CHAPTER 5 - DATA STANDARDS
Table of Contents

		Page
Introduction:	EPA DATA STANDARDS PROGRAM	5-1
Standard #1:	REPRESENTATION OF CALENDAR DATE	5-5
Standard #2:	(Reserved)	5-12
Standard #3:	(Reserved)	
Standard #4:	(Reserved)	
Standard #5:	(Reserved)	
Standard #6:	(Reserved)	
Standard #7:	(Reserved)	
Standard #8:	(Reserved)	
Standard #9:	(Reserved)	
Standard #10:	(Reserved)	

CHAPTER 5 - DATA STANDARDS

1. PURPOSE.

This policy establishes the EPA Data Standards Program. The purpose of this program is to provide consistent definition of data and to facilitate cross-media use of data. This policy sets forth Agency principles on data standards and assigns organizational responsibilities for implementing and administering common data standards.

2. SCOPE AND APPLICABILITY.

This policy applies to all Environmental Protection Agency (EPA) organizations and their employees. It also applies to the facilities and personnel of agents (including contractors and grantees) of EPA who design, develop, operate or maintain Agency information and information systems. This policy applies to automated and manual systems developed for programs or administrative purposes. The requirements of this policy apply to existing data elements as well as new data elements.

3. BACKGROUND.

a. Integration of information and data bases is difficult because program offices use disparate formats and names for similar data elements.

b. There is a need to make and support decisions based on standard information and data collected that cuts across the Agency's programs.

c. Specific programs, such as the Ground-water program, have an increasing need to share data from other programs, other agencies, States and local governments. This adds credence to the need for acceptable data standards to facilitate exchange of information.

d. Information technology has reached a point at which the sharing of data among automated systems is technically feasible.

e. The Agency has implemented standards for hardware and software that facilitate the sharing of data among programs.

f. To support effectively the use of common definitions of environmental data with State programs, EPA must have common definitions for data elements and an intra-agency capability to share data.

g. Organizations outside EPA have been establishing data standards which are accepted nationally or internationally. These pre-existing standards, such as Chemical Abstract Service (CAS) registry numbers, may serve as the best data standard for certain data elements.

h. There is a growing need for agreement on the definition of Agencywide parametric data entities such as "site" and "facility."

i. The Agency has a facilities inventory system that lists facilities regulated by the various programs in EPA. The inventory includes the different names and addresses for a single facility. This system will be a critical part of the Agency data standards effort.

j. At a minimum, there are six major areas which would benefit from the use of data standards: data used in more than one program, facilities and site data, geographic data, measurement data, health and environmental effects data and core office systems data.

4. AUTHORITIES.

a. 15 CFR, Part 6 Subtitle A, Standardization of Data Elements and Representations.

b. OMB Circular A-130, Management of Federal Information Resources.

5. POLICY.

It is EPA policy to create and maintain consistency in the form of data elements that have more than one application within the Agency. This consistency will permit the cross media approach necessary to achieve environmental results. The data standards will reflect the Agency's program priorities.

a. As required by OMB Circular A-130, EPA will adhere to Federal Information Processing Standards (FIPS), except where it can be demonstrated that the costs of using a standard exceed the benefits of the standard or will impede the Agency in accomplishing its mission.

b. All organizational components of EPA, their contractors or grantees will promote the full utilization of Federal and Agency data standards and representations in the design and development of information systems.

c. Data elements, codes and representations already in use by the Agency will be evaluated and adopted as Agency standards wherever practicable.

d. Data elements, codes and representations may be recommended for standardization by any program office within EPA.

e. Geographical information systems developed by the Agency must conform to an established set of appropriate data standards which permit the use of the system by all relevant programs and State agencies.

f. All relevant facilities or sites data must be stored in the Agency's facility or site inventory systems.

6. RESPONSIBILITIES.

a. The Office of Information Resources Management (OIRM) shall:

(1) Provide effective leadership in developing, promulgating and enforcing the policies of the Agency data standards program.

(2) Coordinate the evaluation and approval process of all data standards with the Assistant Administrators, Regional Administrators, Office Directors and Senior Information Resources Management Officers.

