People of the (Digital) Book:

The Metadata of Jewish Text-Based Materials in Two Digital
Repositories

Dev Singer

15 December 2013

Table of Contents

Introduction

Jews have long been known as the "people of the book" due to the fact that the Jewish religion at its most basic is based around the Torah. The history of the Jews is the history of a people which encourages men—and later, women—to study, analyze, and produce scholarship. As far back as 1180,[1] the scholar Maimonides wrote in his legal code the *Mishneh Torah* that "Every Jew is obligated to study Torah, whether he is poor or rich, healthy or ill, young or old."[2] While not every Jew does study Torah, Jewish culture values learning and Jewish scholarship is alive and well in the 21st century. Scholarship related to the Jewish people ranges from in-depth study of Jewish religious texts to sociological studies of Jews and Jewish culture in various areas around the word.

With the rise of digital technologies, scholars of all subjects are better able to communicate with one another and those interested in their work. Digital libraries and repositories are one such tool, as they provide a place where scholarship can be collected and made available to others without requiring users to be within physical proximity of a copy of the work. Through metadata, such works are organized and parsed into information which a computer can understand and use to aid users in discovering resources. Scholars of Jewish Studies have created many digital repositories, from the Judaica Sound Archives at Florida Atlantic University[3] to HebrewBooks.org, which offers a collection of Hebrew books written by American rabbis in the early 20th century.[4] This paper explores two other digital repositories, the Berman Jewish Policy Archive and the Brandeis Institutional Repository, and the way in which they use Dublin Core metadata to make their resources available.

Dublin Core

Dublin Core metadata began as a result of a workshop held between OCLC and the National Center for Supercomputer Applications (NCSA) in Dublin, Ohio in 1995. The objective was to design a metadata system that could be used by authors--not just catalogers—to describe resources to be made available online.[5] The final product was a set of fifteen elements which were designed to "remain as simple and accessible as possible… while providing for optimum retrieval of those resources in an online environment."[6] Qualifiers can be added to these elements to allow for more specific fields. A list of Dublin Core elements and recognized qualifiers as defined by the Dublin Core Metadata Initiative can be found in Appendix A. The list of qualifiers presented here is not comprehensive; rather, it is a list of qualifiers developed by established agencies. The Dublin Core Metadata Initiative recognizes that other qualifiers will be utilized by individual libraries and repositories and makes the following statement in regards to

---

[1] Shraga Simmons, "People of the Book," Aish, http://www.aish.com/h/sh/tat/48969071.html.

[2] Rachel Furst, "The *Mishneh Torah* : Maimonides' Halakhic *Magnum Opus*," My Jewish Learning, http://www.myjewishlearning.com/texts/Rabbinics/Halakhah/Medieval/Mishneh_Torah.shtml?p=0.

[3] http://faujsa.fau.edu/jsa/home.php

[4] "About," HebrewBooks.org, http://hebrewbooks.org/about.

[5] National Information Standards Organization, *Understanding Metadata* (Bathesda, MD: NISO Press, 2004).

[6] CDP Metadata Working Group, Dublin Core Metadata Best Practices, (2006), http://klamathwaterlib.oit.edu/inside/metadatastandards2006.pdf.

qualifiers and interoperability: "The guiding principle for the qualification of Dublin Core elements, colloquially known as the "Dumb-Down Principle," is that a client should be able to ignore any qualifier and use the information as if it were unqualified. While this may result in some loss of specificity, the remaining element value (without the qualifier) should continue to be generally correct and useful for discovery."[7]

Interoperability is important when it comes to sharing resources beyond a library or repository's local community. In a 2010 survey, Park and Tosaka discovered that approximately half of the institutions surveyed allowed search engines and union catalogers to have access to their metadata and more than a third made their data available to Open Access Initiative harvesters.[8] When local metadata is included in these larger databases, systems need to be able to read the local metadata. Crosswalks are available to map elements from one metadata scheme to another, but even when the larger database and a local repository both use Dublin Core there can be some ambiguity surrounding local elements and qualifiers.

Methodology

After conducting a wide search for repositories related to Jewish Studies, the Berman Jewish Policy Archive and the Brandeis Institutional Repository were chosen because both provided easy access to their source metadata. While the Brandeis Institutional Repository is not in itself a Jewish repository, it houses a collection of publications from the Maurice and Marilyn Cohen Center for Modern Jewish Studies. Because this collection does not represent the whole of the Brandeis Institutional Repository, materials related to Judaism from the GSAS Master's Theses collection were selected as well to provide a better picture of the repository's metadata as a whole.

Once the repositories were chosen, 100 metadata records were retrieved: 50 from the Berman Jewish Policy Archive, 25 from the Brandeis Institutional Repository's Maurice and Marilyn Cohen Center for Modern Jewish Studies collection, and 25 from the GSAS Master's Theses collection. Selections from each repository were chosen via a random-ish sample from a search of the records for the word "Jewish." The original text of each record was preserved in a Microsoft Word document in case the records changed or were taken down during the course of this project. Data on the elements used in these records was put into Microsoft Excel to determine element usage and the original records were further analyzed for accuracy, consistency, and controlled vocabulary usage.

Berman Jewish Policy Archives

The Berman Jewish Policy Archive (BJPA) is a central repository for Jewish policy publications and other policy-relevant research and analysis on Jewish life. The Archive aims to

---

[7] Dublin Core Metadata Initiative, "Using Dublin Core - Dublin Core Qualifiers," http://dublincore.org/documents/usageguide/qualifiers.shtml.
[8] Jung-ran Park and Yuji Tosaka, "Metadata Creation Practices in Digital Repositories and Collections: Schemas, Selection Criteria and Interoperability," *Information Technology & Libraries* 29, no. 3 (2010).

"[ensure] that existing research will be easily and instantly accessible to everyone who needs it" in one central place.[9] The website was launched in 2008,[10] and as of this writing, BJPA includes 18,287 publications spanning 1900 to the present, heavily weighted towards more recent works. Materials for inclusion are either sought out by the director of the Berman Jewish Policy Archive or submitted by those who think their work belongs in the archive.[11] Metadata for these all materials is created by student employees under the guidance of the associate director. The Berman Jewish Policy Archives does not have any sort of internal guidelines document used by the student employees; instead, the students are trained by the assistant director in how to create metadata records for the Archives and he remains available for any questions they may have.[12]

The Elements

The Berman Jewish Policy Archive uses the following elements:

citation_title	dc.date
citation_author	dc.related.ispartof
citation_publication_date	dc.citation.volume
citation_journal_title	dc.citation.issue
citation_volume	dc.citation.spage
citation_issue	dc.ciation.epage
citation_firstpage	dc.identifier
citation_lastpage	dc.rights
citation_pdf_url	dc.language
dc.title	dc.abstract
dc.type	dc.subject
dc.creator	dc.format
dc.publisher	dc.coverage
dc.issued	

The elements are broken into two sets: citation elements and Dublin Core elements. If one examines the user side of a record on the BJPA website, it appears that despite being at the beginning of the list of elements, the citation information is formatted into an actual citation at the bottom of the page. BJPA then uses 11 unqualified Dublin Core elements: title, type, creator, publisher, date, identifier, rights, language, subject, format, and coverage. Of the remaining elements, "issued" is a qualifier for the date element, "abstract" is a qualifier for the description element, and related.ispartof is a qualified element. Things get muddled with the dc.citation elements, as there is no citation element in Dublin Core. There is the recognized qualifier bibliographicCitation, but this element contains a full citation, not parts of one broken down into separate elements. The dc.coverage element in the Berman Jewish Policy Archives contains only spatial information; temporal data is never included. It is also important to note that in all the records retrieved for this project, dc.issued and dc.date contained the same information, making one of these elements redundant.

---

[9] "About BJPA," Berman Jewish Policy Archive, http://bjpa.org/About/index.cfm.
[10] Seth Chalmer, telephone interview with author, 22 August 2012.
[11] Ibid.
[12] Seth Chalmer, email exchange with author, 28 October 2013.

