

# Encoded Archival Description: A Structural Overview

JANICE E. RUTH

**Abstract:** Encoded Archival Description (EAD), the SGML Document Type Definition (DTD) for archival finding aids, has been under development for more than three years. Although it has been significantly improved during that time by feedback from early implementers and by new insights from its developers, much of EAD's basic structure dates from the design team's first meeting in July 1995 at the Bentley Historical Library in Ann Arbor, Michigan. During that week-long gathering, the EAD design team articulated its observations about traditional archival finding aids, established design goals and principles, and created the framework for the current version 1.0 DTD structure. This article examines the basic steps in building an SGML DTD and provides a structural analysis of EAD's high-level elements in light of the developers' early design goals and decisions.

*About the author:* Janice E. Ruth is both a writer/editor and an acquisitions archivist in women's history for the Library of Congress Manuscript Division, where she also has held positions as a reference librarian and processing technician. She was a member of the EAD development team convened by Daniel Pitti in July 1995 and continues to monitor and contribute to EAD's progress as a member of SAA's EAD Working Group. She also serves as a member of SAA's Committee on Archival Information Exchange.

## Introduction

AFTER DEVELOPING HIS FIRST Document Type Definition (DTD) for archival finding aids, Daniel Pitti bravely assembled a group of seven archivists and one expert in Standard Generalized Markup Language (SGML) in July 1995 to critique and revise his data content model.<sup>1</sup> Most members of the group had little or no experience with SGML or its derivative, HyperText Markup Language (HTML), but they did possess a strong knowledge of archival descriptive practices and were especially familiar with that particular class of documents (or “document type”) known as archival finding aids. The group shared a common understanding of how and why finding aids are constructed, the kinds of information they contain (or should contain), and the uses to which they have been put. This knowledge and appreciation of traditional paper-based finding aids, coupled with ideas for electronic enhancements, influenced the group’s review and redesign of Pitti’s original model and inspired the development of a new DTD named Encoded Archival Description (EAD).

Other articles in this issue of the *American Archivist* examine the value of SGML and discuss EAD’s development as a potential international standard. Those same themes also surface in this article, but the focus here is on a two-part structural analysis of EAD. The first part begins by examining SGML’s role in structuring documents and then proceeds to explore EAD’s theoretical underpinnings by reviewing the development team’s deliberations, especially its early formulation of design principles and goals. The second part leads readers systematically through EAD’s high-level elements and suggests how the design considerations described in part one influenced the current DTD structure.

## Understanding SGML Document Type Definitions

As a registered international standard, SGML is a public, nonproprietary technique for defining and expressing the logical structure of documents. It is the language used to write Document Type Definitions (DTDs), which are sets of rules for marking up or encoding classes of documents so that the text therein may be searched, retrieved, displayed, and exchanged in a predictable, platform-independent manner. Archivists interested in applying EAD need not become SGML experts, but an awareness of a DTD’s three major functions will promote an increased understanding of the EAD structure and assist beginning encoders in interpreting the EAD tag library and other user documentation.

First, a DTD names and defines all the elements or data fields that may be used to mark up a particular type of document. In the same way that MACHine-Readable Cataloging (MARC) provides a structure for information in a catalog record, the EAD DTD designates the fields or categories of information contained in a finding aid. These data fields in a DTD are called “elements,” and each one is assigned a unique name, abbreviation, and definition. Elements are represented by short alphanumeric words (or “tags”)

---

<sup>1</sup>The group gathered for a week in Ann Arbor, Michigan, in July 1995, to review and revise Pitti’s FindAid DTD under the auspices of the Research Fellowship for Study of Modern Archives, a program supported by the Andrew W. Mellon Foundation, the Division of Preservation and Access of the National Endowment for the Humanities, and the Bentley Historical Library, University of Michigan. The participants were Steven J. DeRose, Jackie M. Dooley, Michael J. Fox, Steven L. Hensen, Kris Kiesling, Daniel V. Pitti, Janice E. Ruth, Sharon Gibbs Thibodeau, and Helena Zinkham. This article draws heavily from the work done by the original design team at that meeting, and by them and Anne J. Gilliland-Swetland, Thomas A. LaPorte, Deborah Lapeyre, and others during the subsequent three years. It incorporates definitions, element descriptions, and other language contributed by this author and other team members to previously issued group documents.

captured as simple ASCII characters that surround the text (or content) being designated. These tags, which are enclosed in angle brackets, indicate to a computer where the text of an element begins and ends. For example, in EAD, the beginning of a paragraph is marked as `<p>` and the end of the same paragraph as `</p>`. Similarly, the beginning of a scope and content note is identified with the element tag `<scopecontent>` and the end of that same note with the close tag `</scopecontent>`.

In developing the EAD DTD, the design team attempted to identify and name elements that reflected both the content and structure of traditional archival finding aids. These included features generic to most text-based products, such as paragraphs, headings, titles, abbreviations, block quotes, lists, and tables, as well as other properties such as scope and content notes, biographical notes, agency histories, and series descriptions, which may be unique to finding aids. Also identified as elements were external and internal pointers, references, and other links that would enable EAD to support hypertext and hypermedia, paving the way for finding aids to become more dynamic in an on-line environment and facilitating the capability to link electronic finding aids to digital representations of the archival materials described therein.

The second function of a DTD is to determine which elements should be further specified through the use of SGML "attributes." For example, an element called `<date>` was established to encode all dates except those associated with the creation of archival materials. (Creation dates are identified by the separately named element `<unitdate>`.) The design team quickly recognized, however, that there may be value in differentiating the many other kinds of dates (e.g., birth dates, flourish dates, publication dates) that appear in a finding aid. To accommodate the potential need to search different kinds of dates separately or to display them in a unique manner, an attribute named `TYPE` was created and linked to the `<date>` element for optional use during finding aid markup. Since it seemed impossible to predict every type of date archivists may want to specify, the value or content of the `<date>` `TYPE` attribute was not limited to a predetermined list of choices but instead was set to accept any character string the encoder enters. Thus if the intention is to identify a birth date within a finding aid, the DTD permits the general designation `<date>1922</date>` or the more precise tagging `<date TYPE="birth">1922 </date>`.

