Application of Dublin Core Metadata in the Description of Digital Primary Sources in Elementary School Classrooms

Anne J. Gilliland-Swetland*
Department of Information Studies, University of California at Los Angeles, 212 GSE&IS Building, Los Angeles, California 90095-1521. E-mail: swetland@ucla.edu

Yasmin B. Kafai
Department of Education, University of California at Los Angeles, 2331 Moore Hall, Los Angeles, California 90095. E-mail: kafai@gseis.ucla.edu

William E. Landis
Department of Special Collections, University of California at Irvine Libraries, P.O. Box 19557, Irvine, California 92623-9557. E-mail. blandis@uci.edu

Researchers evaluated the ability of 4th and 5th grade science and social science students to create Dublin Core metadata to describe their own images for inclusion in Digital Portfolio Archives. The Dublin Core was chosen because it provided a systematic, yet minimal way for students to describe these resources at the item level and relate them to collection-level metadata prepared for digitized primary sources by archivists using Encoded Archival Description (EAD). Researchers found that while students were able to supply simple elements such as title and subject with relative ease, they had difficulty moving between general and progressively more granular or refined descriptive elements. Students performed poorly in distinguishing between and completing related, but distinct metadata elements, such as title, subject, and description. Researchers also found that there are still significant issues that need to be addressed if young users in a variety of learning contexts, especially those who are only recently literate, are to be able to make sense of richer metadata such as EAD that is used to describe collections of primary source material.

1. Introduction

This paper reports on the results of research examining the ability of 4th and 5th grade students to use Dublin Core metadata elements to describe image resources that they have created themselves for inclusion in a digital archive.[1] This research was conducted as part of a project at the University of California, Los Angeles (UCLA) that is seeking to enrich elementary science education through the development of multifunctional networked learning environments known as Digital Portfolio Archives (DPA) (Gilliland-Swetland et al., 1998).

The paper begins with a brief overview of the functionality of the DPA model and the rationale for using historical primary sources[2] in this context. It then reviews issues relating to metadata development for historical primary source materials and discusses the DPA Project research findings. It concludes with some questions arising from the research that indicate areas for further study.

* To whom all correspondence should be addressed.

Received December 11, 1998; revised June 30, 1999; accepted June 30, 1999

[1] The authors recognize that the term “archive” when used in its strictest sense refers either to non-current organizational records that are maintained by archivists because of their enduring value; or to the repository that holds those records for the organization. The term is used in this paper, however, to refer to selected digitized archival materials from one collection that have been brought together with descriptive metadata and additional collections of materials generated by teachers and students in the course of their classroom activities to form a single, thematically related “digital archive.”

[2] Although archival sources are, by definition, primary sources, in the context of the conduct and study of science, the term “primary sources” can also refer to current records and data created during the scientific process. To avoid confusion, the authors have used the terms “historical primary sources” and “archival materials” when referring to existing archival collections.

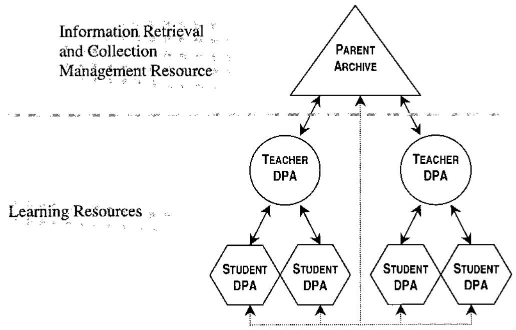

FIG. 1. Relationship between the *parent archive* and *semi-derivative* DPAs in the Digital Portfolio Archives model.

2. The DPA Model

The DPA model delineates an organic, multifunctional, multidirectional, and multiprovenantial resource comprising a parent archive (a digital resource for managing digitized archival resources upon which information retrieval functions can be performed) and unlimited numbers of semiderivative personal or team DPAs developed by teachers and students (digital resources developed by teachers and students to facilitate learning and the management of the resources they have collected as part of their classroom activities). In the model, these semiderivative DPAs hold digitized primary source material *selected* from the parent archive along with materials *created, contributed,* or *collected* by individual teachers and students, or student teams. Semiderivative DPAs might be organized by the creators according to teacher-formulated heuristics or perhaps those of the students themselves after orientation to archival arrangement and description practices. The content of the DPAs would be described using Dublin Core metadata and might include additional user annotation or other commentaries on the content that might be created as a result of teaching, portfolio development, peer commentary, or other learning activities. Creators of the derivative DPAs would have the option of maintaining their own digital portfolio for further review and reflection, and of uploading any of the content that they have created and described into the parent archive (pending review by the teachers and archivists and clearance of student rights issues), where it would be searchable and viewable by other users. All content in the parent archive would be distinguished on the basis of its provenance to avoid intermingling materials deriving from different sources (a potential violation of the archival principle of *respect des fonds*) (see Fig. 1).

