

Standards for Archival Description

External Technical Standards for Data Contents and Data Values: Prospects for Adoption by the Archival Community

SHARON GIBBS THIBODEAU

Abstract: Certain recurrent types of information (e.g., dates, times, geographic locations, languages) are so broadly used in data processing that standard code lists (data values) or standard formats for entering the information (data contents) have been developed and widely adopted. Many of these have direct applications in archival information systems. The author summarizes the issues to be considered by the archival community before adopting these standards. Several specific standards are described as examples and are evaluated, using the criteria developed by the Working Group on Standards for Archival Description.

About the author: Sharon Gibbs Thibodeau is director of the Archival Publications and Accessions Control Staff, National Archives and Records Administration, where she is responsible for the development and application of guidelines for the description of federal records. She prepared this paper for the June 1989 meeting of the Working Group on Standards for Archival Description in response to issues raised at the group's December 1988 meeting.

CERTAIN TYPES OF INFORMATION appear with great frequency in modern communication. Almost every document contains references to dates, times, and geographic locations, to name three examples. Managers of automated data processing systems tend to view such information in a particular way. Wanting to increase the efficiency of their systems and to insure effective manipulation and exchange of data, they establish standard means of entering the data. The standards that they employ involve codes to represent the data (data values) and formats in which the codes are entered (data contents). Effective exchange of data stored in automated systems requires agreement on the codes that are used, hence the development of national standards, the value of adherence to which is apparent to those for whom data exchange is important.

The usefulness of these data value codes to the archival community depends in large part on the role defined for them in the descriptive process. Many archivists are already engaged in data exchange, and they need to evaluate external technical standards for data contents and data values. Because coded information cannot be understood without first being decoded, and because it does not lend itself to easy and reliable decoding by human users, it should be used cautiously, if at all, in finding aids to be read and interpreted by the public.

At its first meeting on 3-4 December 1988, the Working Group on Standards for Archival Description identified several existing external technical standards for data contents and data values and concluded that there might be value in formally adopting some of these standards. The group also discussed criteria that might govern a decision by the archival profession to adopt an existing external standard and identified a number of salient issues.

- A decision to adopt an existing standard would be based on a determination that

the benefits (in terms of increased access to information about records) outweighed the costs (in terms of problems associated with implementation).

- Implementation of a standard could potentially conflict with a previously adopted standard; it could have an adverse impact on existing descriptive information that does not meet the standard; or it could result in loss of information.

A summary of the contents of selected codes that might have particular utility in archival application and a discussion of the applicability of the Working Group's criteria will serve to demonstrate what is at issue for the archival community.

Time and Date Codes

The American National Standards Institute has developed a calendar standard, ANSI X3.30 (Representation for Calendar Date and Ordinal Date for Information Interchange), that provides codes and representations for identifying the specific years, months, and days of the Gregorian Calendar. A date that conforms to the standard would be represented by a numeric code of eight (or six) consecutive positions of the form *yyyymmdd* (or *yymmdd*), where *yyyy* represents the year (or *yy* represents the units and tens identification of the year), *mm* represents the month (from 01 = January to 12 = December), and *dd* represents the day of the month (from 01 to 31). Closely related standards include ANSI X3.43 (Representations of Local Time of Day for Information Interchange) and ANSI X3.51 (Representations of Universal Time, Local Time Differentials, and United States Time Zone References for Information Interchange).

Adoption by the archival profession of this standard for date representation would likely take the form of requiring that all dates represented in automated systems meet the standard's specifications (*yyyymmdd* or

yymmdd). Considered in the context of the Working Group's criteria for standards evaluation, it appears that the ANSI date standard would be widely applicable and practical. It is very likely that the standard already has been implemented in the design of those existing automated systems that support archival description. If not, retrospective implementation of the standard would have minimal impact. Certainly the standard could be implemented in new automated systems with relative ease.

Geographic Location Codes

Two technical standards provide a means of coding the names of states, counties, and county equivalents in the U.S. ANSI X3.38 (Identification of States of the United States, the District of Columbia, and Associated Areas of the United States for Information Interchange) is a set of codes for representing the fifty states of the United States. These include the two-letter code used by the United States Postal Service and a two-digit numeric designator for each state.

ANSI X3.31 (Codes — Structure for the Identification of the Counties and County Equivalents of the United States and Its Outlying and Associated Areas for Information Interchange) is a set of codes for counties within the United States. These are three-digit numeric codes to be used in combination with a state code to uniquely designate a given county. The United States Census Bureau is a principal user of these codes.

Adoption of these standards would entail requiring use of the codes designated in X3.38 or X3.31 whenever information referring to specific states or counties is encoded in an archival information system. This could be accomplished easily for the state codes. The two-letter (postal) codes comprising the standard are familiar to most archivists; it is likely that these codes are already in use in existing automated systems that support archival description.

Adoption of the standard would have minimal retrospective impact, and new system designs could readily accommodate it.