However, a DTD is more than a listing of elements and associated attributes. Its third and perhaps most important function is to specify where and in what sequence elements may be used. For example, a DTD may permit a Title `<title>` element to be used within a Paragraph `<p>` element, but it probably would not permit a `<p>` element to be used within a `<title>` element. One article has suggested that archivists are well suited to writing DTDs because the process is similar to organizing a collection of papers or group of records.<sup>2</sup> In arranging or processing a collection, the archivist analyzes the material, identifies its parts, and determines its logical structure. This process involves recognizing or defining hierarchical arrangements and then developing methodologies to implement those arrangements. Building a DTD also involves analyzing a set of documents (such as finding aids, in the case of EAD), determining how the parts relate to one another, identifying the major structural units, and then subdividing those larger parts into smaller and smaller units or subelements.

---

<sup>2</sup>"Linking the Encoded Archival Description and the TEI," *CETH Newsletter* 4 (Spring 1996): 4-6.

## Document Analysis and the Ann Arbor Accords

In identifying the elements of a finding aid and determining where and in what sequence they should appear, the EAD developers were guided by certain fundamental observations about finding aids and several overriding principles and goals, which the design team articulated at its first meeting in Ann Arbor, Michigan. Named for the meeting site, these principles or accords stated the group's philosophy and outlined the DTD's intent, parameters, and structural features.<sup>3</sup> Responding first to confusion over the meaning and scope of the term "finding aid," the group decided to limit its focus to that class or genre of documents known specifically as archival inventories and registers. This decision narrowed a larger universe of finding aids that typically includes bibliographies, subject guides, repository handbooks, and other descriptive tools, and permitted the developers to optimize the DTD's design for guides to individual collections, record groups, fonds, etc., regardless of the guide's length or the type of materials described therein.

Within this finding aid subset, the group determined that the DTD must handle both the creation of new finding aids and the conversion of existing ones. This requirement posed a slight problem during the development phase, because although a single DTD might be used for many purposes, such as data conversion, interchange, authoring, etc., it could be optimized for only one function.<sup>4</sup> The design team was not prepared to abandon the community's "legacy data" (all those filing cabinets and shelves full of useful finding aids), nor did it want to discard traditional finding aid designs, which were both familiar and functional. Simultaneously, however, the team sought greater structural uniformity across finding aids in the belief that adherence to a consistent data model increased the potential for union databases and document interchange among repositories. Since efforts to ensure structural uniformity meant possible roadblocks to conversion, a compromise had to be reached.

Team members decided that the DTD should not be expected to accommodate all existing practices; they also acknowledged that converting current finding aids to an ideal EAD markup would likely necessitate shifting some text or adding data to conform to the DTD's sequencing of elements and the consignment of certain elements to specific settings. An attempt was made, however, to minimize conversion difficulties by creating a special element called Other Descriptive Data <odd> to encode information, principally in existing finding aids, that may not fit easily into EAD's otherwise distinct categories. The <odd> element could be used when information did not correspond to another element's definition; when the information was of such mixed content as to make a single classification difficult; and when shifting the information to permit more specific content designation would be too costly or burdensome. Despite making the <odd> element widely available throughout the DTD, the team acknowledged that <odd> should be used with restraint and only after carefully considering the consequences that unspecified content designation poses for searching, retrieving, and displaying information in a networked environment.

<sup>3</sup>"Ann Arbor Accords: Principles and Criteria for an SGML Document Type Definition (DTD) for Finding Aids," *Archival Outlook* (January 1996): 12.

<sup>4</sup>SGML expert Deborah Lapeyre, who had been retained by the Library of Congress National Digital Library Program to assist with EAD development, alerted the EAD design team of the conflicting requirements in the Ann Arbor Accords at the group's 1-3 November 1995 meetings.



Having limited the document class to both new and existing registers and inventories, the design team then confined itself to developing an archetypal data structure, resisting efforts to specify or prescribe the intellectual content that would reside inside that structure. The task was not to develop a data content standard, but to create instead a content designation or encoding standard. The group felt that subsequent content guidelines should be developed to address questions of "best practice" and to do for finding aids what the *Anglo-American Cataloguing Rules*, 2d edition (AACR2) and *Archives, Personal Papers, and Manuscripts (APPM)* accomplished for catalog records.

In deciding which elements and attributes to include within the DTD, the group recognized that "while there are certain elements that ought to appear in any finding aid, various intellectual and economic factors influence the depth and detail of analysis employed."<sup>5</sup> Concerned that it not create an overly enforcing, prescriptive, or burdensome DTD, the team created few required elements and allowed for both the nesting and reuse of elements to capture "progressively more detailed and specific levels of description as desired."<sup>6</sup> For example, the DTD contains an element called Administrative Information <admininfo>, which is used to provide descriptive background information concerning an institution's acquisition, processing, and management of a body of archival materials. The <admininfo> element designates facts about acquisition, access and reproduction restrictions, availability of microform and digital surrogates, preferred form of citation, and other descriptive details that help readers of the finding aid know how to approach the archival materials and make use of the information they find. All the specific pieces of information captured in <admininfo> have their own corresponding elements in the DTD—with tag names such as <custodhist>, <accruals>, <acqinfo>, <accessrestrict>, <appraisal>, <userrestrict>, <altformavail>, <prefercite>, and <processinfo>—which may be applied individually if desired. Should such specificity not be needed, however, the archivist may elect to tag the entire body of information at the parent level <admininfo>, and not to encode separately the text relating to each nested subelement.

The design group was cautious about adding to the DTD every element that a team member could identify. Each proposed element was expected to support one of the following functions: description, control, navigation, indexing, or on-line and print presentation. For each element, the team assessed whether staff or researchers would want to search for that particular piece of encoded information, display or print it in a unique way, or take some other specific, definable action on it. If none of these situations was anticipated, then the element was not adopted. If the element passed the "functionality test," it was added to the DTD, often under a language-neutral name designed to enhance broad international application of EAD. Terms such as collection, personal papers, archives, series, fonds, and record group were avoided in favor of "generic terms like unit and component that are not specific to any individual setting or institution."<sup>7</sup>

### Hierarchy, Formatting, and Other Design Considerations

Although not specifically enumerated in the Ann Arbor Accords, several other important observations about paper-based finding aids played a crucial role in the development of EAD. Foremost was the recognition that many archival inventories and registers

<sup>5</sup>"Ann Arbor Accords," 12.

<sup>6</sup>"Ann Arbor Accords," 12.

<sup>7</sup>"Ann Arbor Accords," 12.

describe a unit of records or papers at several different, but related, levels of detail. Within these hierarchical, multilevel views, information about the archival materials is both repeated and inherited. For example, many archival inventories and registers begin by presenting information about the entire body of records or papers described in the finding aid. This information may be conveyed in a provenance statement, a scope and content note, or by means of some other narrative device that describes the unit of materials in its entirety. Certain specific pieces of information about the unit, such as its title, creator, span dates, identification number, extent, location, scope, content, and arrangement, may (and should) be captured in these high-level descriptions, which are intended to give a broad overview of the whole unit. The design team created the Archival Description `<archdesc>` element to identify this description of the whole.