Completeness

According to Park and Tosaka, "The completeness of metadata means that individual objects are described using all metadata elements that are relevant to their full access capacity in digital repositories."[13] This means that a complete record does not necessarily need to include all the elements available in the metadata scheme, but rather only those which are appropriate for the individual resource. For instance, in the Berman Jewish Policy Archive, not all records will contain the DC.relation.ispartof, DC.citation.volume, and DC.citation.issue elements because many resources are not part of a print journal. The following table shows the frequency with which each element is used:

Element	Number of Occurrences	Percentage
citation_title	50	100%
citation_author	45	90%
citation_publication_date	50	100%
citation_journal_title	25	50%
citation_volume	5	10%
citation_issue	8	16%
citation_firstpage	6	12%
citation_lastpage	6	12%
citation_pdf_url	50	100%
dc.title	50	100%
dc.type	50	100%
dc.creator	45	90%
dc.publisher	50	100%
dc.issued	50	100%
dc.date	50	100%
dc.related.ispartof	25	50%
dc.citation.volume	5	10%
dc.citation.issue	8	16%
dc.citation.spage	6	12%
dc.ciation.epage	6	12%
dc.identifier	50	100%
dc.rights	50	100%
dc.language	50	100%
dc.abstract	50	100%
dc.subject	50	100%

---

[13] Jung-ran Park and Yuji Tosaka, "Metadata Quality Control in Digital Repositories: Criteria, Semantics, and Mechanisms," *Cataloging and Classification Quarterly* 48, no. 8 (2010): 699.

dc.format	50	100%
dc.coverage	41	82%

This table indicates some issues in the completeness of Berman Jewish Policy Archive metadata records. While no tags are improperly assigned to a record, there is unfortunately a problem in the metadata of tags being missing when they should properly be included. For 41 of the records retrieved the tag DC.coverage is included, but for nine records it is missing. This is not an indication that the tag is irrelevant, but rather that the person creating the record forgot about it as every single document in the repository covers some area of the world. In addition, for many records the page number elements DC.citation.epage and DC.citation.epage missing page numbers. While not all materials include page numbers, for those that do these fields are used infrequently and irregularly. It is interesting to note that for citation-related elements, both the dc.citation elements and the non-Dublin Core citation elements occur with the exact same frequency and consist of the exact same parts, making the two different sets of elements redundant.

Accuracy

According to Park and Tosaka, "Accuracy concerns the degree to which the data content of metadata elements corresponds to the individual objects being described and the way that it should be represented. It also concerns incorrect or missing data input, such as spelling and typographical errors."[14] For the most part, the metadata created by the Berman Jewish Policy Archive is accurate. There are no glaring spelling or typological errors in the records examined for this project, and most elements are used according to the guidelines written by the Dublin Core Metadata Initiative (see Appendix A). The type and format elements, which are often a problem among those creating Dublin Core metadata,[15] are used correctly.

Consistency

When examining metadata quality, there are two parts to examining consistency: semantic consistency and structural consistency. Semantic consistency involves "the degree to which the same data values or elements are chosen for representing similar concepts" while structural consistency involves "the extent to which the same data structure or format is used to represent information chosen for given metadata elements."[16] The Berman Jewish Policy Archive metadata is consistent in both senses. Dates in the date and issued fields are consistently entered in yyyy/mm/dd format, which is a variation of the ISO 8601 standard.[17] Names are consistently entered in "Last, First" format, and the language is consistently encoded "en-us." In the title element, any titles which include an initial article are entered with the initial article included at the beginning of the title. For other fields, such as publisher, initial articles are left out (i.e. "Jewish Daily Forward" instead of "The Jewish Daily Forward"). Semantically, each

---

[14] Ibid.

[15] Jung-ran Park and Eric Childress, "Dublin Core Metadata Semantics: An Analysis of the Perspectives of Information Professionals," *Journal of Information Science* 35, no. 6 (2009).

[16] Park and Tosaka, "Metadata Quality Control in Digital Repositories: Criteria, Semantics, and Mechanisms," 699.

[17] Misha Wolf and Charles Wicksteed, "Date and Time Formats," W3C, http://www.w3.org/TR/NOTE-datetime.

element contains the same type of information for each record. There is also consistency in the way names of publications are entered; for instance, while The Jewish Daily Forward is also known as The Forward, it is always entered into the relation.ispartof field as "Jewish Daily Forward."

Controlled Vocabularies

The Berman Jewish Policy Archives does not use any recognized subject heading list, though if "a controlled vocabulary is any defined subset of natural language,"[18] the Berman Jewish Policy Archive (BJPA) does employ one for subjects. When the repository first began, the people involved in the project decided that because no controlled vocabulary existed which covered all the issues to be represented in BJPA, they would create their own. The list of topics today includes over 400 terms from which the student employees who create metadata records can choose. If a topic is not represented the students can request that it be added, though this only happens once or twice a year.[19] To understand the reasoning behind this, it is best to look at a sample record:

<meta name="citation_title" content="Jewish Content and Jewish Excitement at Manhattan's JCC" />
<meta name="citation_author" content="Fein, Leonard" />
<meta name="citation_publication_date" content="1997/01/01" />
<meta name="citation_journal_title" content="Jewish Daily Forward" />
<meta name="citation_pdf_url" content="http://www.bjpa.org/Publications/downloadPublication.cfm?PublicationID=15749" />
<meta name="DC.title" content="Jewish Content and Jewish Excitement at Manhattan's JCC" />
<meta name="DC.type" content="text" />
<meta name="DC.creator" content="Fein, Leonard" />
<meta name="DC.publisher" content="Forward Association" />
<meta name="DC.issued" content="1997/01/01" />
<meta name="DC.date" content="1997/01/01" />
<meta name="DC.relation.ispartof" content="Jewish Daily Forward" />
<meta name="DC.identifier" content="http://www.bjpa.org/Publications/downloadPublication.cfm?PublicationID=15749" />
<meta name="DC.rights" content="Download for personal use, freely distribute link" />
<meta name="DC.language" content="en-us" />
<meta name="DC.abstract" content="Leonard Fein profiles the "brimming Jewish excitement" at the New York City Upper West Side JCC as an expression of the expansion of Jewish content in American Jewry. This draft version is archived as part of Leonard Fein's complete papers." />

---

[18] Peter Morville and Louis Rosenfeld, *Information Architecture for the World Wide Web*, 3rd ed.(Sebastopol, CA: O'Reilly, 2006), 194.
[19] Seth Chalmer, electronic exchange with author, 28 October 2013.

```
<meta name="DC.subject" content="Part-Time (Supplementary) Schools" />
<meta name="DC.subject" content="Jewish Identity" />
<meta name="DC.subject" content="Sunday Schools" />
<meta name="DC.subject" content="Jewish Organizations" />
<meta name="DC.subject" content="Community Development" />
<meta name="DC.subject" content="Hebrew Schools" />
<meta name="DC.subject" content="Identity" />
<meta name="DC.format" content="application/pdf" />
<meta name="DC.coverage" content="United States" />
```

The official list of subject headings which would be most likely to contain the required subjects on Jewish topics is the Library of Congress Subject Headings, but this record contains many subjects which would not be found in LCSH. While "Sunday Schools" is indeed a subject, "Hebrew Schools" are actually different from Sunday schools in that they meet multiple times a week; there is no LCSH to indicate this sort of thing. The same is true for "Part-Time (Supplementary) Schools," which is a topic which combines Sunday schools and Hebrew schools and is an important topic for distinguishing between these types of religious education and the education received at Jewish day schools. The presence of both "Part-Time (Supplementary) Schools" and the subjects that make up this topic ("Sunday school") and ("Hebrew schools") indicates a lack of instruction as to the specificity of the subjects, and might be an example of someone giving the wrong subject heading to the article, though perhaps the Manhattan JCC does indeed have both Sunday school and Hebrew school.

In addition to a lack of controlled vocabulary in subjects, the Berman Jewish Policy Archive does not use any sort of name authority to differentiate between authors. This makes it hard to differentiate between different authors, particularly as a repository of this nature is likely to contain authors with common Jewish names who are all writing on similar subjects.