The DPA architecture emphasizes the use of contextual representation to reinforce that primary source materials are not only informational but are also direct by-products, and thus recorded evidence, of such human activities as the conduct of scientific research or of the educational process. In other words, the DPA model extends the scope and conceptualization of existing learning systems and digital archives initiatives by acknowledging distinctions between the functions and scope of information-, evidence-, and learning-based systems.

In studying the feasibility of the DPA model for potential use in elementary science and social science education, researchers worked with teachers to build a parent archive by selecting and digitizing materials created by an early California naturalist, Donald Ryder Dickey, and preparing collection and item-level metadata. Collection-level metadata in the form of an archival finding aid describing the entire Dickey Collection was originally created by researchers using the emerging SGML standard Encoded Archival Description (EAD) for digital archival finding aids. The finding aid was subsequently encoded in HTML to facilitate use in the classroom with a standard Web browser, and although both versions contained the same description, there was a loss of the use of document structure to facilitate flexibility in searching and display. Item-level metadata for archival materials that the researchers had selected and digitized from the Dickey Collection was created by researchers using Dublin Core metadata elements. This item-level metadata was required because the finding aid, while hierarchical in its description of the entire collection and clearly indicating the components of the collection from whence the digitized materials had been drawn, was not sufficiently detailed to capture item-level information. Although the researchers could have added this level of detail to the finding aid, they speculated that this very textual and hierarchical approach would not prove to be intellectually accessible to the 4th and 5th grade students with whom the researchers intended to work. Other reasons for using Dublin Core metadata were to associate item-level metadata integrally with each digitized item so that it could always be identified on its own and so that lateral relationships that existed between each digitized item might be documented. This prototype was then used in two science and social science classroom interventions with 4th and 5th grade elementary students. Students, working in teams of three or four, used these metadata elements to develop minimal descriptions of their own photographs which they had taken on a field trip to one of the same wetlands sites documented by Dickey in the early 1920s, and which they had selected for digitization and inclusion in a team DPA.

The goals underlying this part of the DPA research were:

- To understand better the process by which students select their own materials and build their own archives.
- To examine levels of correct description that students can accomplish with ease such as numbering and dating photographs, indicating the creator of the photograph, and indicating where the photograph was taken.
- To examine how the students applied categorization in description, such as in their ability to progressively refine descriptions and distinguish between related data elements.
- To examine students' ability to make connections between the materials in hand and the other materials to which they were exposed.

3. Use of Digital Historical Primary Sources in Science Learning

The use of historical primary source materials has traditionally found its home more within social science than science classrooms and laboratories, despite repeated calls for making socio-historical understanding part of standard science teaching and learning. Many reform efforts such as Project 2061 sponsored by the American Association for the Advancement of Science (AAAS, 1993) and the recent National Science Standards (NRC, 1996) have stressed that historical knowledge should be an important component of scientific literacy. Related to these reform and standards development efforts are curricula that have been implemented in several states and that integrate aspects of science, technology, and society (Hodson, 1988; Bybee et al., 1991; Solomon, 1992; Linn et al., 1994; Linn & Muilenberg, 1996).

The DPA Project is based on the premise that the wealth of existing, primarily nondigital, primary sources such as official records, manuscripts, photographs, and scientific data can and should contribute to the enrichment of elementary science education. Moreover, because of developments in both digital technology and innovative archival descriptive practices, primary sources that were previously inaccessible due to physical condition or remote location can be brought, at least in some of their aspects, into the classroom through the use of networking technology. Fundamental hypotheses of the DPA research are that 4th and 5th grade students' understanding of the processes of science and the development of scientific knowledge can be enhanced through integration of historical scientific documentation into classroom curricula in the form of digital resources; and that these students are sufficiently sophisticated to be able to use historical scientific documentation if they are presented in an understandable and manipulable way. The researchers also speculated that students and teachers might use such materials in distinct ways that many archivists, concerned about the physical and intellectual integrity of their collections and accustomed to the research practices of major user groups such as historians, administrators, genealogists, and the media, have neither anticipated nor encouraged through their contemporary descriptive and technological practices (Gilliland-Swetland, 1998b).

