Journal of Library Metadata
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wjlm20

Interoperability and Metadata Quality in Digital Video Repositories: A Study of Dublin Core

Julie Weagley [a], Ellen Gelches [a] & Jung-Ran Park [b]
[a] Independent Scholars,
[b] College of Information Science and Technology, Drexel University, Philadelphia, Pennsylvania, USA
Published online: 17 Feb 2010.

To cite this article: Julie Weagley, Ellen Gelches & Jung-Ran Park (2010) Interoperability and Metadata Quality in Digital Video Repositories: A Study of Dublin Core, Journal of Library Metadata, 10:1, 37-57, DOI: 10.1080/19386380903546984

To link to this article: http://dx.doi.org/10.1080/19386380903546984

Journal of Library Metadata, 10:37–57, 2010

ISSN: 1938-6389 print / 1937-5034 online
DOI: 10.1080/19386380903546984

Routledge
Taylor & Francis Group

Interoperability and Metadata Quality in Digital Video Repositories: A Study of Dublin Core

JULIE WEAGLEY and ELLEN GELCHES
Independent Scholars

JUNG-RAN PARK
College of Information Science and Technology, Drexel University, Philadelphia, Pennsylvania, USA

The purpose of this paper is to examine the interoperability of the Dublin Core metadata element set across a selection of digital video repositories. To explore how the metadata was applied and analyze its overall quality, 150 records were selected, 25 from each of 6 collections, and evaluated for use of the original 15 Dublin Core elements. Each record was examined for metadata completeness, accuracy, consistency, and usage of controlled vocabularies. Potential barriers to interoperability are identified, most notably the confusion caused by semantic ambiguity among DC elements.

KEYWORDS digital video repository, Dublin Core, interoperability, controlled vocabularies, metadata quality

INTRODUCTION

As the phenomenon of YouTube suggests, more and more digital video is becoming available on the Internet, both material that is born digital and material that is digitized. The diversity and quantity of the resources is beyond measure—from educational video to stock footage, from news clips to historical documentaries, from personal videos to cultural heritage works. For video, as with all digital resources, the key is finding and retrieving the material—and metadata is central to that goal.

Like digital video itself, the choice in metadata is diverse. Candidate schemes vary in complexity and granularity and their use depends greatly on the needs of the collection and the potential end user. Dublin Core (DC),

Address correspondence to Julie Weagley, 1200 Wilson Drive, West Chester, PA 19380, USA. E-mail: julieannweagley@comcast.net

the most widely used metadata schema for electronic resources, was originally designed to allow high-level searching of textual documents across disciplines, databases, and schemas (Hunter & Iannella, 1998). Its beginnings were in a 1995 workshop held by a group of librarians, digital information specialists, content providers, and text-markup experts who gathered to improve information discovery on the Internet.

At its heart are two versions of Dublin Core. Simple, or unqualified, Dublin Core consists of 15 elements, all optional and repeatable. These 15 elements describe content, intellectual property, and instantiation. Content is tagged in seven elements—Coverage, Description, Type, Relation, Source, Subject, and Title. Four elements—Contributor, Creator, Publisher, and Rights—involve intellectual property. Instantiation is cataloged in four elements—Date, Format, Identifier, and Language (Duff, 2003). In an effort to gather more granular information and to provide extensibility, qualifiers were later devised for the schema. Qualified Dublin Core allows each domain to customize the scheme for its own use. Another characteristic (and advantage over a traditional metadata scheme like MARC 21) is the inclusion of rudimentary structural and administrative metadata in the Dublin Core metadata scheme (mostly in Qualified Dublin Core).

The Dublin Core metadata scheme encompasses the characteristics of simplicity in metadata creation and maintenance, flexibility and extensibility, coarse-grained descriptions for interoperability, and multilingual and multicultural description capability for international scope (Collaborative Digitization Program [CDP], 2006). The initiative lists no restrictions to the types of resources to which DC elements can be applied. Beyond the guidelines established by Dublin Core, local organizations may also develop additional restrictions or rules in applying the Dublin Core elements, as stated in the National Information Standards Organization (NISO) guidelines.

The simplicity of Dublin Core makes it easy to implement, which creates a high degree of compatibility across multiple repositories. This characteristic can be both a strength and a weakness. “Simplicity lowers the cost of creating metadata and promotes interoperability but does not accommodate the semantic and functional richness supported by complex metadata schemes. In effect, the Dublin Core element set trades richness for wide visibility” (NISO, 2007, p. vi). It is ironic that the simplicity and ambiguity of the elements that were intended to improve interoperability can be a hindrance to interoperability as well.

The purpose of this paper is to examine the interoperability of the Dublin Core metadata element set, review the quality of application across a variety of streaming video collections, and analyze the usefulness of the metadata for high-level resource discovery.

Previous studies have addressed the quality of metadata as applied to digital resources and its effect on interoperability. In her overview of the literature examining metadata quality, Park (2009) states that metadata quality can be judged by how well it performs “core bibliographic functions of

discovery, use, provenance, currency, authenticity, and administration"—she identifies completeness, accuracy, and consistency as the most commonly used criteria to evaluate metadata quality. Both Shreeves et al. (2005) and Bui and Park (2006) found that inconsistencies in metadata usage are caused in part by the variety of contributing collections and the resulting variations in quality of metadata application. Bui and Park also point to the contributing collections' use of a number of metadata schemes other than Dublin Core and the inconsistent mapping from those schemes to Dublin Core as a cause of metadata application inconsistency.

