
Regional Research E-Prints Archive Project

Final Report

Paul Schoening, Director

Betsy Kelly, Associate Director

Bernard Becker Medical Library

Washington University School of Medicine

St. Louis MO

April 27, 2005

Contents

Project Summary and Results	1
Introduction.....	2
Selecting and Implementing the Repository Host Software	3
Issues to Consider When Creating an Institutional Repository	4
Faculty Reaction to an Institutional Repository at Washington University	7
Conclusions.....	7
Appendix A: Draft Proposal To Establish An Electronic Review Resource Of Biomedical Research To Support Education	9
Appendix B: Washington University School of Medicine's Institutional Repository	11
Appendix C: Regional Research E-Prints Archive Project Proposal to the MidContinental Regional Medical Library	14
Appendix D: Washington University School of Medicine Dspace Policies and Procedures.....	15
Appendix E: Washington University School of Medicine Dspace Subject WorkGroup	18
Appendix F: Scholarly Publishing and Institutional Repositories Focus Group.....	21
Bibliography.....	33
Technical Guidelines and Issues in Managing a DSpace Repository	38

Regional Research E-Prints Archive Project

Project Summary and Results

The Bernard Becker Medical Library investigated the viability of an institutional repository to store post publications of School of Medicine faculty to

1. Raise Washington University School of Medicine's visibility, status and public value as a research institution by promoting broader access to its research.
2. Create a mechanism by which authors at the school could publish supplemental material.
3. Establish a mechanism for open access to research outside of the traditional publishing community.

The Library purchased a server and staff selected, installed and worked with the DSpace software developed by MIT and Hewlett Packard. The need for developing policies and procedures and a management infrastructure were investigated and guidelines developed. Resource libraries in the MidContinental Region were invited to two teleconferences where the software was discussed and demonstrated. The libraries were invited to use the Washington University server as a testbed for their own exploration of institutional repositories. Faculty of the WU School of Medicine participated in focus groups aimed at eliciting their feedback about the library's concept of an institutional repository.

Several conclusions can be drawn from the work done on this project.

1. While the DSpace software can be used "out of the box" managing a repository requires significant, ongoing commitment of staff time and access to programming and web management expertise
2. Faculty at Washington University School of Medicine have an established communications system for publishing and sharing research and information and see little or no value in contributing to an institutional repository
3. Other applications of repository software will be explored. These include providing space for the MidContinental Regional Medical Library Archives, creating collections of dissertations and theses and working papers and publications of faculty in other University departments and disciplines and further discussions with faculty throughout the University about copyright retention, open access and management of the knowledge created by University faculty and students.

Regional Research E-Prints Archive Project

Introduction

Over the past few years we have been discussing various electronic publishing and digital library ideas with faculty and colleagues in an attempt to better understand potential roles Becker Medical Library could play in these areas. We have the expertise in information technology and were looking for opportunities to creatively leverage it to expand the traditional library mission. In the summer of 2001 our plans began to coalesce around an e-press – eventually leading to a draft proposal to establish an electronic review resource of biomedical research to support education (Appendix A). The proposal reflected the vision of a small group of enthusiastic basic science faculty particularly dedicated to teaching. Ultimately, the idea was not pursued for several reasons: 1) it was not embraced by educational administrators; 2) there was no incentive mechanism to compel faculty to create original content; 3) no champion among the faculty came forward, and 4) no funding was available for the necessary technical infrastructure.

In 2003 we drafted a new proposal to create an institutional repository of scholarly publications produced within the medical school (Appendix B). This proposal did not rely on specially created content, nor did it impinge on other administrative entities within the school. It was tailored to be more closely aligned with the library's traditional role of preserving the medical school's historical record of research and so was less dependent on a faculty champion.

The principal goals established for our repository were:

- Raise Washington University School of Medicine's visibility, status and public value as a research institution by promoting broader access to its research.
- Create a mechanism by which authors at the school could publish supplemental material.
- Establish a mechanism for open access to research outside of the traditional publishing community.

These goals are not unique to Washington University and were closely derived from those identified in Crow, *The Case for Institutional Repositories: A SPARC Position Paper*¹ and Harnand, *For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature. Online Through Author/Institution Self-Archiving, Now*². They were established to resonate with our faculty so they would be encouraged to contribute to the repository. From the library's perspective we also viewed an institutional repository as a mechanism by which we could capture the research history of the institution as well as the contributions of prominent research faculty.

At this point in the process very few faculty members had been involved in the discussion of our proposed repository – it was primarily a library-driven initiative. The few conversations that had taken place with faculty members revealed varying levels of interest in the concept from skepticism to enthusiasm. Most, however, wanted to know details of how the repository would work. Since we didn't have a working system in place we began focusing on the implementation details – including software, hardware, policies, workflow, governance, and staffing. We felt that once a working system was in place it would be easier to demonstrate the concept to our faculty and we would have answers to some of their technical questions.

At about this time we had an opportunity to submit a proposal to the RML requesting funding to establish a regional repository as an educational tool for other libraries in the region to become familiar with institutional repositories and surrounding issues. (Appendix C) The proposal was funded and the resulting work is described in the remainder of this document.

Selecting and Implementing the Repository Host Software

Our first task was to select the software to host the repository. Early in the process we settled on E-Prints (<http://www.eprints.org>), but found that it was difficult to customize and somewhat inflexible for our needs. The software was written in perl and required that the library have access to a programmer proficient in writing perl scripts. The extent of customization required was discouraging and we began to reconsider our choice of eprints for a repository platform. At about the same time MIT's DSpace (<http://www.dspace.org/>) project began to generate interest in the institutional repository community. Among its distinguishing strengths were its ability to better manage the submission process and to aggregate submissions into communities and collections.

Attachment 1 describes in detail the features, workflows and management of a DSpace v1.2 repository. A succinct description of Dspace is readily available from the organization website (<http://www.dspace.org>). In these pages the DSpace Federation has described its intent, provided guides to thinking about, planning for and implementing a DSpace institutional repository. From the site, Dspace is a

“digital repository system that captures, stores, indexes, preserves, and redistributes an organization's research data.

Jointly developed by MIT Libraries and Hewlett-Packard Labs, the DSpace software platform serves a variety of digital archiving needs.

Research institutions worldwide use DSpace to meet a variety of digital archiving needs:

- [Institutional Repositories](#) (IRs)
- Learning Object Repositories (LORs)
- eTheses
- Electronic Records Management (ERM)

- Digital Preservation
- Publishing
- and more

DSpace is freely available as open source software...

DSpace accepts all forms of digital materials including text, images, video, and audio files. Possible content includes the following:

- Articles and preprints
- Technical reports
- Working papers
- Conference papers
- E-theses
- Datasets: statistical, geospatial, matlab, etc.
- Images: visual, scientific, etc.
- Audio files
- Video files
- Learning objects
- Reformatted digital library collections”

The regional test DSpace server went live in November 2003.

Issues to Consider When Creating an Institutional Repository

It is tempting to assume that creating an institutional repository is as simple as setting up a server with the appropriate software and making it available for submissions. However, when one considers how the IR will reflect on the host institution the process becomes much more complicated. In this section we will highlight the issues we encountered during the implementation of the WUSM DSpace.

Library’s Role

Appendix D reflects this library’s considerations of its role, and the policies and procedures it anticipated to manage the repository. Since the IR initiative at Washington University School of Medicine emanated from the medical library one of the first issues to address was what role the library should play throughout the life of the repository. Two obvious and significant roles that the library would play were marketing the IR and faculty education. We believed that an IR would only reach its full potential if the majority of our faculty contributed. Getting faculty to contribute requires making them aware of the IR, of its benefits, how to use it, and of the importance of retaining copyright of their articles. Marketing and faculty education is nothing new to medical libraries, however, so we felt comfortable with these roles for the IR as well.

Marketing is critical to the success of an institutional repository – in fact, the repository goals should be established with marketing in mind so they resonate well with faculty. If faculty, who are the primary contributors to the repository,

cannot relate to the goals it will be very difficult to get their support. Without faculty support an institutional repository will not achieve its goals -- so the two go hand in hand. Besides the usual “get the word out” tactics of marketing it is important to find a few visible proponents of the IR within your institution and bring their collections into the repository. By establishing such a “seed” collection you not only gain some credibility with other researchers, but you also have a working example of how the repository can be used.

The type of content maintained in the repository directly affects what role the library plays. If the IR is to act as an unrestricted, open container for digital submissions from the school’s academic community, the library’s role might be simply to act as system administrators of the underlying servers and software. On the other hand, if the IR is to hold specific content (in our case previously published journal articles for which copyright had been retained) the library’s role might be that of gatekeeper ensuring that only appropriate content makes its way into the repository. During our initial discussions on potential content we carefully considered how the various content types might impact the library’s role while being mindful of “over involving” library staff without additional funding resources.

Other potential library roles include metadata management for submissions, forward migration of digital content, identifying potential material for submission, submitting articles on behalf of authors. Each of these can consume substantial library resources, but it is important to consider them and formulate a plan of action prior to going live with an IR.

Identifying the Audience

The intended audience for an institutional repository is to a large extent determined by the IR’s goals. In Washington University’s instance a principal goal was to support open access to our research so the intended audience is the world. Consideration of the audience is important for determining the type of content and the approval process for submissions. Establishing an open access IR for preprints, for example, could have the undesired effect of preventing submitted articles from being published by commercial journals or could jeopardize institutional intellectual property. Similarly, open collections of documents submitted with no oversight, scrutiny, or vetting process could create opportunities for institutional embarrassment – particularly when the IR is officially supported by the institution.

Audience is also an important factor when discussing the merits of an IR with faculty members. Our experience has shown that researchers are most interested in making their results available to fellow researchers in their field. Many do not see the benefit of general public access to their research and are especially reluctant to add another step – repository submission - to the publication process.

Metadata Issues

The record structure and metadata components must be decided for each institutional repository, and further may be individualized by collection. The extent to which metadata is defined for repository records will affect the ease with which the collection can be searched and the relevancy of retrieval as a result of searches. Appendix E outlines the approach suggested for the WUSM repository and can serve as an example for libraries considering hosting an institutional repository.

Governance

Depending on the goals of an institutional repository a formal governance structure might be appropriate. If the IR will host material that is not peer reviewed such as gray literature, supplements to published journal articles, white papers, course material, or data sets it may be reasonable to create an oversight committee. Since the repository will represent the institution and the credibility of its research it is vital to have the stakeholders involved in the decision making process. An oversight committee can assist in policy creation and ensure compliance with institutional policies. It also offers an opportunity to educate faculty members about the repository and to give them a sense of ownership in it. Regardless of whether an oversight committee is established, it is crucial to have clear policies on allowable content, a well defined submission approval process with clearly indicated responsible parties, and a terms of use agreement for any submitted material. The library, as manager of the IR, can assume the role of gatekeeper to ensure that submitted material conforms to established policies.

The Long Run

If an institutional repository is meant to be a long term archive then, as its manager, the library must be ready to meet that obligation. In particular, this means the library must have a plan – financial and operational – for upgrading the underlying computer and storage systems as necessary to accommodate growth and to keep up with technological change. It is not sufficient, however, to maintain only the systems components of the archive. The digital objects within the repository are subject to changing standards that could leave them unreadable in the future. The library must plan on reformatting the contents as standards evolve – a task that could grow in complexity as more formats are entered. Beyond maintaining over time the integrity of the data, the library must also ensure the integrity of the URLs to specific articles through technology changes and format changes. Ultimately, the library may choose to adopt a selection process for its digital archive in much the same manner it does for its print archives to minimize the burden of maintaining an ever-expanding archive.

Faculty Reaction to an Institutional Repository at Washington University

Before going live with our institutional repository we held two focus groups with faculty representing both clinical and pre-clinical departments in the medical

school. The purpose of these focus groups was to get a better sense from our faculty of their willingness to submit content and whether they see value in an institutional repository. The list of topics and questions is shown in Appendix F, along with a transcript of faculty responses. The results were disappointing. A significant majority saw no value in an institutional repository and would not be inclined to submit articles to it. Some were adamantly opposed to the concept because it undermined the value added by commercial publications. Others questioned its value when everything they publish is online through commercial publisher sites and their research community has ready access to it. Overwhelmingly the concept of the repository as a potential location for supplemental material was denounced for a variety of reasons including lack of peer review, liability, service already provided by commercial publishers, and potential discrepancies with the published article. The small minority that supported the concept did so more for philosophical reasons than for practical access reasons. Many felt the NIH Open Access Policy would accomplish the same purpose with better visibility and economies of scale.

Conclusions

After the focus groups the library IR group reconvened to discuss next steps. It was clear that a critical mass of faculty support was lacking to meet the established goals of our IR. If we were to proceed with establishing the repository in its proposed form it would be difficult to argue for future funding unless a significant number of faculty supported it. So we changed course. Rather than promote the IR as a long-term archive of research we are offering it as an open service to University faculty to use as they see fit with no obligation of future support. The Biology Library is pursuing an interest among some of its faculty to establish a collection, but none has been created to date. It exists as a test bed for those who would like to determine for themselves the value of a general purpose digital repository.

We are left with several questions: Did we set unrealistic goals? Is the organizational structure of our medical school a factor in faculty disinterest? Should we have gone forward with the few faculty members showing interest? Could we justify the expense in that circumstance? What role, if any, does an institutional repository play within the medical school when anyone can establish and maintain their own web-based resource? Now what?

In our particular instance the expense of implementing and maintaining an institutional repository out of the medical library cannot be justified. We will continue to offer the existing hardware and software in its test bed form, but will not invest additional time or money unless a specific need is identified. We are grateful for the RML's funding of this initiative and hope that the other participants were able to gain some experience through our work.

Appendix A

Draft Proposal To Establish An Electronic Review Resource Of Biomedical Research To Support Education

Electronic publishing is changing the way in which scholarly information is created and distributed. Likewise, trends in advanced biomedical research and technology are being led, in large part, by this medical school. An opportunity for WUSM to develop an original resource for the scholarly transfer of knowledge for biomedical research and training is being explored. The Bernard Becker library is proposing to investigate the feasibility of an electronic medical school publication to support the teaching, research, post-doctoral training, public relations and recruiting aims of the Medical School and the Learning-Teaching Center (LTC) initiative.

Aims:

- Create a faculty-written, highly relevant, dynamic, and unique digital biomedical review and research tool for purposes of teaching, on-going training, recruiting, and public relations for Washington University School of Medicine
- Support a 'name-brand,' cutting-edge product to further the goals of the Medical School and to establish a resource with a reputation enjoyed by Washington Manual of Medical Therapeutics

Goals:

- Provide a state-of-the-art review source of biomedical research to support the teaching and research goals of this Medical School and to compliment the goals of the LTC space-planning initiative
- Enhance the Medical School's reputation to national and international communities as a preeminent center of biomedical research with a content-based, digital resource of core and cutting-edge information
- Expand the ability of instructors to train students and fellows by offering a dynamic, locally-authored teaching tool for a variety of learners
- Enhance the established position of the Medical School in the bio-informatics field with a unique review and reference resource for tracking current and future biomedical research trends both at the Medical School and the research community at large

Audience:

- Medical Students
- Graduate students from the Division of Biology and Biomedical Sciences
- Undergraduate Hilltop students
- Clinical and post-doctoral fellows
- Students and fellows from other colleges and universities nation-wide
- Sponsored Research Services/Gifts, Grants and Contracts

- The press
- Lay public

Business Model:

Several models for support are being investigated.

Grant-funded (public and private; e.g. NIH, NSF, SPARC, FIPSE)

Publisher-based partnership

Internally supported as a subscription-based product sold to other institutions

Likewise, several models for incentives are being investigated:

- Bonuses or links to salary
- Promotion/tenure
- Prestige

Administration would be via a central Office staffed with a managing editor, information specialists, and IT personnel. Underlying the product would be an editorial board of WUSM experts. Sections and articles would be based on a standard template maintained by administrative staff, but originally written by WUSM faculty. Links and multi-media would be developed by the Office in consort with the authors and editors.

Content and general description:

One model that is being considered is an interactive text based on core knowledge on a given topic. Conceptually surrounding the 'core,' sit a variety of 'value-added' tools for various purposes. For instance, a venue for peer-reviewed pre-publication research and commentary; WUSM I Selectives; WUSM IV 'back-to-basics;' virtual biochemical processes, histological concepts and laboratory techniques; genetic sequences and other illustrations, charts, and graphs; a press and lay version; and a database of core WUSM research interests expanding upon current RIB and other research interest publications on-campus.

Possible Model:

Diabetes-managing editor

 clinical diabetes-editor 1

 pediatric diabetes-editor 2

 genetic aspects of diabetes-editor 3

 glucose transport-editor 4

 mind/body-editor 5

Appendix B

Washington University School of Medicine's Institutional Repository

This document describes the electronic institutional repository established and supported by the Bernard Becker Medical Library for Washington University School of Medicine.

The WUSM electronic repository is based on the EPrints software (<http://www.eprints.org/>) which provides a web interface to a digital document management system specifically designed for scholarly publications. It incorporates mechanisms for self-submission, peer review, indexing and searching of several document formats. The EPrints software is open source and compliant with the Open Archives Initiative's metadata harvesting protocol to publish information about the contents to central indexing services.

Role of WUSM Repository

The WUSM E-Repository will act as a long-term storage vessel for scholarly publications produced within the School of Medicine. Such publications may include, for example, peer-reviewed published journal articles (copyright permitting), pre-print articles, review articles, book chapters, conference papers, supplemental journal article information, data sets, and electronic theses and dissertations.

With the creation of the repository we hope to:

- Promote broader access to the research coming from Washington University and thereby raise the school's visibility, status and public value as a preeminent research institution.

- Create an opportunity for authors to publish supplemental material (e.g. data sets, multimedia clips, notes, etc.) for previously published articles.

- Establish a mechanism for providing open access to research outside of the traditional pay-for-view journal model in an effort to build momentum for the reform of scholarly communication. The motivating assumption for this goal is that research results are the most beneficial when shared with the broadest audience. The established commercial journal publication business creates barriers to access through high journal subscription fees and limited circulation.

Using and Maintaining the Repository

EPrints' flexible architecture permits a variety of methods for adding content, approving content for publication, and metadata creation. To encourage the repository's use the library will offer substantial support to faculty so that it is as

painless as possible for them to deposit articles. At some point, however, each depositing faculty member will have to interact with the system.

The WUSM Repository is not intended to be the exclusive WUSM repository for scholarly publications – only a consolidation point. As librarians, one of our roles is to act as the institution's historical record keeper; we view the repository as a primary mechanism for performing this task. Authors are free to publish or house their publications anywhere they wish – on their own web servers, in other repositories, etc. We only ask that they deposit a copy in our repository along with any future revisions.

We envision the entire submission and publication process proceeding as follows:

Authors register themselves with the system via a web browser and deposit their documents.

An oversight committee (yet to be established, but most likely consisting of library staff and faculty members) will validate each article's author and determine its suitability for inclusion in the repository. This is not intended to be a formal review of the article's content and scientific merit. It is only meant to ensure that legitimate research material from faculty is deposited.

Librarians will review the article and create additional metadata essential for more accurate indexing.

Once approved and indexed the article will be released for public access and the author will be notified.

The Becker Library is responsible for all archival activities associated with the repository. In particular, library staff will perform nightly database backups and other routine system maintenance procedures to guard against loss of data. The library will also accept responsibility for maintaining the archive in perpetuity – i.e. content reformatting and migration as technology and standards change.