The applicability of county codes within the archival profession may not be as wide as that of state codes, but references to counties are included in many archival descriptions. Few archivists (likely only those who hold machine-readable data files) are as familiar with these three-digit county codes as they are with the two-letter state codes. As with the date and state standards, adoption of the county standard would have minimal retrospective impact. New system designs could readily accommodate it.

Transliteration Standards

The National Information Standards Organization (NISO or Z39), a committee of the American National Standards Institute, has developed a series of standard systems for the Romanization of non-Roman alphabets. They include systems for Japanese (ANSI Z39.11-1989), Arabic (ANSI Z39.12-1989), Slavic Cyrillic [Russian, Ukrainian, Bylorussian, Serbian, Macedonian, and Bulgarian] Characters (ANSI Z39.24-1974), Hebrew (ANSI Z39.25-1975), Lao, Khmer, and Pali (ANSI Z39.35-1979), and Armenian (Z39.37-1979).

Each of these standards has unique characteristics. Z39.12 (Arabic) and Z39.35 (Lao, Khmer, and Pali), for instance, are especially useful for bibliographic applications in rendering personal, corporate, and place names, as well as the titles and other indicia of publications and documents, so that someone using such Romanized material could reconstruct the writing in the original language, if required. The systems presented in Z39.24 for Slavic Cyrillic characters attempts to minimize the number of diacritical marks used. Z39.25 presents four styles for Romanizing the Hebrew alphabet along with guidelines for selecting the one most appropriate for the user's purposes.

Adoption of these standards would commit the archival community to their use in all transliteration activities involving the subject alphabets. Such transliteration standards would have relatively limited application within the American archival profession. However, adherence to the standards would increase access to the records to which they do apply. Few if any of these standards are widely known among archivists, and adoption would require training in their application.

Standards for Numbering Serials and Technical Reports

Two NISO-developed standards, ANSI Z39.9 (International Standard Serial Numbering [ISSN]) and ANSI Z39.23 (Standard Technical Report Number[STRN]) contain rules for formulating identifying numbers for materials held by many archival repositories.

ANSI Z39.9 (ISSN) defines the structure and use of a standard code for the unique identification of serial publications (both print and non-print). It also designates a national code authority for administration of the code. In the United States, the original codes were assigned by Bowker and a network of agencies now controls their assignment. International Standard Serial Numbers have the form ISSN XXXX-XXXX.

ANSI Z39.23 (STRN) is designed to facilitate bibliographic retrieval of technical reports. It outlines a numbering system to be administered by the National Technical Information Service that will coordinate the assignment of code numbers and maintain a registry. The standard specifies the alphanumeric format to be used. Standard Technical Report Numbers have the form AXXXXX/XXXX/XX-YY/XXXX.

Full adoption of these standards would commit the archival community to the use of ISSNs when describing or identifying serials in archival collections and STRNs

whenever individual technical reports are identified. Because collective description is the norm in an archival context, few serials or technical reports have been individually identified. It is true that assignment of these numbers would enable researchers familiar with them to locate information which would otherwise be hidden in archival collections, but it is also true that few archival repositories have the resources to identify individual technical reports comprising large collections. It is unlikely that many members of the archival profession have assigned ISSNs or STRNs or are familiar with the rules to be followed in assigning either number. Training would be necessary to enable members of the profession to implement the standards if adopted.

Symbols of American Libraries

Mnemonic symbols for American libraries and bibliographically related commercial enterprises are maintained by the Catalog Management and Publication Division of the Library of Congress. The National Library of Canada maintains these symbols for Canadian libraries.

Although most of the symbols in the current list indeed refer to libraries, several refer to institutions (e.g., Colgate University) without specifying a library facility. This lack of specificity appears to limit considerably the extent to which these symbols could be used in archival description. Archivists might agree to use the library symbols only to refer to libraries at institutions that maintain separate library and archival facilities, but there is no assurance that other users of the symbols will so limit their use.

The Role of Archivists in Maintaining External Technical Standards

Adoption of an external standard by the archival community should entail the assumption of responsibility for maintaining the standard. In some cases the archival

community may attach enough importance to the area governed by the standard to justify full participation in a standards-setting committee; in other cases it may be appropriate merely to encourage members of the archival community to propose changes to be decided upon by the external standard-setting body. The archival profession is not likely to seek an active role in the maintenance of ANSI time, state or county codes, transliteration conventions, or standard book or report numbers. However, should an influential role be desired, it is conceivable that the Accredited Standards Committee X3 and NISO would each accommodate a

representative of the Society of American Archivists or some other professional archival organization. X3 currently maintains a balanced membership among three categories: producers of information systems, consumers of such systems, and general interest members. The current voting membership of NISO includes a representative of the National Archives and Records Administration who, at the present time, is principally concerned with microform and paper permanence standards. Most other members of NISO represent library interests.