Brandeis Institutional Repository

The Brandeis Institutional Repository "is a resource for the Brandeis community to showcase, promote and preserve research and scholarship."[20] The planning for the repository took place between 2006 and 2007, during which time Brandeis decided to utilize the DSpace software for its repository.[21] Several factors went into the decision to use DSpace. First, DSpace is open-source and thus had low start-up costs. One of the Brandeis University staff members worked on the development of DSpace, so the committee charged with starting the Brandeis Institutional Repository was already familiar with it. Also, since the Library & Techology Services (LTS) department at Brandeis University is fairly small, there was not enough staff to handle creating metadata for every submission during the busy period at the end of the academic year and there are not enough submissions the rest of the year to warrant a staff member devoted solely to the repository. DSpace works on a self-submission model which means that authors

---

[20] "About the Brandeis Institutional Repository," Brandeis University, http://bir.brandeis.edu/about.

[21] Digital Initiative Planning Group Library and Technology Services, "Brandeis Institutional Repository Plannin Documents, 2006-2007," Brandeis University, http://hdl.handle.net/10192/21866.

submit their own metadata along with their materials using an online form available within DSpace.[22]

The Brandeis Institutional Repository contains 17 different collections, though not all collections are handled equally. For most collections, submissions go through a process before being posted. First Erin Jerome, a member of LTS, gives a person submission rights to the repository. That person submits a work to a collection in the repository, then the collection's administrator checks the metadata and other technical issues. Only after this is done does the material appear in the repository for general consumption. However, the three collections categorized under "Centers & Institutes" are separate and do not go through this same process. There is no staff member who double-checks these submissions; what the person who submits the work puts in as metadata is what shows up in the repository.[23]

For submitters, Brandeis University provides multiple sets of instruction. First, and most obvious, is the help document attached to the Brandeis Institutional Repository. This can be accessed by clicking the "Help" link on the left-hand side of any Brandeis Institutional Repository page, and a copy of the relevant section appears in Appendix B. However, there are also tutorials which have been created by the LTS staff which walk submitters through the submission process using screenshots. The screenshots in these tutorials look very different from the images included in the DSpace help file. It is unclear whether Brandeis used DSpace's default help file or whether it was modified at all for Brandeis's use, but the images in the tutorials show the correct view of the repository's submission process.[24] There are two different tutorials: one for dissertation submission, linked to from the Graduate School of Arts and Sciences Dissertation Guide,[25] and one for senior thesis submission, attached to the "Guide for Thesis Writers at Brandeis" LibGuide. Despite the name of the guide, its intended audience is writers of senior theses, not master's theses.[26] There is no tutorial pdf for writers of master's theses.

GSAS Master's Theses Collection: The Elements

The GSAS (Graduate School of Arts and Sciences) Master's Theses collection uses the following elements:

> Title
> contributor.advisor
> contributor.author
> contributor.department
> description.abstract
> description.sponsorship

---

[22] Erin Jerome, electronic exchange with author, 13 December 2013.

[23] Ibid.

[24] Ibid.

[25] "Online Submission to the Brandeis Institutional Repository (Bir)," Brandeis University, http://www.brandeis.edu/gsas/current/dissertation-guide/bir.html.

[26] "Guide for Thesis Writers at Brandeis," Brandeis University, http://brandeis.libguides.com/content.php?pid=425184&sid=3679783.

identifier.uri
Subject
date.issued
Rights
date.copyright
language.iso

Of these elements, three are unqualified: title, subject, and rights. The rest are qualified, though only three of these are recognized qualifiers. Author, advisor, and department are not recognized qualifiers for contributor, uri is not a recognized qualifier for identifier, and iso is a specification of controlled vocabulary for language, not a proper qualifier. The sponsorship qualifier is used for description, but since the sponsorship element always contains "Brandeis University, Graduate School of Arts and Sciences" it seems that it should be a qualifier on the contributor element, not the description element.

GSAS Master's Theses Collection: Completeness

The following table shows the frequency with which each element is used:

Element	Number of Occurrences	Percentage
Title	25	100%
contributor.advisor	24	96%
contributor.author	25	100%
contributor.department	25	100%
description.abstract	24	96%
description.sponsorship	24	96%
identifier.uri	25	100%
Subject	17	68%
date.issued	25	100%
Rights	21	84%
date.copyright	1	4%
language.iso	24	96%

As this table shows, the metadata for records in this collection is fairly complete, with most instances of missing elements appearing only once. The single instance of date.copyright is an example of someone using the wrong element; that particular submitter put the rights information in this field and the staff member reviewing the record didn't catch it. The biggest problem is the number of records which are missing the subject element. This is possibly a case of people following the wrong instruction guide. The instructions for senior theses indicate that the subject field should not be used because the Archives and Special Collections department wants to use Library of Congress Subject Headings for senior theses, but this is not the case for

the Master's Theses collection.[27] However, even if this is the case, the staff member reviewing the records should have caught the omission.

GSAS Master's Theses Collection: Accuracy

As mentioned above, there is one instance in which the date.copyright field is used inaccurately; rather than being a copyright date, this element contains the entire copyright statement that properly belongs in the rights element. There is also one case in which two departments are listed in one contributor.department field when they should be split into two separate instances of the element. These are the only cases of elements being used incorrectly in the collection. There are also no glaring spelling or typological errors.

GSAS Master's Theses Collection: Consistency

The records in this collection are very consistent. Contributors are consistently listed in "Last name, first name" format and dates are consistently listed in the ISO 8601 standard format of YYYY-MM-DD.[28] This is likely aided by the fact that DSpace had separate fields in the web form for each part of these elements, so the formatting is generated by the software. For the language element, which DSpace provides a drop-down menu for, "en" and "en-us" are used interchangeably. This is the only consistency issue shown in the examined records.

GSAS Master's Theses Collection: Controlled Vocabularies

In the help file located within the Brandeis Institutional Repository itself, users are instructed to "enter as many subject keywords as are appropriate to describe this item, from the general to the specific. The more words you provide, the more likely it is that users will find this item in their searches… Your community may suggest the use of specific vocabulary, taxonomy, or thesaurus. If this is the case, please select your subject words from that list. Future versions of DSpace will provide links to those lists."[29] However, the GSAS Master's Theses collection does not utilize controlled vocabulary, rather relying on keywords developed by authors. An example of this can be found in the following record:

> title   Inside-Out, Outside-In: Yeshivat Chovevei Torah's Open Orthodoxy Transmitted, Absorbed, and Applied
> contributor.advisor   Fishman, Sylvia
> contributor.advisor   Sarna, Jonathan
> contributor.author   Light, Katherine
> contributor.department   Hornstein Jewish Professional Leadership Program & Department of Near Eastern and Judaic Studies
> description.sponsorship   Brandeis University, Graduate School of Arts and Sciences

---

[27] Erin Jerome.
[28] Wolf and Wicksteed, "Date and Time Formats".
[29] "DSpace Help," DSpace, https://bir.brandeis.edu/help/index.html.

identifier.uri   http://hdl.handle.net/10192/22927
subject Open Orthodoxy
subject Clergy
subjehct   Orthodox Judaism
subject Judaism
subject rabbinical school
subject American Judaism
subject Yeshivat Chovevei Torah
subject theological seminary
subject Sociology
subject public religion
subject pastoral imagination
subject clergy education
subject sociology of religion
date.issued   2008-09-15T15:54:00Z
language.iso   en_US

Because there is no controlled vocabulary, several of the subjects attached to this record overlap, such as "theological seminary," "rabbinical school," and "clergy education." Three subjects, "Judaism," "Orthodox Judaism," and "Open Orthodoxy" are increasingly specific terms to describe the same thing. However, the use of uncontrolled vocabulary allows terms which have not been adopted by a recognized controlled vocabulary to be used. In this case, "Open Orthodoxy" and "pastoral imagination" are topics which are too new to be included in the Library of Congress Subject Headings. People who are searching for materials on these topics will have an easier time finding this record due to its lack of controlled vocabulary.

Like the Berman Jewish Policy Archive, this collection does utilize any name authority to differentiate between authors of the same name. While it is less likely that this will be a problem in a small institutional repository, there is still the chance that two people with the same name will write on similar subjects and it would be hard to identify who wrote what.