Within the `<archdesc>` element, other more detailed descriptions of the subordinate parts may be presented, which the EAD designers designated by the element Description of Subordinate Components `<dsc>`. For example, an archivist may elect to describe separately and in greater detail all the series within a manuscript collection. This description of the series presents another view of the entire unit, this time in terms of its major components. Similarly, the archivist may elect to prepare lists of files or items within each series. Again, these contents lists are views of the entire collection, but they usually contain an even greater level of detail than was captured in either the first collection-level description or the second series-level description. The EAD design team recognized that successive levels of description inherit information from the preceding component- or unit-level summary, and that every level repeats or reuses some of the same basic data elements, such as title, creator, span dates, identification number, extent, location, scope, content, and arrangement. This recursive, repeatable character of finding aids is reflected in the EAD structure.

Also noted by the design team, but perhaps less elegantly resolved, was the problem that although SGML permits almost complete separation of format from content, archival finding aids do not. The tabular formatting of finding aids is a mechanism for imparting information about the organization and arrangement of the materials being described. Archival inventories and registers often contain parallel structures—one that conveys the intellectual arrangement of the materials, and the other representing the physical arrangement. In many paper-based finding aids, these dual structures are presented through the use of columns: The intellectual hierarchy runs down one side of the page, and a listing of container numbers or microfilm locations runs down the other side. Often these dual structures shift or break at different points. Since SGML does not simultaneously accommodate dual structures effectively, the design team had to choose which structure it would optimize the DTD to handle.

It was agreed that the intellectual arrangement of the archival materials was more important and more permanent than the physical order, and the DTD was designed accordingly. The team also decided that the DTD need not slavishly recreate the exact appearance of every finding aid; in other words, fidelity to the printed page would not be supported. Nevertheless, the designers sought to enable DTD users to replicate columnar layouts in two ways. First, SGML stylesheets may be used to manipulate intellectual content elements for basic columnar output. Secondly, for more precise columnar layouts, including greater control of indentations, encoders may overlay the intellectual markup with a special set of tabular display elements, Table Specification `<tspec>`, Display Row `<drow>`, and Display Entry `<dentry>`, created specifically for EAD. The tabular elements serve as outer wrappers when manipulating groupings of intellectual content ele-

ments in order to achieve desired on-line and print presentations. Utilizing EAD's special tabular superstructure has been problematic, however, and many early implementers have achieved satisfactory display results without it. Reacting to early implementers' general lack of interest in the display elements, and concerned about the confusion those elements have caused for some archivists just beginning to learn the DTD structure, the design team decided during version 1.0 development to reduce the presence of the display elements within the DTD. Under version 1.0, if encoders wish to use the tabular display elements, they must first make a minor modification to their copies of the DTD in order to access the elements.

A final consideration in designing EAD was the DTD's relationship to other standards, especially the USMARC Archival and Manuscripts Control (AMC) format and the General International Standard Archival Description (ISAD(G)). After much discussion on the degree to which MARC-formatted data should be extractable from an EAD-compliant finding aid, the group decided that it would not design EAD for exporting MARC fields down to the indicator and subfield levels. The team acknowledged that finding aids are the chief source of information for creating cataloging records in the MARC AMC format but felt that it would be burdensome and unwieldy for EAD to be structured so that a complete MARC record could be harvested automatically from the SGML markup. A compromise was reached: With the exception of the Controlled Access Headings <control-access> element (discussed in part two of this article), no elements were added to the DTD simply for the sake of providing a corresponding data structure to MARC, but for those MARC-like elements already represented in EAD, the team added an optional ENCODINGANALOG attribute, which permits the designation of the applicable MARC field or subfield together with the authoritative form of the data. By using these ENCODINGANALOG attributes, archivists can generate skeletal MARC records automatically from EAD finding aids.

Although EAD began its development in the United States based on archival descriptive practices used in this country, its progress has been closely followed by members of the international archival community. From the beginning, the EAD design team was hopeful of creating a model with worldwide appeal, and the international community was represented on the team by a longtime member of the International Council on Archives (ICA) Ad Hoc Commission on Descriptive Standards, which developed the ISAD(G).<sup>8</sup> As mentioned earlier, the EAD design team adopted neutral language throughout the DTD as a means of facilitating global acceptance, and as the EAD structure evolved, a separate special effort was made to test its international applicability by comparing it to the ISAD(G), which had been approved by the ICA in 1994.

Issued as a guideline to be followed in the preparation of archival finding aids, the ISAD(G) consists of two major segments: 1) a segment that provides rules for systematic, multilevel presentation in a single finding aid of descriptive information about a whole unit of records or papers and its component parts or divisions; and 2) a segment that specifies the individual elements that may be presented about the whole archival unit or any component part in accordance with the multilevel rules.<sup>9</sup> Like EAD, the ISAD(G)

<sup>8</sup>EAD team member Sharon Gibbs Thibodeau served on the ICA's Ad Hoc Commission on Descriptive Standards from its inception in 1990 until 1997. The author is grateful to her for providing a summary of ISAD(G) for use in this article and for conducting the EAD-ISAD(G) comparison for the team.

<sup>9</sup>For the full text of ISAD(G), see International Council on Archives, *ISAD(G): General International Standard Archival Description*, adopted by the Ad Hoc Commission on Descriptive Standards, Stockholm, Sweden, 21–23 January 1993 (Ottawa: International Council on Archives, 1994) or the ICA's homepage at <http://www.archives.ca/ica/>.

assumes the following three things: 1) a finding aid consists of hierarchically organized information describing a unit of records or papers along with its component parts; 2) information in the finding aid is inherited from one descriptive level to another; and 3) descriptions of both the whole and the parts each comprise essential data elements. A direct comparison of EAD with ISAD(G)<sup>10</sup> conducted by an EAD team member found that EAD can accommodate a finding aid that complies with ISAD(G). The EAD DTD provides for the tagging of all the essential elements specified in the ISAD(G), and the direct parallels can be indicated by invoking the same optional `ENCODINGANALOG` attribute used for MARC comparisons. This compatibility between EAD and the ISAD(G) affirms the validity of the EAD structure and suggests a role for EAD in fostering the international exchange of archival descriptions.