(3) Exercise final approval authority for the adoption of data standards. Grant waivers to the implementation of approved Agency data standards.

(4) Support other EPA data administration efforts, e.g., encourage cross reference files for non-standard information. Encourage the use of data element dictionaries.

(5) Propose and apply effectively data elements or representations for use by more than one organizational component of EPA as Agency standards.

(6) Publish and promulgate approved Agency standards in an EPA Data Standards Catalog.

b. Assistant Administrators, Associate Administrators, Regional Administrators, Laboratory Directors, Headquarters Staff Office Directors, General Counsel, Inspector General, and SIRMOs shall:

(1) Implement approved Agency data standards that are published under the provisions of this policy.

(2) Establish an organization-wide data standards work group which reviews and provides information and comments on proposed data standards.

(3) Propose the adoption of data standard for Agency use within the environmental community.

(4) Submit requests for waiver or deferments to the use of Agency data standards to OIRM.

7. DEFINITIONS.

a. "Data Element is a unit of information used to describe data characteristics and attributes, e.g., eyes - blue or BL.

b. "Data Standards are standards used generally, but not exclusively, for automated systems to ensue that one type of data is defined the same way in all systems. A similar definition means having the same name, the same number of maximum characters the same type and content of data in all systems where a specific data item appears.

c. "Information Technology" refers to the hardware and software used in connection with government information, regardless of the technology involved, whether computers telecommunications, micrographics or others.

d. "Media" means Water, Air,- Hazardous Waste and Pesticides and Toxic Substances program offices.

e. "System" is the organized set of procedures used to collect, transmit and disseminate information whether automated or manual.

8. PROCEDURES AND GUIDELINES.

Procedures and guidelines for the Agency data standards program will be issued under separate cover.

STANDARD #1

EPA DATA STANDARD FOR REPRESENTATION OF CALENDAR DATE

1. PURPOSE.

The purposes of this standard are to provide for the accurate use of dates during systems operations and to provide a consistent numeric representation of calendar date to facilitate interchange of date data among information systems.

2. SCOPE AND APPLICABILITY.

This standard is applicable wherever dates are included in an Agency information system and wherever data containing dates is interchanged among information systems. The scope of this standard for date includes the processing of date and date-dependent data, including but not limited to calculating, comparing, and sequencing. All software used for data entry, storage, computation, and retrieval of date in Agency, information systems must provide the intelligence to process date in compliance with this standard.

This standard does not assign any particular meaning or interpretation to any data element (e.g., a fiscal year) that uses date representations in accordance with this standard. Such meaning will be determined by the context of the application.

This standard applies to all new data processing applications, data management systems, databases, or manual information management systems developed by the Agency or its agents. The standard also applies to all existing Agency software applications where such applications perform computation of dates before and beyond December 31, 1999, when such operations would otherwise yield incorrect results because of the lack of a date format that includes Century year characters.

This standard is applicable to Agency contracts for acquisition of commercial off-the-shelf (COTS) software and development of custom software. Where relevant, Agency contracts will be written to require vendors to warrant performance by their product in processing of date and date-dependent data across the Year 2000 boundary.

This standard also is to be utilized by writers of Agency regulations and by persons commenting on, or developing proposed legislation that will result in information gathering, processing, or dissemination by EPA.

This date standard is *not* intended to prescribe the format for date on printed reports, correspondence, or other appearance of date when presented to users.

3. REFERENCES**.**

The following references were used as a basis for preparation of this standard:

a. American National Standard Representation for Calendar Date and Ordinal Date for Information Interchange, **X3.30-1985** (R1991), American National Standards Institute (ANSI).

b. Federal Information Processing Standards Publications **(FIPS PUB) 4-1**, Representation of Calendar Date and Ordinal Date for Information Interchange, 1988, January 27. This publication adopts ANSI X.30-1985 (R1991). Change number one of March 25, 1996, highly recommends that four-digit year time elements be used, and that two-digit year time elements should not be used for the purposes of any data interchange among U.S. Government agencies.

c. **ISO 8601**, Data elements and interchange formats -- Information interchange -- Representation of dates and times. International Organization for Standards, First Edition, 1988, June 15.