Bearman, Miller, Rust, Trant, and Wibel (1999) and Godby, Smith, and Childress (2003) both point out the semantic overlap in several Dublin Core elements, which can allow them to be used interchangeably. Studies by both Park (2005) and Zeng, Subrahmanyam, and Shreve (2004) have discussed the confusion in the use of the Format and Type elements. The confusion stems from the fact that the DCMI definitions for both elements include references to physical or digital manifestations [representations]. Consequently, either could be mapped to the MARC physical description field. As Park states, "boundaries among these elements are fuzzy and not clear cut" (2006). Both studies also surfaced confusion between the Source and Relation elements. Two qualified Relation elements—Relation, is Part of, and Relation, is Version of—drift into the meaning of the unqualified element, Source, "a reference to a resource from which the present resource is derived" (NISO, 2007, p. 4). Source can be seen as a type of Relation (Park, 2006). Difficulties with the application of Creator, Contributor, and Publisher elements also result from these elements' definitions not being mutually exclusive. "According to the . . . definitions, Creator can be seen as both a particular type of Contributor and Publisher" (Park, 2006).

This paper continues the research into metadata quality by focusing on the application of Dublin Core to streaming video. To do so, the researchers investigated a wide range of digital video repositories to identify those that used Dublin Core. From the small number identified, a choice was made based on the type of content as well as the method by which the metadata was harvested. Repositories selected run the gamut from a federated catalog to a small, single-source collection designed to highlight the works of its constituents and local artists. The means by which the repositories gather metadata are equally as diverse ranging from sophisticated content management systems like CONTENTdm to more proprietary methods.

OVERVIEW OF COLLECTIONS

Australian Centre for the Moving Image (ACMI). According to its site, ACMI "celebrates, explores and promotes the cultural and creative richness of the moving image in all its forms" (ACMI, n.d., Our Story section, ¶ 1). ACMI offers a lending library of films, special events, and resource materials about

film and television, as well as classes in filmmaking. The organization's site also features short films created by students in ACMI production workshops, videos of talks and programs, and 30-second works created by contemporary Australian artists (ACMI, n.d.). While it does not have quantities of video, it has a decent selection that can benefit from the use of metadata both for search on the site and to ensure that these works are findable if the collection were to merge with another repository or archive. ACMI currently codes the pages and resources on its site with Dublin Core; this information is available in the source code.

Ball State University Libraries' Digital Media Repository. The Digital Media Repository of Ball State University "provides a centralized, coordinated, and user-focused resource to serve the teaching, learning, and research needs of students, faculty, and researchers at Ball State and beyond" (Ball State University Libraries, n.d., About the Ball State University Libraries' Digital Media Repository section, ¶ 1). The repository unifies all of the libraries' digital assets as well as their digital activities so that they are searchable and accessible from one Web site. Currently the repository comprises 28 collections, which can be browsed by department, by collection, or alphabetically. Simple Search and Advanced Search use the typical union catalog approach of providing the fields available within those collections selected. CONTENTdm is the content management system for the Digital Media Repository.

The film and video collections within this repository chosen for this project are Ball State Student Filmmakers, Commencement Collection, NewsLink Indiana Videos, WIPB-TV Documentary Video Collection, and WWII Historical Films.

Ball State University Libraries' Digital Media Repository site has a page providing information for prospective submitters of resources to the repository. The site lists Dublin Core as its metadata schema and provides a link to DCMI. James Bradley, head of Metadata & Digital Initiatives, supplied the researchers with Ball State's Cardinal Core DMR Metadata Schema, which lists required fields, conditionally required fields, recommended fields, and optional fields and maps those fields to Dublin Core. As Mr. Bradley explains, "The Cardinal Core elements were developed over time to meet the growing/changing needs of OCLC's CONTENTdm product (and our modifications of it), adaptations to our traditional Archival and Special Collections practices, OAI-PMH and other aggregators or Web searching providers, and the desire to remain consistent with the DC standards and best practices" (personal communication, June 26, 2009).

Folkstreams.net. Folkstreams.net is a national preserve of close to 100 documentary films on the American folk, or roots, culture, spanning the time period from the 1960s to the present (Folkstreams, Inc., 2006). One of the repository's goals is to make these films accessible to a new audience. Films can be browsed by title, filmmaker, distributor, subject, and region.

The Advanced Search feature searches 11 fields. Extensive information on each film is available, information that was compiled before Dublin Core metadata was applied. Kristina M. Spurgin, a PhD student who served as the ibiblio.org Folkstreams grant manager from 2004 to 2007, explained that there are no internal guidelines for applying Dublin Core metadata. "We have a behind-the-scenes database that keeps everything organized and generates the Web site. This database and its design were in use before any information professionals were brought on the project and is designed to meet the needs and interests of the filmmakers. The fields and values do not adhere to any standard metadata scheme, and it was decided that instead of changing the database structure, we would map values from various fields to standard metadata schemes when necessary" (personal communication, May 30, 2007).

Folkstreams lists as its institutional partners ibiblio.org at the University of North Carolina at Chapel Hill and that university's School of Information and Library Science. In addition, Folkstreams.net is a contributor to OAIster, although no mention of that is made on the Folkstreams site.

NJVid. The New Jersey Digital Video Repository was launched February 23, 2009, as an online digital video portal and centralized repository that houses and archives video material from a wide range of educational and cultural institutions in the state of New Jersey. The repository supports three collections:

- locally owned material that is submitted by an organization and can be viewed by all audiences with no restrictions
- permission-restricted, commercially licensed or distributed video
- non-commercial educational materials intended for a specific audience and access restricted

The NJVid platform allows participating institutions to provide its audience access to streaming video even if they are unable to support the technology locally. In addition, NJVid offers "preservation quality management and archiving" of digital content (NJVid, n.d., Frequently Asked Questions section, ¶ 6).

