute
     values nearest those of the data element under development.
     This procedure should result in no more than one candidate
     standard data element. Prepare and submit the additional
     attributes required to register a new application of an
     existing data element.

g   If Unmatched, Complete Definition. If no existing data
     element will fulfill the requirements of the data element

under development, continue development of the new data element. The remaining data element definition attributes metadata are:

        Data element integrity rules
        Data element unit of measure reference
        Data element sensitivity
        Data steward
        Data element authority
        Data element authoritative source
        Data element model reference
        Data element system reference
        Data element Mission Area reference
        Data element status

(1)   Complete Data Element Integrity Rules. Define data element integrity rules by documenting business rules that constrain instances of the data element that depend on other data elements and specify the nature of the constraints. For example, a business rule might require that customer accounts be permitted only for those customers whose addresses were known. In this case, an integrity rule could require that every instance of the data element Customer Account Identifier must correspond to a customer whose address is known.

(2)   Complete Data Element Unit of Measure Reference. Identify the unit of measure that applies to this data element (e.g., square feet, acres, cwt).

(3)   Complete Data Element Sensitivity. Identify the sensitivity level for the data element. Examples of possible sensitivity levels are as follows:

        P   Subject to the Privacy Act
        S   Sensitive
        N   Non-sensitive
        U   Undefined, pending further analysis

(4)   Complete Data Steward. Identify by organization and title the person or group responsible for approving the definition of a new data element and changes to the definition of an existing data element.

(5)   Complete Data Element Authority. Identify any document that establishes the data element and precludes change.

(6)   Complete Data Element Authoritative Source. Identify the

        most reliable source for this data element usually, the point of entry.

(7)   Complete Data Element Model Reference. Identify models that use this data element.

(8)   Complete Data Element System Reference. Identify systems

that use this data element.

(9) Complete Data Element Mission Area Reference. Identify the Mission Areas that use this data element.

(10) Document Data Element Status. For data elements progressing through the standardization process, this status indicator will describe the progress (e.g. submitted, reviewed, or approved).

h   Submit for Approval. Submit candidate standard data element for approval according to procedures that will be provided in a separate manual. Data element status is updated during the approval process.

Signed by:

JOHN L. OKAY

APPENDIX A

DEFINITIONS

The following selected definitions will aid in the understanding of this
manual. Many of the definitions were taken from the ANSI publication American National Standard for Information Systems-- Dictionary of Information Systems, FIPS PUB 11-3, ANSI X3.172- 1990. The ANSI definitions frequently include references to other definitions in the ANSI publication. When such references were relevant to the subject of data element standardization, they were included here, even though the referenced term is not always included here. Definitions preceded by the symbol (ISO) were reprinted (in the ANSI publication) from the International Organization for Standardization's (ISO) Vocabulary--Information Processing, ISO 2382. Definitions preceded by the symbol (USDA DR 3400-4) were taken from the USDA Departmental Regulation 3400- 4, Departmental Data Administration Program, dated August 2, 1994.

Access Name. In an information resource dictionary, the name by which an entity is known to the user interfaces. It is the combination of an assigned access name and version identifier that together serve as the primary identifier of each entity. 1

Accuracy. (1) (ISO) A quality of that which is free of error. (2) (ISO) A qualitative assessment of freedom from error, with a high assessment corresponding to a small error.1

ADRS. Agriculture Data Repository System. The ADRS does not exist at this time, however, the Office of Information Resources Management intends to acquire a repository for department wide use.

Agency, Staff Office, Special Interagency Project Data Administrator.
(USDA DR 3400-4) The official who develops and manages the respective
Data Administration Program, within the guidelines established in
Departmental Regulation 3400-4.

Agriculture Data Repository System. A software tool used to control,
describe, protect, document, and facilitate the use of USDA's
information resources. (ADAPTED FROM IRDS DESCRIPTION, FIPS PUB 156)

Alias. (1) An alternate label; for example, a label and one or more
aliases may be used to refer to the same data element or point in a
computer program. (2) Synonym for alternate name. 1

Alternate Name. (1) An alternate label; for example, a label and one
or
more alternate names may be used to refer to the same data element or
point in a computer program. (2) In an Information resource
dictionary,
any name by which an entity is known and that may be associated with
more than one entity. (3) Synonymous with alias. 1

Application Data Element. A data element used in an automated
information system (i.e., in an application system). Registration of
application data elements in the USDA Data Repository System (ADRS)
identifies the various uses of data elements throughout the USDA.
Application data elements registered in the ADRS for specific systems
may have characteristics that are different than those of the standard
data element.