Cohen Center Publications Collection: The Elements

The Cohen Center Publications collection uses the following elements:

Title	Subject
contributor.author	subject.other
contributor.department	date.issued
Description	Publisher
description.abstract	relation.ispartof
description.sponsorship	relation.ispartofseries
identifier.citation	Rights
Identifier	language.iso
identifier.uri	
identifier.doi	

Many of these elements were also included in the GSAS Master's Theses collection and were discussed above, but the Cohen Center Publications collection includes additional elements, including the unqualified elements description, identifier, and publisher. The collection also includes the recognized qualified element relation.ispartof and several additional qualified elements: identifier.citation, identifier.doi, subject.other, relation.ispartof, relation.ispartofseries. The identifier.citation element is an unrecognized form of a recognized qualified element; according to the Dublin Core Metadata Initiative, this element should be labeled bibliographicCitation.[30]

Cohen Center Publications Collection: Completeness

The following table shows the frequency with which each element is used:

Element	Number of Occurrences	Percentage
Title	25	100%
contributor.author	25	100%
contributor.department	6	24%
Description	7	28%
description.abstract	18	72%
description.sponsorship	1	4%
identifier.citation	4	16%
Identifier	5	20%
identifier.uri	25	100%
identifier.doi	1	4%
Subject	16	64%
subject.other	6	24%
date.issued	23	92%
Publisher	7	28%
relation.ispartof	5	20%
relation.ispartofseries	1	4%
Rights	16	64%
language.iso	19	76%

The metadata for this collection is very incomplete, as evidenced by the fact that only three elements are used 100%. Of these three, identifier.uri is generated by DSpace while only title and contributor.author are created by the submitter. While some elements, such as citation and

---

[30] Dublin Core Metadata Initiative, "Using Dublin Core - Dublin Core Qualifiers".

publisher, are only relevant to a record if the material being described has been published previously, many others—including rights, language, and subject—should be present in all records.

Cohen Center Publications Collection: Accuracy

There are many accuracy issues in this collection's metadata. The following table shows the issues discovered in the 25 records pulled for this project:

Problem	Number of Occurrences
Multiple subjects are listed in a single instance of the subject element	5
Journal volume and date are included in the publisher element	1
Non-rights information is put in the rights element	2
Only the copyright date is included in the rights element	2
Non-citation information is put in the citation element	3
The citation element consists of an incomplete citation	1
Subject information is put in the related.ispartofseries element	1
Subject information is put in the related.ispartof element	3
Rights information is put in the description element	1

The number of errors is likely due to two issues. First, there is no Brandeis-designed tutorial to help staff at the Cohen Center create their metadata and the help file graphics do not match the layout of the actual repository submission screens. Second, there is no LTS staff member checking the metadata for this collection. It is likely that the people who are submitting the metadata for this collection are not aware that they are doing so incorrectly since there is no one to tell them that there is a problem. On the positive side, there are no glaring spelling or typological errors in the metadata. The problems surround understanding the proper use of Dublin Core metadata.

Cohen Center Publications Collection: Consistency

Like the GSAS Master's Theses collection, title, contributor, and date information is all formatted consistently while "en" and "en_US" are used interchangeably in the language element. However, the Cohen Center Publications collection also shows inconsistency is the use of elements. For some records, a general description field is used where other records use description.abstract; only one record uses both, and in that one the description field contains extra information about the event the abstract mentions. The subject and subject.other elements are also used interchangeably.

Cohen Center Publications Collection: Controlled Vocabularies

Like the GSAS Master's Theses collection, the Cohen Center Publications collection does not utilize controlled vocabulary for its subjects. However, this collection's subject heading usage is more problematic than that of its sister collection, as evidenced in records like the following:

title   Limmud NY 2006: The Event and Volunteers Who Made It Happen

contributor.author   Koren, Annette

contributor.author   Androphy, Aviva

contributor.department   Maurice and Marilyn Cohen Center for Modern Jewish Studies

description   This report is available at http://cmjs.org/files/filewcount.cfm?file=Limmud2006_3.pdf. In January 2006, over 750 Jewish New Yorkers of all ages and many different backgrounds came together for the second annual Limmud NY event at Kutshers Resort in the Catskill Mountains. This diverse group of Jews celebrated their Jewish identity through a myriad of activities from serious text study to challah baking and rocking musical concerts. Initiated by a group of lay leaders who had attended Limmud UK, funded in part by the UJA Federation and the Picower Foundation, and organized by a group of dedicated volunteers, Limmud NY offered Jews a variety of access points to Judaism: traditional study, culture and arts, history and politics, and worship services. Much as in 2005, Limmud NY in 2006 again engaged and excited participants of all ages. Participants and volunteers left Limmud NY 2006 feeling charged and energized about the event and their experience. They are eager to tell their friends about Limmud NY and plan to encourage them to attend next year. A third of the 2006 participants had attended the 2005 event. Most of them intend to become "regular" Limmudnyks with 60% of them already sure they will return in 2007. In the survey, a third of 2006 participants wrote about interests stimulated by Limmud NY that they are pursuing. Some of them are already involved in planning for 2007. Whether the event and the planning experience have long-term impact is a question for future research. For now, participants say it helped them make new friends and generated new interests and connections to Judaism. As this report makes clear, the excitement surrounding Limmud NY and its celebration of Jewish life continues.

identifier   120

identifier.uri   http://hdl.handle.net/10192/22965

subject.other   Family/Intermarriage/Jewish Education

date.issued   2006-05-01

relation.ispartof   Limmud

This is one example where the accuracy of records affects other topics related to metadata quality. Here, the fact that multiple subjects are grouped together in one instance of the subject.other element means that the combined schedule is unlikely to exist in any other records, nor does it make any sense. One of the issues of uncontrolled vocabularies in general is also represented here, as the same concept could be expressed by either "family" or "families."

Like the other collections studied in this project, the Cohen Center Publications collection does not utilize any name authority to differentiate between authors. However, of the collections studied here, this is of the least concern here because it is not very likely that two people with the same name will work for and publish with the Cohen Center.

Interoperability

The collections examined here all use Dublin Core metadata, but the elements they utilize are very different. The following table[31] gives a side-by-side comparison of the elements used:

Berman Jewish Policy Archive	BIR Cohen Center Publications Collection	BIR GSAS Master's Theses Collection
dc.title	Title	Title
dc.type	contributor.author	contributor.advisor
dc.creator	contributor.department	contributor.author
dc.publisher	Description	contributor.department
dc.issued	description.abstract	description.abstract
dc.date	description.sponsorship	description.sponsorship
dc.related.ispartof	identifier.citation	identifier.uri
dc.citation.volume	Identifier	Subject
dc.citation.issue	identifier.uri	date.issued
dc.citation.spage	identifier.doi	Rights
dc.ciation.epage	Subject	date.copyright
dc.identifier	subject.other	language.iso
dc.rights	date.issued	
dc.language	Publisher	
dc.abstract	relation.ispartof	
dc.subject	relation.ispartofseries	
dc.format	Rights	
dc.coverage	language.iso	

Many of the same elements are used between these collections, but they are expressed differently. For instance, all three collections used the issued qualified element, but the Berman

---

[31] For the purposes of this comparison, only Dublin Core elements are displayed for the Berman Jewish Policy Archive.

Jewish Policy Archive expresses it as dc.issued while the Brandeis Institutional Repository expresses it as date.issued. There are also different elements used to describe the same thing, such as the Berman Jewish Policy Archive's dc.creator and the Brandeis Institutional Repository's contributor.author. Looking at the list of elements here, it is easy to see the difficulty of interoperability between any two collections. The following is a crosswalk between the two repositories and the recognized Dublin Core elements:

Dublin Core	Berman Jewish Policy Archives	Brandeis Institutional Repository
title	dc.title	Title
type	dc.type	-
creator	dc.creator	contributor.author
contributor	-	contributor.advisor
contributor	-	contributor.department
issued	dc.date	dc.issued
dateCopyrighted	dc.date	date.copyright
bibliographicCitation	dc.citation.volume	identifier.citation
bibliographicCitation	dc.citation.issue	identifier.citation
bibliographicCitation	dc.citation.spage	identifier.citation
bibliographicCitation	dc.ciation.epage	identifier.citation
identifier	dc.identifier	identifier.uri
identifier	dc.identifier	identifier.doi
identifier	dc.identifier	Identifier
publisher	dc.publisher	Publisher
isPartOf	dc.related.ispartof	relation.ispartof
isPartOf	dc.related.ispartof	relation.ispartofseries
rights	dc.rights	Rights
language	dc.language	language.iso
description	-	Description
abstract	dc.abstract	description.abstract
contributor	-	description.sponsorship
subject	dc.subject	Subject
subject	dc.subject	subject.other
format	dc.format	-
coverage	dc.coverage	-

The variation between the metadata used in the two repositories and the official Dublin Core elements underscores the difficulty in translating elements between different repositories. This becomes even harder when one considers inaccurate use of elements. While the crosswalk maps relation.ispartofseries onto dc.related.ispartof and isPartOf, in reality the one instance of this

element's use has subject data in the field. Using a crosswalk or a computer programmed to translate metadata elements between repositories or metadata schemes preserves all existing inaccuracies as well as any inaccuracies which may occur due to faulty crosswalks or errors in crosswalk usage. This decreases metadata quality as a whole.