### Overview of the EAD Data Model

Through an examination of a DTD's major functions, a description of the Ann Arbor Accords, and a look at other design considerations, the first part of this article attempted to highlight aspects of EAD's theoretical framework preparatory to exploring the DTD's high-level elements in greater detail. The intent of the second part of the article is not to describe every element and the circumstances surrounding its use (a tag library and application guidelines<sup>11</sup> are designed for that task), but instead to provide a general overview of some of the major elements and their relationship both to one another and to the design considerations previously discussed. Since EAD is an ongoing project, changes to the version 1.0 DTD structure outlined here are inevitable, and readers are advised to monitor new developments by subscribing to the EAD listserv and consulting the official EAD website maintained by the Network Development and MARC Standards Office of the Library of Congress, where the most recent version of the DTD and related documentation are posted.<sup>12</sup> Since some readers of this article may have been familiar with the beta version of the DTD, footnotes provide commentary about significant changes between that version and version 1.0, scheduled for release in summer 1998.

### Designating Bibliographic Information About Finding Aids

Although not universally employed, title pages, prefaces, and use instructions are often the first pieces of information found in archival inventories and registers. Observing this fact led the EAD design team to deduce that at the most basic level, archival finding

---

<sup>10</sup>See Sharon Gibbs Thibodeau, "A Mapping of EAD Tags to the Elements of Description Incorporated in the International Standard Archival Description, ISAD(G)," in "Encoded Archival Description Document Type Definition (DTD) Applications Guidelines," by Anne J. Gilliland-Swetland, edited by Thomas A. LaPorte, unpublished draft disseminated electronically, December 1996, 70–73 at <<http://scriptorium.lib.duke.edu/findaids/ead/guidelines/index.html>>.

<sup>11</sup>[Deborah Lapeyre and Janice E. Ruth, eds.], "Draft Tag Library for EAD Alpha DTDs," unpublished draft distributed in limited paper copies and disseminated electronically, February 1996; Anne J. Gilliland-Swetland and Thomas A. LaPorte, eds., "Encoded Archival Description Document Type Definition (DTD) Beta Version, Tag Library," unpublished draft disseminated electronically, October 1996; and Gilliland-Swetland and LaPorte, "Encoded Archival Description Document Type Definition (DTD) Applications Guidelines." The URL for both the beta tag library and applications guidelines is <http://scriptorium.lib.duke.edu/findaids/ead/guidelines/index.html>. A version 1.0 tag library will be published by SAA at roughly the same time that the revised DTD is released in summer 1998.

<sup>12</sup>Version 1.0 of the DTD is scheduled for release in summer 1998. The URL for the official EAD website maintained by Library of Congress, Network Development and MARC Standards Office is <http://lcweb.loc.gov/ead>. The site contains information on subscribing to the EAD listserv.

aids consist of two segments: 1) a segment that provides information about the finding aid itself (its title, compiler, compilation date, etc.); and 2) a segment that provides information about a body of archival materials (a collection, a record group, a fonds, or a series).<sup>13</sup> As shown in Figure 1, the EAD DTD splits the first segment into two high-level elements known as EAD Header `<eadheader>` and Front Matter `<frontmatter>`. The second segment, consisting of information about the archival materials, is contained within the third high-level element named Archival Description `<archdesc>`.<sup>14</sup> All three of these high-level elements are contained within the outermost element named Encoded Archival Description `<ead>`. The beginning `<ead>` and closing `</ead>` tags wrap around the entire document.

The `<eadheader>`, outlined in Figure 2, is modeled on the header element in the Text Encoding Initiative (TEI), an international humanities-based effort to develop a suite of DTDs for encoding literary texts or other objects of study. In an attempt to encourage as much uniformity as possible in the provision of metadata across document types, the design team elected to use a TEI-like header to capture information about the creation, revision, publication, and distribution of finding aid instances. The resulting `<eadheader>` consists of four subelements, some of which are further subdivided: EAD Identifier `<eadid>` provides a unique identification number or code for the finding aid and can indicate the location, source, and type of the identifier. File Description `<filedesc>` contains much of the bibliographic information about the finding aid, including the name of the author, title, subtitle, and sponsor (all contained in the Title Statement `<titlestmt>` element), as well as the edition, publisher, series, and related notes encoded separately. Profile Description `<profiledesc>` is used to record the language of the finding aid and information about who created the encoded version of the document, and when. Revision Description `<revisiondesc>` summarizes any revisions made to the EAD document.<sup>15</sup>

The sequence of elements and subelements in the `<eadheader>` is specified by the DTD, with the expectation that searches across repositories will be more predictable if the elements are uniformly ordered. Such searches may help filter large bodies of machine-readable finding aids by specific categories, such as title, date, repository, language, etc. Required use of the `<eadheader>` compels archivists to include essential information about their machine-readable finding aids that often went unrecorded in paper form. In addition, elements in the `<eadheader>` may be used to generate electronic and printed title pages for finding aids.

Because the elements within the `<eadheader>` must follow a prescribed order to ensure uniformity across finding aids, the team also created an optional `<frontmatter>` element, which can be used to generate a title page that follows local preferences for the sequencing of information. The Title Page `<titlepage>` subelement within `<frontmatter>`

<sup>13</sup>"Encoding Standard for Electronic Aids: A Report by the Bentley Team for Encoded Archival Description Development," *Archival Outlook* (January 1996): 10.

<sup>14</sup>Those familiar with the beta version of the DTD should note that the `<findaid>` element no longer exists. In response to input from early implementers, the Adjunct Descriptive Data `<add>` element was made available directly within `<archdesc>`, eliminating any need for the `<findaid>` element, which had served as a wrapper for the `<archdesc>` and `<add>` elements. For a fuller discussion of the change to `<add>`, see footnote 20.

<sup>15</sup>The beta version of the DTD also included a `<footer>` element in the `<eadheader>`. In version 1.0, the `<footer>` element became the `<runner>` element, which is available only within `<archdesc>`, immediately before `<did>`. The `<runner>` element provides for a header, footer, or digital watermark to appear on every page of the finding aid.

Figure 1. Overview of the Encoded Archival Description (EAD) DTD, Version 1.0



reuses many of the same subelements designated in <filedesc>. The <frontmatter> element also can be used to encode structures such as prefaces, dedications, or other text concerning the creation, publication, or use of the finding aid. The design team did not create a specific element for each of these structures, opting instead for a single generic Text Division <div> element.