Atomic Datum. An instance of an atomic fact.

Atomic Fact. The smallest component of information. A single concept.

Attribute. (1) (ISO) A property or characteristic of one or more
entities; for example, color, weight, sex. (2) A property inherent in
an entity or associated with that entity for database purposes. (3)
See
data attribute. 1

CASE Tools. Computer aided software engineering tools.

Computer Aided Software Engineering Tools. Software tools used to
automate software development activities.

Conceptual Schema. A representation of Data Administration's unified
view of the total information resources of the organization. (NIST
Special Publication 500-173, Guide to Data Administration)

Data. (1) (ISO) A representation of facts, concepts, or instructions
in
a formalized manner suitable for communication, interpretation, or
processing by humans or by automatic means. (2) Any representations
such as characters or analog quantities to which meaning is or might be
assigned. 1

Data Administration. The function of controlling the acquisition,

analysis, storage, retrieval, and distribution of data. Synonymous
with
data management. 1

Data Administrator. The person who defines, organizes, manages,
controls, and protects data. 1

Data Attribute. A characteristic of a data element such as length,
value, or method of representation. 1

Database. (1) A collection of interrelated data, often with controlled
redundancy, organized according to a schema to serve one or more
applications; the data are stored so that they can be used by different
programs without concern for the data structure or organization. A
common approach is used to add new data and to modify and retrieve
existing data. (2) See archival database, distributed database. 1

Database Administrator. The person who defines, organizes, manages,
controls, and protects a database. 1

Database Management System (DBMS). (1) An integrated set of computer
programs that collectively provide all of the capabilities required for
centralized management, organization, and control of access to a
database that is shared by many users. (2) A computer-based system
used
to establish, make available, and maintain the integrity of a database,
that may be invoked by nonprogrammers or by application programs to
define, create, revise, retire, interrogate, and process transactions;
and to update, back up, recover, validate, secure, and monitor the
database. 1

Data Code. (1) A structured set of characters that may be used to
represent data items; for example, the use of 01, 02, ...12 to
represent
the months of the year. (2) A representation of data that is produced
by a code. (3) Contrast with abbreviation. 1

Data Dictionary. (1) A database used for data that refers to the use
and structure of other data; that is, a database for the storage of
metadata. (2) An inventory that describes, defines, and lists all of
the data elements that are stored in a database. (3) A subset of a
data
dictionary/directory that provides definitions for each data element.
(4) Loosely, a data dictionary system. (5) Loosely, a data directory.
(6) Loosely, a data directory system. (7) Loosely, a data
dictionary/directory. (8) Loosely, a data dictionary/directory system.
(9) Synonymous with data element dictionary. 1

Data Dictionary System. A computer software system that maintains and
manages a data dictionary. 1

Data Element. (1) (ISO) A named unit of data that, in some contexts,
is
considered indivisible and in other contexts may consist of data items.
(2) A named identifier of each of the entities and their attributes
that
are represented in a database. (3) See derived data element. 1

Data Element Definition. A data element is defined by its association
with a specific entity and by all its attributes.

Data Element Description. A text statement explaining what a data
element represents and distinguishing it from other similar data
elements.

Data Element Design. A preliminary step in the data element
standardization process that produces a data element description and a
data element name.

Data Element Standardization. The process of designing and defining a
data element and taking it through the review/coordination/approval
process.