While archivists in some settings, especially the National Archives and states archives and historical societies, have experience working with K-12 through outreach programs to local schools, little is known about how students as young as those in the 4th and 5th grade might use digitized historical primary source materials in the classroom, or how metadata might serve as an enhancement or impediment for such use. Borgman et al. conducted significant work relating to the development of a Science Library Catalog for elementary school students that provided insight into the intellectual and physical representation of bibliographic metadata in an online catalog. The extent to which these findings are transferable to digital archives systems, however, is unclear (Borgman et al. 1991, 1995). A number of recent research projects in science education have recognized that students themselves can be creators of primary sources in the form of raw scientific data, annotations, field notes, or specialized reports. In the University of Michigan MYDL project for middle schools, student-generated value-added contributions such as critiques and abstracts of scientific reports become part of a database to be used by future classes (Bos, 1997; Bos et al., 1997; Lyons et al., 1997; Wallace et al., 1997). In the CSILE project (Scardamalia & Bereiter, 1996), students were contributing science content to a communal database that was enriched by a second layer of annotations made by their classmates. Several other projects such as "Kids as Global Scientists" (Songer, 1997) or microlabs (Berenfeld, 1994) have students contribute original weather data that they have collected for a global database which can be used to graph and analyze weather conditions. In the Collaborative Visualization Project, students use original weather data collected by the National Weather Service to represent, or "visualize," weather conditions together with other students and experts (Gordin et al., 1994). In all of these projects, however, there is no explicit reference to, or use of, historical primary sources. One of the reasons for this apparent absence in K-12 education is that most historical scientific documentation is currently not available in digital format. One of the few research projects to date to look at this issue has been the user needs analysis conducted by the Human-Computer Interaction Laboratory (HCIL) at the University of Maryland for the Library of Congress' American Memory Project. The analysis was conducted to provide data for creating interface prototypes and accompanying tools that would serve a variety of users, including K-12 users (Marchionini et al., 1996). Marchionini et al. looked at a variety of user groups, including K-12 teachers, and encountered both a lack and a diversity in types of descriptive metadata relating to objects and concluded that additional metadata would be needed to support their design aims. In particular, they found that negotiating several levels of hierarchy within a site was not conducive to uncovering "treasures" and that there needed to be ways to get users more directly to the actual digital materials.

4. Applying Dublin Core Metadata to Historical Primary Sources

Metadata applications in archival settings can be broken down into at least five categories: administrative, technical, preservation, descriptive, and use, all of which are important for maintaining the integrity and evidential value of historical collections (Gilliland-Swetland, 1998a). The principle tool for gaining access to archival collections is the finding aid, which gives detailed narrative information about the content, context, structure, and management of the materials being described, and which is designed primarily for use in a setting where access to collections is mediated by a reference archivist. A typical finding aid will document the provenance and original order of the collection (i.e., the way in which the archival materials were arranged by the individual, office, or function responsible for their creation). It may contain several components that fall within the above categories of metadata, including a title page and other front matter; administrative information such as acquisition and ownership data, preservation status, and details regarding any access restrictions; an historical or biographical note providing information about the circumstances of creation of the materials being described; a scope and content note, often including an elucidation of collection arrangement; series and subseries descriptions; container lists, sometimes providing item-level detail; indexes; and information about related collections. In 1995, researchers at the University of California, Berkeley began to develop an SGML Document Type Definition (DTD) that would provide a standardized data structure for archival finding aids that could then be made accessible over the Web with links to existing MARC records. The result was Encoded Archival Description (EAD), a metadata structure standard which allows for hierarchical description at varying levels of aggregation from the collection to the item level. EAD also permits the embedding of pointers to relevant digital objects at any level within that descriptive hierarchy (Hensen, 1995; McClung, 1995).

Archival description—for reasons of both contextualization and the efficient management of voluminous materials—is generally conducted at the collection level, providing primary access points through the provenance, title, and dates of the collection, as well as through summary notes regarding the collection. Image materials, however, frequently challenge this practice since they lend themselves to item-level description and users tend to approach them through subject- rather than provenantially-based queries. When describing image or other material in the form of digital objects, the need for item-level metadata to distinguish between the digital copy and the original object and in some way to certify that copy becomes even more acute and there is arguably a need to invoke an additional metadata standard such as the Dublin Core that can interface with EAD. Moreover, in an environment such as a DPA, where teachers and students are expected to prepare at least minimal descriptions of their own resources for inclusion in the DPA (this does not, of course, rule out the possibility of subsequent value-added description being supplied by archivists), it is unrealistic to expect that they will be able to create metadata using a highly complex standard designed for expert users such as archivists (i.e., EAD).

Again, what is needed is a simpler data structure that can later be mapped onto EAD. This need is also emphasized by Dempsey and Heery: "it seems likely that specialist users will want to search such [expert] data directly, but that to make data more visible to more general "discovery" tools, there may be export of data in some of the simpler formats . . . Indeed, the Dublin Core has been explicitly positioned as a basis for semantic interoperability across richer formats, although it has not yet been widely used in this context" (Dempsey & Heery, 1998, p. 161).

The Dublin Core is conceived of as a simple descriptive format that could be used in the identification and retrieval of a wide range of resources (Dempsey & Weibel, 1996; Weibel & Miller, 1997; Weibel et al., 1998). It allows for enhanced semantic metadata to be embedded into online resources such as HTML, and would also provide a format that might be able to be used to map between different, more complex metadata formats such as EAD (Dempsey & Heery, 1998). Among the principles underlying the Dublin Core metadata set are that any given element should be considered both optional and repeatable and that some of these elements might take on different labels according to the language of different user communities, but would still map onto the same definition of the element. Most research to date has focused on enhancing the use of Dublin Core for resource discovery, particularly on the Internet, and less on how lay users, such as those being studies in the DPA Project, might use it to create descriptive metadata for their own digital objects. The Gateway to Educational Materials (GEM) project, for example, seeks to build on the Dublin Core by creating an enriched metadata set, prototype interfaces, and harvesting tools to facilitate access by K-12 teachers to Internet-based educational resources (Sutton & Oh, 1997).