Conclusions

The metadata used in different repositories can vary greatly, even when the repositories all use Dublin Core. Different repositories may use different terminology to express the same elements and local elements may be added, making it hard to translate from one scheme to another. There is also great variation in the quality of metadata in different repositories. The quality of the metadata in the Berman Jewish Policy Archives is high, despite the lack of an internal guidelines document, due to the training received by metadata creators and the availability of an expert to answer question. Metadata quality in the Brandeis Institutional Repository's GSAS Master's Theses' collection is also high due to the fact that a staff member reviews all submitted metadata before it goes live, though there is a noticeable lack of subject element usage on many of the records. The quality of the metadata in the Brandeis Institutional Repository's Cohen Center Publications collection is poor due to the fact that it is neither created nor checked by a staff member trained in metadata.

Somewhat specific to the analysis of Jewish repositories is the issue of controlled vocabularies. While controlled vocabularies are the ideal, many topics covered by the repositories studied here are not listed in a recognized controlled vocabulary. The Berman Jewish Policy Archive has created its own controlled vocabulary, but this does not help any other Jewish repositories whose materials cover specific topics that are not reflected in established vocabularies. While this is not the reason that the Brandeis Institutional Repository collections examined here forego controlled vocabularies, the 50 records retrieved for this project contain multiple topics which would be benefited by the creation of a recognized Judaic studies-focused controlled vocabulary.

Bibliography

"About." HebrewBooks.org, http://hebrewbooks.org/about.

"About BJPA." Berman Jewish Policy Archive, http://bjpa.org/About/index.cfm.

"About the Brandeis Institutional Repository." Brandeis University, http://bir.brandeis.edu/about.

"DSpace Help." DSpace, https://bir.brandeis.edu/help/index.html.

Furst, Rachel. "The *Mishneh Torah* : Maimonides' Halakhic *Magnum Opus*." My Jewish
     Learning, http://www.myjewishlearning.com/texts/Rabbinics/Halakhah/Medieval/
     Mishneh_Torah.shtml?p=0.

Group, CDP Metadata Working. *Dublin Core Metadata Best Practices*. 2006.
     http://klamathwaterlib.oit.edu/inside/metadatastandards2006.pdf.

"Guide for Thesis Writers at Brandeis." Brandeis University,
     http://brandeis.libguides.com/content.php?pid=425184&sid=3679783.

Initiative, Dublin Core Metadata. "Dublin Core Metadata Element Set, Version 1.1."
     http://dublincore.org/documents/dces/.

———. "Using Dublin Core - Dublin Core Qualifiers."
     http://dublincore.org/documents/usageguide/qualifiers.shtml.

Library and Technology Services, Digital Initiative Planning Group. "Brandeis Institutional
     Repository Plannin Documents, 2006-2007." Brandeis University,
     http://hdl.handle.net/10192/21866.

Morville, Peter, and Louis Rosenfeld. *Information Architecture for the World Wide Web*. 3rd ed.
     Sebastopol, CA: O'Reilly, 2006.

"Online Submission to the Brandeis Institutional Repository (Bir)." Brandeis University,
     http://www.brandeis.edu/gsas/current/dissertation-guide/bir.html.

Organization, National Information Standards. *Understanding Metadata* Bathesda, MD: NISO
     Press, 2004.

Park, Jung-ran, and Eric Childress. "Dublin Core Metadata Semantics: An Analysis of the
     Perspectives of Information Professionals." *Journal of Information Science* 35, no. 6
     (2009): 727-39.

Park, Jung-ran, and Yuji Tosaka. "Metadata Creation Practices in Digital Repositories and
     Collections: Schemas, Selection Criteria and Interoperability." *Information Technology &*
     Libraries 29, no. 3 (September 2010): 104-16.

———. "Metadata Quality Control in Digital Repositories: Criteria, Semantics, and
     Mechanisms." *Cataloging and Classification Quarterly* 48, no. 8 (2010): 696-715.

Simmons, Shraga. "People of the Book." Aish, http://www.aish.com/h/sh/tat/48969071.html.

Wolf, Misha, and Charles Wicksteed. "Date and Time Formats." W3C,
     http://www.w3.org/TR/NOTE-datetime.

Appendix A: Dublin Core Elements and Qualifiers

1. Title: A name given to the resource. Typically, a Title will be a name by which the resource is formally known.[32]
   a. Alternative: Any form of the title used as a substitute or alternative to the formal title of the resource. This qualifier can include Title abbreviations as well as translations. An alternative title can be used to provide access to secondary titles, but should only be used when a value is present in the Title element.[33]
2. Creator: An entity primarily responsible for making the resource. Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.
3. Contributor: An entity responsible for making contributions to the resource. Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.
4. Publisher: An entity responsible for making the resource available. Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.
5. Description: An account of the resource. Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.
   a. tableOfContents: A list of subunits of the content of the resource. When a description of a resource consists of a list of the contents, whether from a menu or other mechanism, tableOfContents can be used to differentiate this list from descriptive text that is written in sentence form. This allows more options for display and indexing.
   b. Abstract: A summary of the content of the resource. Used when a description of a resource consists of a formal abstract. For implementations where formal abstracts are preferred, using the specific term allows the label to better reflect the level of the description.
6. Subject: The topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.
7. Date: A point or period of time associated with an event in the lifecycle of the resource. Date may be used to express temporal information at any level of granularity. Recommended best practice is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].
   a. Created: Date of creation of the resource. If the date of creation of the resource is known, and that date is important to note specifically (e.g., there are other relevant dates to record), use the term Created for the creation date of the resource. Note that the "one-to-one" rule requires that the creation date be that of the resource being described, not any early version from which the current resource is derived.

---

[32] "Dublin Core Metadata Element Set, Version 1.1," http://dublincore.org/documents/dces/. All element descriptions are direct quotations from this page.
[33] "Using Dublin Core - Dublin Core Qualifiers". All qualifier descriptions are direct quotations from this page.

b. Valid: Date (often a range) of validity of a resource. If the resource is only valid or relevant for a particular date or range of dates, the term Valid may be used to express those dates. This may be particularly important if the resource will be retained over time but its use is valid only during a particular period or until a particular date.

c. Available: Date (often a range) that the resource will become or did become available. In general, the term Available should be used in the case of a resource for which the date of availability may be distinct from the date of creation, and the date of availability is relevant to the use of the resource.

d. Issued: Date of formal issuance (e.g., publication) of the resource. The term Issued should be applied when a formal date of issuance or publication is relevant to the resource, and is distinct from other dates that may be used with the resource.

e. Modified: Date on which the resource was changed. Modified dates may be used to record either all instances of modification or only the latest. When only one modified date is recorded, it is assumed to be the latest.

f. dateAccepted: Date of acceptance of the resource (e.g. of thesis by university department, of article by journal, etc.). If, in the lifecycle of a resource, the date of acceptance by a formal body or entity is relevant to the use of the resource, dateAccepted may be used.

g. dateCopyrighted: Date of a statement of copyright. If, in the lifecycle of a resource, the date of copyright is relevant to the use of the resource, dateCopyrighted may be used.

h. DateSubmited: Date of submission of the resource (e.g. thesis, articles, etc.). If, in the lifecycle of a resource, the date of submission to a body or entity is relevant to the use of the resource, dateSubmitted may be used.