Data Independence. (1) The organization, storage, or retrieval of data
or programs in a way not dependent upon the manner specified in user
programs. (2) The property of a database management system that
insulates application programs from the complexities of the data
structure. 1

Data Integrity. (1) The state that exists when data is handled as
intended and is not exposed to accidental or malicious modification,
destruction, or disclosure. (2) The preservation of data for its
intended use. (3) See also data security, system integrity. 1

Data Item. A named component of a data element; usually the smallest
component. 1

Data Model. (1) In a database, the user's logical view of the data in
contrast to the physically stored data, or storage structure. (2) A
description of the organization of data in a manner that reflects the
information structure of an enterprise. (3) See entity-relationship
data model. 1

Data Quality. (ISO) The correctness, timeliness, accuracy,
completeness, relevance, and accessibility that make data appropriate
for use. 1

Data Resource. Any data created manually or by automatic means, used
by
a system or enterprise to represent its information. 1

Data Resource Management. The responsibility for planning, organizing,
and controlling data resources consistent with the overall goals and
objectives of an enterprise. See also information resource management.
1

Data Security. The protection of data from accidental or intentional
modification or destruction and from accidental or intentional
disclosure to unauthorized personnel. See also data integrity. 1

Data Steward. The person or group that manages the development,
approval, and use of data within a specified mission area, ensuring
that
the data can be used to satisfy data requirements throughout the USDA.

Data Type. (1) (ISO) A set of values and a set of allowable operations on those values. (2) The characteristics and attributes of data; for example, length, precision, alphanumeric representation. 1

Data Validation. (1) (ISO) A process used to determine if data are inaccurate, incomplete, or unreasonable; the process may include format checks, completeness checks, check key tests, reasonableness checks and limit checks. (2) The checking of data for correctness or compliance with applicable standards, rules, and conventions.

Data Value. An instance of a data item. Synonymous with value.1

Departmental Data Administrator. (USDA DR 3400-4) The official who manages the USDA Data Administration Program.

Derived Data. Data values that are derived from the values of other data by a specified algorithm. 1

Derived Data Element. A data element that has a domain identical to that of a specified general data element; for example, country of citizenship is derived from the element, countries of the world. 1

Designated System. A system selected according to USDA policy by OIRM or other appropriate authority for application of data element standards
as defined in this manual. Selection of a designated system will take into consideration such factors as the degree to which the system's data
are or will be shared or exchanged and the cost to apply the procedures.

Domain. (1) The set of possible data values of an attribute. (2) The set of permissible data values from which actual values are taken for a particular attribute or specific data element. (3) In a relational database, all of the permissible tuples for a given relation. (4) In distributed data processing, that part of a network at which data processing resources are under common control. 1

Edit. (ISO) To prepare data for a later operation. Editing may include
the rearrangement or the addition of data, the deletion of unwanted data, format control, code conversion, and the application of standard processes such as zero suppression. 1

Entity. (1) Anything, such as a person, place, process, object, concept, association, or event. (2) Anything about which information is
stored in a database. (3) In a conceptual schema language, any concrete
or abstract thing of interest, including associations among things. 1

Entity-Relationship Data Model. A data model based on the concept of entities and relationships among entities, and of the attributes of entities and relationships. 1

External Schema. A representation of the ways in which different groups
of users view the portions of the corporate information resource
relevant to them. (NIST Special Publication 500-173, Guide to Data
Administration)

Identifier. (1) (ISO) One or more characters used to identify or name a
data element and possibly to indicate certain properties of that data
element. (2) In programming languages, a token that names a data object
such as a variable, an array, a record, a subprogram, or a function. 1

Information. (1) (ISO) The meaning that is currently assigned to data
by means of the conventions applied to that data. (2) In a conceptual
schema language, any kind of knowledge about things, facts, or concepts
of a universe of discourse that is exchangeable among users. (3) See
formatted information, narrative information. 1

Information Resource Dictionary System (IRDS). (1) A computer software
system that provides facilities for recording, storing, and processing
descriptions of an organization's significant information and
information processing resources. (2) A computer software system that
maintains and manages an information resource dictionary. (3) In a
conceptual schema language, an information system that deals with the
information about a universe of discourse and that consists of another
information system, its environment, and its implementation in yet
another, not necessarily disjoint, information system. (4) Loosely,
synonymous with data dictionary/directory system. 1

Information Resource Dictionary System Extensibility. The capability to
create new functionality in an information resource dictionary system.
1