5. Methodology and Data Analysis

The DPA researchers opted to use the Dublin Core because it provided a systematic yet minimal way for lay creators, in this case elementary school students, to describe their own resources in a way that could later be integrated with EAD. The researchers used the thirteen metadata elements that at the time made up the Dublin Core: *Subject*, *Title*, *Author*, *Publisher*, *OtherAgent*, *Date*, *ObjectType*, *Form*, *Identifier*, *Relation*, *Source*, *Language*, and *Coverage.* In developing the metadata for the digitized materials selected from the Dickey Collection, the researchers' aims were to:

- identify the digitized facsimile and distinguish it from the original item and other similar items (for example, nitrate and glass negatives of the same image) in the Dickey collection;
- to document the place of the original item within the collection;
- to document any relationship between the digitized facsimile and other materials (for example, p. 2 of a 3-page field notebook entry);
- to document the digitization process and the creator of the digitized facsimile; and
- to assign subject access points in addition to the classification scheme used by Dickey (since the specialized scientific terminology used by Dickey was only minimally intellectually accessible to the students and also employed terms that are not in use today).

Table 1. Picture selection for team notebooks for both classes (percentage of total pictures).

	Both classes	
	(%)	Freq.
Selected	79.7	192
Excluded: Technical reasons	6.2	15
Excluded: Information value	7.1	17
Excluded: Subjective criteria	5.8	14
Excluded: Other	1.2	3
	$n = 241$	

The research activities took place in two integrated 4th and 5th grade classrooms in a local urban elementary school associated with the University of California, Los Angeles. Class A and Class B each consisted of 29 students. Each class participated in the research activities during separate 2-week periods. Over each 2-week period, each class met for seven sessions of 45–60 min, participating in both classroom and field trip research activities.

The researchers were interested in looking at 4th and 5th grade students' abilities to document their activities as naturalists by keeping notebooks, in which they recorded observations and taking photographs while they were on researcher-organized class field trips; to select from among that documentation those aspects that they felt were most worthy of retention; to make a written justification for why they did not select certain materials for retention; and then to describe their own photographic images using the Dublin Core elements. Each team of three or four students received a team notebook in which to collect all of the notes and documentation that were part of the research project. The researchers provided students with the following set of worksheets on which they could record this documentation: title page, photo log page, description/annotation pages, and field note pages.

The analysis of the team notebooks took place in three phases. In preparation for digitization each team recorded data regarding their selection or exclusion of photographs taken during the wetlands field trip. In the first phase of data analysis, researchers examined the reasons that students gave for the selection or exclusion of photographs. During the second phase, researchers analyzed the description/annotation worksheets, which each team completed for each photograph they selected for digitization. Finally, three of the descriptive elements from the description/annotation worksheets completed by each team were assessed together in order to explore the students' ability to engage successfully in progressively more detailed description of single images through the use of descriptor fields mapped to Dublin Core metadata elements.

Students selected approximately 80% of all the photographs they had taken for retention and inclusion in their DPAs. Table 1 indicates, by percentage and frequency, the selection and categories of exclusion for photographs across both classes combined. Reasons that the students gave for not including the remaining 20% included technical reasons (for example, the students indicated the poor quality of the photographic image, especially when compared to those of Dickey); lack of information value (for example, the intended subject of their photograph did not show up in as much detail as they had hoped it would); and subjective criteria (for example, the photographic content did not match students' perceptions of what a naturalist should be photographing, particularly housing development or people).

The description/annotation worksheets contained nine categories of descriptive information which students were asked to complete for each picture. Each category was labeled and some contained additional parenthetical information supplied at the request of the teachers in order to give students more help in determining how the categories would be used. Eight of the nine categories were chosen because they corresponded to elements being recommended in the Dublin Core as metadata that would create basic descriptions by novices or experts of digital resources in a networked environment (Dempsey & Weibel, 1996). Since all the Dublin Core metadata elements can be optional and repeatable, the researchers chose to use only those that seemed most likely to be useful to the students. Table 2 contains the mapping of these eight descriptive categories to their seven Dublin Core element equivalents, as well as how that element would be manifested in EAD. The category "who else?" was designed to indicate secondary individuals who were instrumental in the creation of each team's photographs (thus it corresponded to the Dublin Core metadata element *OtherAgent*). However, it was used primarily by the

Table 2. A mapping of item-level Dublin Core metadata elements used in DPA image descriptions to elements in Encoded Archival Description (EAD) available for item-level descriptions in archival finding aids.