Essential Elements for a Successful Repository

Two elements are essential for the success of the WUSM Repository – (1) faculty acceptance and (2) retention of copyrights. The repository will only reach its full potential if a large majority of faculty buy in to the idea and submit their articles for publication. This will only be possible if authors retain appropriate copyrights for their articles published in peer-reviewed journals. Many journals have already adopted less stringent copyright agreements that allow simultaneous or delayed publication on an institutional electronic archive. However, there are still a significant number of prestigious journals that refuse to grant such rights. The library will undertake an educational campaign to assist faculty in retaining their rights when submitting articles for peer-reviewed publication.

Building the Repository

During the first several months of the repository's life our focus will be on archiving peer-reviewed research articles and electronic theses and

dissertations. The library will undertake a campaign to raise awareness of and support for the repository among faculty and perform document conversions and metadata creation for submitted articles. Policies on use and acceptable content will be drafted based on input from faculty and the oversight committee as we become more familiar with maintaining the repository.

Appendix C

Regional Research E-Prints Archive Project Proposal to the MidContinental Regional Medical Library

Description

The open access movement (<http://www.soros.org/openaccess/read.shtml>) and high subscription costs are causing many libraries to consider establishing electronic archives of research publications produced by their faculty. Although free software to implement such systems (e.g. E-Prints at <http://www.e-prints.org>) is readily available, the technical, policy and procedural issues necessary to make an electronic archive successful require significant resources to resolve. Furthermore, it is difficult to evaluate many of these issues without a working electronic archives environment in place. Our goal in this project is to reduce the entrance barrier to electronic publishing by libraries in our region. We intend to do this by establishing a regional E-Prints Archive at Washington University School of Medicine and making available technical training, template policies and educational materials to assist other libraries in establishing their own E-Prints archives. During the course of the project we also intend to assess the potential for a regional E-Prints archive to be shared by several institutions.

Specifically, during the course of this year-long project we will accomplish the following:

1. Establish an E-Prints server at Washington University to act as a regional test bed for other libraries to evaluate electronic publishing.
2. Hold on-line workshops on issues specific to implementing an institutional e-archive, such as:
 - Policies and procedures necessary for managing an archive.
 - Impact on library staff and delegation of responsibilities.
 - Tactics for garnishing faculty support for an institutional e-archive.
 - Educating faculty on copyright retention.
3. Produce a white paper which details required technical staff skills, hardware specifications, appropriate infrastructure architectures, implementation guidelines, and approximate costs of establishing a robust e-publishing environment based on the E-Prints software.
4. Attend sessions on institutional repositories at the SPARC conference.

We are confident that this project will dramatically reduce the cost of entry for libraries to get into the electronic publishing business and, therefore, result in more libraries establishing such archives. This, in turn, will enable free and unrestricted access to the research coming from those institutions and add momentum to the open access movement.

Appendix D

Washington University School of Medicine Dspace Policies and Procedures

Any effort to establish an electronic repository – the long-term storehouse for scholarly publications of Washington University in St. Louis – requires the establishment of policies and procedural guidelines that govern ownership, goals, registration, content, format, license, distribution, and privacy.

Ownership

This electronic repository is owned and maintained by the Bernard Becker Medical Library for Washington University in St. Louis. All submissions and activities associated with the archive will be centralized in the archive, and housed on Becker Library servers. The official name of the e-repository is *Washington University School of Medicine DSpace*. The system administrator is Ms. Betsy Kelly, Associate Director, Information Technologies & Library Systems.

Archive information is available at:
<http://Dspace.wustl.edu>

Goals

- Promote broader access to the research coming from Washington University in St. Louis; raising the school's visibility, status and value as a preeminent research institution.
- Create an opportunity for authors to publish supplemental material (e.g. multimedia clips, notes, etc.) for previously published articles.
- Establish a mechanism for providing open access to research outside of the traditional pay-for-view journal model and encouraging the reform of scholarly communication.

The DSpace Community

A DSpace “Community” is an administrative unit at Washington University in St. Louis that produces research, has a defined leader, has long-term stability, and can assume responsibility for setting Community policies. Each Community must be able to assign a coordinator to work with DSpace staff. Groups wishing to establish a DSpace Community that do not fall into this definition will be considered on a case-by –case basis. Individuals may not submit items without belonging to an established Community in DSpace.

Collections

Communities can maintain an unlimited number of collections in the archive. Collections can be organized around a topic, type of information, or any other sorting method a Community finds useful.

Content

The following will be accepted: published peer- reviewed materials for which copyright has been retained, and related supplemental materials (data sets, tables, lab notes, graphics, etc.) to the published work. The submitted work must be produced or sponsored by WU faculty who give Becker Library/WU School of Medicine the right to preserve and distribute the work. It must be research-oriented or scholarly and should be in digital form, complete and ready for publication in the archive.

Refusal

Becker Library has the right to refuse any paper that is not consistent with the above goals, or which is in violation of normal academic standards of discourse, or which violates the U.S. copyright agreement, or is not a scholarly publication.

Withdrawal

A petition for withdrawal of content can be submitted. The petition may be granted on the grounds of plagiarism, or if application of the data could result in harm. If a withdrawal is granted, then global access to the document's content will be removed, but the record will remain in the archive- with a reason for withdrawal – to avoid loss of the historical record.

Format

The archive will support a variety of formats (*marc, doc, pdf, sgml, html, xls, vsd, ppt, psd, aif, mpeg, ram, wav, gif, jpg, rtf, xml, mov, qt, etc.*), to ensure that everything in the repository can be retrieved. A combination of techniques (digital, migration, and emulation) will be applied to preserve submissions for future use. The system administrator would appreciate notification of new formats.)

Registration

All users may register to become subscribers. For areas of DSpace that require authorization, the system will prompt for a login. Some restricted functions, such as content submission, require authorization from the appropriate Washington University DSpace Community.

License and Distribution

In order for the Becker Library, through the archive, to reproduce, translate and distribute any submission, authors must sign and electronically submit a license/agreement. The agreement grants the Becker Library the non-exclusive right to reproduce, translate and distribute, in print and electronic format, the submission to the world.

Privacy

All submissions, once the agreements are signed, will be available globally. The archive will collect personal information for anyone who submits content and data. Personal information –such as name, address, email and telephone number - will not be shared or disclosed.

Responsibilities

Becker Library responsibilities

- Retain and maintain content submitted to the archive

- Preserve and distribute content
- Provide access to content
- Notify Communities of any changes to content or format
- Return content should the repository cease to exist

Washington University School of Medicine responsibilities

- Determine school-wide institutional policy regarding the archive
- Support the functions of the archive

Author responsibilities/rights

- Review material to be signed away
- Prepare content
- Clear copyright for items submitted
- Petition for removal of content

Community Coordinator responsibilities

- Arrange for submission and description of content
- Make decisions for submission and description of content
- Notify DSpace of organizational changes affecting submission
- Reply to annual reconfirmation of Community information
- Understand and observe policies relevant to DSpace, and educate Community submitters regarding these policies

Appendix E

Washington University School of Medicine Dspace Subject WorkGroup

The subject workgroup is charged with defining the structure and metadata components of the physical archive. In particular, determine the indexes or subject headings to include and to identify any fields unique to WUMS.

DISCUSSION TOPICS

1. Will authors placing material in the e-print archive be responsible for their own indexing, will library staff do the indexing, or will indexing be done by a combination of author and librarians?
 - Type of indexing and subject heading will depend in part on who does the indexing.
 - MeSH headings are complex and change frequently. MeSH is probably too much to scroll through as a dropdown. Would work better if librarian did the indexing.
 - Scientists are used to assigning keywords to their articles and will use all terms possible.

1. The test archive has LC classification headings in a drop-down box. NLM classification terms with selected LC terms seem appropriate for us. How many levels should we go down? When does a dropdown become too much to scroll through?
 - If go down only 2-3 levels, NLM classification with added LC as appropriate will probably work as a dropdown.

1. As best we can tell, only the subject headings (keywords and LC class terms), title words and the abstract are searchable. Christina to make sure abstract is searchable. The body of the work is not searchable. Does this mean more care needs to be taken in assigning subject headings, index terms?
 - Is there a way to have the entire work searchable?

1. Based on the above, it seems reasonable to have scientists/authors assign keyword subjects and select a broad subject term from NLM classification.

- If the body of paper is not searchable may be necessary to have librarian review indexing or subject heading assignment and amend as necessary. Experience based on review of many works will probably be needed to make this decision.
1. What types and/or formats of works will be housed in the database? We can assume that the types and formats will expand over time. Some types/formats discussed in the committee meeting were dissertations and post-publication research articles. Other suggestions were student papers, pre-prints, technical reports, conference reports, supporting data for published articles.
- The committee meeting mentioned mainly text type archives. What about posters, power-point presentations, videos, digital recordings from confocal microscopes, software, etc.
 - Since we understand that it may be difficult to make changes and additions to the e-print archive software after it is established, descriptive metadata elements for most formats and types should be included.
 - Item level metadata that should be included are:
 - Title
 - Creator
 - Subject (controlled and uncontrolled)
 - Description
 - Date
 - Format
 - Identifier
 - Relation
 - Optional metadata elements to consider are?
 - Publisher
 - Contributor
 - Type
 - Source
 - Language
 - Coverage
 - Rights
 - Examples of metadata format elements
 - Text/html
 - Text/msword
 - Text/rft
 - Text/plain

Application/powerpoint
Application/pdf
Image/GIF
Image/Jpeg

- Examples of metadata type elements

Data, numeric
Data, statistical
Data, structured text
Image, graphic
Image, moving
Software
Text, manuscript
Text, minutes
Text, monograph
Text, proceedings
Text, serial
Text, thesis

1. Depending on what the various e-print archive committees decide, should status of material tags be included? Examples:
 - Published
 - In-Press
 - Unpublished
 - Peer reviewed
 - Revision

CONCLUSIONS

1. Use MeSH and appropriate LC classification headings as the controlled subject heading. The number of levels to go down is still to be decided.
2. Include as many metadata elements as possible because it is likely the database will expand over time.
3. Other decisions and recommendations of the Subject Workgroup will depend on decisions of other committees and on further meeting.

Appendix F

Scholarly Publishing and Institutional Repositories Focus Group

Questions/Prompts

The goals of the focus group are:

- 1) to better understand the factors that influence our faculty members' choices of journals for their publications;
- 2) to get a sense whether our institutional research is reaching the right audiences; and
- 3) to explore the merits and viability of an medical school digital research archive.

The final summary of our focus groups will be emailed to each participant.

Questions

1. What factors influence your decision when selecting a journal for your research publications? Which are the most important? Which are the least important?
 - a. Are publications in high-ranking journals a significant requirement for promotion or tenure in your department?
 - b. Does it matter to you how widely your research is made available by the publisher? For example, Brain Research with it's very high subscription rate is probably less generally accessible than other journals. Similarly, some journals do not provide electronic access making them only available via paper.
2. What mechanisms do you employ to keep up with trends in your research area?
 - a. Do you subscribe to TOC services?
 - b. Do you use automated Medline searches that email results periodically?
 - c. Do you browse paper journals?
3. The library is considering a complimentary mechanism for electronically publishing previously published peer-reviewed research coming from Washington University School of Medicine with three principal goals.

- a. Promote broader access to the research coming from WU; raising the school's visibility, status and value as a preeminent research institution.
 - i. Is this important to you?
 - ii. Do you see any value in achieving this goal?
 - iii. From an outsider's perspective, do you perceive any value? Would you ever visit such a site at another institution?
 - b. Create an opportunity for authors to publish supplemental material (e.g. multimedia clips, notes, etc.) for previously published articles.
 - i. Is this important to you?
 - ii. Do other publishers offer similar mechanisms that would make this less attractive?
 - iii. What other issues could this create?
 - c. Establish a mechanism for providing open access to research outside of the traditional pay-for-view journal model and encouraging the reform of scholarly publishing.
 - i. Is this important to you?
 - ii. Are you aware of the *scholarly publishing crisis*?
 - iii. Are you familiar with the recent NIH proposal requesting the submission within 6 months to Pubmed Central of all research publications resulting from NIH funded research?
 - iv. Does this diminish the value of a WU research repository?
4. How many of you currently post your published articles on a personal or departmental website?
- a. Do you know whether this is consistent with the copyright agreement you signed?
 - b. Do you consider such behavior as a copyright violation?
5. If an institutional research repository/archive is established, what barriers do you think would have to be overcome for our faculty to participate?
- a. Would the library have to solicit papers from our faculty?

6. Can you think of other areas in which a digital repository maintained by the library could prove useful? Theses & dissertations, digital images, etc.?

Meeting Notes
November 3, 2004

ATTENDEES:

Gen (Genetics)
P/I (Pathology/Immunology)
Oto (Otolaryngology)

Q. Selecting a Journal for research publications:

Oto: Impact factor (if considering a choice), knowledge of the Journal, appropriateness

P/I: Broad enough general interest in Journal or should target – rely on experience and where found most useful articles.

Gen: Is “best Journal” not important? He now goes to “solid paper” articles. If a “cool” paper, goes for impact factor

Impact factor for importance in grant review process and in the future. Journal factor influenced by mentor.

Major factor is are you funded or does funding largely depend on impact factor

P/I: if want to get out, send to paper with good reviewer for short turn around time, get comments and better paper.

Gen: won't submit to Journal that doesn't have electronic access

Paul asked about impact factor. P/I indicated that she was not aware of impact factor when submitting an article to a journal. Her decision to submit was based on her area of study, her experience in the field, and which journals are most suitable or most useful as per article topic.

P/I indicated that a good review process on the part of journals makes a difference whether to submit an article. Gen agreed and added that if a journal takes more than nine months to review and approve an article for publishing he will not submit his work to that journal.

Gen indicated that convenience was a factor and that all faculty especially junior faculties from his department are aware of the impact factor.

Publishing history is a factor when submitting an article. [Did not get the name of the person who said this]

Gen indicated that for NIH grant reviewers the citation history of the applicant makes a difference. Gen added that he will not submit an article to a journal that offers a Print only version.

Q. How to keep up:

Paul asked about the open access concept. P/I was not familiar with this concept but Gen was.

Paul asked about what methods are used to keep up to date. P/I has saved searches on a number of databases including one on CD-ROM. [Did not get the title of this] Oto uses automatic saved searches on Medline and downloads onto Reference Manager. Gen relies on works of others in the same field.

Oto: automated Medline search

P/I: has set up on searches and browse a handful of Journals

Gen: use PubMed to track certain authors

Q. Guest electronic archives:

Promote Washington University important?

P/I: She wouldn't use – only a few people doing her work

Q. Is it important to archive?

Not convinced would add prestige – all agreed

Paul asked about DSpace. P/I did not know about DSpace and asked about visibility—who would be using and searching in DSpace. The general consensus among all three participants is that the DSpace concept was too broad and not discipline specific enough for their needs. It is not relevant for their information and research needs but perhaps relevant for an institutional corpus and public relations efforts. DSpace would be too shallow for research and teaching efforts and needs.

Paul explained further by adding that all WUMS staff could submit their work as a means of safe storage and archiving. Gen indicated that PDF precludes full text searching and natural language searching.

Gen also mentioned the Columbia University model that is a corpus of work based on subject communities.

P/I said that she could not foresee using a Harvard DSpace.

Paul added that DSpace could hold aggregate data such as images, multimedia content, audio, charts, etc. Gen replied that departmental home pages are able to do this as well.

Oto agreed and said that he posts supplemental data on his home page.

Question was asked about supplemental data being stored on journal web sites. Paul asked if this was an area of concern for the participants. All said no.

P/I asked how would we attract users to DSpace? Paul explained that all documents would have metadata attached to allow for searching and that there would be indexing from Google.

Paul asked about Open Access and the scholarly publishing crisis and if the participants were aware of such. P/I and Oto replied yes.

Paul asked about the new NIH proposal and P/I and Oto were both in favor of this.

Paul asked the participants if they saw any value for DSpace. P/I and Oto are comfortable with their established means of research and did not see a need for DSpace.

Paul asked about copyright and asked the participants if they posted their articles on a web site or departmental page. P/I does not but Oto posts his work on a web site. Paul informed Oto that he was in violation of copyright rules.

Paul asked what barriers there would be for joining DSpace. Oto indicated that it would be a hassle and a low priority. P/I asked if Becker could “capture” works done by WUMS staff instead of relying on submission by authors.

Paul asked if the participants saw any use for DSpace for dissertations or theses. P/I replied that these could be available as a resource. Oto indicated that digital images of radiological slides for teaching efforts or course related efforts would be handy.

Oto also added that it is hard to imagine a general purpose for DSpace.

Q. Any value to broader repository of Washington University work in 5-10 years?

Gen: would want to do full text searching. Could NOT be in PDF. Have information in consistent format, maybe a fragment of literature to start. (Newspapers are full text searchable). Good to get started as a research project and development of standards.

P/I: as an investigator, wouldn't go to an institutional repository. To University as a whole, might be of some value. NIH – looking how work they fund gets translated to publications – maybe show that?

Supplementary Material: question to publish work:

Gen: might be interesting to ask departments. His department has own web site. Might be his department doesn't track

Oto: put on his own web page.

P/I: a lot of work for little value. Would go in there and never be touched again.

Once Paul talked about metadata identified by Google; solves problem of transient web sites, persistent URL – all more receptive.

Gen: many publishers insist on their web sites

Oto: don't have intellectual property issues shown in department

Open Access – Reform

Oto/P/I: Don't know, don't care.

Crisis dollar squeeze – less fortunate don't have resources – limiting access to research

Paul explained PLoS – shifting costs back to author – open access model, NIH proposal

Oto: Google and PubMed – permanent online library that is with him everywhere.

Barbara Halbrook's perspective: do non-research staff thesis, posters, PowerPoint, lecture

P/I: uses Easy Article – software to search her article database

Oto: when Journal publishes his article, he downloads PDF and puts on his web site. Ignores copyright.

Does post some figures, etc that are edited out.

P/I: has own community

Q: Barriers

Oto: would have to be totally hassle free. If we could “update” or have CVI

P/I: Can log on to WU articles in other databases; way to manage CV instead of hard copy now.

Other areas where repository might be useful:

- images, thesis, dissertation
- radiology as partner to pathology

Oto/P/I: More for teaching

Q. Turf issues/ownership issues:

P/I: need for efficient teaching would drive, but not for investigators

Meeting Notes

November 11, 2004

Attending:

Anat (Anatomy)
Peds1 (Pediatrics)
Psych (Psychiatry)
Gen (Genetics)
IM (Medicine)
Neuro (Neurology)
Peds2 (Pediatrics)

1. **What factors influence your decision...**

Psych – perceived strength of the journal based on own experience in the field; also topic.

Anat – prestige, institutional pressure (bean counter). *In response to question from Paul* – bean counter is pressure from department or weight given within department to particular journals, impact factor.

Psych – generally prefers his own perception of journal, but does consider promotion and tenure

Gen – Don't feel impact factor is a major factor in promotion and tenure

Anat – Perception varies from department to department. Department chair sets the direction.

Gen – Some people on tenure and promotion committees (or some committees) don't consider impact factor – they don't read the articles. Consider only *Cell*, *Science*, and *Nature* as worth publishing in.

IM – have tenure, so liberated from promotion and tenure committees. Can concentrate on what he really wants to accomplish. Story of publishing a paper on tobacco that after publication was displayed (unknown to him) on a anti-tobacco web site. Caused copyright problems, but made him realize that there is a problem with materials being expensive publications – need to be available. Now tries to get where people will read. Wants to retain copyright and everyone in the world can read.