Internal Schema. A representation of the actual physical storage of
information resources, which may be stored on multiple, dispersed
databases. (NIST Special Publication 500-173, Guide to Data
Administration)

Interoperability. The capability of two or more systems to exchange and
use information. 1

Key. (1) (ISO) An identifier within a set of data elements. (2) One or
more characters, within a set of data, that contains information about
the set, including its identification. (3) In a record, a data element
whose value is unique for each occurrence of the record and is used to
identify or locate the record in a database management system. (4) On a
keyboard, a manually actuated mechanism that performs a specific
operation or causes the printing of a particular character. (5) See
access key, command key, database key, foreign key, function key,
privacy key, search key, sort key, typing key. 1

Logical. (1) Pertaining to content or meaning as opposed to location
or
actual implementation. (2) Pertaining to a view or description of data
that does not depend on the characteristics of the computer system or
of
the physical storage. (3) Contrast with physical. 1

Metadata. In database management systems, information about an
organization's information and data activities. 1

Precision. (1) (ISO) A measure of the ability to distinguish between
nearly equal values; for example, four-place numerals are less precise
than six-place numerals; nevertheless, a properly computed four-place
numeral may be more accurate than an improperly computed six-place
numeral. (2) The degree of discrimination with which a quantity is
stated; for example, a three-digit numeral discriminates among 1000
possibilities. (3) See double-precision, multiple-precision,
single-precision, triple-precision. 1

Primitive Data. Data values that are not derived from the values of
other data.

Quality Assurance. (1) (ISO) The planned systematic activities
necessary to ensure that a component, module, or system conforms to
established technical requirements. (2) All actions that are taken to
ensure that a development or organization delivers products that meet
performance requirements and adhere to standards and procedures. (3)
The policy, procedures, and systematic actions established in an
enterprise for the purpose of providing and maintaining some degree of
confidence in data integrity and accuracy throughout the life cycle of
the data, which includes input, update, manipulation, and output. 1

Relation. In a relational database, a named table that identifies the
set of occurrences of entities that have the same attributes. 1

Relational Model. (1) A data model whose pattern of organization is
based on a set of relations defined in the form of tables whose rows of
data items are ordered by the attributes of the associated data
elements. (2) A data model that provides for the expression of
relationships among data elements as formal mathematical relations. 1

Relationship. A special type of entity that is used to indicate a
dependency, an association, or a link that may be inherent between two
entities or among attributes of the same entity, and that is
represented
or recorded in a database. Synonymous with association. 1

Relationship Type. A specified class of relationships, each of which
is
associated in the same way with a member of one class of entities. See
also attribute type, entity type. 1

Standard Mission Area Database. A database of standardized data that
are shared and/or exchanged by multiple systems or databases within a
Mission Area.

Tuple. In a relation, the part that identifies an entity and its

attributes. A tuple is one row of a relation table. 1

Under/Assistant Secretary (Mission Area) Data Administrator. (USDA DR
3400-4) The official who coordinates the Data Administration Program
for
all agencies within the Mission Area, and/or all agencies in Special
Interagency Projects.

APPENDIX B

DATA ELEMENT ATTRIBUTES

Data Element Alias(es)1: An alternate label used to refer to the
data element.

Data Element Authoritative Source: The most reliable source for this
data element usually, the point of entry.

Data Element Authority: Any document that establishes the data element
and precludes change.

Data Element Description1: A text statement explaining what the data
element represents and distinguishing it from other similar data
elements.

Data Element Domain Description1: General description of the overall
meaning or general characteristics of the data element domain: CODE,
NAME, etc.

Data Element Domain Values1: For specific domains, the list of all
allowable values that can be assigned to the data element.

Data Element Domain Range1: For quantitative general domains,
the beginning and ending values of the allowable range.

Data Element Integrity Rules: List of business rules that constrain
instances of the data element that are dependent on other data elements
and a specific description of the nature of the constraints.

Data Element Length1: The standard number of characters or digits used
to accommodate the data element values.

Data Element Mission Area Reference: Mission areas that use this data
element.

Data Element Model Reference: Identity of models that use a data
element.

Data Element Name1: A label by which the data element is referenced.