DPA description content	Dublin core element	EAD element
Unique picture identifier	Identifier	⟨unitid⟩ in ⟨c0x⟩⟨did⟩
Date on which picture was taken	Date	⟨unitdate⟩ in ⟨c0x⟩⟨did⟩
Description of where picture was taken	Source	⟨note⟩ in ⟨c0x⟩⟨did⟩
Who took the picture?	Author	⟨creator⟩ used at appropriate level within the finding aid
Who else?	Author	⟨creator⟩ used at appropriate level within the finding aid
Relation to other pictures	Relation	In EAD, relationships between items are established by the hierarchical tree structure enforced by the DTD
Subject category	Subject	⟨controlaccess⟩⟨subject⟩ at appropriate level within the finding aid
Title of this picture	Title	⟨unittitle⟩ in ⟨c0x⟩⟨did⟩

students to denote additional team members and, therefore, functioned as a repeated *Author* element.

Each category of descriptive or annotative information on the description/annotation worksheets was coded for analysis. The following is an explanation of the coding for each of these nine categories of information:

The first category, *picture number,* was one in which students were asked to supply a number to each photograph consisting of their team number and a consecutive picture number based on the order in which each team took the pictures. Team responses on the description/annotation sheets for this category were coded as either *blank, contains accurate information,* or *contains inaccurate information.* In all 241 description/annotation sheets, only one picture number was coded as inaccurate, and that because the team members gave two photographs the same number, when only one of them actually matched the information on the team's photo log sheet entry for that picture.

In the second category, *date (picture was taken on),* students were asked to record the date on which the photographs were taken. Team responses were coded as either *blank, contains accurate information,* or *contains inaccurate information.* No inaccurate responses were coded during analysis for this category.

In the third category, *source (describe the place where you took the picture),* students were asked to provide a description of the place in which each photograph was taken. Team responses were coded as either *blank, contains accurate information,* or *contains inaccurate information.* Of the 241 description/annotation sheets, only three contained responses that were coded as inaccurate. In each of these cases, team members provided information that could not be characterized as descriptive of the physical location in which the photograph was taken. Examples of these inaccurate responses are "big," "dry," and "lots of plants."

The fourth category, *author (picture taken by),* students were asked to provide the name of the person who actually took the picture. As previously mentioned, a related question on the description/annotation sheets asked who else assisted in taking the picture. Because the primary research objective of these related questions was to see if students could accurately record the person who had chief responsibility for the creation of each picture, the *who else?* category was not analyzed for the purposes of this report. Team responses were coded as either *blank, contains accurate information,* or *contains inaccurate information.* No inaccurate responses were coded during analysis for this category.

In the fifth category, *relation to other pictures (say if this picture has any other pictures related to it),* students were asked to state if the photograph had others that were related to it. Team responses were coded as either *blank, contains accurate information,* or *contains inaccurate information.* Accurate answers ranged from a simple "yes" or "no," to actual lists of the picture numbers of the related photographs and descriptions of the relationships. Inaccurate responses were ones in which descriptive information did not address the relationship between photographs. Examples of inaccurate responses are "in the water" and "a picture of no one."

The sixth category, *subject category,* was one in which students were asked to provide a subject descriptor for each photograph. A parenthetical list of four examples—plant, animal, landscape, and other—was provided, though students were not instructed that their response had to be from this list. Team responses were coded as either *blank, contains one of the subject categories provided, contains some other accurate subject category,* or *contains an inaccurate subject category.* Inaccurate responses were those which contained a information which was not a subject category. Examples of responses coded as inaccurate are an unintelligible response and "where nests are," which was the subject category provided for a photograph of a tree.

In the seventh category, *title of this picture,* students were asked to provide a descriptive title for each picture. Team responses were coded as either *blank, contains accurate information that is the same as the subject, contains accurate information that is more detailed than the subject, contains accurate information that is less detailed than the subject,* or *contains inaccurate information.* No inaccurate responses were coded during analysis for this category. An example of an accurate response that was less detailed than the subject is the title "plants" where the subject category was "dune lupine."

In the eighth category, *picture (describe what the picture is),* students were asked to describe in detail what each photograph was. Team responses were coded as either *blank, contains accurate information that is the same as the title, contains accurate information that is more detailed than the title, contains accurate information that is less detailed than the title,* or *contains inaccurate information.* Examples of responses that were coded as inaccurate are "it is a yellow flower," when the flower in the photograph was purple, and "11–12." An example of a response that was coded as accurate, but less detailed than the title, was "plants" where the title was "doron loopin" (dune lupine).

The ninth and final category was *describe in detail why you took the picture.* Team responses were coded as either *blank, contains information about why the picture was taken, contains descriptive information that is the same as either the title or picture categories, contains descriptive information that is more detailed than the title or picture categories, contains descriptive information that is less detailed than the title or picture categories,* or *contains inaccurate information.* It should be noted that students were asked to describe *why* instead of *what,* so any responses containing descriptive information about the photograph rather than information about why team members chose to take it were ultimately grouped together as inaccurate for the overall descriptive statistics. There was only one instance where team members provided a response in this category that was coded outright as containing inaccurate information. In this instance, the photograph was of ice plant and the team's reason for taking the picture was "because keeps food cool."