Peds1 – advocate of open access, does matter to him, but most colleagues he talks to are clueless. Open access still not widely considered. Peer group is a small community – don't need to scour open access in his field.

Psych – open access is not that important to him. Won't send paper to a journal where he has to pay to get reviewed (J Neuroscience, JCI (Journal of Clinical Investigation)). Won't send if there is a submission charge.

Many of group – discussion of author fees – basically, journals more and more seem to charge submission fees (\$50-\$100 handling) and some are now charging just to have an article peer reviewed (\$500). Publishers are beginning to charge authors for more things – in addition to page charges which have been around for a while.

Paul – Is there value in having scientific information reach everyone through open access?

Psych – open access overblown

IM – depends on article

Anat – it is important, especially when she wants to read something and doesn't want to pay for it on ILL.

2. What mechanisms do you employ to keep up....

Gen – automated services (weekly), PubMed. Use PubMed when there is a topic of interest. Usually just scans articles or TOC, doesn't read.

Peds1 – gets some TOC and some journals. Scans.

Neuro – uses PubMed just when he needs information. No regular routine or TOC service.

Anat – uses Forum on Alzheimer's – there is a paper of the week plus abstracts of other papers.

Psych – Looks at high impact journals in his area – goes to journal from library link, downloads what he wants – it's free.

Peds1 – Not free, but can't imagine not having.

Psych – absolutely useless unless can get on line – absolutely worth \$2 million library pays for access

Gen and IM – should do study – see if outline has greater impact factor (quick discussion of many – how to do, get numbers, etc.)

Gen – regarding open access, why would library pay for published journal if could get articles free?

Paul – danger of cuts are real – discussion of Big Deal and publisher pricing vs. single publisher journal – single journal gets cut. Effect is limiting access.

Peds1 – why should publisher have exclusive rights to article paid for by public written by researcher, reviewed by colleagues?

Neuro – have to keep this (peer reviewed publishing) going. Who pays if articles are free, why join society?

3. The Library is considering... DSpace...

All – bad idea not to have final PDF of published paper. Word Document, agPeds2ation or supplemental material – must be peer reviewed.

Any value to school?

Peds1 – how would we find out about an article in some university's DSpace?

Neuro – only makes sense if lots of institutions do this

Psych – researchers won't bother. Use PubMed only.

Gen – maybe if a journal wasn't available and wanted paper.

Anat – would be helpful in recruiting students. Also for handling reprint requests.

Peds1 – would be useful to “put in a safe” as PDF that could send when get a reprint request.

All – we ignore copyright

Gen – see DSpace as subversive – trying to be a publisher without paying – stealing from publisher. Put society publishers out of business

IM – journal publisher not doing all the work.

Peds1 – peer review and editing not necessarily a publisher function – main cost was actual print.

Gen – have to maintain infrastructure for publishing. Cost \$2500-3000 per paper to publish. Lot of printing. Society has role in publishing, members join to get journal No longer depend on society for meetings. The DSpace and open access debate is more than “free is good”

Peds2 – Is DSpace an alternative to going through a commercial publisher?

Neuro – in any case, someone has to pay.

Gen – subversive, scientific societies will die

Peds1 – too much research is now in hands of publishers. If electronic, they can turn off and it's gone.

Neuro – need way for journals to make money for publishers and still have small charge

Gen – Gen Bank, for example – when do I put my sequence in a public database? People wanted to put on own web page first. Chaos. Now, mostly in public databank PubMed Central a good idea as long as also published. Cannot have less than the final version of paper as published.

4. How many of you currently post your published articles...

Peds1 – it's illegal. But he does retain right to publish (inks into publishing contract/agreement) on his or institutional web site

Gen – doesn't worry about copyright – ignores it

Neuro – don't understand copyright – isn't there some rights for scholarly use?

Paul - fair use...

Peds1 – 3-6 month embargo makes sense. Compromise. Publisher still makes money and keeps research available

Psych – would need mirror servers. University or PubMed can't be only one.

5. If an institutional research repository/archive is established...

Psych – would have to be useful and easy. Has to be automatic. What would be useful would be to update CVs now. Neither department or division handles

Paul – define supplemental data. Is raw data?

All – must be peer reviewed. Could be changed, link could be broken, code might not work, who would be responsible.

Neuro – this sort of happening in neuroimaging. Putting on web. Others can come in and look at back end data. There is a risk.

Gen and Neuro – who will police, judge? Way to risky.

Bibliography: Introduction to Issues in Scholarly Communication, Open Access & Institutional Repositories

(Compiled by Erin Ferguson, Washington University School of Medicine)

Organizations and Initiatives:

Association of College & Research Libraries - Scholarly Communication Initiative
http://www.ala.org/Content/NavigationMenu/ACRL/Issues_and_Advocacy1/Scholarly_Communication/Scholarly_Communication.htm

Association of Research Libraries: Issues in Scholarly Communication
<http://www.arl.org/scomm/>

BioMed Central
<http://www.biomedcentral.com>

Budapest Open Access Initiative
<http://www.soros.org/openaccess/>

CARL Institutional Repository Pilot Project: Online Resource Portal
<http://www.carl-abrc.ca/projects/ir/index.htm>

Coalition for Networked Information
<http://www.cni.org/>

Create Change
www.createchange.org

Digital Library Federation
<http://www.diglib.org/>

Information Access Alliance
<http://www.informationaccess.org>

Open Archives Forum
<http://www.oaforum.org>

Public Library of Science
<http://www.plos.org/>

SPARC: The Scholarly Publishing and Academic Resources Coalition
<http://www.arl.org/sparc>

Standards Issues

Dublin Core Metadata Initiative
<http://dublincore.org/>

Open Archives Initiative
<http://www.openarchives.org/>

Open Source Initiative
<http://opensource.org>

Intellectual Property and Copyright Issues

Copyown: A Resource on Copyright Ownership for the Higher Education Community
<http://www.inform.umd.edu/copyown/>

Copyright Management for Scholarship
<http://www.surf.nl/copyright/>

Creative Commons
<http://creativecommons.org/>

Project RoMEO
<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html>

Other Useful Sources

Directory of Open Access Journals
<http://www.doaj.org/>

List of Repositories from the Open Archives Forum
http://www.oaforum.org/oaf_db/list_db/list_repositories.php

Selected Articles & Publications

Atkinson, Richard C. "A New World of Scholarly Communication." The Chronicle of Higher Education 50.11 (2003): B16.

Available: <http://chronicle.com/free/v50/i11/11b01601.htm>

Bide, Mark. Open Archives and Intellectual Property: Incompatible Worldviews? A Report for the Open Archives Forum. November 2002.

Available: http://www.oaforum.org/otherfiles/oaf_d42_cser1_bide.pdf.

Case, Mary M, and Judith Matz. "Framing the Issue: Open Access." ARL Bimonthly Report. 226 (2003): 8-11.

Available: http://www.arl.org/scomm/open_access/framing.html

Case, Mary M., and Prudence S. Adler. "Promoting Open Access: Developing New Strategies for Managing Copyright and Intellectual Property." ARL Bimonthly Report. 220 (2002): 1-5.

Available: <http://www.arl.org/newsltr/220/access.html>

Crow, Raym. "The Case for Institutional Repositories: A SPARC Position Paper." Washington D.C.: The Scholarly Publishing & Academic Resources Coalition. (2002).

Available: <http://www.arl.org/sparc/IR/ir.html>

Crow, Raym. Open Society Initiative Guide to Institutional Repository Software v1. New York: Open Society Initiative. (2003).

Available: <http://www.soros.org/openaccess/software/>

Crow, Raym. Sparc Institutional Repository Checklist & Resource Guide. Washington D.C.: The Scholarly Publishing & Academic Resources Coalition. (2002).

Available: http://www.arl.org/sparc/IR/IR_Guide.html

Fletcher, Gordon. "Averting the Crisis in Medical Publishing - Open Access Journals." He@lth Information on the Internet.30 (2002): 6-7.

Available: <http://www.biomedcentral.com/html/info/about/FletcherHOITI.pdf>

Gadd, Elizabeth, Charles Oppenheim, and Steve Proberts. "The Intellectual Property Rights Issues Facing Self-Archiving: Key Findings of the Romeo Project." D-Lib Magazine. 9.9 (2003).

Available: <http://www.dlib.org/dlib/september03/gadd/09gadd.html>

Harnad, Stevan. "For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature." Online Through Author/Institution Self-Archiving, Now. (2001).

Available: <http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm>

Hitchcock, Steve. "Metalist of Open Access E-Print Archives: The Genesis of Institutional Archives and Independent Services." ARL Bimonthly Report. 227 (2003): 4-11.

Available: <http://www.arl.org/newsltr/227/metalist.html>

Johnson, Richard K. "Partnering with Faculty to Enhance Scholarly Communication." D-Lib Magazine 8.11 (2002).

Available: <http://www.dlib.org/dlib/november02/johnson/11johnson.html>

Kyrillidou, Martha. "Journal Costs: Current Trends & Future Scenarios for 2020." ARL Bimonthly Report. 210 (2000): 10-11.

Available: <http://www.arl.org/newsltr/210/costs.html>

Lynch, Clifford A. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." ARL Bimonthly Report. 226: (2003): 1-7.

Available: <http://www.arl.org/newsltr/226/ir.html>

Lynch, Clifford A. "Metadata Harvesting and the Open Archives Initiative." ARL Bimonthly Report. 217 (2001): 1-9.

Available: <http://www.arl.org/newsltr/217/mhp.html>

McCabe, Mark J. "The Impact of Publisher Mergers on Journal Prices: An Update." ARL Bimonthly Report. 207 (1999).

Available: <http://www.arl.org/newsltr/207/jrnprices.html>

Nixon, William. "The Evolution of an Institutional E-Prints Archive at the University of Glasgow." Ariadne. 32 (2002).

Available: <http://www.ariadne.ac.uk/issue32/eprint-archives/intro.html>

Pinfield, Stephen, Mike Gardner, and John MacColl. "Setting up an Institutional E-Print Archive." Ariadne. 31 (2002).

Available: <http://www.ariadne.ac.uk/issue31/eprint-archives/>

Prosser, David. "Information Revolution: Can Institutional Repositories and Open Access Transform Scholarly Communications?" The ELSO Gazette. 15 (2003).

Available: <http://www.the-also-gazette.org/magazines/issue15/features/features1.pdf>

Research Libraries Group. "Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report." Research Libraries Group. Mountain View, CA: May 2002.

<http://www.rlg.org/longterm/repositories.pdf>

Shearer, Kathleen. A Guide to Setting-up an Institutional Repository. September 27, 2002.

Available: <http://www.carl-abrc.ca/projects/ir/setting-up.htm>.

Smith, MacKenzie, et al. "Dspace: An Open Source Dynamic Digital Repository" D-Lib Magazine 9.1. (2003).

Available: <http://www.dlib.org/dlib/january03/smith/01smith.html>

Young, Jeffrey. "'Superarchives' Could Hold All Scholarly Output." The Chronicle of Higher Education 48.43 (2002): A29.

Available: <http://chronicle.com/free/v48/i43/43a02901.htm>

Keeping Current

D-Lib Magazine

<http://www.dlib.org/>

An electronic publication with a focus on digital library research and development, including but not limited to new technologies, applications, and contextual social and economic issues.

Ariadne Magazine

<http://www.ariadne.ac.uk>

An electronic magazine which covers topics on information service developments, information networking issues, and current digital library initiatives.

ARL Bimonthly Report – Scholarly Communication Topic Link

<http://www.arl.org/newsltr/osc.html>

The bimonthly report actions from ARL, CNI and SPARC on research library issues and actions specifically related to scholarly communication.

Open Access News Blog

<http://www.earlham.edu/~peters/fos/fosblog.html>

A news blog administered by Peter Suber (of Earlham University and senior researcher at SPARC) which disseminates news about the [open access movement](#).

Scholarly Electronic Publishing Bibliography

<http://info.lib.uh.edu/sepb/sepb.html>

A continually updated bibliography published by Charles W. Bailey, Jr. of the University of Houston Libraries.

The SPARC Open Access Newsletter

<http://www.earlham.edu/~peters/fos/index.htm>

A monthly newsletter maintained by [Peter Suber](#) (of Earlham University and senior researcher at SPARC) which offers news and analysis of the open access movement.

Technical Guidelines and Issues in Managing a DSpace Repository

Table of Contents

- Administration
 - Creating Communities and Collections
 - Workflows
 - Creating Epeople and Groups
 - Authorization Policies
- Web Interface
 - Brief Overview
 - Files That Make Up the Interface
 - Copy Files to /Local Directory
 - Style Sheet-Styles.css.jsp
 - Images - /local/images
- Item Importer and Exporter
- Appendix A – Editing styles.css.jsp
- Appendix B – DSpace Administration Files
- Appendix C – DSpace Files
- Appendix D – DSpace Flowchart
- Appendix E - Customizing the Web User Interface

Administration

Creating Communities and Collections

A DSpace installation is divided into Communities, Sub-communities, and Collections. A **Community** is the major division of DSpace. A Community would be administered by a major department or division of the organization sponsoring the DSpace installation. In our installation, our Communities are the Becker Medical Library and the MidContinental Regional Medical Library. A DSpace installation can have two or more Communities.

A Community is made up of Sub-communities and Collections. A **Sub-community** is a smaller division of a community. A Sub-community must have collections to have any content. A **Collection** is the secondary division of DSpace. It is where the content of DSpace is stored. A Collection has users, groups, and records. A Collection cannot exist without a Community. There can be no content without a Collection. A Community can have two or more Collections.

Creating a Community

To create a Community, you must be an administrator. The URL to login as an administrator is <http://linpub1.wustl.edu:8080/dspace-admin>.

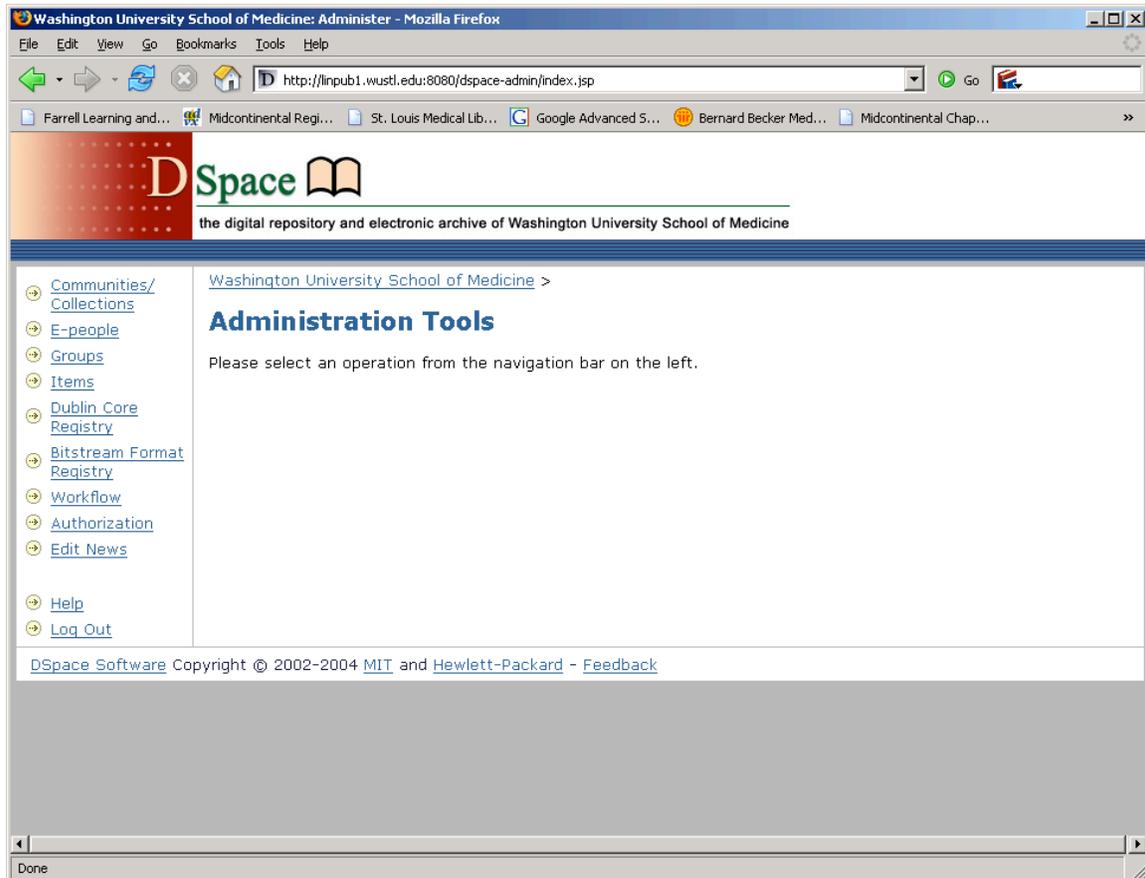


Image 1 – Administration Main Page

Click on the **Communities/Collections** link

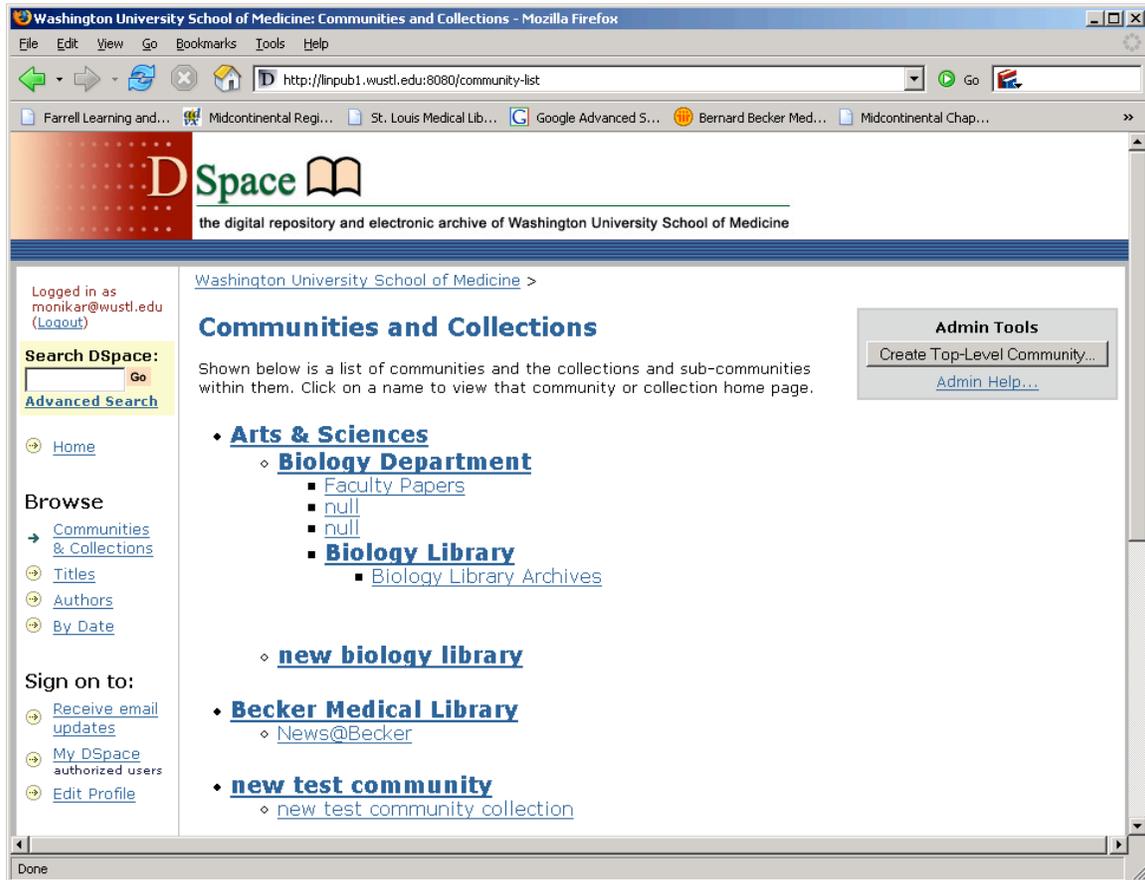


Image 2 – Edit Communities/Collections Page

From this page, the user can create Communities and Sub-communities.