Currently the open-access video on the site can be browsed by sixteen broad subject categories and eight contributing organizations. Records are viewable in MARC, Dublin Core, and MODS. Rights and technical metadata for the video are also available. Metadata for new videos are either entered by NJVid personnel based on a Metadata Deposit Form filled out by the video submitter or entered directly by the video submitter through the Workflow Management System (WMS) (NJVid, 2010).

Shannon County Film. Shannon County Film is a collection of raw footage shot as part of a documentary project in the late 1970s. Robert Moore, the director/producer of the project, approached Missouri State

University about preserving the original film rolls created in producing the documentary. The film was archived and also digitized so that the materials could be made accessible as streaming video on the Missouri State University Special Collections and Archives site (McCroskey, 2006). Additionally, plans were made with the Missouri Digitization Planning Project to offer the footage as streaming video through the Virtually Missouri Web site, a centralized, searchable database of digital collections from cultural and scientific heritage institutions across the state of Missouri. (Virtually Missouri has been superseded by Digital Missouri.) As of April 2009, 106 rolls of the raw footage have been digitized.

The metadata records for the original 45 rolls were first created in MARC based on the production notes of Robert Moore. These original MARC records were then used as a template for professional MARC catalogers (McCroskey, 2006). Since the records were required to be in Dublin Core for the Virtually Missouri project, Moore then created the Dublin Core records based on the MARC to Dublin Core crosswalk (after additional research) and after reviewing the Virtually Missouri guidelines. When Missouri State University purchased CONTENTdm, Moore then input the metadata using that system (R. Moore, personal communication, May 25, 2007).

Currently the records appear in the Digital Collections section of Missouri State University. Each record displays cataloger assigned CONTENTdm labels that may or may not clearly represent the Dublin Core elements. The records are browseable by thumbnail image, title, subject, and summary. The records are searchable by keyword by means of a basic and an advanced search.

Windows on Maine. Windows on Maine calls itself a searchable Web database of video programs and clips, and other primary and secondary digital resources (Windows on Maine, n.d.). It is designed especially for use by students and teachers interested in the history, culture, and legacy of the state of Maine, aggregating resources that have subjects pertinent to the issues of that state. Contributors to the collection include Maine Public Broadcasting Network, Fogler Library at the University of Maine, Maine State Museum, Maine State Archive, Maine Historical Society, Northeast Historic Film, and Maine Folklife Center (Windows on Maine, n.d.). The site makes its content searchable by subject via a drop-down menu and keyword in the title, subject, description, and transcript. The material can also be located by resource type or date and by town, city, or reservation name. It can also be browsed by collection.

Windows on Maine acts as a union catalog, with an interactive cataloging interface to facilitate delivery of metadata from contributing collections. The metadata can originate in a variety of schemes and are first mapped to MARC and then to Dublin Core (Library Application Profile), the metadata profile for the collection. At that point, collection curators can revise the metadata as needed (Windows on Maine, n.d).

METHODOLOGY

To analyze how the metadata were applied and the overall quality, 25 records were selected from each of the collections and evaluated for use of the original 15 Dublin Core elements. Each record was examined for metadata completeness, accuracy, consistency, and usage of controlled vocabularies. The elements were considered used if they could be inferred from the content on the site or from the URI, if the metadata existed in the source code, or if the repository forwarded documentation that indicated the elements were used. For comparison purposes, if the collection used qualifiers, the elements were rolled up to the 15 in Simple Dublin Core. The usage of elements in each repository is illustrated in Table 1.

In only two repositories, Folkstreams and ACMI, are the DC elements visible in the source code. The records on the NJVid site are displayed in MARC; however, a dialog box offers the Dublin Core, MODS, Rights, and Technical records as well. For the purposes of this paper, the NJVid records were reviewed in DC Format.

Ball State and Shannon County use CONTENTdm, so assumptions had to be made about some of the elements. Ball State forwarded to the researchers its Cardinal Core documentation mapping Ball State elements to DC. For Shannon County, a sample record with MARC to Dublin Core mapping was used as a reference. Other MARC records were reviewed on the Missouri State University Library site and the MARC to Dublin Core crosswalk consulted to confirm mapping of the data. If elements were harvested by the repository, but did not display on the site, they were not counted as used, because metadata quality could not be assessed.

In response to the researchers' request, Windows on Maine forwarded Dublin Core records for a random selection of videos. Since the elements as listed on the exported records were not the only elements harvested as Dublin Core by the repository, other elements were extrapolated from the site. As documented by the MARC to Dublin Core crosswalk, the values listed under the label Contributor on the site have been mapped to Creator; this strategy was confirmed by the staff at Windows on Maine. As with Shannon County, if elements were collected but did not display on the site, they were not counted as used.

FINDINGS AND DISCUSSION

Once the elements from the individual collections were mapped to Dublin Core, usage of the terms was tallied, as illustrated in Table 2.

Completeness

According to Park (2009), the completeness of metadata description can be measured by full access capacity to individual local objects and connection to

TABLE 1 Use of Dublin Core Elements Across Selected Repositories

	ACMI	Ball State	Folkstreams	NJ Vid	Shannon County	Windows on Maine
DC.Title	dc.title	Title	Title	Title	Title	Title
DC.Title						
DC.Creator	dc.creator	Creator	Filmmaker		Creator	Creator
DC.Creator		Created by				Annotator
DC.Creator						Arranger
DC.Creator						Artist
DC.Creator						Associated name
DC.Creator						Author of dialog
DC.Creator						Author of introduction
DC.Creator						. . .
DC.Subject	dc.subject	Subject	Film subjects	Subject	Subject	Subject
DC.Subject		Geographic Subject				
DC.Subject						
DC.Subject						
DC.Description	dc.description	Summary	Description	Description	Summary	Description
DC.Description		Segments			Table of contents	
DC.Description		Description			Description	
DC.Description		Run time				
DC.Description		Length				
DC.Description		Special thanks				
DC.Publisher		Publisher	Not mapped from display label/hard-coded from internal database	Publisher	Publisher	Publisher
DC.Contributor	dc.publisher	Director		Contributor	Contributor	
DC.Contributor		Screenwriter				
DC.Contributor		Writer				
DC.Contributor		Cast				
DC.Contributor		Producer				
DC.Contributor		Executive Producer				
DC.Contributor		Associate Producer				
DC.Contributor		Narrator				
DC.Contributor		Directory of photography				