**Figure 1.** Continued

```

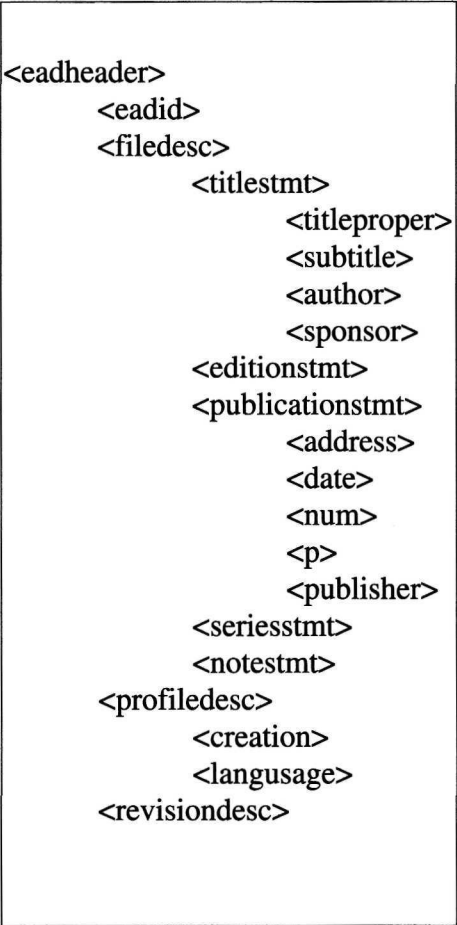
<ead> (continued)
  <archdesc> (continued)
    <bioghist>
    <controlaccess>
      <corpname>
      <famname>
      <function>
      <genreform>
      <geogname>
      <name>
      <occupation>
      <persname>
      <subject>
      <title>
    <dao> and <daogrp>
    <note>
    <odd>
    <organization>
    <scopecontent>
    <dsc> (See Figures 3-5)
      <c01>
        <did>
        <add>
        <admininfo>
        <arrangement>
        <bioghist>
        <controlaccess>
        <dao> and <daogrp>
        <note>
        <odd>
        <organization>
        <scopecontent>
        <c02>
          <did>
          (and so forth)

```

### Encoding the Heart of the Finding Aid

As noted in Figure 1, the third high-level element in <ead> is Archival Description <archdesc>, which consists of information about a body of archival materials. Within this element is found hierarchically organized information that describes a unit of records or papers along with its component parts or divisions. It includes information about the

Figure 2. Model for EAD Header <eadheader> Element



content, context, and extent of the archival materials as well as optional supplemental information that facilitates their use by researchers.<sup>16</sup>

When most of us think about archival finding aids, we envision the kinds of hierarchical, multilevel descriptions discussed earlier. Those descriptions form the heart of archival inventories and registers, which generally describe a unit of records or papers at several different, but related, levels of detail. As noted in the design team's first progress report, <archdesc> encompasses these unfolding, hierarchical levels by first allowing for a descriptive overview of the whole, followed by more detailed views of the parts, designated by the element Description of Subordinate Components <dsc>. Data elements available at the <archdesc> or unit level are repeated at the various component levels within <dsc>, and information is inherited from one hierarchical level to the next. As further explained in a set of remarks first prepared for the EAD alpha tag library, <archdesc> not only serves as a wrapper for all the descriptive information about an entire

<sup>16</sup>"Encoding Standard for Electronic Aids: A Report by the Bentley Team for Encoded Archival Description Development," 11.

body of archival materials, it also, through a LEVEL attribute, identifies the highest tier of the materials being described.<sup>17</sup>

Imagine a typical scenario: An archivist begins encoding a finding aid by first opening the <ead> element and creating the required <eadheader>. He or she may then add some optional <frontmatter> before opening the <archdesc> element and setting its required LEVEL attribute to the value "collection," "record group," "fonds," or "series," depending on which term best reflects the character of the whole unit being described in the finding aid. What then follows are data elements that describe that whole unit, including a special subset of core data elements that are gathered together under a parent element called Descriptive Identification <did>. These <did> subelements are among the most important for ensuring a good basic description of an archival unit or component. Grouping these elements together serves several purposes. It insures that the same data elements and structure are available at every level of description within the EAD hierarchy. It facilitates the retrieval or other output of a cohesive body of elements for resource discovery and recognition. And, because the elements appear together in the tag library and on software menus and templates, it helps to remind encoders to capture descriptive information that they may otherwise overlook.

As Figure 1 shows, the <did> element may contain, in any order, one or more of the following descriptive subelements, which are familiar mainstays of archival cataloging: Container <container>, identifying the number of the carton, box, folder, or other holding unit in which the archival materials are arranged and stored;<sup>18</sup> Origination <origination>, denoting the individuals or organizations responsible for the creation or assembly of the archival materials; Physical Description <physdesc>, identifying the extent, dimensions, genre, form, and other physical characteristics; Physical Location <physloc>, identifying the stack number, shelf designation, or other storage location; Repository <repository>, designating the institution responsible for providing intellectual access; Date of the Unit <unitdate>, designating the creation dates of the archival materials; Identification of the Unit <unitid>, containing an accession number, classification number, lot number, or other such unique and permanent identifier; and Title of the Unit <unittitle>, containing the title of the archival materials at whatever level they are being described, such as collection title, series title, subseries title, file title, or item title. The <did> element also provides for the use of both an Abstract <abstract> and a general Note <note> element, as well as for Digital Archival Object <dao> and Digital Archival Object Group <daogrp> elements, which may link to digital surrogates of the material being described in the finding aid. Attributes also are available for all <did> subelements to specify their content further.

Having used the <did> elements to capture a basic description at the <archdesc> level, the archivist may proceed directly to a description of the unit's component parts. More likely, however, the finding aid creator will provide additional narrative information about the content, context, or extent of the whole unit. This description usually appears in prose form within elements with tag names such as <admininfo>, <bioghist>, <scopecontent>, <organization>, and <arrangement>, which are suggestive of the cat-

<sup>17</sup>Janice E. Ruth, "Introductory Remarks About High-Level Elements," in [Lapeyre and Ruth, eds.], "Draft Tag Library for EAD Alpha DTDs," 11-12; reproduced in Gilliland-Swetland and LaPorte, eds., "Encoded Archival Description Document Type Definition (DTD) Beta Version, Tag Library," 6-7.