To create a Community, click on the **Create Top-Level Community...** button to create a new Community. The Create Community page is pictured on the next page.

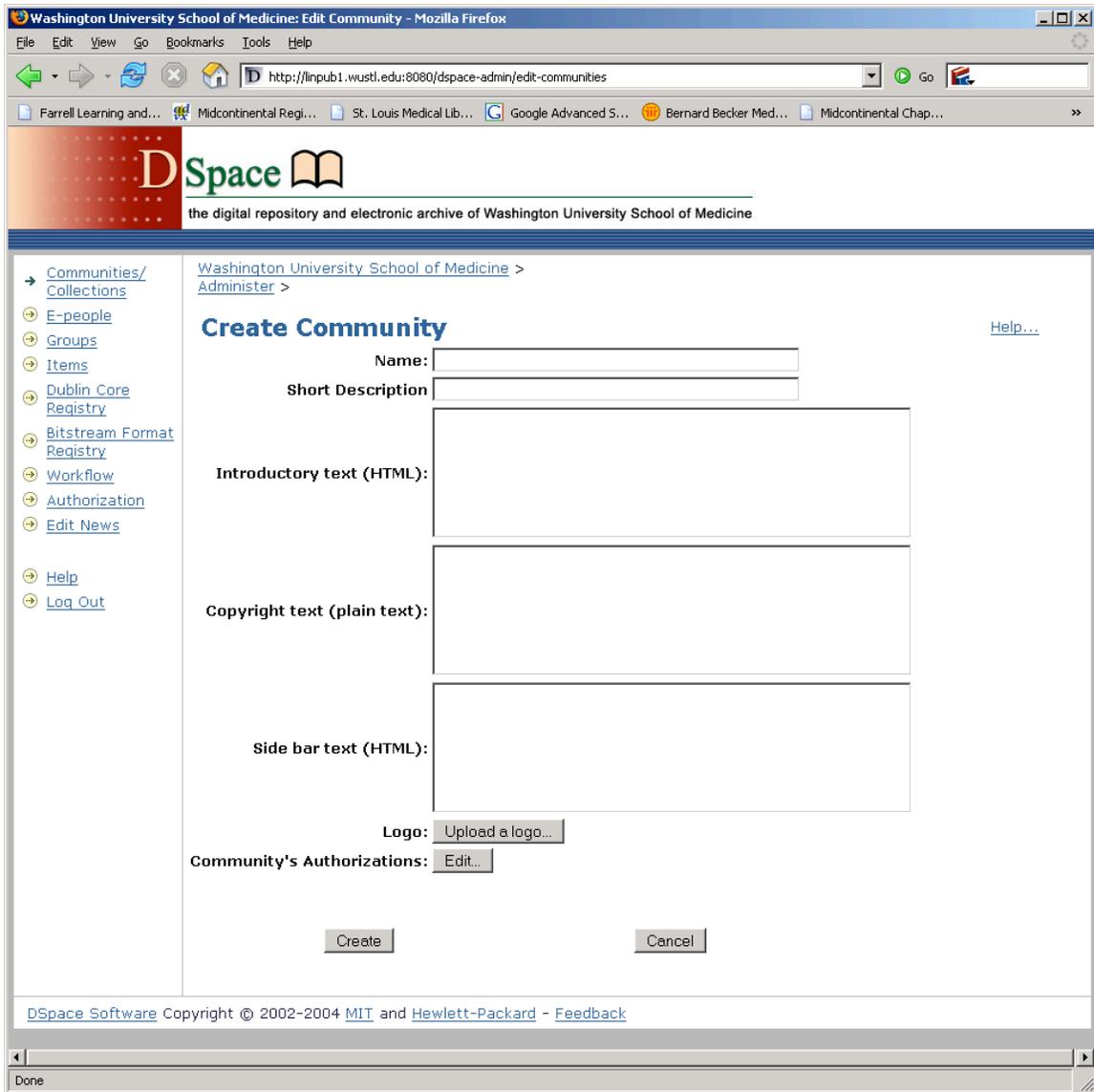


Image 3 – Create Community Page

Fields in Create Community Page:

- **Name:** Enter the name of your community here. Shows up in the top left corner of the Community page. **See Image 4 for all of the examples.**
- **Short Description:** Enter a short description of your community. Shows up under the name of the community.
- **Introductory text:** A short paragraph describing your community and its contents. Shows up between search box and list of collections. This text is written in HTML.
- **Copyright text:** Copyright statement for the community. Shows up when adding a record to a collection in this community.
- **Side bar text:** New about the community. Shows up below the list of new records to the community. This is also written in HTML.
- **Logo button:** Click here to add a logo to the community's main page.
- **Community's Authorizations button:** Click here to change the community's authorizations. See section about authorizations later in document.

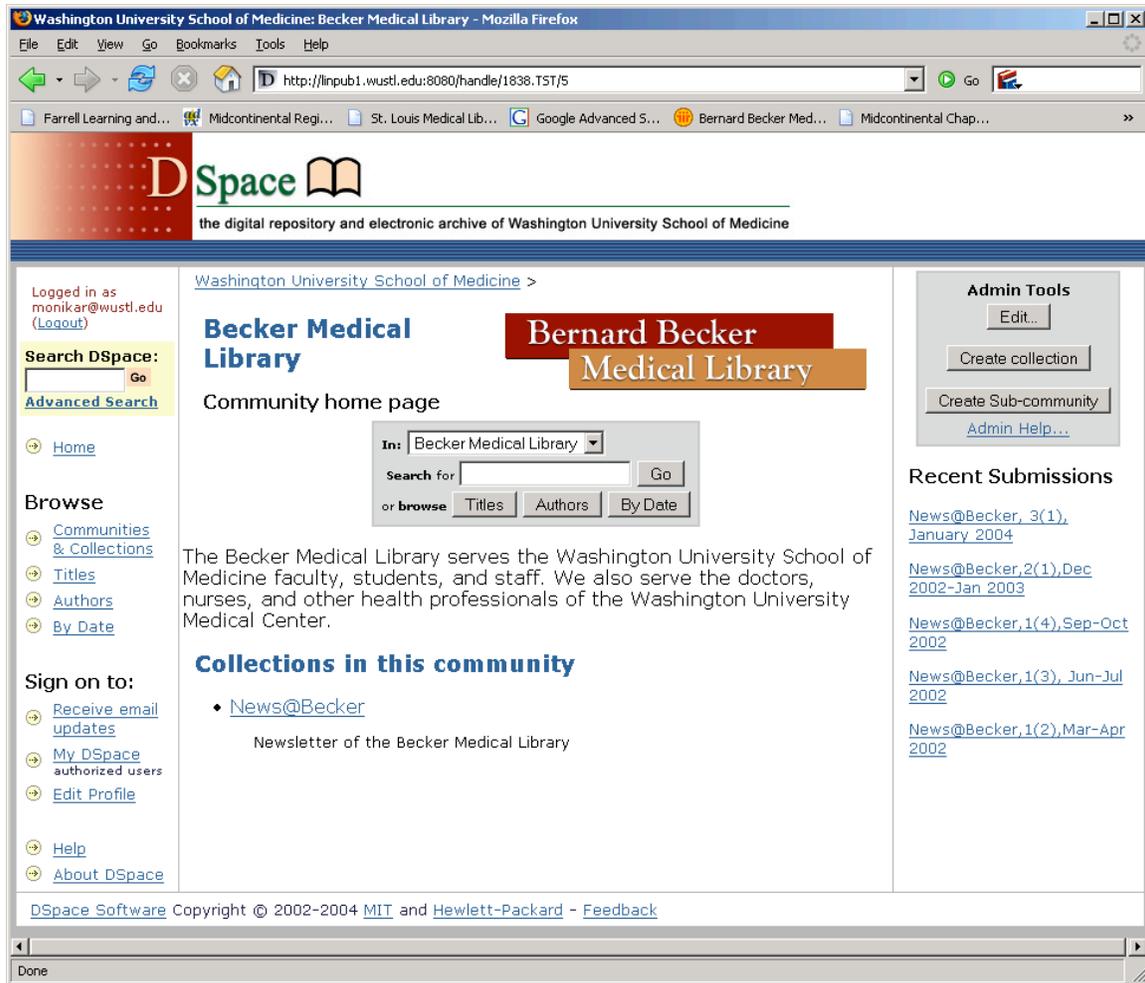


Image 4 – Public Community Page

After filling out the fields, click on the **Create** button to create a new community. To cancel, click on the **Cancel** button.

Once the user has created a Community, the community can be edited by clicking on the Edit button next to the community to be edited on the Edit Communities/Collections page (Image 2). Here is what the Edit a Community page looks like.

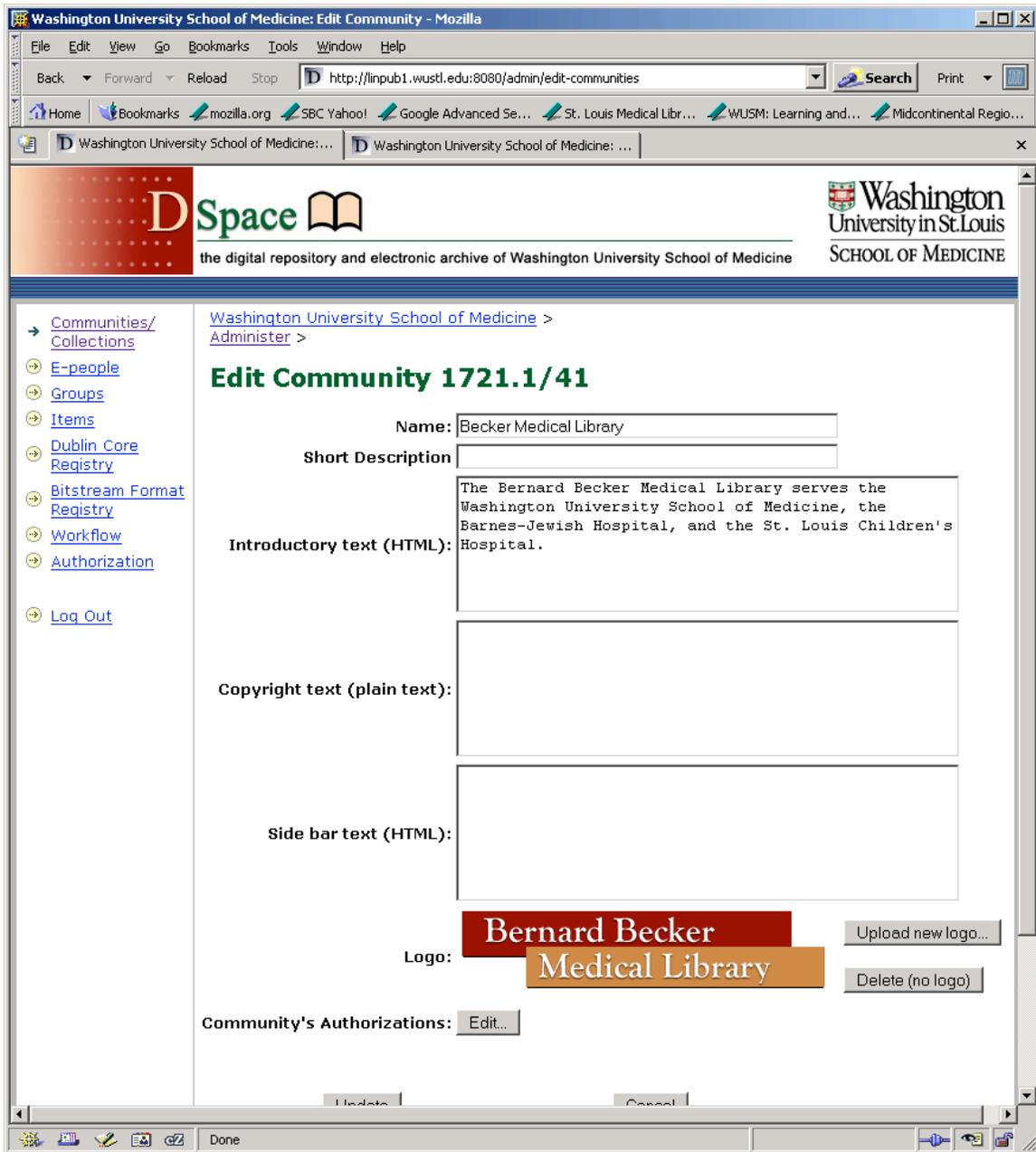


Image 5 – Edit Community Page

Here the text can be edited or new text can be entered. The user can change or delete the logo and change the policies for the community.

Creating a Sub-community

Back on the Communities/Collections page (Image 2), once a top-level Community is created, Sub-communities or Collections can be created.

To create a Sub-community, click on the link to the top-level Community.

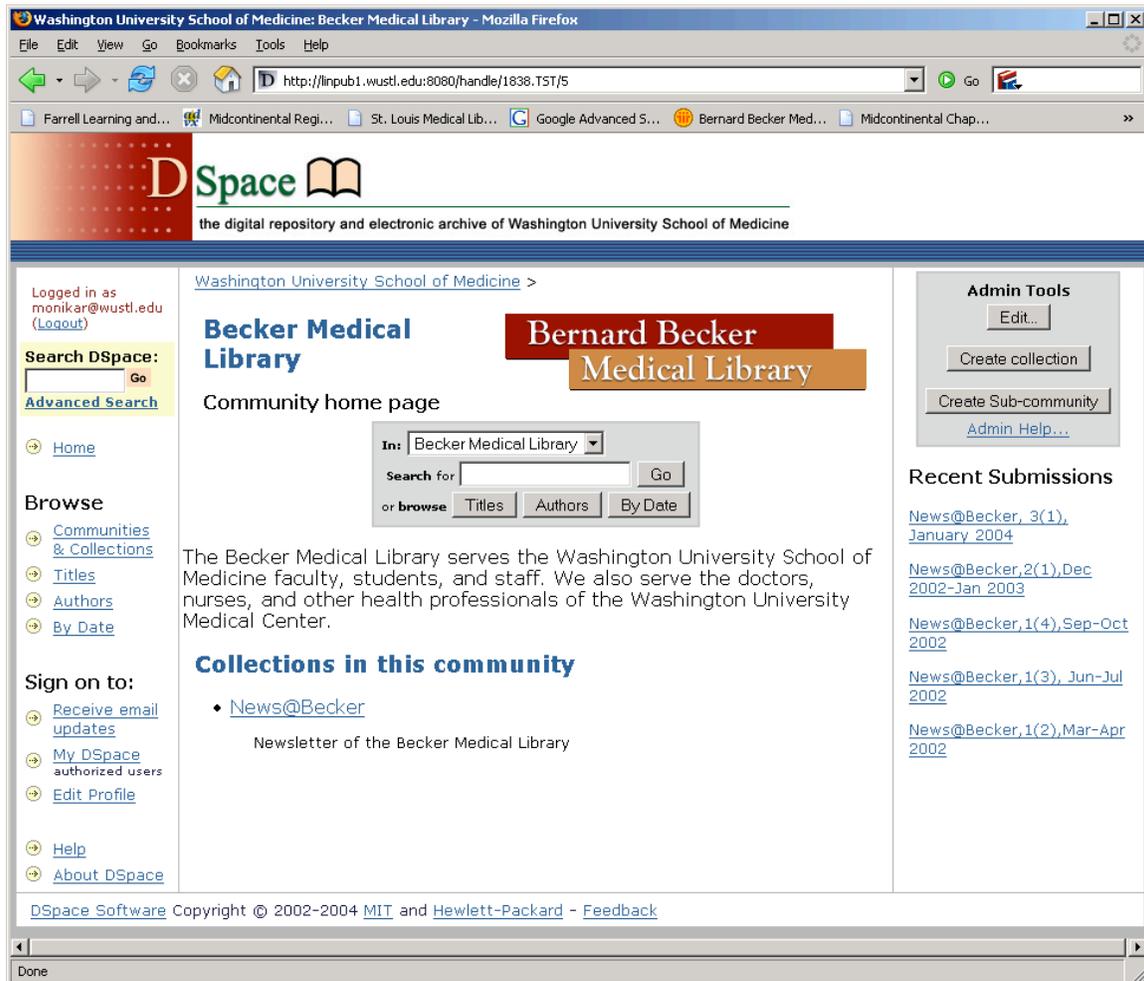


Image 6 – Community Page

Along the right side, you would click on the Create Sub-community button to create a Sub-community.

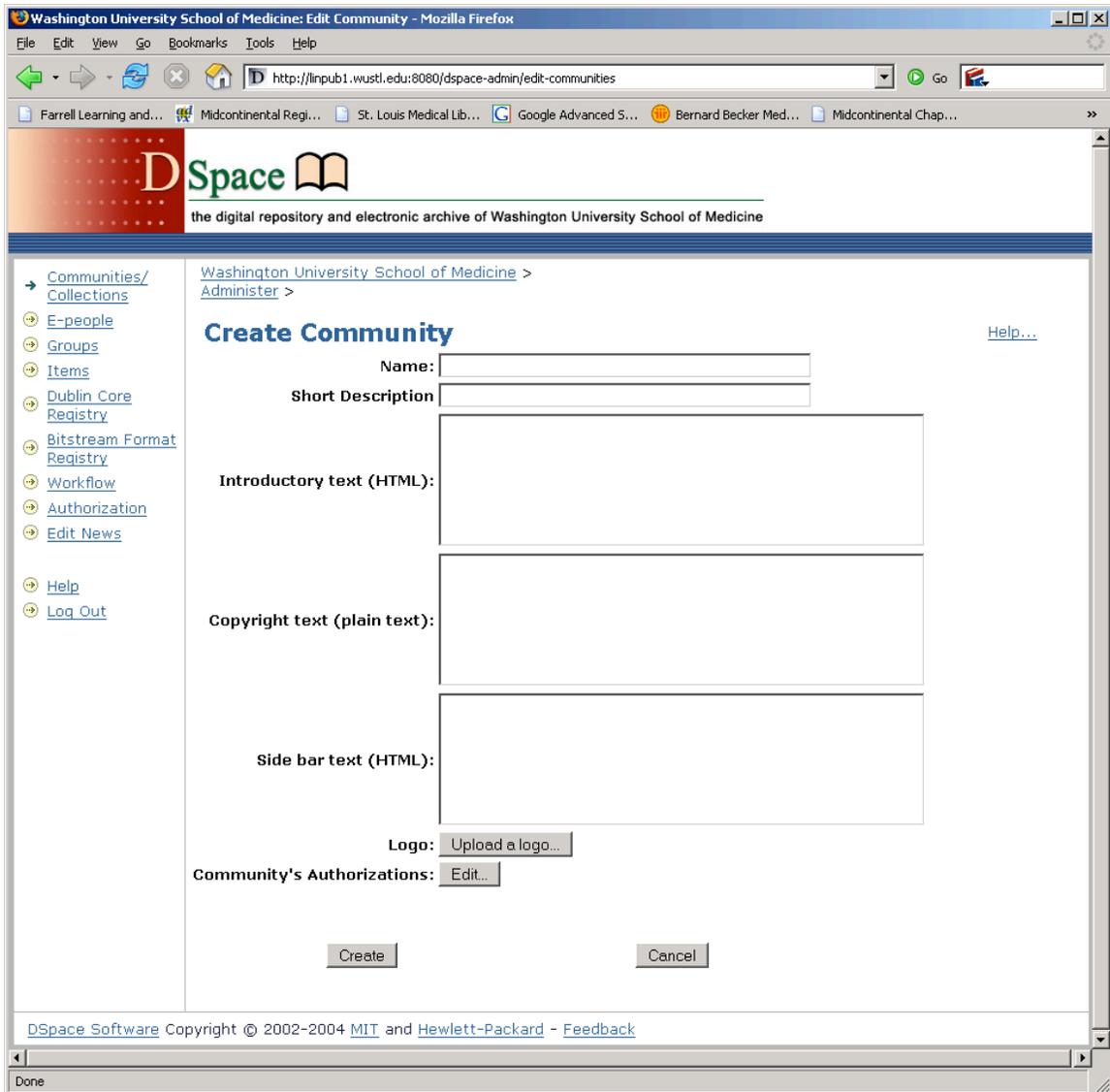


Image 7 – Create Sub-community Page

As you see, Image 7, the Create Sub-community Page is the same as Image 3, the Create Community Page. However, this sub-community will be under the top-level community of the Becker Medical Library. It won't have any contents itself but will provide a way for the Becker Library community to break up its community into sub-communities.