DC.Contributor		Videographer				
DC.Contributor		Photography				
DC.Contributor		Production Coordinator				
DC.Contributor		Editor				
DC.Contributor		Art director				
DC.Contributor		Graphics				
DC.Contributor		Lighting				
DC.Contributor		Sound				
DC.Contributor		Research				
DC.Contributor		Script Consultants				
DC.Contributor		Music				
DC.Contributor		Persons interviewed				
DC.Contributor		Newscasters				
DC.Contributor						
DC.Contributor		Partners				
DC.Date	dc.date.modified	Publication Date	Release date	Date	Date.original	Date
DC.Date		Date			Date.digital	Date published
DC.Date		Original date				Copyright date
DC.Date		Digital date				
DC.Type		Media type	Not mapped from display label/hard-coded from internal database	Type	Type	Type
DC.Type		Original Physical Format				
DC.Format		Digitization specifications	Not mapped from display label/hard-coded from internal database	Format	Format	Duration
DC.Format					Requirements	Color
DC.Format						Sound
DC.Format						Size
DC.Format						

(Continued on next page)

TABLE 1 Use of Dublin Core Elements Across Selected Repositories *(Continued)*

	ACMI	Ball State	Folkstreams	NJ Vid	Shannon County	Windows on Maine
DC.Identifier		Link to Streaming Video	Film number	Identifier	Identifier	Identifier
DC.Identifier		Digital identifier				
DC.Identifier		URL				
DC.Identifier		Item identifier				
DC.Identifier						
DC.Source		Repository			Source	Source
DC.Source						
DC.Source						
DC.Language	dc.language	Language	Not mapped from display label/hard-coded from internal database	Language		
DC.Relation		Archival identification		Relation	Relation	Find similar resources
DC.Relation		Collection				Related Resources
DC.Relation		Collection URL				This resource is part of
		Subcollection				
		Subcollection URL				
DC.Coverage		Decade		Coverage		Geographic area
DC.Coverage		Geographic Location				Coverage
DC.Rights	dc.rights	Copyright		Rights		Rights
Locally added						
Locally added						
Locally added						
Locally added						

TABLE 2 Usage Totals for Dublin Core Elements Across Selected Repositories

Element Name	ACMI	BSU	Folkstreams	NJVid	Shannon County	Windows on Maine	% Used
Title	25	25	25	25	25	25	100%
Description	25	25	25	25	24	25	99%
Date	25	25	25	25	25	19	96%
Identifier	0	25	25	25	25	25	83%
Type	0	25	25	25	25	25	83%
Relation	25	25	0	25	25	25	67%
Format	0	0	25	25	25	24	66%
Subject	0	22	25	25	20	25	65%
Rights	25	25	0	25	25	20	63%
Language	25	10	25	25	0	0	57%
Publisher	25	8	25	5	25	14	51%
Creator	25	14	0	0	25	18	38%
Contributor	0	11	24	22	25	0	38%
Source	0	23	0	0	25	24	31%
Coverage	0	1	0	21	0	10	21%

the parent local collection(s). This reflects the functional purpose of metadata in resource discovery and use. The completeness of metadata description seems to be conditioned by characteristics of the resource type within a given domain and specifically by local metadata guidelines and best practices.

Of the fifteen elements suggested by Dublin Core, only Title is supplied across the repositories 100% of the time. Following Title are Description (99%), Date (96%), Identifier (83%), Type (83%), and Relation (83%). Each of these is collected by five of six repositories at the 100% level, but was harvested in varying degrees by the sixth repository. Four elements are collected at less than 50%, Creator, Contributor, Source, and Coverage. All elements except Coverage are gathered at 100% by at least one repository.

ACMI has the biggest impact on frequency of usage. Of the original 15 Dublin Core elements, ACMI collects only eight, which it does at 100%: Title, Description, Date, Relation, Rights, Publisher, Language, and Creator. The site does not collect Dublin Core elements Subject, Type, Source, Coverage, Contributor, Format, and Identifier. Similarly, Folkstreams chose not to harvest five DC elements, Relation, Rights, Creator, Source, and Coverage.

In the 25 records examined, Ball State collects all elements except Format. The only field Ball State maps to the Format element is Digitization Specifications, a field not utilized in the video records. Shannon County, NJVid, and Windows on Maine harvest all but two elements, although not the same two. Shannon County does not display Language or Coverage in CONTENTdm; while it was collected in the original phase of the project, use of the field cannot be verified. NJVid elected to populate Contributor rather than Creator and did not use Source. Windows on Maine follows the MARC to DC crosswalk in mapping Contributor values to Creator. It does

not display Language on its site although the cataloging interface supports collection of that element. Again, use of that element cannot be verified.

Accuracy

Accuracy concerns the accurate description and representation of data and resource content. It also concerns accurate data input (Bruce & Hillman, 2004). Zeng (2006) describes the accuracy of metadata in terms of three elements: the correctness of the data element's content, intellectual property, and instantiation.

Across the repositories, many of the elements have been harvested with little or no errors. Where applied, Language appears without error. Similarly, Title has been collected without issue across all repositories. Relation has no observable errors in the collections that displayed that information. However, naming convention is not consistently observed. In Ball State's records, the same surname-forename order (surname followed by forename) is not used across all the records examined. Over 25% of the records used the reverse (forename followed by surname).