4. BACKGROUND

Many data collections present date with two digits for “year” in a “date” field, assuming that all dates pertain to the 20th century. Whenever a computer system or software application uses two digits to represent the year, a change of date from 1999 to 2000 and beyond may produce inaccurate date computations. A two-digit representation of date is not adequate to meet the needs for most hardware operating systems and data

processing applications into the next century and beyond. This standard will help to facilitate exchange of information between systems and to mitigate problems created by the turn of the Century, which occurs at the start of the Year 2000.

EPA has not yet adopted an Agency-wide standard for representation of date for data processing or for data interchange. Established standards organizations (e.g., the American National Standards Organization [ANSI], the International Standards Organization [ISO] and the National Institute for Standards and Technology [NIST]) provide accepted standards for date representation to meet the needs of government and industry to manage and communicate data. International standards specify the representation of dates in the Gregorian calendar, including both calendar and ordinal date formats. Calendar date is a representation composed of the time elements: year, month of year, and day of month. Ordinal date is a representation composed of the time elements: year and day of year.

Many Agency information systems use commercial database management systems to store and manage data. These products employ specific algorithms to store and manipulate date information. There is no need to establish standards for the way dates are physically stored and managed *within* such operating environments. There ***is,*** however, a definite need for the Agency to adopt a standard for representation of date in data files to avoid errors during data interchange, particularly where one application directly accesses data files created by another application which may have been stored in a non-compliant format.

There also is a need to ensure that acquisitions of commercial off-the-shelf (COTS) software and custom-designed software products require suppliers to verify their products will be capable of importing, exporting and computing dates using a four-digit year representation that complies with this standard.

5. DEFINITIONS.

a. Agency Information System: Refers to the organized collection, processing, maintenance, transmission, and dissemination of information in accordance with defined procedures, whether manual or automated. Information Models are included in this definition.

b. Date, calendar: A particular day of a Gregorian calendar year, identified by a single eight-digit numeric data element composed of the ordinal number for its calendar day, the ordinal number for its calendar month, and the ordinal number for its calendar

year.

c. Day, calendar: A particular day within a month of a Gregorian calendar year, identified by a two-digit number with a leading zero where the number representing the day has only one digit. The first day of the month is represented by the ordinal number "01," and subsequent days are numbered in ascending sequence from "02" to the end of the month.

d. Gregorian calendar: A calendar in general use introduced in 1582 to correct an error in the Julian calendar. In the Gregorian calendar, common years have 365 days and leap years 366 days, divided into 12 sequential months.

e. Month, calendar: A period of time resulting from the division of a calendar year into 12 sequential periods of time, each with a specific name and containing a specified number of days. Month of the year is represented by two digits, with a leading zero where the number representing the month has only one digit. January is represented by the ordinal number "01," and subsequent months are numbered in ascending sequence from "02" to "12."

f. Year, calendar: A cyclic period of time in a calendar that is required for one revolution of the earth around the sun. In the Gregorian calendar, a calendar year is either a common year or a leap year.

g. Year, common: In the Gregorian calendar, a year that has 365 days.

h. Year, leap: In the Gregorian calendar, a year that has 366 days. A leap year is a year whose number is divisible by four an integral number of times, except that if it is a centennial year, it shall also be divisible by 400 an integral number of times. The year 2000 is a leap year.

6. STANDARD.

EPA's standard for representation of calendar date in Agency information systems requires compliance with the following, for interchange of calendar date data:

a. EPA's standard for representation of calendar date is an eight-digit sequence composed of numeric characters, in the format YYYYMMDD, where:

(1) “YYYY” represents the calendar year. The allowed values for calendar year are “0001" through “9999". Calendar year is presumed to be anno Domini (A.D.), or Common Era (CE), unless otherwise specified.

(2) “MM” represents the calendar month of the year. The allowed values for calendar month are “01" through “12", with January represented by the ordinal number “01" and subsequent months numbered in ascending sequence from “02" to “12".

(3) “DD” represents the calendar day of the month. The allowed values for calendar day are “01" through “31", depending on the number of days in a calendar month. The first day of the month is represented by the ordinal number “01" and subsequent days are numbered in ascending sequence from “02" to the end of the month.