Creating a Collection

Just as for a Community, only an administrator can create a Collection. This is done at the Edit Communities/Collections page (Image 2).

Once you have created a Community, the only way it can have any content is inside a Collection. In the right side column of the Community or Sub-community home page, are buttons to perform the following functions:

- Edit the current Community
- Create Collection
- Create Sub-community

Here is what the page looks like:

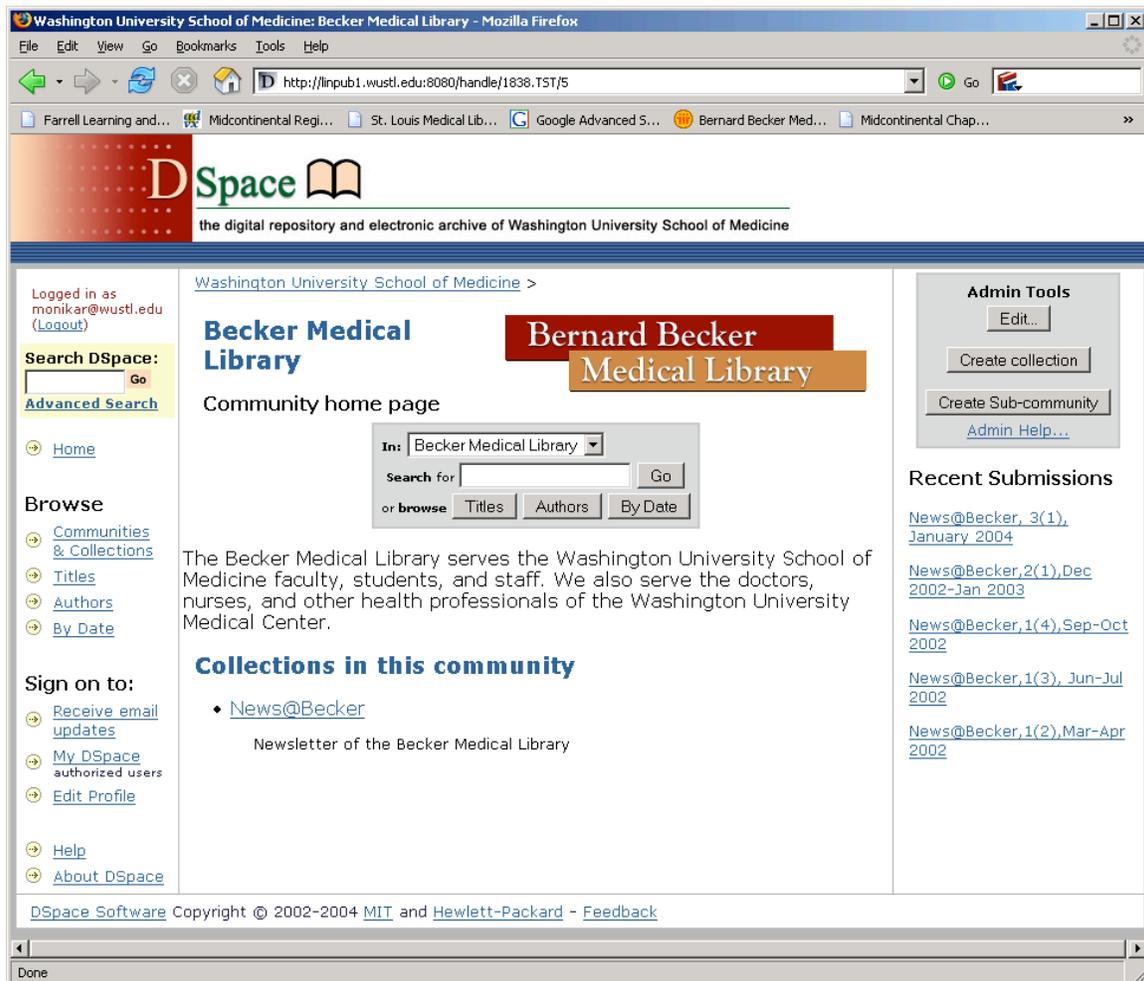


Image 8 – Create a Collection

To Create a Collection, you click on the **Create collection** button in the right side column. This is the first screen you will get in the process.

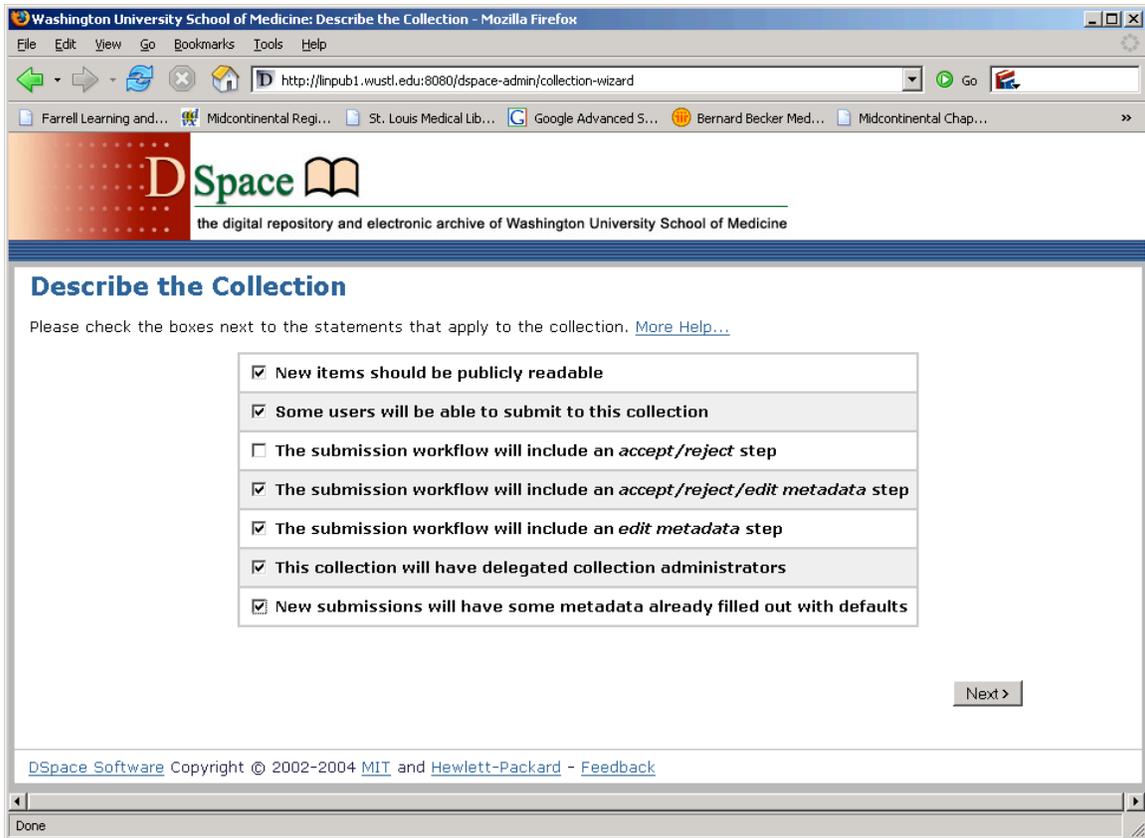


Image 9 – Describe the Collection

This step allows you to create the workflow steps that will be needed by this collection. It also sets up whether the collection will have collection administrators and whether the submission form will have some metadata fields filled out by default.

When finished selecting your choices for this screen, click on the **Next** button.

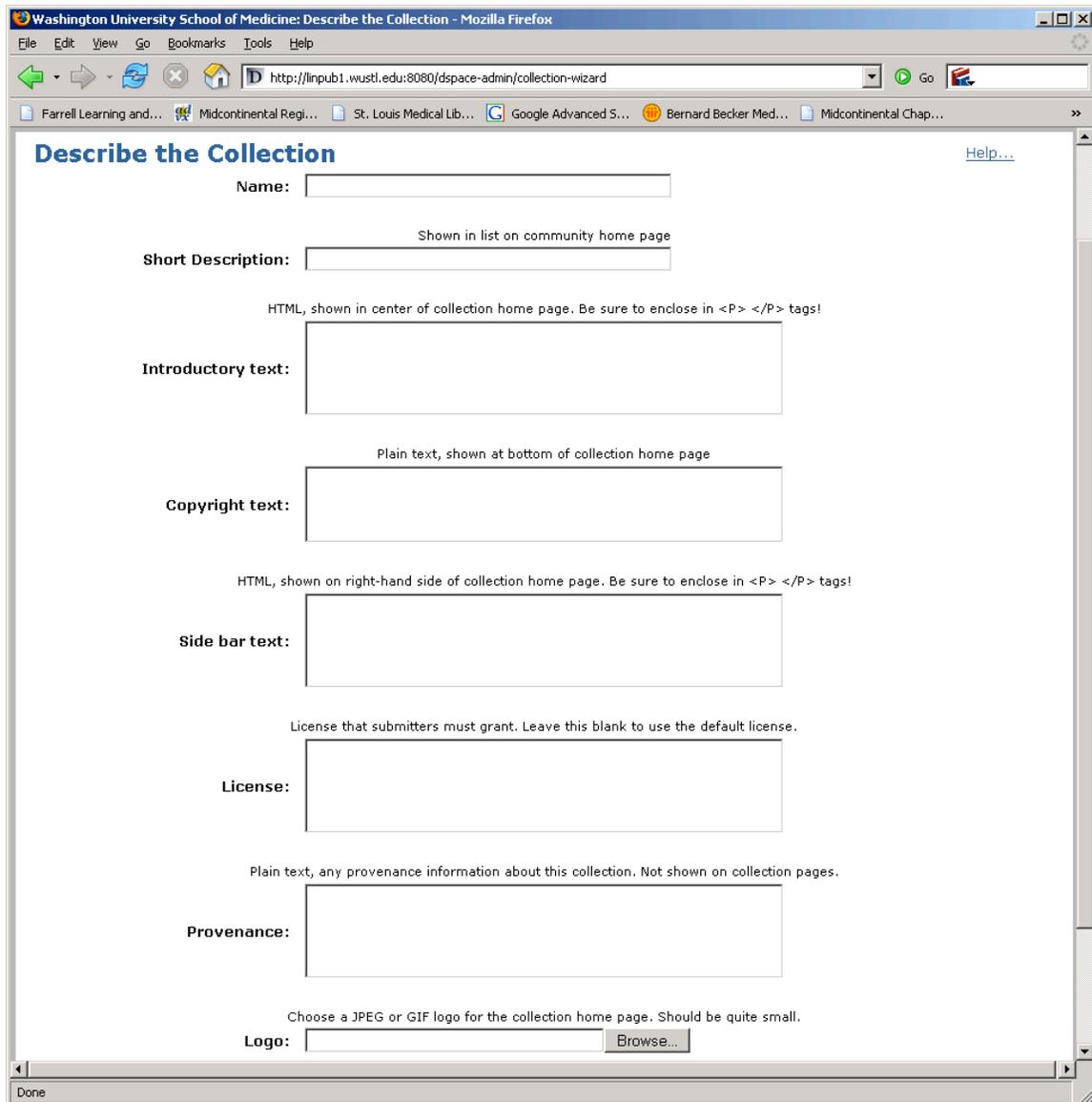


Image 10 – Describe the Collection 2

This screen lets you describe the collection in full. Here are the fields in the Describe the Collection Page

- **Name:** Enter the name of your community here. Shows up in the top left corner of the Community page. **See Image 7 for all of the examples.**
- **Short Description:** Enter a short description of your community. Shows up under the name of the community.
- **Introductory text:** A short paragraph describing your community and its contents. Shows up between search box and list of collections.
- **Copyright text:** Copyright statement for the community. Shows up when adding a record to a collection in this community.

- **Side bar text:** New about the community. Shows up below the list of new records to the community.
- **License text:** This text grants DSpace license to store and distribute the author's work. The author or submitter grants this license when they submit their document to DSpace.
- **Provenance:** This field is for you to use to add any text that you feel is relevant.
- **Logo button:** Click here to add a logo to the community's main page.

After filling out these fields, click the **Next** button. The next screen is the Authorization to Submit page where you can select individuals or groups that are allowed to submit items to this collection.

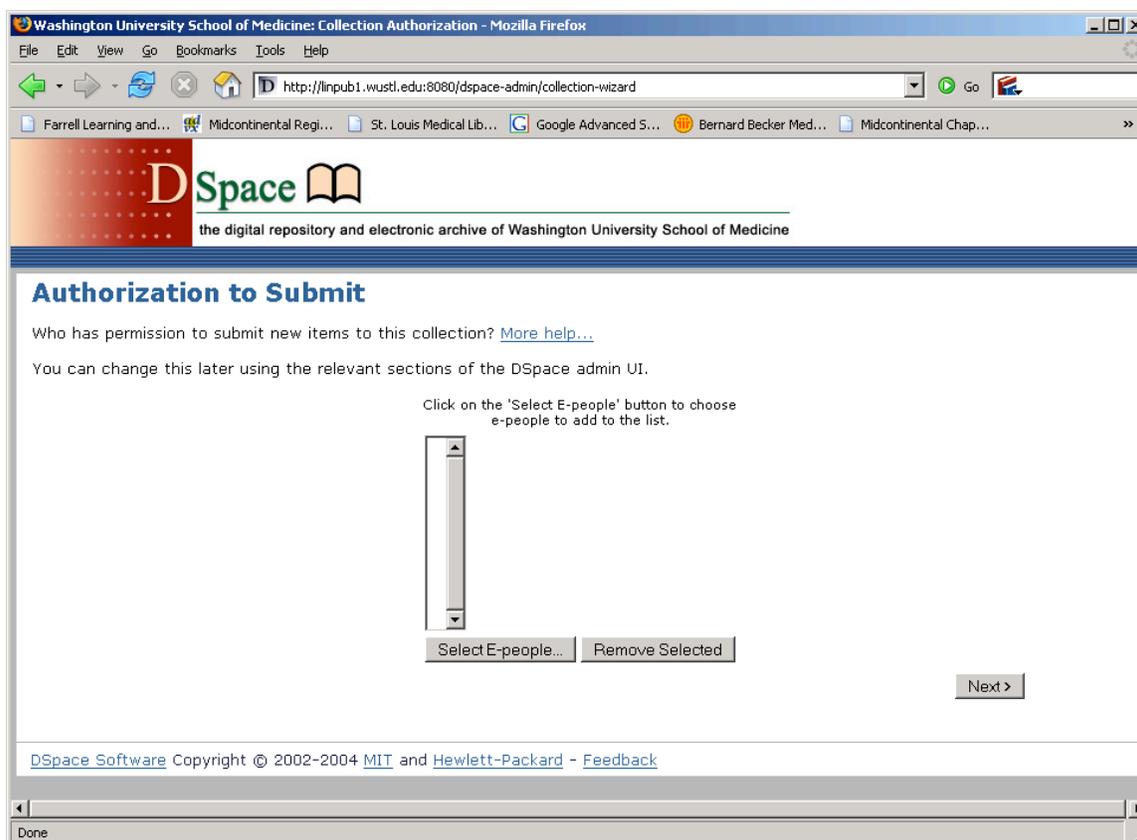


Image 11 – Authorization to Submit 1

Click on the **Select E-People** to make your selections. A new window will open and you can select e-people who will be allowed to submit items to this collection.

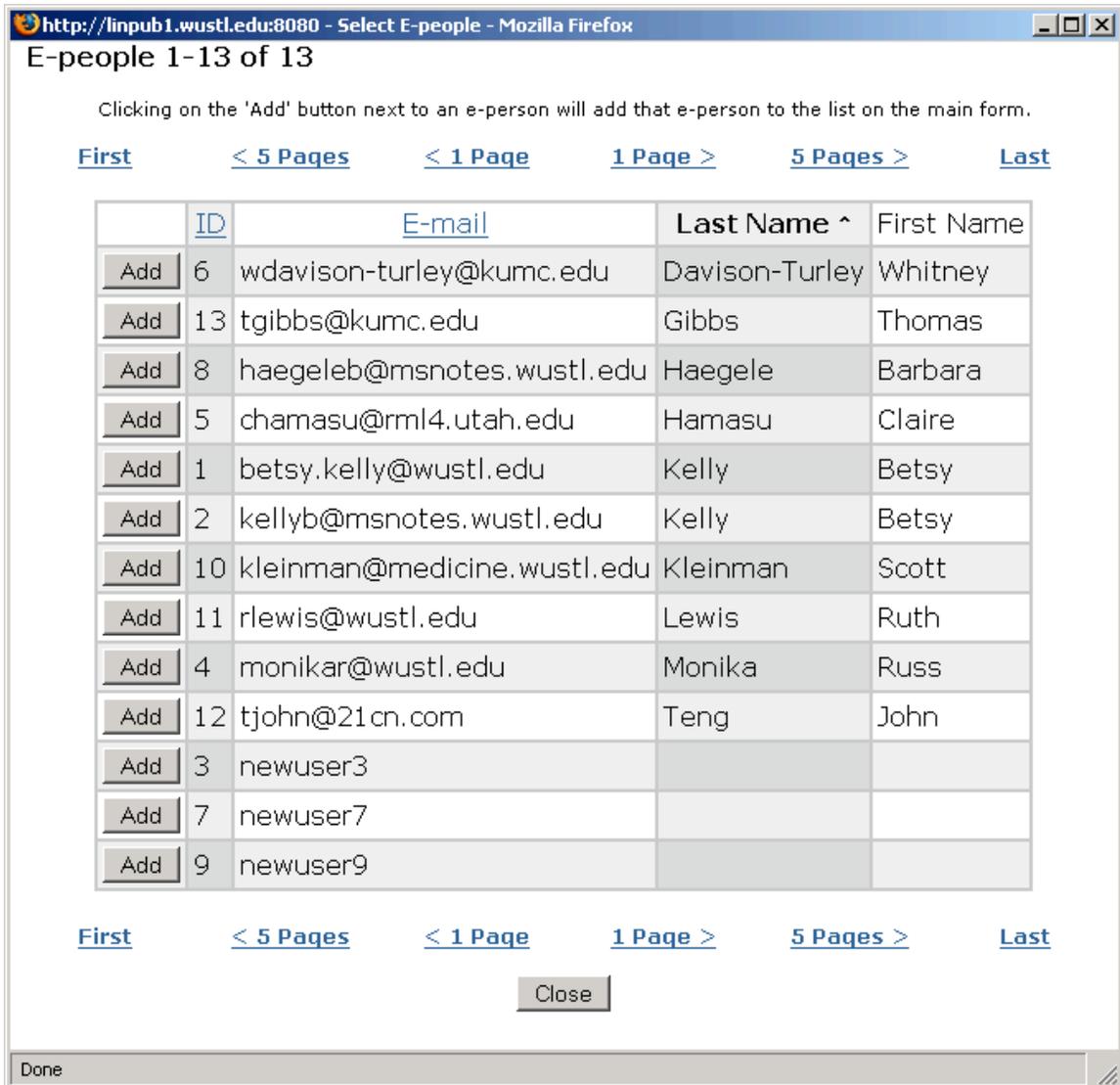


Image 12 – Selecting E-People

Once you have selected e-people by clicking on the Add button by their names, click on the **Close** button to close the window and now the Authorization to Submit page has people in the window.