Data Element Precision1: For numerical data elements, the number of

decimal places required.

Data Element System Reference: Identity of systems that use this data
element.

Data Element Unit of Measure Reference: Identity of the unit of
measure
that applies to this data element.

Data Element Sensitivity: Identification of the level of sensitivity
(relative to Privacy Act, etc.) of a data element.

Data Element Status: The data element status in terms of the USDA
standardization process.

Data Steward: Organization and title of the person or group
responsible
for approving the definition of a new data element and changes to the
definition of an existing data element.

Data Type 1: Basic category of data element values; i.e., numeric,
string, date or time. Numeric data may be integer, decimal, or
floating
point. String data may be fixed or variable length character.

APPENDIX C

AUTHORIZED CLASS WORDS

Class Word Name Definitions

Class Word	Abbreviation	Description and/ or Definition Structure
Amount	AM	A monetary value. (Includes average, balance deviation, factor, index, level, mean, mode, scale, and yield.)
		The standard data element definition should begin: "The (modifiers) amount of"
Angle	AN	The rotation measurement between two lines and/or planes diverging from a common point and/or line. (Includes azimuth and heading.)
		The standard data element definition should begin: "The (modifiers) angle

		between (modifiers) for a"
Area	AR	The measurement of a surface expressed in unit squares (2-dimensional). The standard data element definition should begin: "The (modifiers) area of"
Code	CD	A combination of one or more numbers, letters, or special characters substituted for a specific meaning. Represents finite, predetermined values. (Must have a specific domain.) (Includes category and status.) The standard data element definition should begin: "The (modifiers) code that represents and/or denotes a"
Coordinate	CN	Designation of the location of a line or plane. (Includes latitude and longitude.) The standard data element definition should begin: "The coordinate identifying the (modifiers) location of"
Date	DT	The designation of a specific 24 hour period of time. The standard data element definition should begin: "The (modifiers) date of and/or when and/or on which a"
Dimension	DM	A measured linear distance (1-dimensional). (Includes altitude, depth, diameter, elevation, height, length, radius, vertex, and width.) The standard data element definition should begin: "The dimension (length, width, height, radius, or elevation, etc.) of and/or from"
Identifier	ID	A combination of one or more numbers, letters, or special characters that designate a specific object/entity, but that have no readily definable meaning. (Must have a general domain.) (Includes designator, key, number.) The standard data element definition should begin: "The (modifiers) identifier that represents"

Mass	MS	The measure of inertia of a body.
		The standard data element definition should begin: "The (modifiers) mass of"
Name	NM	A designation of an object and/or entity expressed in a word or phrase.
		The standard data element definition should begin: "The name of"
Number	NR	A numeric value, often used in the business world as a synonym for Identifier. (e.g. Social Security Number, Phone Number, etc.)
Quantity	QY	A nonmonetary numeric value. (Includes average balance, count, deviation, factor, index, level, mean, median, mode, and scale).
		The standard data element definition should begin: "The (modifiers) quantity of"
Rate	RT	A quantity or degree of something in relation to units of something else (e.g., miles per gallon). (Includes acceleration, density, factor, flow, force, frequency, humidity, impedance, inductance, intensity, magnitude, moment, percent, power, pressure, resistance, scale, speed, tension, torque, velocity, viscosity, and voltage.)
		The standard data element definition should begin: "The rate of"
Temperature	TP	The measure of heat in an object or space.
		The standard data element definition should begin: "The temperature of"
Text	TX	An unformatted character string, generally in the form of words. (Includes category and comments.)
		The standard data element definition should begin: "The text of"
Time	TM	A designation of a specified chronological point within a period.
		The standard data element definition

```
                         should begin:  "The time of"

Volume        VL         Measurement of space occupied by a
                         three-dimensional figure as measured in
                         cubic units.

                         The standard data element definition
                         should begin:  "The volume of"

Weight        WT         The force with which an object is
                         attracted toward the earth and/or
                         another celestial body by gravitation.

                         The standard data element definition
                         should begin:  "The weight of"
```

Additional class words may be added later by proposing standard data elements for which none of the existing class words are appropriate.