8. Type: The nature or genre of the resource. Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.

9. Format: The file format, physical medium, or dimensions of the resource. Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].

a. Extent: The size or duration of the resource. Because the refinement Extent is used in a variety of situations, it generally consists of both a numeric value and a caption that is needed to interpret the numeric value. Best practice is to separate the numeric value and the caption with a space, whether the caption appears before or after the value.

b. Medium: The material or physical carrier of the resource. Medium is generally used when the resource is of a physical nature, for instance a painting or model, where the physical carrier or material used is relevant to the user. For instance, if the resource is a movie on DVD, and is only available as a physical object, it should be described as such. If it is available digitally, for download or presentation on a website, its format would be reflected in the Format element. Note that, because of the physical nature of materials described with this refinement, the encoding scheme "IMT" is not valid for use with Medium.

10. Source: A related resource from which the described resource is derived. The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.

11. Language: A language of the resource. Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].

12. Relation: A related resource. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.

    a. isVersionOf: The described resource is a version, edition, or adaptation of the referenced resource. Changes in version imply substantive changes in content rather than differences in format. Use only in cases where the relationship expressed is at the content level. Relationships need not be close for the relationship to be relevant. "West Side Story" is a version of "Romeo and Juliet" and that may be important enough in the context of the resource description to be expressed using isVersionOf. The Broadway Show and the movie of "West Side Story" also relate at a similar level, but the video and DVD of the movie are more usefully expressed at the level of format, the content being essentially the same.

    b. hasVersion: The described resource is a version, edition, or adaptation of the referenced resource. Changes in version imply substantive changes in content rather than differences in format.

    c. isReplacedBy: The described resource is supplanted, displaced, or superseded by the referenced resource. When establishing a chain of versions, where only one version is valid, the use of isReplacedBy and Replaces allows the relationship to be expressed and the user directed to the appropriate version. In this case, the reciprocal relationships are quite important.

    d. Replaces: The described resource supplants, displaces, or supersedes the referenced resource.

    e. isRequiredBy: The described resource is required by the referenced resource, either physically or logically. In the case of IsRequiredBy and Requires, there is a clearer need to express the Requires relationship than the IsRequiredBy, though both can be useful. This relationship is most often seen in relationships between software and documents or applications and hardware and/or software requirements.

    f. Requires: The described resource requires the referenced resource to support its function, delivery, or coherence of content.

    g. isPartOf: The described resource is a physical or logical part of the referenced resource. The isPartOf and hasPart relationships are essentially "parent/child" relationships--hierarchical in nature. With them can be expressed both one-to-one and one-to-many types of relationships.

    h. hasPart: The described resource includes the referenced resource either physically or logically.

    i. isReferencedBy: The described resource is referenced, cited, or otherwise pointed to by the referenced resource. The isReferencedBy and References refinements enable the expression of relationships that aid the user but are not necessary tied to the lifecycle or necessary for the intended use of the resource. This relationship

might be used to link an article critical of a resource to that resource, a satire of a speech to the original speech, etc.

    j.   References: The described resource references, cites, or otherwise points to the referenced resource.

    k.   isFormatOf: The described resource is the same intellectual content of the referenced resource, but presented in another format. This relationship is explicitly for the expression of relationships between resources for which format is the primary variable. Because Dublin Core maintains the principle of "one-to-one," each resource is expected to have its own description.

    l.   hasFormat: The described resource pre-existed the referenced resource, which is essentially the same intellectual content presented in another format.

    m. conformsTo: A reference to an established standard to which the resource conforms. The standards referenced might be educational standards, accessibility standards, or any other established standard that is relevant to the use of the resource.

13. Coverage: The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.

    a.   Spatial: Spatial characteristics of the intellectual content of the resource. Spatial characteristics may include geographic names, latitude/longitude, or other established georeferenced values. Clearly, this refinement does not allow complex or sophisticated georeferencing, but attention to standard schemes and controlled vocabularies should provide useful results. Controlled vocabulary terms can be drawn from recommended vocabularies, or standard labelling within the value can provide useful assistance to users and applications.

    b.   Temporal: Temporal characteristics of the intellectual content of the resource. Temporal characteristics include those aspects of time that relate to the intellectual content of a resource and not its lifecycle. Examples might include a resource describing some aspect of the 19th century but itself created this year. In that case, the Temporal Coverage would be the 19th century, and the Date (or Date Created) would be 2003. Values can be text strings or encoded values.

14. Rights: Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.

    a.   accessRights: nformation about who can access the resource or an indication of its security status. Access Rights may include information regarding access or restrictions based on privacy, security or other regulations. Access rights is intended to allow the characterization of restrictions to view, search or use resources, based on attributes of the resource itself or the class or category of user. An example would be a resource that was restricted to users holding a particular

security clearance, or one that required login or authentication at a particular website.

b. License: A legal document giving official permission to do something with the resource. Recommended best practice is to identify the license using a URI. Examples of such licenses can be found at http://creativecommons.org/licenses/. License is designed to allow the inclusion of specific licensed uses to be specified. An example would be a resource that was available to be used freely but not for reproduction within commercial applications.

15. Identifier: An unambiguous reference to the resource within a given context. Recommended best practice is to identify the resource by means of a string conforming to a formal identification system.

a. bibliographicCitation: A bibliographic reference for the resource. Recommended practice is to include sufficient bibliographic detail to identify the resource as unambiguously as possible, whether or not the citation is in a standard form. Because this term is not describing a relationship to another resource, it should be limited to citations to the resource described in the remainder of the record. For instance, if the resource is an article for a journal, it is appropriate to include very specific information about the article, even page references, if such information is used to cite the article in a standard format for reference by other resources, even if the article being described is in a digital format.

Appendix B: Submission Guidelines Imbedded in DSpace[34]

SUBMIT

Stopping during the Submission Process:

At any point in the submission process you can stop and save your work for a later date by clicking on the "cancel/save" button at the bottom of the page. The data you have already entered will be stored until you come back to the submission, and you will be reminded on your "My DSpace" page that you have a submission in process. If somehow you accidentally exit from the submit process, you can always resume from your "My DSpace" page. You can also cancel your submission at any point.

Choose Collection

Progress Bar - Oval Buttons at Top of Page:

At the top of the submit pages you will find 7 oval buttons representing each step in the submission process. As you move through the process these ovals will change color. Once you have started you can also use these buttons to move back and forth within the submission process by clicking on them. You will not lose data by moving back and forth.

Select Collection:

Click on the arrow at the right of the drop-down box to see a list of Collections. Move your mouse to the collection into which you wish to add your item and click.

(If you are denied permission to submit to the collection you choose, please contact your DSpace Administrator for more information.)

You must be authorized by a community to submit items to a collection. If you would like to submit an item to DSpace, but don't see an appropriate community, please contact your DSpace Administrator to find out how you can get your community set up in DSpace.

---

[34] This section is copied directly from "Dspace Help".

Click on the "next" button to proceed, or "cancel/save" button to stop and save or cancel your submission.

SUBMIT: Describe Your Item - Page 1

If you respond "yes" to any of the questions on this page, you will be presented with a modified input form tailored to capture extra information. Otherwise you will get the "regular" input form.

* More than one title - Sometimes an item has more than one title, perhaps an abbreviation, acronym, or a title in another language. If this is the case, and you want this information captured, click in the "yes" box.

* Previously issued - New items that have NOT been previously published or distributed will be assigned an issue date by the system upon DSpace distribution. If you are entering older items that have already been distributed or published, click in the "yes" box. You will receive a form prompting you for several pieces of information relating to publication.

* Multiple files - An item can consist of more than one file in DSpace. A common example of this would be an HTML file with references to image files (such as JPG or GIF files). Another example of this would be an article supplemented with a video simulation and a data file. If you are submitting more than one file for this item, click in the "yes" box.

Click on the "next" button to proceed, or "cancel/save" button to stop and save or cancel your submission.

SUBMIT: Describe Your Item - Page 2

The information you fill in on these two screens will form the metadata record that will enable users to retrieve your item using search engines. The richer the metadata, the more "findable" your item will be, so please take the time to fill in as many fields as are applicable to your item.

Author:

This can be a person, organization or service responsible for creating or contributing to the content of the item. By clicking on the "Add More" button you can add as many authors as needed. Examples:

If the author is an organization, use the last name input box for the organization name:

Title:

Enter the full and proper name by which this item should be known. All DSpace items must have a title!

Other Title: (note - this input box appears only if you indicated on the first page that the item has more than one title.)