<sup>18</sup>The <container> and <physloc> elements are new to version 1.0 of the DTD. In the beta version, they were attributes on the <unitloc> element, which no longer exists.

egories of information typically present in traditional paper-based finding aids.<sup>19</sup> For each of these categories of information, the encoder may use the Heading `<head>` element to provide a heading based on local preferences, which may or may not correspond to the element name. For example, the DTD permits encoders to identify a biographical note or agency history by any heading they choose (e.g., Biographical Summary, Biography, Jane Doe's Key Dates) as long as the content is correctly tagged as `<bioghist>`. Structurally, from an SGML perspective, the content models for these narrative-based elements are "heads" and "text," with the latter generally composed of paragraphs or various types of List(s) `<list>`, including the specially created Chronology List `<chronlist>`, consisting of Chronology List Item(s) `<chronitem>` that pair a `<date>` with its corresponding `<event>` to enable linking and tabular display. By comparison, the information within the `<did>` subelements is often presented as a short labeled phrase, or several subelements are pieced together to form a simple uniform data string.

Once an archivist has completed the description of the records or papers at the highest (or unit) level, the `<dsc>` element may be opened, and the focus shifts to describing one or more of the unit's component parts. As explained in the tag library, the `<dsc>` can assume several different forms, which are identified by the element's `TYPE` attribute. The `TYPE` attribute can be set to a value of analytic overview ("analyticover"), to identify a series or subseries description; "in-depth," to identify a listing of containers or folders, a calendar, or a listing of items; "combined," to identify instances in which the description of each series is followed immediately by a listing of containers or folders for that series; and "othertype," to identify models that do not follow any of the above-mentioned formats.

After the form of the `<dsc>` has been selected, the archival components are identified, and a `LEVEL` attribute may be assigned. For example, as shown in Figure 3, an archivist who wishes to provide a summary listing of all the series in a collection may open a `<dsc>`, set the `TYPE` attribute to "analyticover," open a Component `<c>` or `<c01>` (components may be numbered `<c01>` through `<c12>` to keep better track of the hierarchical levels), set the `LEVEL` attribute to "series," and proceed to describe the first series-level component by utilizing the same extensive set of elements that previously were available for describing the whole unit at the `<archdesc>` level. The same procedure would be followed again for the second and all subsequent series-level components, after which point the `<dsc>` element would be closed. In general, certain `<did>` subelements, such as `<repository>` and `<origination>`, are unlikely to be used within a `<c>` because the information they contain has been encoded at the `<archdesc>` level and inherited by the `<c>`. Other `<did>` subelements, such as `<container>`, `<unitdate>`, and `<unittitle>`, will frequently be used within a `<c>` to encode new information or more detailed descriptions at a lower hierarchical level.

As Figure 4 shows, a second `<dsc>` might then be opened with a `TYPE` attribute set to "in-depth" so that a container list can be presented. Each series, subseries, file, or item represented in the container list would be tagged as recursive, nested components, possibly with optional `LEVEL` attributes set to identify their hierarchical order within the collection or record group. As in the series description, information about each component `<c>` may be identified, if desired, by utilizing the full complement of descriptive elements.

<sup>19</sup>The discussion in this and the following paragraphs is based on Ruth, "Introductory Remarks About High-Level Elements."

Figure 3. Tagged Example of <dsc type = "analyticcover"> (Analytic Overview Model)

Description of Series	
Container Nos.	Series
1	Diary and Diary Notes, 1932-34, n.d. A high-school diary and an undated, single-page diary fragment kept by Jackson. Arranged chronologically.
2	Family Papers, 1938-65, n.d. Letters received, notes, and cards. Organized alphabetically by family member and arranged chronologically therein.
3-12	Correspondence, 1936-70, n.d. Letters received and occasional copies of letters sent, telegrams, postcards, and miscellaneous enclosures. Organized alphabetically by correspondent and arranged chronologically therein.
13-19	Literary File, 1943-70, n.d. Correspondence, manuscript drafts, royalty statements, printed matter, notes, outlines, research material, screenplays, and miscellaneous items and enclosures relating to books and short stories by Jackson. Organized alphabetically by type of material and arranged alphabetically by title or topic therein. Publication dates of books are given in parentheses.

Tagged Example	
<dsc type="analyticcover"><head>Description of Series</head><thead><row valign="top"><entry colname="1">Container Nos.</entry><entry colname="2">Series</entry></row></thead>	
<c01 level="series"><did><container>1</container><untitled>Diary and Diary Notes, <unitdate>1932-34, n.d.</unitdate></untitled></did><scopecontent><p>A high-school diary and an undated, single-page diary fragment kept by Jackson.</p><arrangement><p>Arranged chronologically.</p></arrangement></scopecontent></c01>	
<c01 level="series"><did><container>2</container><untitled>Family Papers, <unitdate> 1938-65, n.d.</unitdate></untitled></did><scopecontent><p>Letters received, notes, and cards.</p><arrangement><p>Organized alphabetically by family member and arranged chronologically therein.</p></arrangement></scopecontent></c01>	
<c01 level="series"><did><container>3-12</container><untitled> Correspondence, <unitdate> 1936-70, n.d.</unitdate></untitled></did><scopecontent><p>Letters received and occasional copies of letters sent, telegrams, postcards, and miscellaneous enclosures.</p><arrangement><p>Organized alphabetically by correspondent and arranged chronologically therein.</p></arrangement></scopecontent></c01>	
<c01 level="series"><did><container>13-19</container><untitled>Literary File, <unitdate>1943-70, n.d.</unitdate></untitled></did><scopecontent><p>Correspondence, manuscript drafts, royalty statements, printed matter, notes, outlines, research material, screenplays, and miscellaneous items and enclosures relating to books and short stories by Jackson.</p><arrangement><p>Organized alphabetically by type of material and arranged alphabetically by title or topic therein. Publication dates of books are given in parentheses.</p></arrangement></scopecontent></c01></dsc>	

It is also possible to use within each <c> the <drow> <dentry> display elements mentioned earlier in this article. This structure of endlessly nested components inside a <dsc>, and further, inside <archdesc>, addresses the design team's desire to provide for descriptive information that is inherited from one level to another and that shares or repeats the same essential data elements.

Figure 4. Tagged Example of <dsc type = “in depth”> (In-depth Model)

Container List	
Container Nos.	Contents
LITERARY FILE, 1943-70, n.d.	
46	Bibliographies and publishing lists, 1951-66 Books <i>Raising Demons</i> (1957) Reviews, 1956-57, n.d. Royalty statements, 1956-69
47	<i>The Road Through the Wall</i> (1948), 1947-70, n.d. Short stories and other writings "The Lottery" Dramatic adaptations Correspondence, 1949-53, 1967-70 Scripts and screenplays, n.d. Royalty statements, 1950-53, 1964-70 "Lover's Meeting," n.d.