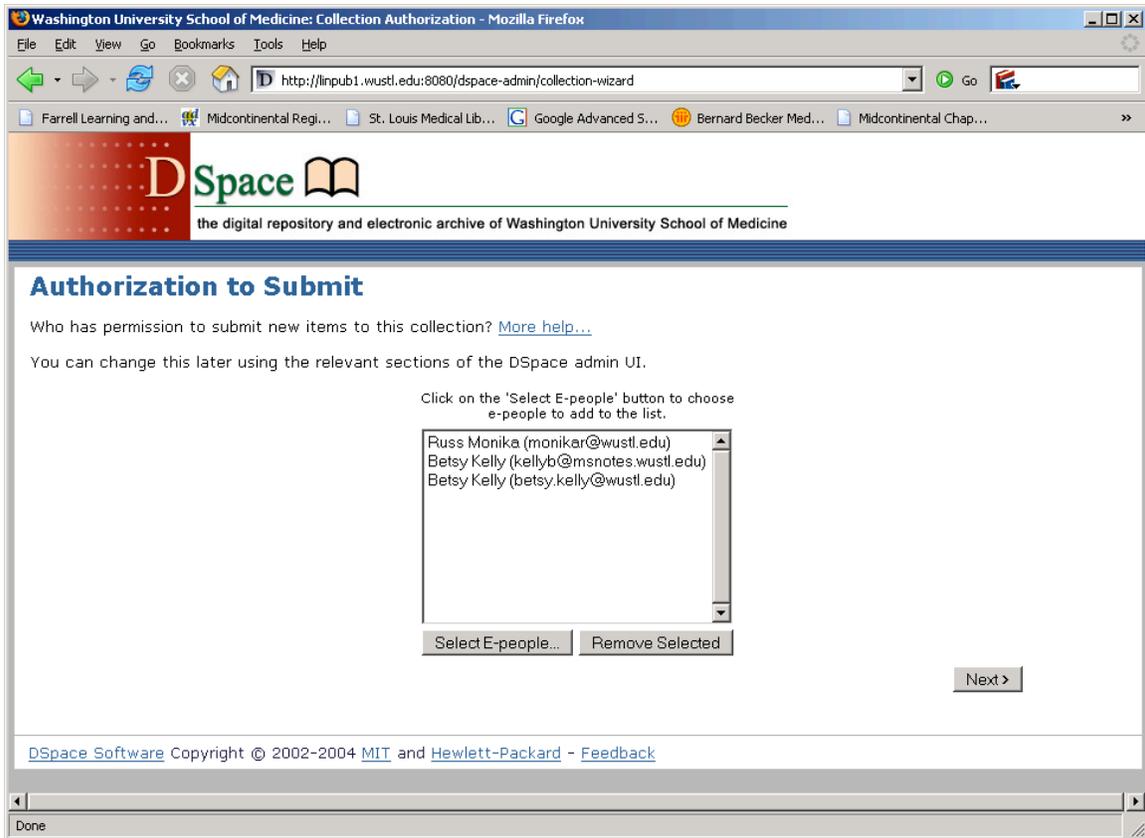


Image 13 – Filled out Authorization to Submit page

Here is the Authorization to Submit page filled out with e-people that have been selected to submit items to this collection. After this has been completed, click on the **Next** button to go to the next step in the process.

The next step is setting up the **Workflow Accept/Reject Step**. This used to be called Workflow 1 in previous versions of DSpace. This is the step that can only accept or reject a submission.

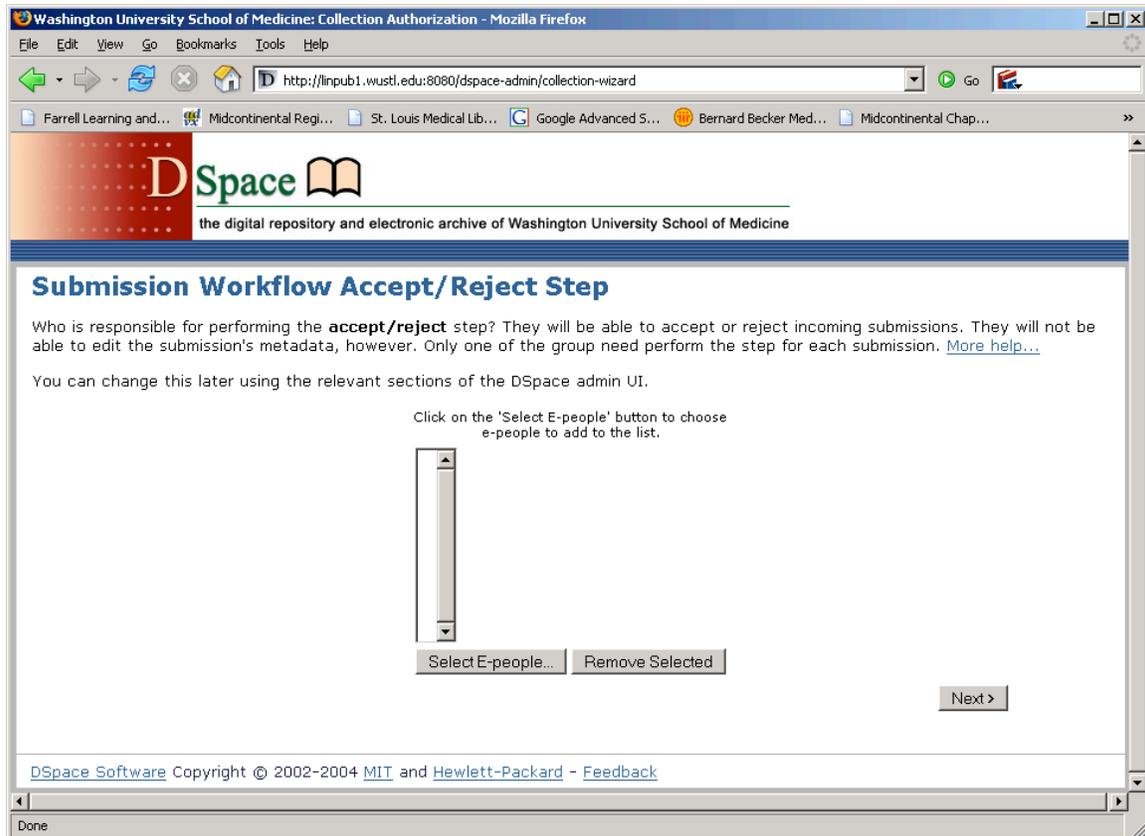


Image 14 – Workflow Accept/Reject Step

It works similarly to the Submission step. You click on the **Select E-People** button and a new window opens. From that list, you select people by click on the **Add** button next to their names to add those people to this step. When finished, click on the **Close** button to close that window. When you are finished with this step, click on the **Next** button to move on to the next workflow step.

The other two workflow steps are:

- **Workflow Accept/Reject/Edit Metadata Step** – This workflow step can accept or reject a submission as well as edit the metadata for the submission.
- **Workflow Edit Metadata Step** – This step is the final one in the workflow process. This step can no longer reject the submission. It can edit the metadata. Once this step, if is part of the submission process, OKs the submission, the item becomes part of the collection.

After you have set up the workflows that you selected on the first screen of the collection creation process, the next step is to set up the collection administrators. This is if you set up collection administrators other than the system administrators.

Workflow Notes

If no Workflows are assigned to a Collection, the submitted items are archived in DSpace without any review.

Accept/Reject Step should be assigned to a subject matter expert who can assess the quality and accuracy of the work. The only point of the Accept/Reject Step is to accept or reject the submission based on appropriateness to the collection and community. If there is no Accept/Reject/Edit Metadata Step or Edit Metadata Step, it is archived in DSpace without any opportunity to look at the metadata entered by the submitter.

Accept/Reject/Edit Metadata Step is the best one to use if you only want one workflow for the collection. It can be used accept or reject the submission and it can also edit the metadata submitted with the item or items. The submitter is probably not well versed in the metadata concept, which is why having a workflow step that has someone who is versed in metadata analysis is very important to the accuracy of the information in the DSpace records. If there is no Edit Metadata Step, the record is archived in DSpace.

Edit Metadata has no power to reject the submission. It is used to edit the metadata. Once it is edited by WF 3, it is archived in DSpace.

See Appendix D – DSpace Flowchart for more information.

The **Delegated Collection Administrators** page will show up if you made that selection at the beginning of the process after you click **Next** on the last Workflow step.

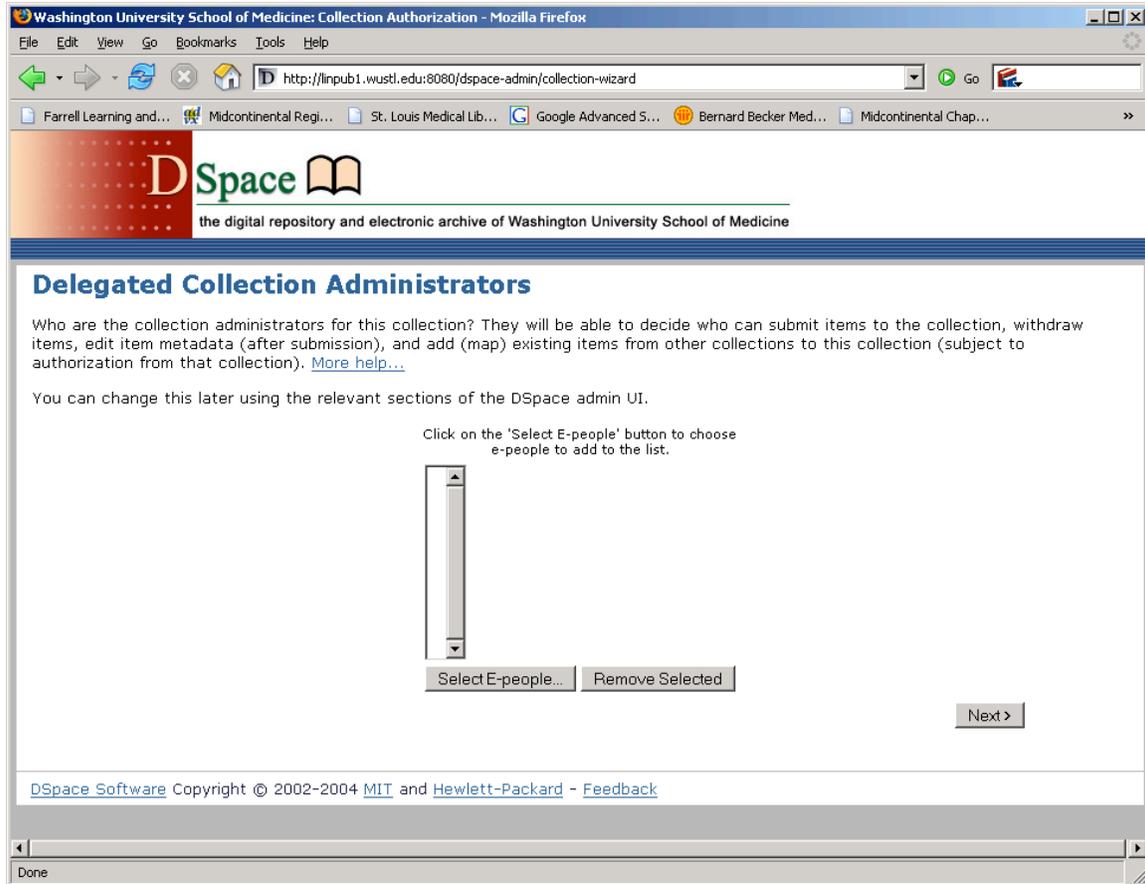


Image 15 – Delegated Collection Administrators

This page works like the Workflow steps. You click on the **Select E-People** and click on the **Add** button next to people's names on the resulting list in the new window. When you are finished, click on the **Close** button to close the new window. Click on the **Next** button to move to the next step in the Create Collections process.

The next step would be to identify and fill out the default metadata fields for each item.

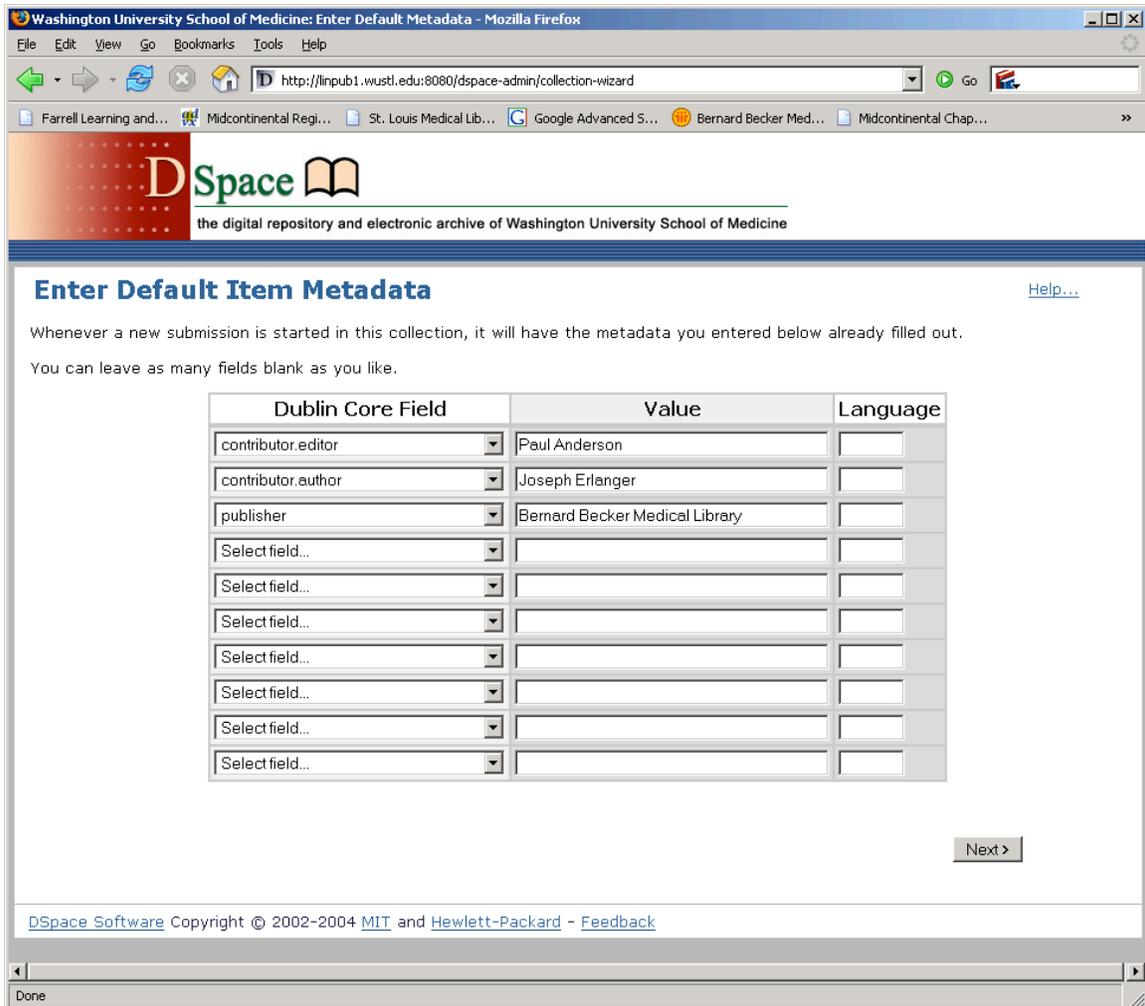


Image 16 – Enter Default Item Metadata

Here you can chose which metadata fields will have default data in it and what that data will be. This will make the submissions easier and quicker.

Once you have finished adding default values for metadata fields, click on the Next button. This will take you to the Edit Collection page where all items will be filled out. This is the place where you can review all of the selections you have just gone through.