Table 3. Descriptive elements supplied for each picture by team members.

	Total		
	Accurate (%)	Inaccurate (%)	Blank (%)
Picture number	92.5	0.4	7.1
Date	80.9	0.0	19.1
Source	76.3	1.2	22.4
Author	88.4	0.0	11.6
Relation to other pictures	59.0	1.2	39.8
Subject category	95.8	0.8	3.3
Title of this picture	91.3	0.0	8.7
Picture	79.7	1.2	19.1
Describe in detail why you took this picture	37.8	9.2	53.1

$n = 241$

The following are some examples of the more complicated coding of this final category of information from the description/annotation worksheets. Some examples of responses that contain information about *why* the picture was taken are "it was going over a river and it looked nice" (for a photograph of old trolley tracks crossing the wetlands) and "we liked the way the water looked, we thought it was neat that it had a mini forest in it" (for a photograph of dense plant growth underwater in a shallow area of the wetlands). An example of a response in which team members provided *descriptive* information that was more detailed than either the title or picture categories is "gray, white plant that lives in bushes," where the title supplied by the students was "dune luban" (dune lupine) and the information in the picture category was "plant that lives on sand." Table 3 contains percentages for these nine categories with the two classes combined. One can see that in no instances were all descriptive elements supplied for all photographs, although, with the exception of "describe in detail why you took this picture," those elements that were completed were for the most part accurate. The elements which students appeared to find most difficult to complete were those that required them to describe more of the the creative context of the photograph—"relation to other pictures" and "describe in detail why you took this picture."

Finally, an analysis of a combination of three of the categories of information on the description/annotation worksheets provided researchers with an opportunity to focus on students' capabilities to engage in progressively more detailed descriptions (i.e., descriptions which proceed from broad to more detailed information) of photographs taken during the wetlands field trip. Taken as a unit in this order, the categories *subject category, title of this picture,* and *picture (describe what the picture is)* provide some insight into the cognitive abilities of 4th and 5th graders to use descriptive categories in combination in order to document an item in increasing detail.

Table 4 provides data for a consideration of each team's capabilities for engaging in progressively more detailed description and discrimination between different Dublin Core metadata elements, as seen through team members' completion of the description/annotation worksheets for the photographs taken during their field trip. Column I records the percentage of photographs taken by each team for which that team was able to provide an accurate subject category. This analysis ignores the differentiation in coding between a team supplying a subject category from the four examples provided parenthetically on the description/annotation worksheet and an accurate subject category of their own invention; either of these means for supplying subject information were considered by the researchers to be accurate. Column II records the percentage of photographs for which team members were able to provide both an accurate subject category and a more descriptive title. This analysis includes only includes *accurate information that is more detailed than the subject* code for the category of information *title of this picture* as described above. Column III records the percentage of photographs for which team members were able to provide an accurate subject category, a more descriptive title, and a description of the picture that contained even more detail than the title. This analysis includes *accurate information that is more detailed than the title* code for the category of information *picture (describe what the picture is)* as described above.

6. Discussion

Dempsey and Heery have pointed out that those using the Dublin Core will need to be able to perform such actions as progressive refinement of description, distinguishing between categories, assigning names to resources, describing relationships between resources, and distinguishing between the physical and digital manifestation of an object (Dempsey & Heery, 1998, p. 150). In learning contexts, teachers may also need to be able to evaluate the quality of the metadata created by students as an indicator of student learning, progress, or achievement (Bos et al., 1997, p. 2). In terms of assigning metadata to their photographs, students appeared to have few difficulties in supplying simple indicators (for example, the unique identification number, the title, and the subject of the picture were all assigned with above 90% accuracy). Some of the success rate in supplying these metadata elements can be attributed to the researchers instructing the students how to number their photographs sequentially; and also providing the students with a list of sample subject headings and an explanation of what each subject category meant. Other metadata elements, however, proved to be more problematic for these 4th and 5th grade students, including supplying a date, source, author, and description for each image. In these cases, students were being asked to supply the relevant information themselves, rather than identify the correct information derived from options provided by the researchers.

Students appeared to find most problematic a non-Dublin Core element in which they were asked to provide a brief rationale for why their team had taken each photograph.

Table 4. Progressively more detailed description by student teams.