Tagged Example

```
<dsc type="in-depth"><head>Container List</head><thead><row valign="top"><entry colname="1">Container Nos.</entry><entry colname="2">Contents</entry></row></thead>

<c01 level="series"><did><unittitle>LITERARY FILE, <unitdate type="inclusive">1943-70, n.d.</unitdate></unittitle></did>
  <c02><did><container>46</container><unittitle>Bibliographies and publishing lists, 1951-66</unittitle></did></c02>
  <c02><did><unittitle>Books</unittitle></did>
    <c03><did><unittitle><title render="italic">Raising Demons</title> (1957) </unittitle></did>
      <c04><did><unittitle>Reviews, 1956-57, n.d.</unittitle></did></c04>
      <c04><did><unittitle>Royalty statements, 1956-69</unittitle></did></c04></c03>
    <c03><did><container>47</container><unittitle><title render="italic">The Road Through the Wall</title> (1948), 1947-70, n.d.</unittitle></did></c03></c02>
  <c02><did><unittitle>Short stories and other writings</unittitle></did>
    <c03><did><unittitle><title render="quoted">The Lottery</title></unittitle></did>
      <c04><did><unittitle>Dramatic adaptations</unittitle></did>
        <c05><did><unittitle>Correspondence, 1949-53, 1967-70</unittitle></did></c05>
        <c05><did><unittitle>Scripts and screenplays, n.d.</unittitle></did></c05></c04>
      <c04><did><unittitle>Royalty statements, 1950-53, 1964-70</unittitle></did></c04></c03>
    <c03><did><unittitle><title render="quoted">Lover's Meeting,</title> n.d. </unittitle></did></c03></c02></c01> . . . </dsc>
```

The approach described in the previous paragraph may be termed the “two-<dsc>” model. An alternative to encoding a <dsc TYPE = “analyticover”> followed by a <dsc TYPE = “in-depth”> would be to simply use the <dsc TYPE = “combined”>. This may be referred to as the “combined” model. As Figure 5 shows, the combined model is perhaps a purer manifestation of an unfolding hierarchical description, in that the first component <c01> (in this case, a series) is encoded only once, followed immediately by a fuller description of its nested parts (subseries, files, and items). The combined model avoids the potential confusion of machine-processing identical information that has been encoded twice in the same document, a situation that occurs in the two-<dsc> approach.



Figure 5. Tagged Example of <dsc type = "combined"> (Combined Model)

Container List	
Container Nos.	Contents
LITERARY FILE, 1943-70, n.d.	Correspondence, manuscript drafts, royalty statements, printed matter, notes, outlines, research material, screenplays, and miscellaneous items and enclosures relating to books and short stories by Jackson. Organized alphabetically by type of material and arranged alphabetically by title or topic therein. Publication dates of books are given in parentheses.
46	Bibliographies and publishing lists, 1951-66 Books <i>Raising Demons</i> (1957) Reviews, 1956-57, n.d. Royalty statements, 1956-69
47	<i>The Road Through the Wall</i> (1948), 1947-70, n.d. Short stories and other writings "The Lottery" Dramatic adaptations Correspondence, 1949-53, 1967-70 Scripts and screenplays, n.d. Royalty statements, 1950-53, 1964-70 "Lover's Meeting," n.d.

Tagged Example

```
<dsc type="combined"><head>Container List</head><thead><row valign="top"><entry colname="1">Container  
Nos.</entry><entry colname="2">Contents</entry></row></thead>  
  
<c01 level="series"><did><unittitle>Literary File, <unitdate>1943-70, n.d.</unitdate></unittitle></did>  
<scopecontent><p>Correspondence, manuscript drafts, royalty statements, printed matter, notes, outlines,  
research material, screenplays, and miscellaneous items and enclosures relating to books and short stories by  
Jackson.</p><arrangement><p>Organized alphabetically by type of material and arranged alphabetically by title  
or topic therein. Publication dates of books are given in parentheses.</p></arrangement></scopecontent>  
  
    <c02><did><container>46</container><unittitle>Bibliographies and publishing lists,  
    1951-66</unittitle></did></c02>  
    <c02><did><unittitle>Books</unittitle></did>  
        <c03><did><unittitle><title render="italic">Raising Demons</title> (1957) </unittitle></did>  
            <c04><did><unittitle>Reviews, 1956-57, n.d.</unittitle></did></c04>  
            <c04><did><unittitle>Royalty statements, 1956-69</unittitle></did> </c04></c03>  
        <c03><did><container>47</container><unittitle><title render="italic">The Road Through the Wall</title>  
        (1948), 1947-70, n.d.</unittitle></did></c03></c02>  
    <c02><did><unittitle>Short stories and other writings</unittitle></did>  
    <c03><did><unittitle><title render="quoted">The Lottery</title></unittitle> </did>  
        <c04><did><unittitle>Dramatic adaptations</unittitle></did>  
            <c05><did><unittitle>Correspondence, 1949-53, 1967-70</unittitle> </did></c05>  
            <c05><did><unittitle>Scripts and screenplays, n.d.</unittitle> </did></c05></c04>  
        <c04><did><unittitle>Royalty statements, 1950-53, 1964-70</unittitle> </did></c04></c03>  
    <c03><did><unittitle><title render="quoted">Lover's Meeting,</title> n.d.  
    </unittitle></did></c03></c02></c01> ... </dsc>
```

Depending on the sophistication of a system's searching and processing capabilities, the two-<dsc> approach may hamper the ability to show a relationship between the description of the <c01> and the description of its parts. On the other hand, using the two-<dsc> approach not only readily accommodates a legacy data structure found in many existing

finding aids, it also replicates the functionality which that structure provided. For example, many archivists have found it extremely helpful to assemble in one spot all the first-level component descriptions to provide researchers with a quick overview of the archival unit's content and organization and to permit ready comparisons between components. Flipping through a long paper guide or scrolling and jumping through an electronic finding aid to locate all the first-level summaries is a drawback of the combined model, a problem which an on-line delivery system would need to address.