Other elements have more serious issues. The Identifier for the Windows on Maine collection is the URL for the launch page, not the streaming video itself. This practice does not adhere to DCMI's Using Dublin Core (1995–2009, The Elements section, ¶ 4.14.), which defines the Identifier element as "an unambiguous reference to the resource within a given context" and, in the Guidelines for content creation for the Identifier element, specifies that "it should not be used for identification of the metadata record itself."

Alone of all the repositories, Folkstreams uses the value "text/html" for format even though the metadata are referring to video. Similarly, it uses "text" as a value for Type. Since this is a repository of documentary films, the use of "text" is inaccurate. These errors occur in 100% of the records examined. The format values are on the Film Facts display of the record (http://www.folkstreams.net/filmfacts,135) but have not been transferred to the DC metadata correctly.

The Description field has inaccuracies in a handful of records across the repositories. In ACMI, the descriptions of some of the videos have extra characters that are not part of the title, such as for the work, *North*: 'Hit the North. The North is where they put most of the wogs.' Utilizing excerpts from Christos Tsiolkas' novel Loaded, this is a lyrical and haunting work about space and place in Melbourne's inner-city northern suburbs." *Short Band Video* has an error in the Description field in relaying the name of another artist, which could impact search on that name. In the Description field on NJVid, apostrophes are consistently transformed into a series of characters. On a positive note, Shannon County is extraordinarily accurate with its content in the Description field, with only one typographical error, the misspelling of mechanical. Ball State, too, is extremely accurate in its

population of the Description field. There are only two minor typographical errors and one misuse of a suffix to a number: *131th* should be *131st*.

Windows on Maine has a similar problem with extraneous characters in its Source element for *The Flooding of Flagstaff*. It also has some display issues with duplicate values, a 20-year date range, and two dissimilar dates listed under the Publication Date (Date element) label on the site.

ACMI chooses to name itself the Creator of each video. As defined by the DCMI Using Dublin Core, the creator is the "entity primarily responsible for making the content of the resource" (1995–2009, The Elements section, ¶ 4.8). While ACMI may have sponsored the creation of some of these videos, most have an individual or other entity named as a creator on the launch page. Use of the artist as creator is particularly important for those videos listed in the Video Art section. ACMI is already designated as Publisher for these videos and indicated as the Relation value; the redundancy as the Creator lessens potential search value.

Consistency

To offer the most value to the end user, there is "a need to ensure that elements are conceived in a way that is consistent with standard definitions and concepts used in the subject or related domains and presented to the user in consistent ways" (Bruce & Hillman, 2004). Conceptual/semantic consistency entails the degree to which the same data values or elements are used for delivering similar concepts in the description of a resource. On the other hand, structural consistency concerns the extent to which the same structure or format is used for presenting similar data attributes and elements of a resource (Park, 2009; Gasser & Stvilia, 2001).

Consistency within a repository is often dependent on the type of collection it is. Generally speaking, repositories mapped to Dublin Core from MARC seem to be more consistent. For Shannon County, the materials in the collection are all from the same source with the same type of content. In addition, for the first phase of the project, the metadata was essentially produced by the creator of the materials, first in a rough MARC format, then as mapped MARC to Dublin Core, and finally as the CONTENTdm version. Folkstreams, with its metadata applied in house, is the most consistent of the six repositories. A repository like Windows on Maine, as an aggregate of seven distinct archives, has the greatest challenge in terms of consistency, relying on data submitted by contributing organizations.

Both Shannon County and Windows on Maine have issues with the format of the Date values. In Shannon County, Date.Original is in the format 1978-11-26. Date.Published is in the format 2007. Both are in accordance with the Dublin Core user guide, but inconsistent within the repository. For Windows on Maine, date format appears inconsistent even within the

same field. In some cases, such as the record for Harrie B. Coe, the Date Published information on the site appears as 1/1/1930. On other records, such as Fishing for the Future, it appears as only 2001. While both are correct according to Dublin Core, the inconsistency can cause confusion for the user and create doubt about the accuracy of the information. Ball State and NJVid have similar inconsistencies. Most of Ball State's records supply the year only, but 3 of the 25 records use the format mm/dd/yyyy. Likewise, the majority of NJVid's 25 records supply the year only, but 5 provide the more specific mm/dd/yyyy.

Ball State has some consistency issues with the Type element and with Run Time, which is mapped to the Description element. In the Type element, Ball State uses the values Digital Video, Streaming Video, Streaming Video (WMV), Moving Image, and QuickTime Video to mean virtually the same thing. The usage varies by collection within the repository. In describing duration, the configuration mm:ss is used in 60% of the records; the other 40% use the configuration 00 Minutes.

On Windows on Maine, the title is sometimes displayed with the initial article in the beginning of the title or at the end (Maine Lobster, The). For the record The Great Bays of Maine (Penobscot Bay) the article remains in the front. A similar lack of uniformity is apparent in the Description element on Windows on Maine and Ball State. For some videos, the descriptions are written in full sentences and even paragraphs, and in others, fragments. There is a similar disparity among the descriptions in NJVid, although the disparity is not as pronounced.

In Shannon County, the values in the Contributor field are formatted as follows: Role, First Name Last Name, i.e. "Executive producer, Robert Flanders." Within the same collection, values for Creator follow the format Last Name, First, that is, "Moore, Robert." Windows on Maine has a similar issue with their values for Creator. The record The Maine Lobster lists creator as John W. Agnew, Jr., Director; Live Lobsters lists as Creator Dice, Peg, Director. The same element, the same role, yet the format of the value is different. This inconsistency is apparent in other records as well. In the Harrie B. Coe record, the collection name is actually inverted–Coe, Harrie B. Collection, while the value for Coe as a Contributor is not–Other: Harrie B. Coe.