		Classes					
		I: % of *n* assigned an accurate subject description by team members		II: % of *n* assigned an accurate subject description *and* then given a more descriptive title		III: % of *n* assigned an accurate subject description, given a more descriptive title *and* then given a more detailed description of what picture was	
Class A	Class B	Class A	Class B	Class A	Class B	Class A	Class B
TEAM 1 (n = 22)	TEAM 1 (n = 18)	95.5	88.9	90.9	0.0	40.9	0.0
TEAM 2 (n = 13)	TEAM 2 (n = 6)	100.0	100.0	7.7	33.3	7.7	33.3
TEAM 3 (n = 19)	TEAM 3 (n = 12)	100.0	58.3	100.0	58.3	42.1	0.0
TEAM 4 (n = 18)	TEAM 4 (n = 11)	100.0	100.0	66.7	100.0	27.8	27.3
TEAM 5 (n = 9)	TEAM 5 (n = 18)	88.9	100.0	0.0	38.9	0.0	16.7
TEAM 6 (n = 10)	TEAM 6 (n = 12)	100.0	100.0	40.0	100.0	10.0	100.0
TEAM 7 (n = 10)	TEAM 8 (n = 11)	100.0	100.0	80.0	100.0	40.0	72.7
TEAM 8 (n = 16)	TEAM 9 (n = 13)	100.0	100.0	50.0	84.6	12.5	61.5
TEAM 9 (n = 23)		95.7		52.2		17.4	

Note. n = number of pictures.

Also problematic was the metadata element in which they were asked to indicate the relationship between the image in hand and other images generated by the same activity (i.e., indicating that they were aware of the collectivity of the documentation). Two possible explanations for these increasing levels of difficulty evidenced by students might be that some of the students are not yet very comfortable with writing, especially when it does not involve copying from supplied texts, and also the students' lack of familiarity with working with primary sources. However, a developmental issue may be indicated by the fact that the students had difficulty moving between general and progressively more granular or refined descriptive elements (although it should be noted that there is little research evidence that adult lay people would be any more capable of such progressive refinement of description using the Dublin Core). Students performed poorly in distinguishing between and completing related, but distinct metadata elements, for example, title/subject/and description. In cases such as that of Team A5, students simply repeated the same metadata in all three elements, and in others, such as that of Team B1, students assigned an accurate subject description to their photographs, but failed to complete any of the title and detailed description elements. However, as Table 4 also indicates, the task was accomplished by some students in most instances. Team B6, for example, showed 100% accuracy in differentiating between these three elements and giving progressively more detailed descriptions of their photographs.

Students' understanding of *archival presentation* of the primary source materials was an important aspect in this evaluation. Students were given collection-level metadata in the form of an archival finding aid and digital copies of materials selected from the Dickey Collection described using item-level Dublin Core metadata. The researchers found that most students, when given the task of using the Dublin Core metadata in simple resource discovery were able to do so (e.g., find all examples of images taken in Ventura County). The researchers found that few students had the interest or the stamina to read the actual finding aid provided online and did not use it as a means to locate materials. While this aspect of evaluation was not directly part of this research project, researchers speculated that this might be due to the general difficulties of 4th and 5th grade students in reading or in negotiating lengthy text on the computer screen.

7. Conclusion and Further Research Questions

Despite new descriptive standards such as EAD and the Dublin Core that are making archival metadata considerably more standardized and Web-accessible, this research indicates that, while young students are indeed able to work with historical primary sources in their learning activities, there are still significant aspects of metadata development that need to be addressed in order to assist those users in understanding the nature and context of existing primary source content, as well as in describing the content they are themselves developing. Two aspects in particular present sets of questions that deserve further research:

(1) Young children cannot always distinguish between related Dublin Core data elements for the purposes of either metadata creation or resource discovery; and when describing their own materials, children may not be able to supply sufficiently detailed information within an element to make it useful for information retrieval purposes. Can the Dublin Core, therefore, play a useful role in the description and location of digital resources by young children, without becoming too generic (in terms of granularity of description) or limited (in terms of the number of data elements applied or applicable)?

(2) Finding aids may have too much text for young readers and may be written using the technical language associated with the activity of the creator of the collection (in Dickey's case, contemporary naturalism) as well as the professional jargon of archival practice. Could the language used by the professionals who prepare collection and item-level metadata be mapped automatically to children's vocabularies to facilitate metadata creation as well as resource discovery? Could the hierarchical design of archival finding aids be further exploited by information visualization systems to render alternative spatial ways of navigating primary source collections that might circumvent a dependence upon student literacy skills and also benefit users who are less textually oriented?

The Digital Portfolio Archives in Learning Project was funded by the National Science Foundation Collarborative Research in Learning Technologies, NSF CRLT #96-16396.

References

American Association for the Advancement of Science (1993). Benchmarks for science literacy. New York: Oxford University Press.

Berenfeld, B. (1994). Technology and the new model of science education: The global lab experience. Machine-Mediated Learning 4 (2 and 3), 203–228.

Borgman, C. L., Walter, V. A., Rosenberg, J. B., & Gallagher, A. L. (1991). Children's use of a direct manipulation library catalog. ACM SIGCHI Bulletin, 23, 69–70.

Borgman, C. L., Gallagher, A. L., Hirsh, S. G., & Walter, V. A. (1995). Children's searching behavior on browsing and keyword online catalogs: The science library catalog project. Journal of the American Society for Information Science.