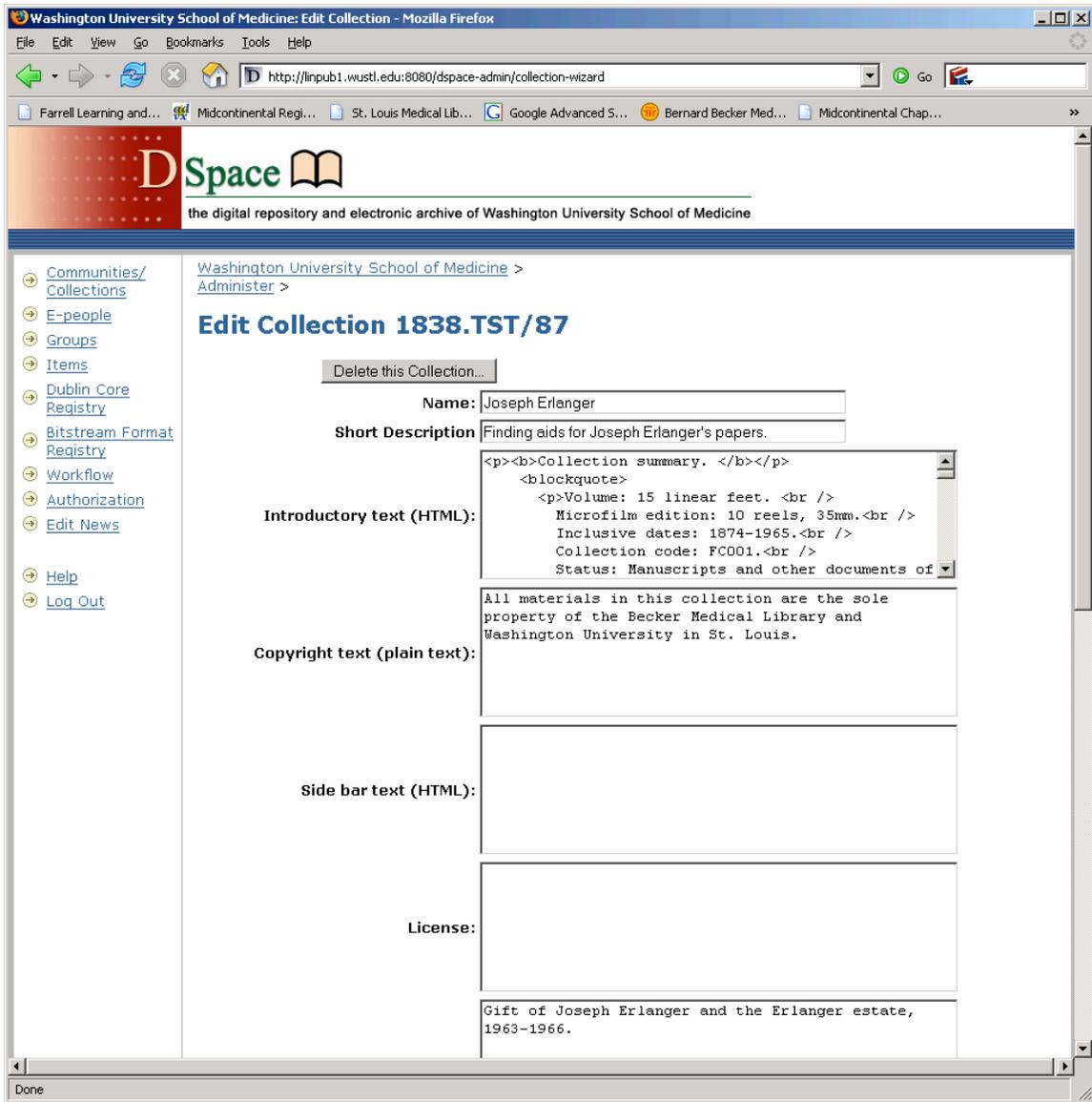


Image 17a – Edit Collection Page – Part 1

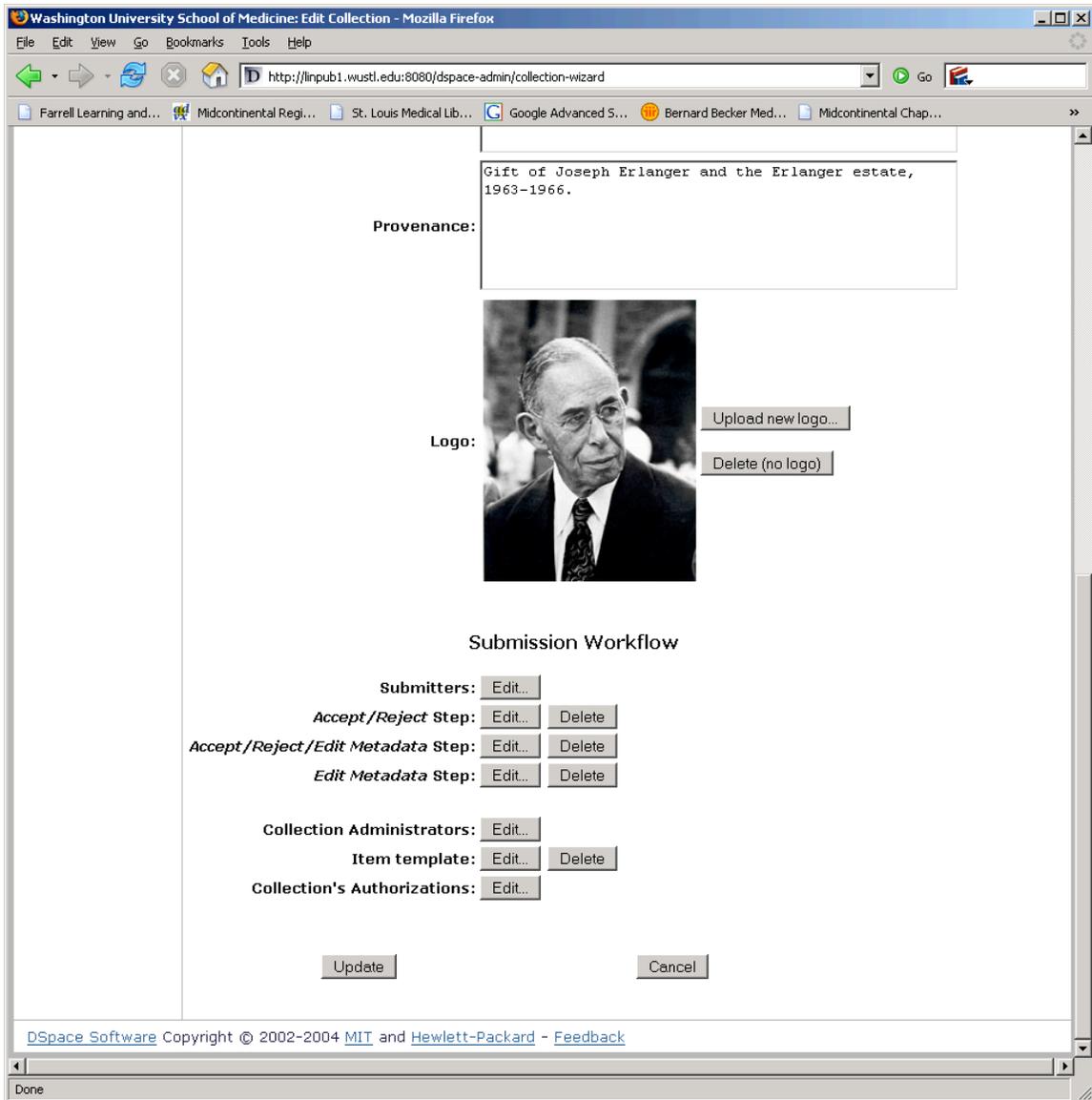


Image 17b – Edit Collection Page – Part 2

To confirm all of the selections you have made, click on the Update button at the bottom of the page. Now the collection is ready to accept submissions.

Creating Epeople and Groups

Individual users in DSpace are called **EPeople** in the Administration section. **Groups** can be created and two or more epeople can be added to the group to give multiple users the same authorizations or assign multiple users to the same workflow.

EPeople

EPeople are added to DSpace by creating an account from the DSpace Main Page (Image 40). They click on the MyDSpace link in the left column and go to the Log in to DSpace page.

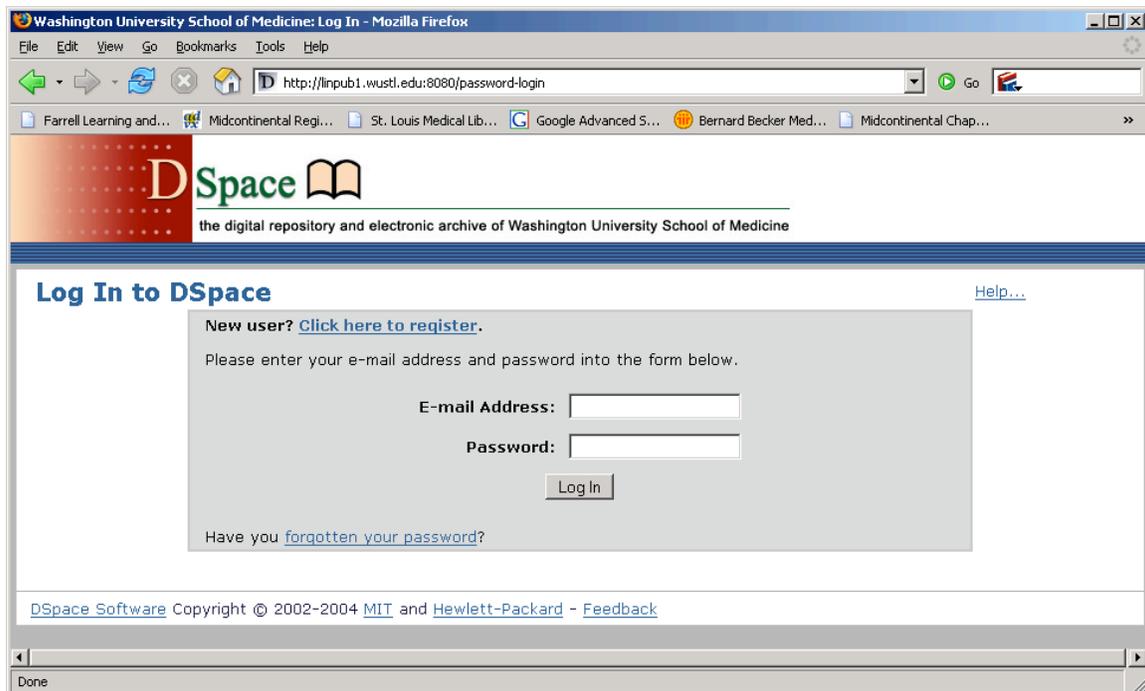


Image 18 - Login to DSpace Page

If you are already a registered user, you would enter your **e-mail address** and **password** and log in to DSpace. If you are a new user, you would click on the **Click Here to Register** link to register to use DSpace.

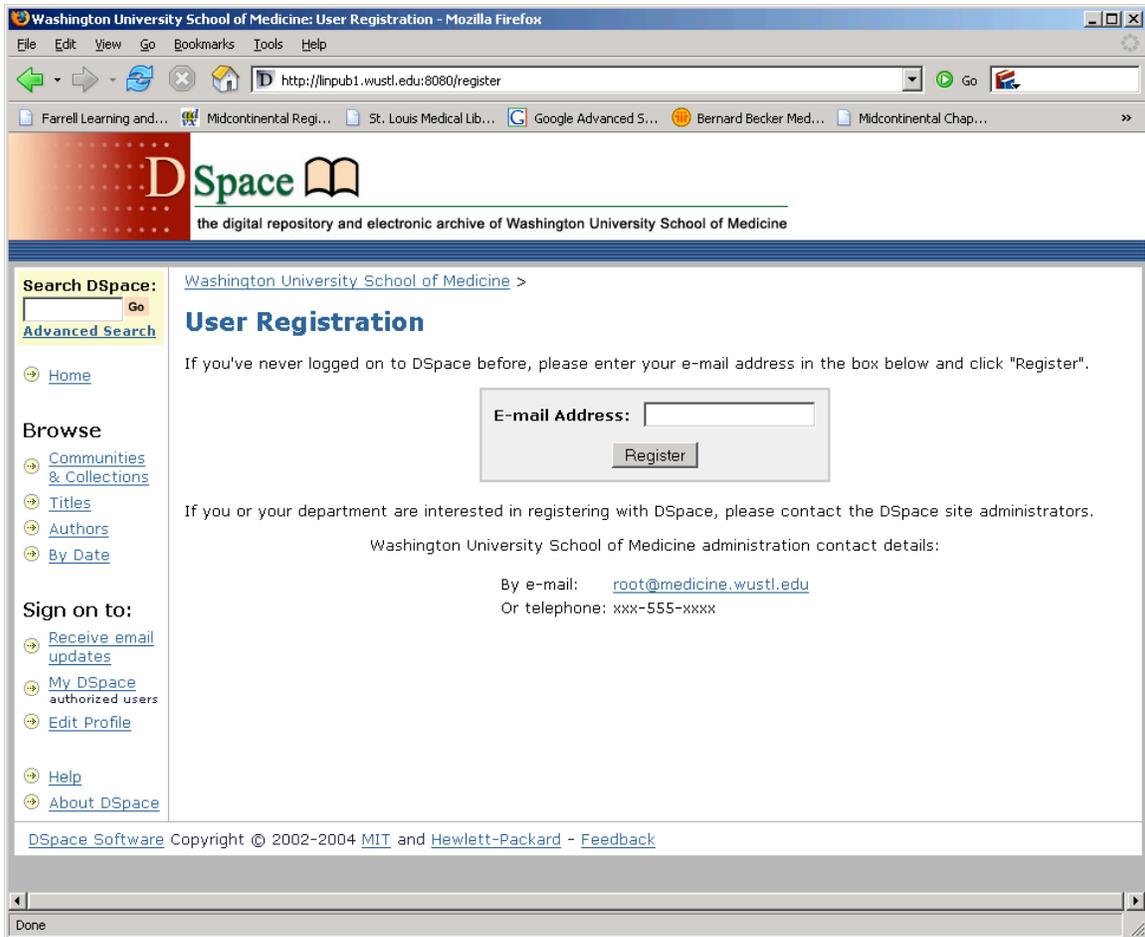


Image 19 – User Registration

You will enter your e-mail address in the text box and click on the Register button. This will send a registration request to the DSpace server administrator. The administrator will send you an email with your registration information shortly. Once you receive this information, you can go to the Log in to DSpace page and log in to MyDSpace.

MyDSpace

MyDSpace is an Eperson's portal to DSpace. Here is what the MyDSpace looks like.

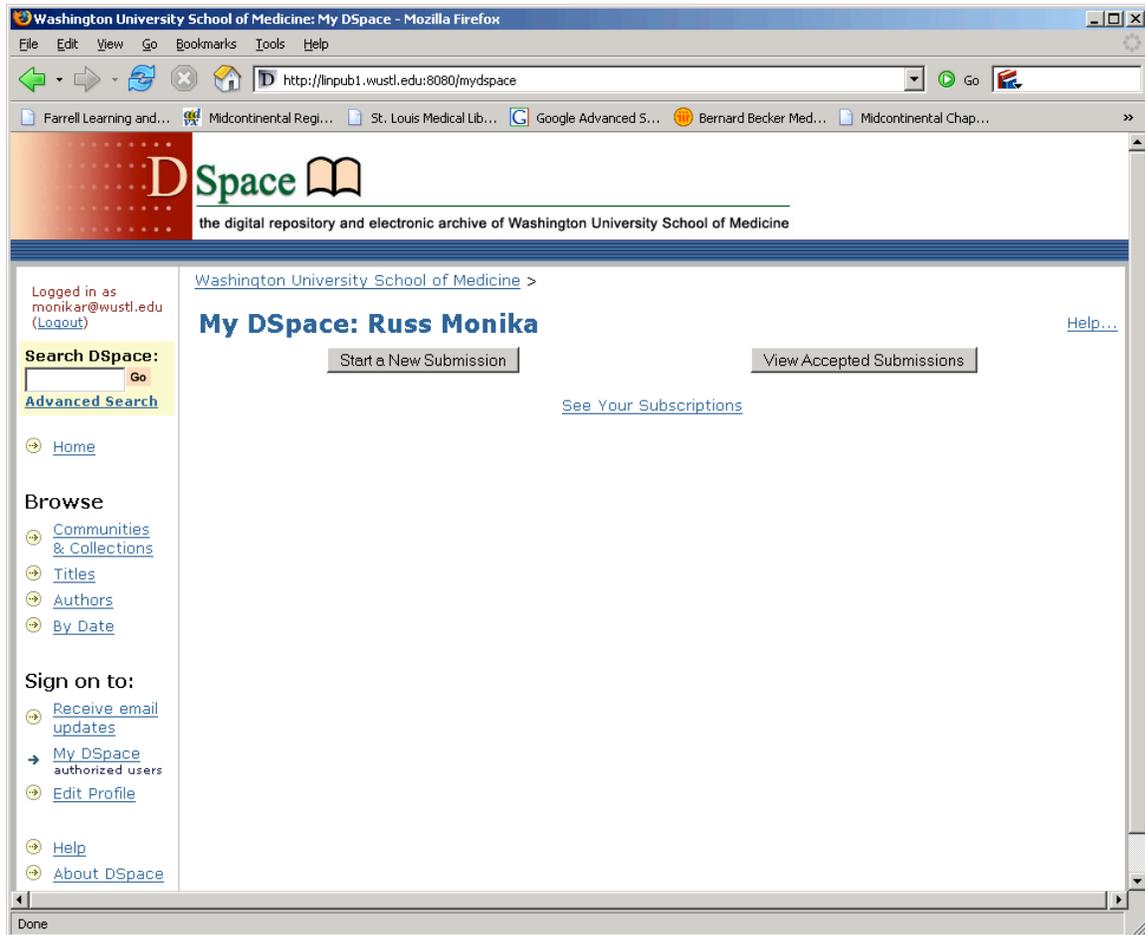


Image 21 – MyDSpace

It has the standard window as does the general main page. But here you can click on the **Start a New Submission** to submit an item to DSpace. You must be an Eperson to submit items to DSpace. We will go over the Submission process later in this report.

You can also click on the **View Accepted Submissions** to see all of the items that you have submitted that have been accepted into DSpace.

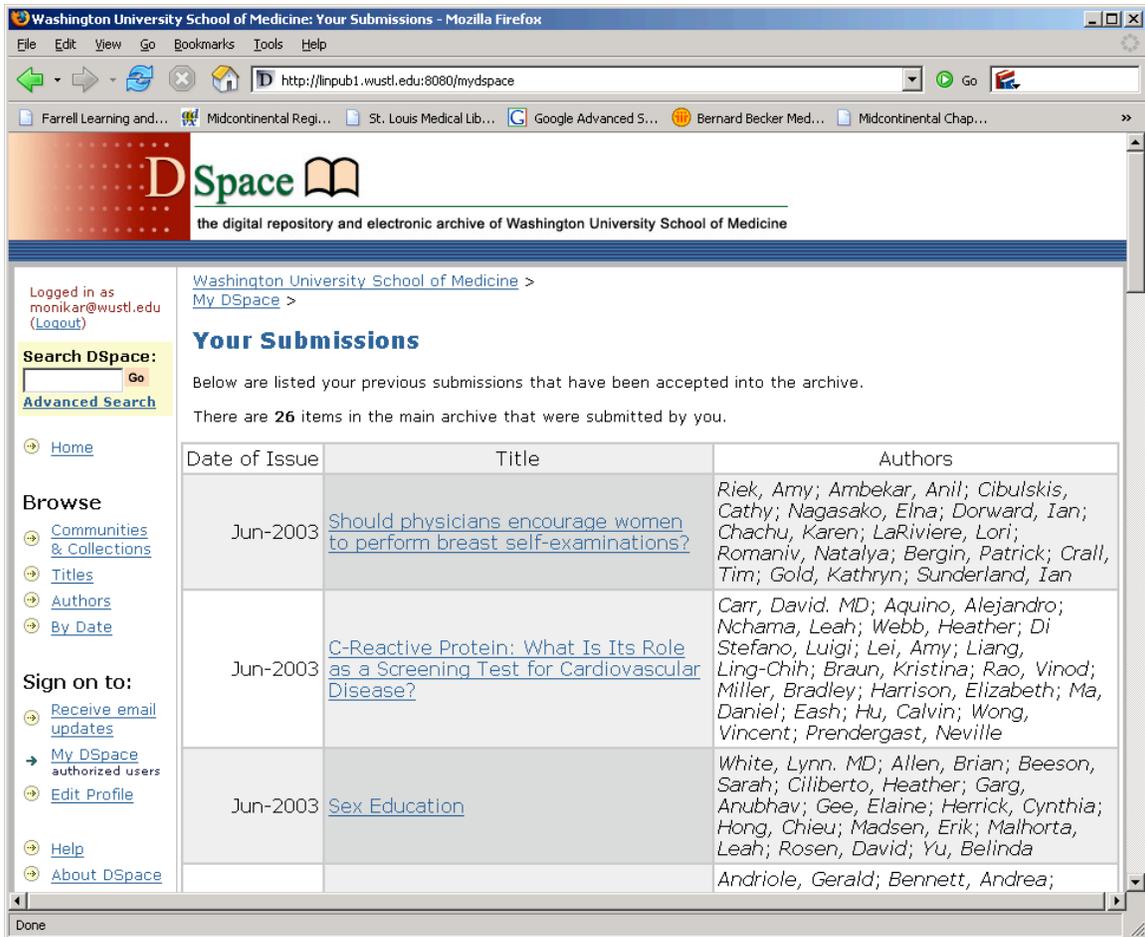


Image 21 – Accepted Submissions

You can also click on the **See Your Subscriptions** link to see to which Collections you are subscribed. As a DSpace user, you can **subscribe to a collection** and be notified when a new item has been submitted to the collection.