(4) No characters may be added to or deleted from the eight character representation of YYYYMMDD.

And:

(5) The order of the time elements shall be high order to low order: Century/year; month; day of month (e.g. CCYYMMDD).

(6) No alphabetic characters are used to represent “month”.

(7) The numbers that represent month of year and day of the month shall include leading zeros whenever their respective values contain only one digit. For example, the month of May is “05" and the 7th day of the Month is “07".

(8) No separators are used between the time elements (i.e. no slashes, hyphens, or spaces) for the interchange of date. The numbers shall be contiguous.

b. Examples of the standard date format are:

<u>Required Standard Date Format</u>: <u>The Alphanumeric Equivalent</u>:

19970101	January 1, 1997
19980704	July 4, 1998
20001225	December 25, 2000

c. Agency information management systems must include the intelligence to accurately process date according to the following criteria:

(1) Calculations must execute, using dates with a four-digit year.

(2) Functionality, both online and batch, including but not limited to entry, inquiry, maintenance and update, must support four-digit year interfaces.

(3) Interfaces and reports must support processing of four-digit year.

(4) Successful transition into the year 2000, using the correct system date, must occur by the system without human intervention.

(5) After transition to the year 2000, processing with a four-digit year must occur without human intervention.

(6) Calculations that determine and process leap years must be accurate.

(7) Correct results in forward and backward data calculations that span century boundaries must be provided, including the conversion of previous years currently stored as two digits.

(8) Date data that are output for data interchange must be in the format specified above for standard representation of calendar date.

d. EPA's standard for representation of calendar date conforms to the appropriate subset of each of the standards referenced in paragraph three, titled REFERENCES, above.

e. The format for date, where it is represented as a character string for data interchange, shall be binary coded character(s) using the American Standard Code for Information Interchange (ASCII).

7. RESPONSIBILITIES.

The **Chief Information Officer** (CIO) of the Agency shall be responsible for issuing waivers for compliance consistent with Paragraph. 9, titled PROVISON FOR WAIVER, below.

The **Office of Information Resources Management** (OIRM) shall:

a. Lead the efforts to develop, implement, and ensure adherence to this data standard.

b. Lead the efforts to develop a management plan describing steps for implementation of the standard.

c. Provide guidance and technical assistance in meeting the requirements of this standard.

d. Oversee resolution of conflicts regarding applicability or other issues relating to this standard.

e. **Senior Information Resources Management Officials** (**SIRMOs) and Regional IRM Branch Chiefs** shall be responsible for assuring compliance with this standard within their information management environments.

8. COMPLIANCE DATES -- NEW ACQUISITIONS AND EXISTING APPLICATIONS

Beginning January 1, 1998, all **new** applications software acquired by the Agency will comply with this standard and must be certified by vendors to be in compliance with Year 2000 requirements.

The standard also applies to **existing** Agency applications by the Year 2000 or earlier whenever such operations would otherwise yield incorrect results because of the lack of a date format that includes Century year characters. Wherever an application uses projected dates for accelerated computation of interest, retirement, or other values, more immediate attention to compliance with this standard may be required to avoid errors. It is the responsibility of the program operating the application to make such a determination.

9. PROVISION FOR WAIVER.

A waiver may be granted by the Agency's Chief Information Officer (CIO) for legacy applications that achieve compliance by means other than the use of this standard, or for systems or applications that will be terminated and retired before the Year 2000, or for systems where the costs of implementing the standard are significantly higher than the benefits. The guiding principles for a waiver will involve:

a. Developing an application for waiver to the CIO outlining the reasons why the date data standard should not be implemented in the information collection. The application will contain a complete risk assessment and cost-effectiveness evaluation of continued operation in a non-compliant mode.

b. Obtaining approval by the decision officials in the requesting office, as defined by EPA's System Lifecycle Management Policy and the organization's Senior Information Resources Management Official (SIRMO).

c. Submitting the application to the CIO, who has responsibility for final disposition.

d. The CIO notifying the applying office in writing of the disposition of the waiver.

Appeals may be brought to the Executive Steering Committee for IRM (ESC-IRM) if outstanding issues cannot be resolved. The ESC-IRM may make recommendations to the CIO for final decision.