If your item has a valid alternative title, for instance, a title in another language or an abbreviation, then enter it here. Example:

Date of Issue: (note - this input box appears only if you indicated on the first page that the item has been previously published or distributed. If DSpace is the first means

of distribution of this item, a date will be assigned by the system when the item becomes a part of the repository.)

If your item was previously published or made public, enter the date of that event here. If you don't know the month, leave the default "no month"; otherwise select a month from the drop-down box. If you don't know the exact day, leave that box empty.

Date of Issue *Month:* February   *Day:* 18   *Year:* 2000

Publisher: (note - this input box appears only if you indicated on the first page that the item has been previously published or distributed.)

Enter the name of the publisher of this item.

Citation: (note - this input box appears only if you indicated on the first page that the item has been previously published or distributed.)

Enter citation information for this item if it was a journal article or part of a larger work, such as a book chapter. For **journal articles**, include the journal title, volume number, date and paging.
For **book chapters**, include the book title, place of publication, publisher name, date and paging.

Series/Report No.:

Some of the collections in DSpace are numbered series such as technical reports or working papers. If this collection falls into that category, then there should be a default value in the **Series Name** box which you should not change, but you will have to fill in the assigned number in the **Report or Paper No.** input box. Examples:

               Series Name   *Report or Paper No.*
Series/Report No. Software Technical Report   STR-133   Add More

Identifiers:

If you know of a unique number or code that identifies this item in some system, please enter it here. Click on the arrow to the right of the input box, and select from one of the choices in the drop down menu. The choices refer to:

Govt.doc # - Government Document Number - e.g. NASA SP 8084
ISBN - International Standard Book Number - e.g. 0-1234-5678-9
ISSN - International Standard Serial Number - e.g. 1234-5678
ISMN - International Standard Music Number - e.g. M-53001-001-3
URI - Universal Resource Identifier - e.g.. http://www.dspace.org/help/submit.html
Other - An unique identifier assigned to the item using a system other than the above

Type:

Select the type of work (or genre) that best fits your item. To select more than one value in the list, you may have to hold down the "ctrl" or "shift" key.

Language:

Select the language of the intellectual content of your item. If the default (English - United States) is not appropriate, click on the arrow on the right of the drop down box to see a list of languages commonly used for publications, e.g.

Language [French ▼]

If your item is not a text document and language is not applicable as description, then select the N/A choice.

Click on the "next" button to proceed, or "cancel/save" button to stop and save or cancel your submission.

SUBMIT: Describe Your Item - Page 3

Subject/Keywords:

Please enter as many subject keywords as are appropriate to describe this item, from the general to the specific. The more words you provide, the more likely it is that users will find this item in their searches. Use one input box for each subject word or phrase. You can get more input boxes by clicking on the "add more" button. Examples:

Your community may suggest the use of a specific vocabulary, taxonomy, or thesaurus. If this is the case, please select your subject words from that list. Future versions of DSpace will provide links to those lists.

Abstract:

You can either cut and paste an abstract into this box, or you can type in the abstract. There is no limit to the length of the abstract. We urge you to include an abstract for the convenience of end-users and to enhance search and retrieval capabilities.

Sponsors:

If your item is the product of sponsored research, you can provide information about the sponsor(s) here. This is a freeform field where you can enter any note you like. Example:

Description:

Here you can enter any other information describing the item you are submitting or comments that may be of interest to users of the item.

Click on the "next" button to proceed, or "cancel/save" button to stop and save or cancel your submission.

SUBMIT: Controlled Vocabulary

A controlled vocabulary is a set of terms which form a dictionary of descriptions of particular types of content or subject matter. These are maintained by standards bodies in order to standardise the way that similar materials are categorised in archives.

Accurately categorising material using a controlled vocabulary increases the likelihood that relevant results will be returned to users when searching individual or multiple archives.

To enter a controlled vocabulary term in the form, select "Subject Categories" from underneath the input field. This will open a window containing the available vocabularies. You may filter the vocabulary lists as described above in order to find the terms most relevant to your submission. Once you have found the term that you wish to enter, simply click on it, and it will be automatically entered into the submission form and the popup window will close. You may add as many subject category terms as you like into the form. Use "Add More" on the right to generate more input boxes.

Filtering the category list will remove from the list any terms which do not match the filter. The remaining terms are any category or sub category which contains the filter term anywhere in the heirarchy. Expanding each category will show you which terms (or sub terms) did match the filter.


SUBMIT: Upload a File

There are two methods of entering the name of the file you wish to upload:

1. Type the full path and file name into the input box and then click on the "next" button in the lower right hand corner of the screen.
2. Click on the "browse" button and a window showing your files will appear. You can navigate through your directories and folders until you find the correct file to upload. Double-click on the file name you wish to upload, and the name will be entered into the input box.

Note: If you are using the Netscape browser, the default file type will be set to "HTML files", so you may not see all of your files. Here is how to see all your files:
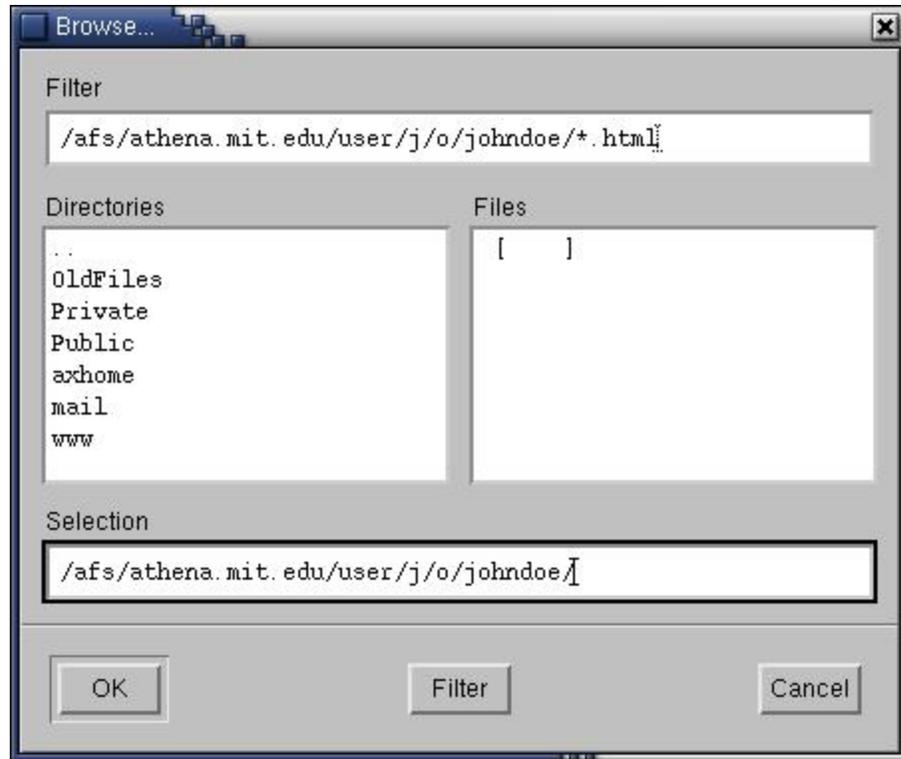
- o **On Windows and Macintosh computers**, the window initially looks like this:
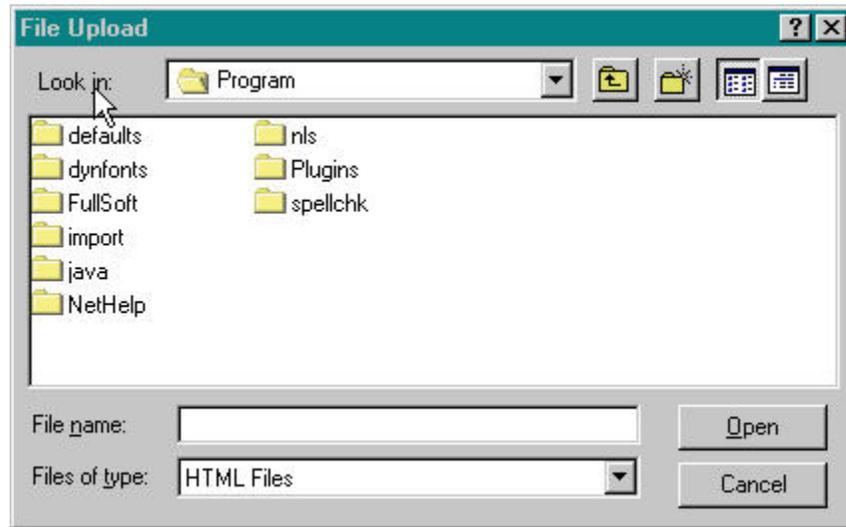
Click on the drop down arrow next to the "HTML Files" and select "All Files (*.*)":
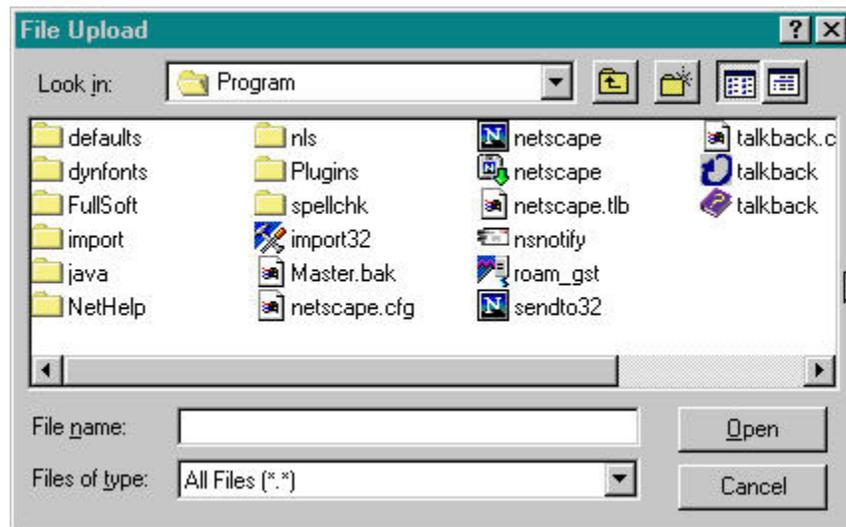
- o **On UNIX systems,** the window initially looks like this:

Click to the right of the "*.html" in the top text box, and remove the ".html" with the backspace key, but leave the "*". Then, click on the "Filter" button at the bottom of the window. You will then see all of your files, like this:
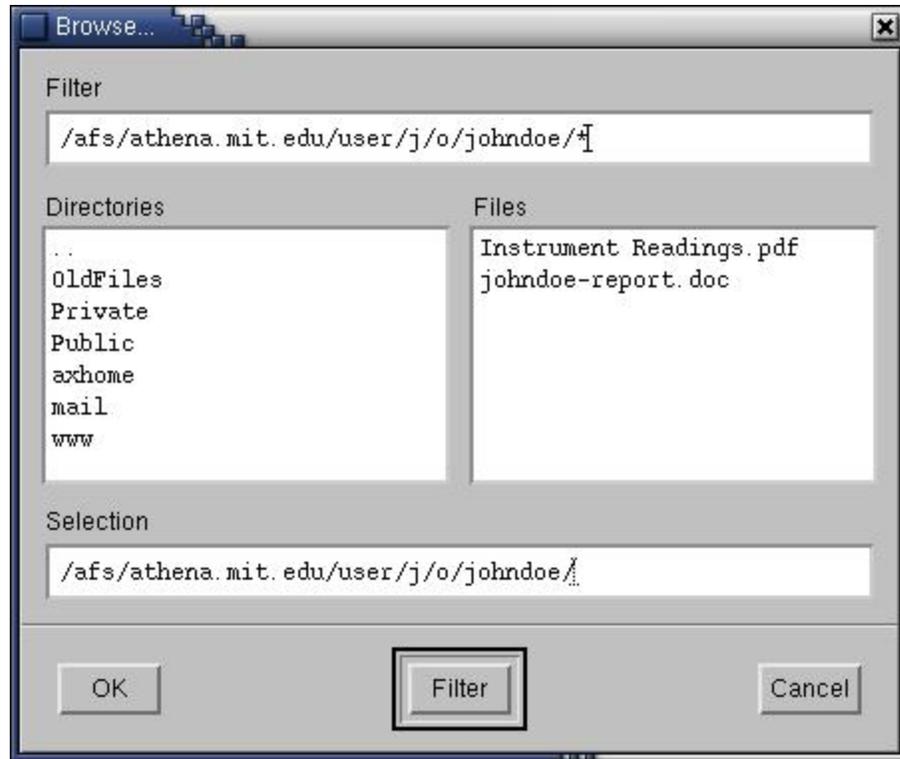
Once the correct file name is in the input box, click on the "next" button to proceed.

File Description

If you specified at the beginning of the submit process that you had more than one file to upload for this item, you will see an input box marked "**File Description**". The information you provide here will help users to understand what information is in each file, for instance, "main article" or "images" or "computer program" or "data set". Enter file descriptions for each item, and click on the "next" button to proceed.

SUBMIT: File Formats

To properly archive and give access to a file, we need to know what *format* it is, for example "PDF", "HTML", or "Microsoft Word". If the system does not automatically recognize the format of the file you have uploaded, you will be asked to describe it. If the format of the file appears in the list offered, click on it and then on "Submit". If you can't see the format in the list, click on "format not in list" and describe the format in the text box lower down on the page. Be sure to give the name of the application you used to create the file and the version of that application, for example "Autodesk AutoCAD R20 for UNIX".

For more information about file formats, see DSpace Supported Formats.

Uploaded File

After you have uploaded a file, check the information in the table to make sure it is correct. There are two further ways to verify that your files have been uploaded correctly:

- Click on the filename. This will download the file in a new browser window, so that you can check the contents.
- Compare the file checksum displayed here with the checksum you calculate.

If you're only uploading one file, click on "Next" when you're happy that the file has been uploaded correctly.

If you're uploading more than one file, click on the "Add Another File" button (this will appear if you checked "The item consists of more than one file" on the "Submit: Describe Your Item" page). When you are satisfied that all files for this item have been successfully uploaded, click on the "Next" button.

If you're uploading an HTML page with embedded files, click on the "Add Another File" button, and upload all files or bitstreams referenced in the html page. After all the are uploaded, in the column marked "Primary Bitstream", select the bitstream or file that is the index page or the top page for the web page. This will ensure that all of your embedded files will display properly on the HTML page. Then click on the "Next" button.

Checksums

DSpace generates an MD5 checksum for every file it stores; we use this checksum internally to verify the integrity of files over time (a file's checksum shouldn't change). You can use this checksum to be sure what we've received is indeed the file you've uploaded.

If you wish to verify the file using checksums, click "Show checksums" on the "Uploaded File" page. The DSpace-generated MD5 checksum for every file we've received from you will show to the right of the filename. You will then need to use a local program to generate your own checksum for these files, and verify that your results match ours. On most UNIX-like systems (including Mac OS X), use md5sum. For instance, type "md5sum MYFILE" for every file you want to check; the summary should print on your screen. For Windows machines, MD5 tools are freely available: try md5 (from http://www.fourmilab.ch/md5/), or md5sum, available via the textutils

package in Cygwin (http://www.cygwin.com/). All of these utilities will need to be run from a command-line, or terminal, window. The entire digest printed out when you run the md5 tool on your local copy of the file you're uploading should be exactly equal to what DSpace reports.

SUBMIT: Verify Submission

This page lets you review the information you have entered to describe the item. To correct or edit information, click on the corresponding button on the right, or use the oval buttons in the progress bar at the top of the page to move around the submission pages. When you are satisfied that the submission is in order, click on the "Next" button to continue.

Click on the "Cancel/Save" button to stop and save your data, or to cancel your submission.

SUBMIT: License

DSpace requires agreement to this non-exclusive distribution license before your item can appear on DSpace. Please read the license carefully. If you have any questions, please contact your DSpace Administrator.

SUBMIT: Submission Complete

Now that your submission has been successfully entered into the DSpace system, it will go through the workflow process designated for the collection to which you are submitting. Some collections require the submission to go through editing or review steps, while others may immediately accept the submission. You will receive e-mail notification as soon as your item has become a part of the collection, or if for some reason there is a problem with your submission. If you have questions about the workflow procedures for a particular collection, please contact the community responsible for the collection directly. You can check on the status of your submission by going to the My DSpace page.