### Adjunct Descriptive Data <add>

In addition to providing information about the content, context, and extent of the archival materials, the <archdesc> element also includes optional supplemental information that facilitates use of the materials by researchers. This supplemental information, bundled within the Adjunct Descriptive Data <add> element, includes additional access tools to the materials, such as indexes, file plans, and other finding aids, as well as descriptions or lists of materials separated from or related to those described in the finding aid. The <add> element reflects one of the design principles mentioned earlier in this article in connection with the <admininfo> element, namely that EAD accommodates both detailed and "lite" approaches to tagging. Archivists may elect to tag all the adjunct information simply as an <add> containing a series of paragraphs <p>, or they may open the <add> element and encode each piece of information with its specific corresponding tag, such as <bibliography>, <fileplan>, <index>, <otherfindaid>, <relatedmaterial>, and <separatedmaterial>. As a subelement of both <archdesc> and <c>, <add> may appear throughout a finding aid in whatever information sequence best suits the repository's needs. For many encoders, the best sequence will likely be to group all the <add> information together near the end of the finding aid.<sup>20</sup>

### Enhanced Searching Capability Through Access Terms

Aside from encoding the major structural parts of a finding aid and designating the core descriptive data about the unit and its components, users of EAD also have the option of identifying character strings throughout the finding aid that are likely to be the objects of searches, such as personal, corporate, family, and geographic names; occupations; func-

<sup>20</sup>In response to input from early implementers, the EAD developers decided for version 1.0 of the DTD to make the <add> element part of the <archdesc> element rather than maintaining its separate existence under <findaid>. Early implementers of EAD, especially those familiar with using ISAD(G), demonstrated that adjunct descriptive data is not always supplemental "appended" data, as originally conceived by the EAD developers, but instead may be necessary information that is inseparable from other elements of description in a finding aid. Attempting to tease apart, for example, information about related or separated materials from information about the materials featured in the finding aid may not always be possible or desirable. Similarly, finding aid creators may determine that their researchers are better served if a file plan or bibliography relating to a specific unit or component of material appears with other elements of description for that unit or component as opposed to surfacing elsewhere in the finding aid, such as in an appendix. By subsuming <add> as a subelement under <archdesc>, the DTD may become more flexible in handling legacy data and more accommodating of European finding aid practices. The <add> element is now available at various places within <archdesc> to enable finding aid creators to select an information sequence that best suits their needs. The change does not preclude EAD users from following the beta model of gathering at the end of the finding aid all adjunct descriptive data; such users would simply insert the <add> just before the close </archdesc> tag rather than immediately after </archdesc> as was done under the beta DTD. With the change to subsume <add> under <archdesc>, there was no longer any need for the <findaid> element to serve the function of wrapping the <archdesc> and <add> elements. Consequently, <findaid> was eliminated from the DTD.

tions; form and genre terms; subjects; and titles. All of these elements (<persname>, <corpname>, <famname>, <geogname>, <name>, <occupation>, <function>, <genreform>, <subject>, and <title>) permit, through the use of attributes, the designation of encoding analogs and authorized forms as mentioned in the earlier discussion of MARC and ISAD(G). Additional optional attributes allow for specifying the role or relationship of persons and corporate bodies (e.g., author, editor, photographer) and the source of the controlled vocabulary terms used (e.g., *AACR2*, *Library of Congress Subject Headings*, *Library of Congress Name Authority Files*, *Art and Architecture Thesaurus*, *Dictionary of Occupational Titles*). Although the DTD permits liberal access to these elements throughout the finding aid, especially within the <p> and <unittitle> elements, special mention should be made of the ability to bundle them together under the parent element Controlled Access Headings <controlaccess>.

The design team created <controlaccess> specifically to enable authority-controlled searching across finding aids on a computer network. The developers envisioned that users may approach on-line finding aids via a variety of avenues. Some may search a repository's on-line catalog, locate relevant entries, and follow links from those entries to on-line versions of finding aids. Others may start by searching the finding aids directly, bypassing the catalog and losing the advantage of the authority-controlled search terms contained therein. The <controlaccess> element is designed to replicate in a finding aid the collection-level search terms found in the 1xx, 6xx, and 7xx fields of MARC catalog records. Finding aid searches limited to the <controlaccess> element will improve the likelihood of locating strong sources of information on a desired subject, because access terms will have been entered in a consistent and authorized form across finding aids, and also because only the most significant terms are likely to have been selected for encoding.

## Summary

Although EAD has been under development for more than three years and has been significantly improved during that time by feedback from early implementers and by new insights from the design team, much of the DTD's basic structure and approach remains unchanged from the blueprint created during the team's first meeting in July 1995. At the week-long gathering in Ann Arbor, the EAD developers acquired a working knowledge of SGML, articulated their ideas about finding aids, established a set of guiding principles and goals, and created a rudimentary high-level model. Led by the team members most knowledgeable about SGML, the group began tackling three of the most important steps in building a DTD: naming and defining the elements; naming and defining the attributes; and determining where and in what sequence elements may appear. To accomplish these tasks, group members analyzed the structure and functionality of traditional finding aids and made a series of choices about the scope and purpose of their endeavor.

They agreed to create a DTD optimized for authoring new archival inventories and registers, which are a subset of all finding aids, but they also sought to make the structure flexible enough to accommodate existing legacy data. They created only those elements that served a function, and they nested and repeated the elements in an order that reflected the hierarchical, recursive structure of finding aids. The resulting DTD successfully addressed other design goals as well, such as enabling both "lite" and detailed levels of tagging, giving preference to intellectual order over physical arrangement, preserving tabular display features, and accommodating the ISAD(G) and MARC standards.

The model that emerged from the first week's deliberations underwent extensive fine-tuning during the next fourteen months before culminating in the release of a beta version of EAD in September 1996. More than a year of beta testing followed, after which proposals for change were considered by SAA's EAD Working Group, which had been charged with intellectual maintenance of the DTD. The working group will issue version 1.0 of the DTD by late summer 1998. Like earlier versions, it separates information about the finding aid from information about the body of archival materials being described. It encodes the rich, hierarchical descriptions we most closely associate with inventories and registers, while also accommodating the markup of optional supplemental access tools and lists. The DTD's recursive structure allows for multilevel descriptions that begin with a summary of the whole unit and proceed to description of the component parts. Essential descriptive elements are repeated at each level, and linking, display, and search term elements are widely available.

Whether the version 1.0 DTD structure will withstand the test of time is not yet known. After both the alpha and beta test periods, important changes were made to the DTD, and additional revisions are still likely as more archivists, situated in a greater variety of settings, begin to work with version 1.0 of the DTD and contribute to shared or linked databases of finding aids. Only through the continued input and assistance of the entire archival community can EAD develop into a dynamic new tool for accessing and exchanging the wealth of information contained in archival finding aids.