Ball State has inconsistencies among its aggregate collections concerning the entities it labels Creator, those it labels Publisher, and those it labels by the many roles that collectively makeup Contributor. In the student films, no Creator is identified, yet clearly, the film student who wrote, produced, and directed the film should be its Creator. Several collections within Ball State's repository seem to reserve the element Creator for corporate entities. United States; The War Dept. is the creator of a World War II propaganda film. Ball State University is the Creator of a news broadcast.

Use of Controlled Vocabularies

Of all the repositories reviewed for this paper, ACMI alone uses no controlled vocabulary. In the source code, the subject scheme is described as "ACMI." Upon review, however, none of the source code for the streaming video on the site reveals any usage of controlled vocabulary in describing the subject of the video, nor are the resources tagged with natural language keywords, although there is opportunity in the coding to do so. Perhaps the argument could be made that these short works do not have enough content to index, but these videos warrant metadata just as an image does. One of the spots is called "The Irrepressible Jack Stone." The description on the site states, "Jack's life was an amazing 14 years. He had achieved a lot and the quality of his life was fantastic for someone who suffered from a life threatening heart condition" (ACMI, n.d.). This piece could easily be tagged with a subject heading from the Library of Congress Subject Heading Authority (LCSH) such as "Heart Diseases Patients Family relationships." If and when this item were archived into a larger collection, this information would be valuable in finding this resource.

Ball State University, Windows on Maine, Shannon County, and NJVid all take advantage of standard controlled vocabularies for subject headings. In the Shannon County collection, each roll is described with multiple subject headings. In some instances, the use of multiple headings occurs because the rolls contain brief segments not related to each other but which had been filmed on the same day. It is likely as well that multiple subject headings could be applied to differentiate the content. Certainly, all of the rolls could be described with "Shannon County (Mo.)" and most are. The additional subject headings should help the end user determine what other topics appear on each specific roll. This collection also uses more subdivisions, perhaps to offer more specificity or to differentiate one segment from another. In addition, the number of subdivisions could also be due to the fact that the collection's metadata are mapped from MARC. The majority of the subjects on Windows on Maine are LCSH. The Ball State repository, according to its Cardinal Core DMR Metadata Schema mapping to Dublin Core, has a required field for subject headings using LCSH. The Thesaurus for Graphic Materials is required for images but the guidelines do not specify moving images. Ball State has optional fields for additional headings using other controlled subject headings including locally derived controlled subject headings and the Art and Architecture Thesaurus. Nevertheless, the majority of subject headings on the moving images in the Ball State repository use LCSH.

All repositories (except ACMI) include assigned subjects that are not found in a search of any of the authorities. Some are personal names that have been constructed seemingly in accordance with name authority file format. Others are constructed like LCSH headings but are not found in a search of LCSH or in a search of the Thesaurus for Graphic Materials, the Art and

Architecture Thesaurus, or the Getty Thesaurus of Geographic Names (TGN). In the Shannon County collection, one example is "Munsell Chapel (Shannon County, Mo.)." These terms are assumed to be locally assigned controlled vocabularies. In Windows on Maine, the subject field often also includes a string of keywords as a value, for example "agriculture, fishing industry, mills and mill-work, transportation, tourist trade, paper industry." This string seems to consist of a series of LCSH terms strung together. Other terms, such as Passamaquoddy Tribe and Farming/Agriculture, are not in any of the controlled vocabulary sites searched. The drop-down of "subjects/topics" on the Windows on Maine site also includes "teacher resources" and topics such as "Maliseet Tribe," which are locally assigned subject headings.

The subject headings in use on Folkstreams appear to be a combination of two headings from LCSH. For example, Narrative & Verbal Arts is a combination of Narrative Art and Verbal Art from LCSH. Since this is a collection based on a relatively narrow topic, the catalogers might have chosen to create their own subject headings tailored to the collection. There is no mention of controlled-vocabulary recommendations in the best practices guide published on the Folkstreams Web site. NJVid recommends the use of the Moving Image Collection's listing of subjects, a list of only 15 terms to which NJVid has added two terms, "Biography" and "Languages and Literature." LCSH subject headings are also visible in the Subject element on the records in NJVid and many records have terms from both lists. For example, the film, "No Place to Call Home," is assigned the subject headings "Country or Region Specific" and "Society" both from the MIC list and "Homeless women" from LCSH. Subject headings on the 25 records investigated are used in accordance with the glossaries of each list, but, because the metadata can be assigned by film contributors, the number of subject headings per film varies widely. Some records have up to 10 subject headings, taken from both lists; others have as few as 2, taken from just the MIC list.

Windows on Maine also utilizes controlled vocabulary for the DC element coverage.spatial. The site allows the user to search for resources by Town/City/Reservation name. When available, this information is also displayed on the record as geographic area and as "geocodes" on a clickable map. The About Windows on Maine page indicates that the geocodes are mapped from the metadata produced by the cataloger to ensure consistency (Windows on Maine, n.d.). The values for data are derived from a controlled vocabulary, in this case, the Geographic Names Information System. Folkstreams also indexes its films by region. None of the five regions is in the LC Name Authority File but several are in the TGN, including Appalachia, Middle Atlantic, Midwest, and Southwest. The Web site seems to have either accidentally used some of the TGN's names or chosen the ones it wanted. The geographic names do not make it into the DC metadata for this repository. Shannon County also supplements its LCSH heavy subject headings with some use of TGN terms. NJVid maps three hierarchical levels

(country, state, county) of TGN terms from MARC 752: subfields a,b,&c to three separate Coverage fields. Geographic names are not heavily indexed in the records from Ball State, but when they are, they follow the same practice as NJVid.