Bos, N. (1997). Student publishing of value-added contributions to a digital library. In E. Soloway (chair), Using online digital resources to support sustained inquiry learning in K-12 science. AERA. http://mydl.soe.umich.edu/papers/.

Bos, N., Krajcik, J., & Soloway, E. (1997). Student publishing in a WWW digital library—goals and instructional support. In P. Bell (chair), Artifact-building in computer learning environments: Supporting students' scientific inquiry. AERA. http://mydl.soe.umich.edu/papers/.

Bybee, R. W., Powell, J. C., Ellis, J. D., Giese, J. R., Parisi, L., & Singleton, L. (1991). Integrating the history and nature of science and technology in science and social studies curriculum. Science Education, 75, 143–155.

Dempsey, L., & Heery, R. (1998). Metadata: A current view of practice and issues. Journal of Documentation, 54, 145–172.

Dempsey, L., & Weibel, S. (1996). The Warwick Metadata Workshop: A framework for the deployment of resource description. D-Lib Magazine (July/August). http://www.dlib.org/dlib/july96/07weibel.html.

Gilliland-Swetland, A. J. (1998a). Defining metadata. In Introduction to metadata: Pathways to digital information. (pp. 1–8). Los Angeles: Getty Information Institute.

Gilliland-Swetland, A. J. (1998b). An exploration of K-12 user needs for digital primary source materials. American Archivist, 61, 136–157.

Gilliland-Swetland, A., Kafai, Y., & Maddox, A. (1998). Digital Portfolio Archives in Learning: Modeling Primary Content Transformation for Elementary Science Education. Final Report to the National Science Foundation Collaborative research on Learning Technologies Program, August 1998. Award number 96-16396.

Gordin, L., Gomez, L. M., Pea, R. D., & Fishman, B. J. (1994). Using the World Wide Web: Learning communities in K-12. http://www.covis.nwu.edu/Papers/k12web.html#biblio.

Hensen, S. L. (1995). NISTF II: The Berkeley Finding Aids Project and New Paradigms of Archival Description and Access. Paper presented at the Berkeley Finding Aids Conference, April 4–6, Berkeley, California. http://sunsite.berkeley.edu/FindingAids/.

Hodson, D. (1988). Toward a philosophically more valid science curriculum. Science Education, 72, 19–40.

Linn, M. C., diSessa, A., Pea, R. D., & Songer, N. B. (1994). Can research on science learning and instruction inform standards for science education? Journal of Science Education and Technology, 3, 7–15.

Linn, M. C., & Muilenburg, L. (1996). Creating lifelong science learners: What models form a firm foundation. Educational Researcher, 25, 18–23.

Lyons, D., Hoffman, J., Krajcik, J., & Soloway, E. (1997). An Investigation of the Use of the World Wide Web for On-Line Inquiry in a Science Classroom. Paper presented at the Meeting of the National Association for Research in Science Teaching, Chicago, IL.

Marchionini, G., Plaisant, C., & Komlodi, A. (1996). User Needs Assessment for the Library of Congress National Digital Library (1996). ftp://ftp.cs.umd.edu/pub/hcil/Reports-AbstractsBibliography/3640html/3640.html.

McClung, P. (1995). Access to Primary Sources: During and After the Digital Revolution. Keynote Address for the Berkeley Finding Aids Conference, April 4–6, Berkeley, California. http://sunsite.berkeley.edu/FindingAids/.

National Research Council (1996). National Science Education Standards. Washington, DC: National Academy Press.

Scardamalia, M., & Bereiter, C. (1996). Student communities for the advancement of knowledge. CACM, 39, 36–37.

Solomon, J., Duveen, J., & Scot, L. (1992). Teaching about the nature of science through history: Action research in the classroom. Journal of Research in Science Teaching, 29, 409–421.

Songer, N.B. (1996). Knowledge construction through global exchange and dialog: A case of kids as global scientists. Journal of Learning Sciences 5(4), 297–328.

Sutton, S., & Oh, S. (1997). GEM: Using metadata to enhance Internet retrieval of educational materials by K-12 teachers. Bulletin of the American Society for Information Science, 24, 21–24.

Wallace, R., Krajcik, J., & Soloway, E. (1997). Digital libraries in the science classroom: An opportunity for inquiry. D-Lib Magazine. http://www.dlib.org/dlib/september96/umdll/09walace.html.

Weibel, S., & Miller, E. (1997). Image description on the Internet: A summary of the CNI/OCLC Image Metadata Workshop, September 24–25, Dublin, Ohio, D-Lib Magazine. http://www.dlib.org/dlib/january97/oclc/01weibel.html.

Weibel, S., Kunze, J., & Lagoze, C. (1998). Dublin Core metadata for simple resource discovery, Internet draft, February 10 Work in progress. ftp://ftp.ietf.org/internet-drafts/draft-kunze-dc-01.txt.