For the Type element, Windows on Maine, Shannon County, and NJVid all use the value Moving Image consistently, which is the correct term in the DCMI Type Vocabulary. Ball State at times uses Moving Image and at other times uses Digital Video and Streaming Video, terms not in the DCMI Type Vocabulary. Also mapped to the Type element in the Ball State repository is a field for original physical format. Folkstreams uses one term from the DCMI Type Vocabulary—Text—but should use Moving Image. ACMI does not use the element Type.

Almost none of the repositories use the IANA Video Media Types vocabulary, one vocabulary recommended by DCMI for the Format element.

Interoperability

There are several barriers to interoperability among the collections. Most obvious are the inconsistencies, errors, and semantic ambiguities involving the elements Format and Type, Contributor and Creator. Ball State uses terms for Type that are not in the DCMI Type vocabulary and are not used by the other repositories. Folkstreams mistakenly tags Type and Format as text. Not only does Folkstreams get the value wrong, since the record is describing film, but it gives almost the same value to two different elements, thereby using these elements interchangeably. Any search across these repositories for Moving Image would miss most of the resources in these repositories—the search would find Ball State's commencement videos, which do use Moving Image as the value for Type.

Semantic ambiguities are also exhibited with the elements Creator, Contributor, and Publisher. Ball State appears to use corporate entities like the War Department or Ball State University for Creator. Folkstreams only populates the Creator element with the filmmaker. ACMI uses itself as the Creator, Publisher, and Rights Holder. All of Windows on Maine Contributors/Creators are mapped to Creator, although they may be more appropriate in Contributor. So, while these repositories all use the element Creator, they are not all populating it with the same type of information. NJVid does not use the Creator element at all but maps creators and contributors alike to the Contributor element. Searching for creators across these repositories would provide unreliable results.

Inconsistencies in Format values are rampant. Shannon County uses very specific data about the streaming video itself; NJVid uses streaming video formats although not as specific as Shannon County. Folkstreams has several Beta formats listed in Film Facts on its site, but in the DC element Format it lists them as text/html, because the format values on the site are

not mapped to the Format element. So a search for specific formats among the repositories would not produce the correct results.

Finally, while Shannon County, Windows on Maine, Ball State, and NJVid to some extent use LCSH for the Subject element, Folkstreams uses a locally edited one and ACMI uses none at all. Inconsistent use of other controlled vocabularies across the repositories is another issue. Quite obviously, any subject search across all collections would miss many resources, most of them in Folkstreams and ACMI.

Overall, consistency within the individual collections in these repositories appears to be high. It is when they become part of an aggregate that interoperability becomes a problem. Even mapping from MARC does not guarantee interoperability among repositories because of local cataloging decisions and interpretations of the MARC to DC Crosswalk. Intercollection interoperability issues are easily visible on the Missouri State University Special Collections and Archives Digital Collections site where the Shannon County collection is hosted. The Advanced Search functionality allows the user to select by Collection and by named field. The drop down in the named field includes the various elements discussed in this paper: Title, Subject, Description, Creator, Publisher, Contributors, Date, Type Format, Identifier, Source, Language, Relation, Coverage, Rights, and Audience. However, it also includes an additional 26 elements, which appear to be the actual Dublin Core element names, some of which are used in the Shannon County collection. These elements have populated the drop down due to the other collections in the Digital Collection. The problem arises when a user wants to specifically search for a keyword in the relation field. The drop down reveals 14 qualified elements from which to choose, requiring the user to search 14 times or broaden his or her search to include all fields.

CONCLUSION

Overall, the review of the six collections illustrates just the state of flux that metadata is in–particularly in collections that are not under any sort of library-type jurisdiction by which quality standards are maintained and monitored. The overview of the collections demonstrates that a wide variety of metadata is not collected on a consistent basis, especially in collections that are an aggregate of archives or repositories. In some cases, the implementation itself seems to be incomplete, as with ACMI, where data are simply not collected or Folkstreams where elements are deliberately not mapped. In others, like Windows on Maine, the information is not collected because it is just not available from the participating archives. No matter what metadata scheme is used, this presents a very real problem. In other cases, execution seems to limit the effectiveness of the efforts. Even in a case such as Shannon County,

where the metadata have been hand selected by the creator of the materials and proofed by professional catalogers, the interoperability of the data was hindered by the translation to a site such as the Missouri State University Digital Collection. Inconsistency in what is gathered across databases, where some things are deemed necessary by one collection but not by another and confusion in the way metadata are represented or labeled on a site may be considered universal problems not specific to one metadata schema in particular.

More of an issue is the inconsistency and inaccuracy of some of the elements as applied. There is a real potential for major barriers to interoperability across these six repositories because of the confusion caused by the semantic ambiguity of Creator, Contributor, and Publisher. Development of local guidelines, agreement and adherence to controlled vocabularies, training staff in the best practices of Dublin Core metadata creation, and, in general, giving metadata creation a higher priority would remove many of these barriers.

Further work could outline what content management system each repository uses, whether or not each has Dublin Core documentation available, how many collections are contained in each, what search capabilities are offered to the end user, and what fields are searchable. Such a comparison could further identify barriers to interoperability. In addition, an exploration of how Folkstreams maps its records to OAIster could offer valuable insight into how one collection's records are integrated into a large federated repository.

Despite the issues identified in this paper, the level of metadata supported by Dublin Core proves effective as a crosswalk between multiple schemas if the mapping guidelines are followed. It achieves its primary purpose of increasing findability *within* a repository if only because of its recommended use of controlled vocabularies. The Dublin Core metadata as harvested for these collections serves adequately as a high-level resource discovery mechanism, despite its ambiguities.

REFERENCES

Australian Centre for the Moving Image. (n.d.). Retrieved October 20, 2009, from http://www.acmi.net.au/about.htm

Ball State University Libraries. (n.d.). Digital media repository: A project of the university libraries. Retrieved June 28, 2009, at http://libx.bsu.edu/

Bearman, D., Miller, E., Rust, G., Trant, J., & Weibel, S. (1999, January). A common model to support interoperable metadata: Progress report on reconciling metadata requirements from the Dublin Core and INDECS/DOI communities. *D-Lib Magazine*. Retrieved June 28, 2009, from http://www.dlib.org/dlib/january99/bearman/01bearman.html

Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D. Hillmann & E. Westbrooks (Eds.), *Metadata in practice*. Retrieved June 28, 2009, from http://hdl.handle.net/1813/7895

Bui, A. & Park, J.-R. (2006). An assessment of metadata quality: A case study of the National Science Digital Library Metadata Repository. In H. Moukdad (Ed.), *Information science revisited: Approaches to innovation, Proceedings of the 2006 annual conference of the Canadian Association for Information Science*. Retrieved December 7, 2009, from http://www.cais-acsi.ca/proceedings/2006/bui_2006.pdf

Collaborative Digitization Program. (2006, September). Dublin Core metadata best practices, Version 2.1.1. Retrieved June 28, 2009, from http://www.bcr.org/cdp/best/dublin‑core‑bp.pdf

Dublin Core Metadata Initiative. (1995–2009). Using Dublin Core. Retrieved June 28, 2009, from http://dublincore.org/documents/usageguide/

Duff, W. M. (2003). The Dublin Core and its limitations. *DigiCULT.Info* (6), 31–32. Retrieved June 28, 2009, from http://www.digicult.info/downloads/dc_info_issue6_december_20031.pdf

Folkstreams, Inc. (2006). Folkstreams.net. Retrieved June 28, 2009, from http://www.folkstreams.net/

Hunter, J., & Iannella, R. (1998). *The application of metadata standards to video indexing*. Retrieved June 28, 2009, from http://www.itee.uq.edu.au/~eresearch/papers/1998/ECDL2.pdf

Gasser, L., & Stvilia, B. (2001). A new framework for information quality. *Technical Report*. Champaign, IL: University of Illinois at Urbana-Champaign.

Godby, C. J., Smith, D., & Childress, E. (2003). Two paths to interoperable metadata. DC-2003: Supporting communities of discourse and practice—Metadata research & applications. Retrieved December 7, 2009, from http://www.oclc.org/research/publications/archive/2003/godby-dc2003.pdf

McCroskey, M. (2006, October). *Cataloging streaming video on the Web: Collaboration between catalogers, an archivist, and a documentary filmmaker*. OLAC Poster session presented at the meeting of the Online Audiovisual Catalogers, Phoenix, AZ. Retrieved October 21, 2009, from http://www.tnla.org/displaycommon.cfm?an=1

National Information Standards Organization (NISO). (2007). *The Dublin Core metadata element set*. ANSI/NISO Z39.85-2007. American National Standards Institute. Retrieved June 28, 2009, from http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=9b7bffcd2daeca6198b4ee5a848f9beec2f600e5

NJVid. (2010). *About NJVid*. Retrieved June 13, 2009, from http://www.wpunj.edu/njvid/about/default.html

Park, J.-R. (2006). Semantic interoperability and metadata quality: An analysis of metadata item records of digital image collections. *Knowledge Organization*, *33*(1), 20–34.

Park, J.-R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. In M. S. Babinec & H. Mercer (Eds.), Special issue on Metadata and Open Access Repositories. *Cataloging and Classification Quarterly, 47*(3), 213–228.

Shreeves, S. L., Knutson, E. M., Stvilia, B., Palmer, C. L., Twidale, M. B., & Cole, T. (2005). Is "quality" metadata "shareable" metadata? The implications of local

metadata practices for federated collections. Retrieved December 7, 2009, from http://mailer.fsu.edu/~bstvilia/papers/ACRL.pdf

Windows on Maine (n.d.). *About Windows on Maine*. Retrieved June 28, 2009, from http://www.windowsonmaine.org/about.aspx

Zeng, M. L., Subrahmanyam, B., & Shreve, G. M. (2004). Metadata quality study for the National Science Digital Library (NDSL) metadata repository. In *Digital Libraries: International Collaboration and Cross-Fertilization. 7th Conference on Asian Digital Libraries*, ICADL 2004 Shanghai, China, December 2004

Zeng, M. L. (2006). Metadata quality study for the National Science Digital Library (NSDL) Metadata Repository. *Presented paper at the Research and Teaching Talk Series*, Information Science and Technology, Drexel University.

APPENDIX A

URL Addresses for Subject Repositories

ACMI: http://www.acmi.net.au/
Ball State University Libraries: http://libx.bsu.edu/
Folkstreams, Inc.: http://www.folkstreams.net/
NJVid: http://fdr.njedge.net/njvid/
Shannon County Film: http://faculty.missouristate.edu/r/RobertMoore/sc-ilot/index.html
Windows on Maine: http://www.windowsonmaine.org/

APPENDIX B

URL Addresses for Sample Resources From Subject Repositories

Australian Centre for the Moving Image: http://www.acmi.net.au/FD93DB9E0A3643E7AFACA83C5AA47827.aspx
Ball State University Libraries: http://libx.bsu.edu/u?/newslink,503
Folkstreams, Inc.: http://www.folkstreams.net/film,69
NJVid: http://fdr.njedge.net/njvid/showvideo.php?pid=njcore:16572 /
Shannon County Film: http://faculty.missouristate.edu/r/RobertMoore/sc-pilot/cr165/cr165.html
Windows on Maine: http://windowsonmaine.library.umaine.edu/fullrecord.aspx?
ObjectID=6-1000 /