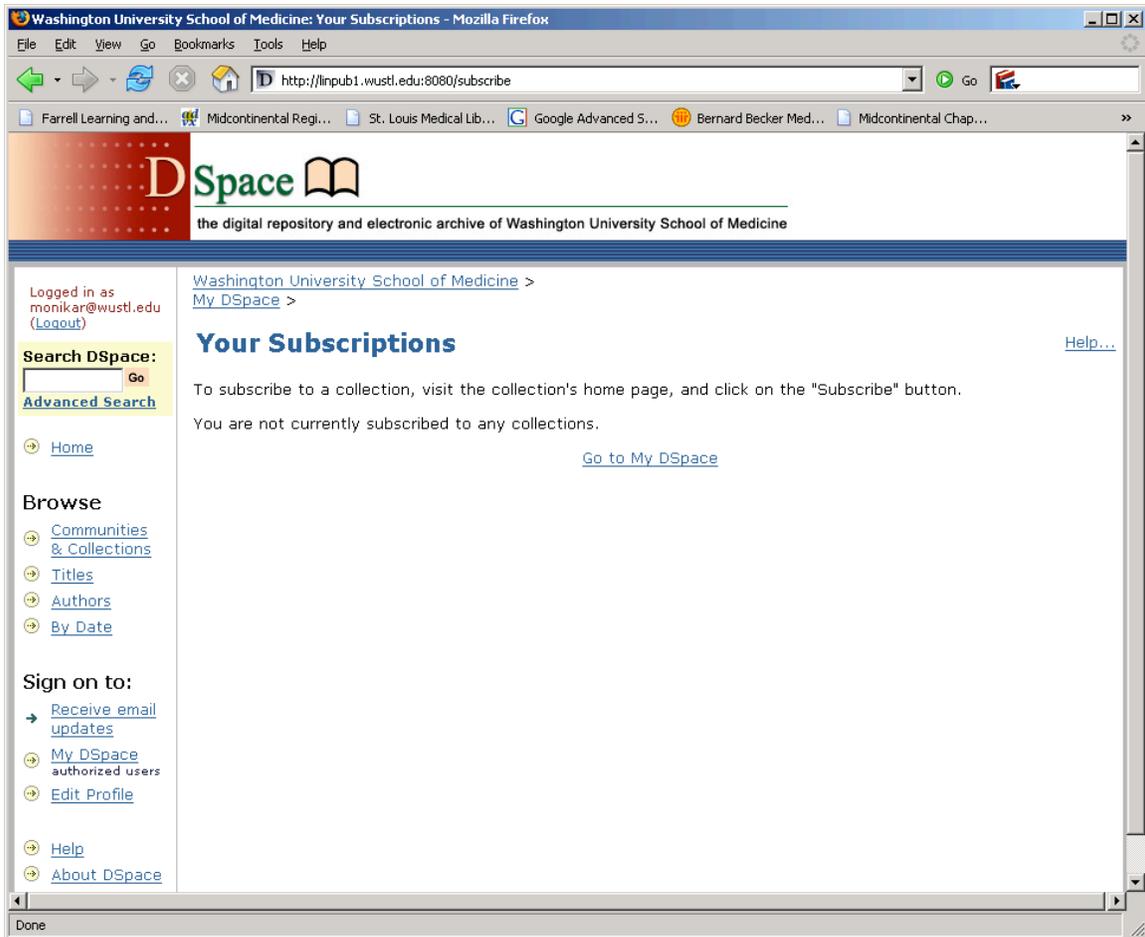


Image 22 – Your Subscriptions

Administration

From the Administration Main Page (Image 1) you can get the list of EPeople in the WUSTL DSpace by clicking on the **EPeople** link in the left side menu.

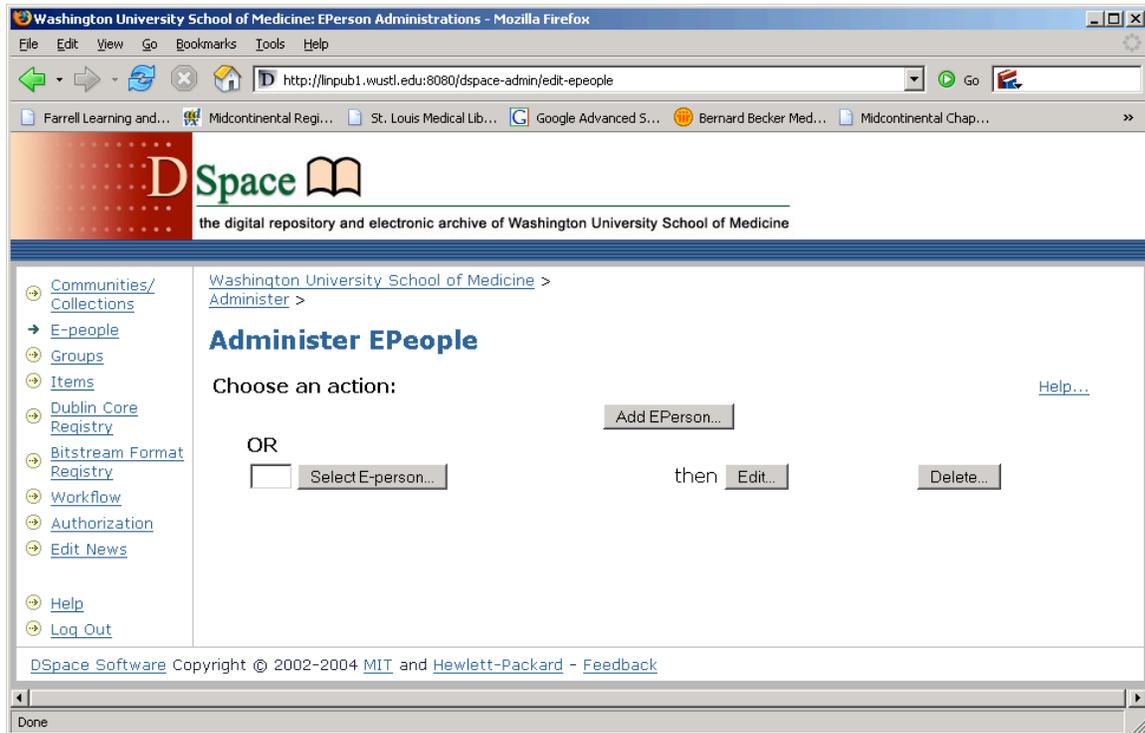


Image 23 – Administer EPeople Page

On this page, you can add an e-person or select an e-person to edit or delete.

To add an e-person, click on the Add E-person button.

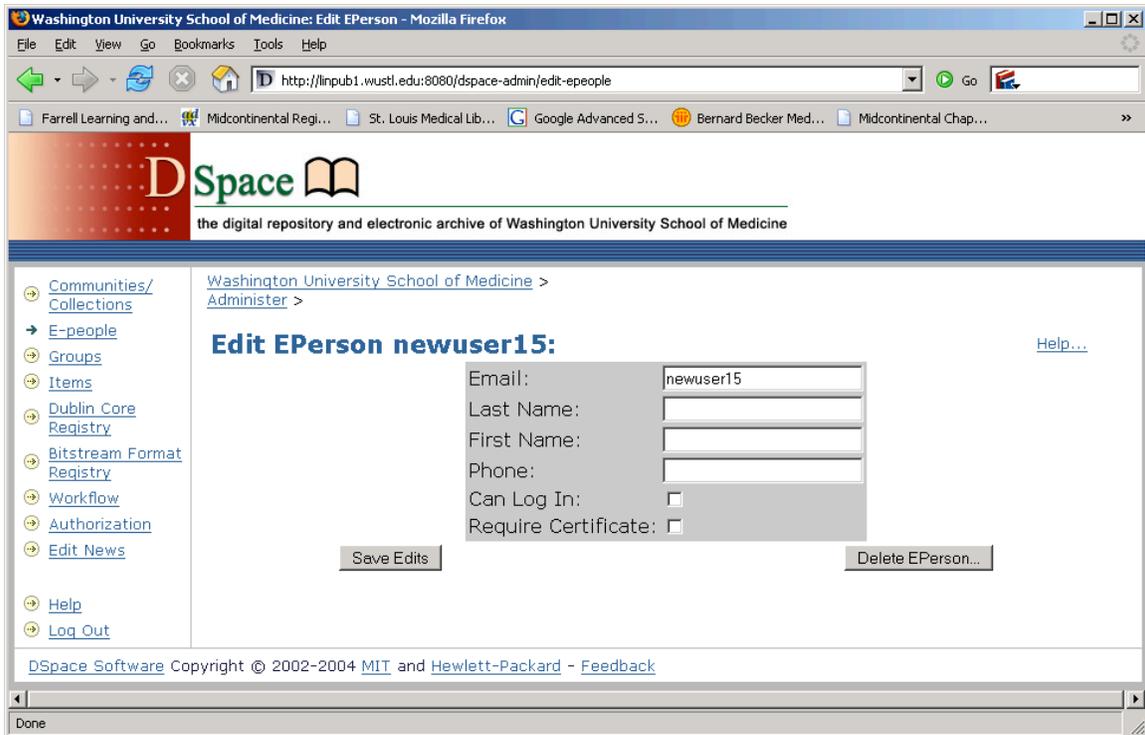


Image 24 – Edit Eperson Page

Here you can edit the fields in the Eperson's record and click on the Save Edits button to save the record. You can also delete the Eperson on this page as well.

To edit or delete an e-person, click on the **Select E-person** button and a new window will pop up with a list of all of the e-people. Click on the **Add** button next to the person you wish to edit, and the window will close and the e-person will be added to the text box next to the button.

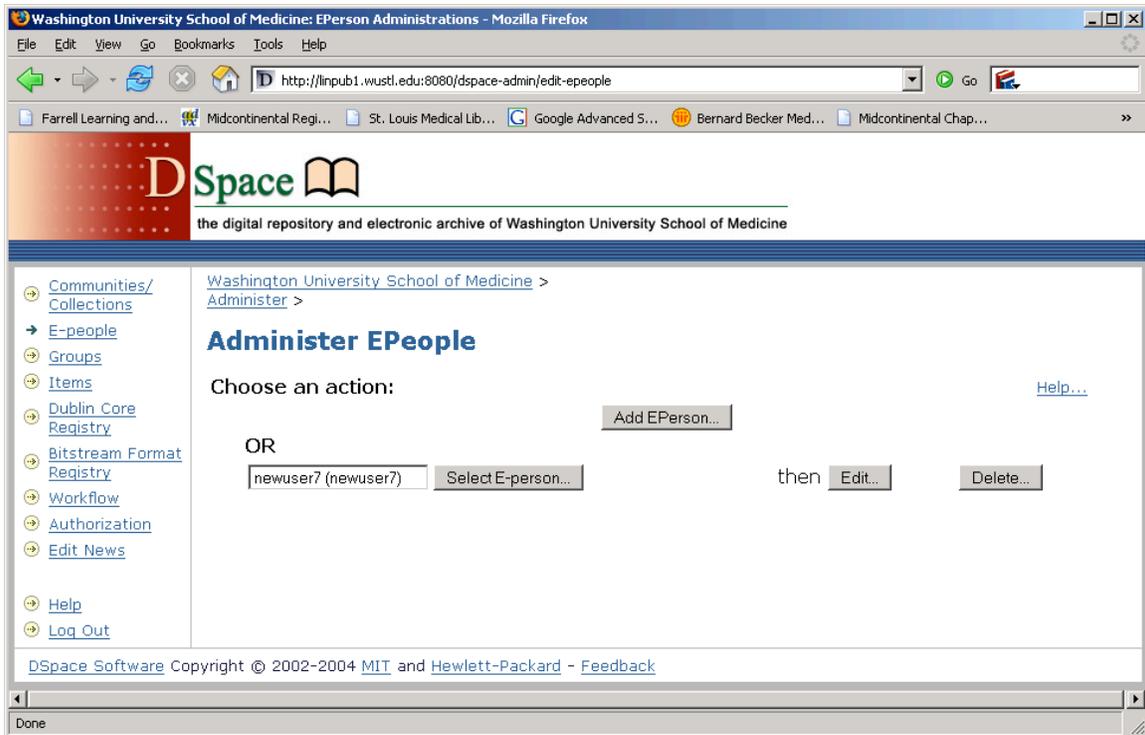


Image 25 – Selecting an E-person

To edit the e-person, click on the **Edit...** button.

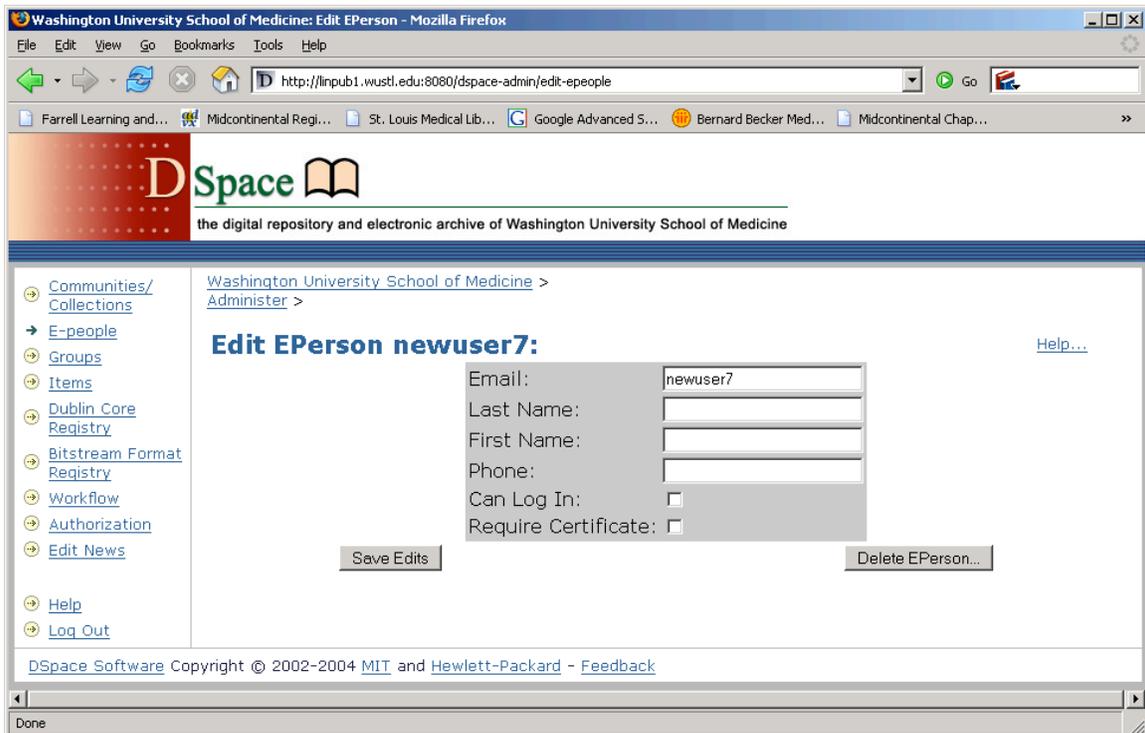


Image 26 – Edit an E-person

Here you can fill out or edit the fields in the e-person's record. Click **Save Edits** to save the record. Here you can also delete the record by clicking on the **Delete Eperson** button. A confirmation page will load to confirm the deletion of the e-person's record.

To delete an e-person, click on the **Delete** button and the following page will load.

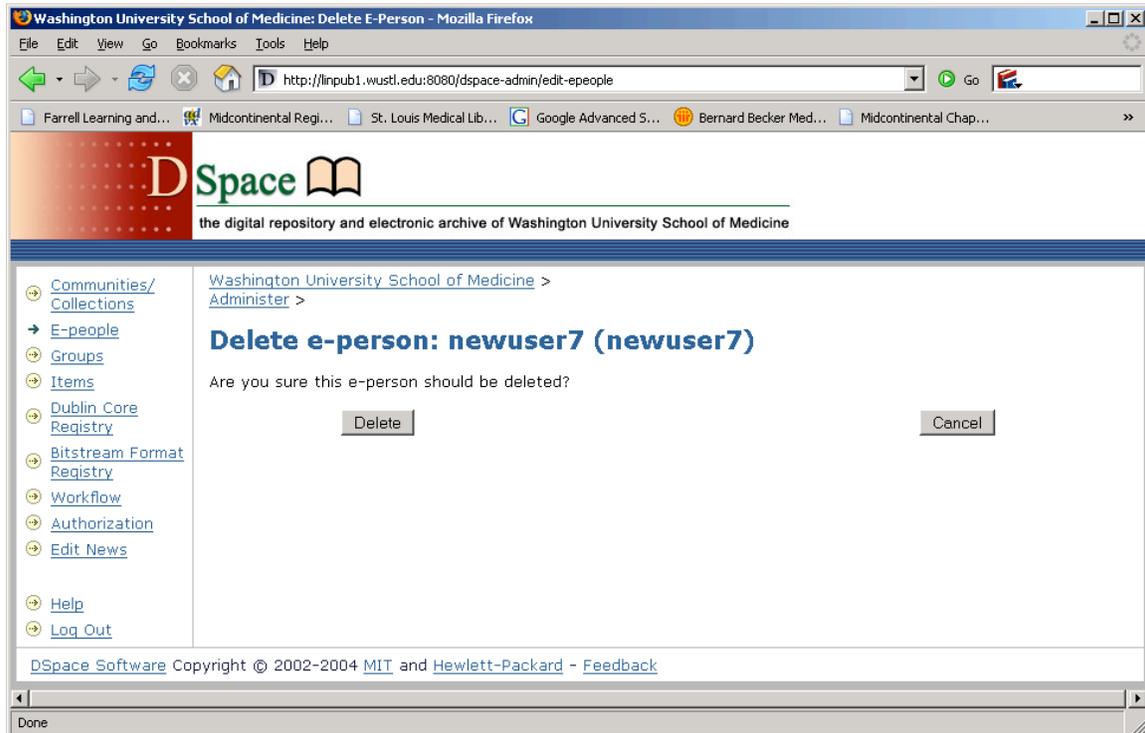


Image 27 – Deleting an E-person

Click on the **Delete** button do confirm the deletion of the e-person.

Groups

Groups are created to apply authorizations or assign duties to multiple epeople. Administrators assign groups. Users can't assign themselves to groups.

Administration

Groups are administrated from the Administration Main Page (Image 1). By clicking on the Groups link in the left side column, you get the following page.

Washington University School of Medicine >
Administer >

Group Editor

Note that you do not need to manually add users to the "anonymous" group - all users are members implicitly.

Warning - if you try to delete a group that is referred to by an authorization policy or is a workflow group you will get an internal server error.

Create New Group

ID	Name	Edit...	Delete
1	Administrator	Edit...	Delete
0	Anonymous	Edit...	Delete
49	Biology Admin	Edit...	Delete
48	Biology Department	Edit...	Delete
56	COLLECTION_10_ADMIN	Edit...	Delete
26	COLLECTION_10_DEFAULT_ITEM_READ	Edit...	Delete
25	COLLECTION_10_SUBMIT	Edit...	Delete
54	COLLECTION_10_WFSTEP_2	Edit...	Delete
28	COLLECTION_11_DEFAULT_ITEM_READ	Edit...	Delete
27	COLLECTION_11_SUBMIT	Edit...	Delete
30	COLLECTION_12_DEFAULT_ITEM_READ	Edit...	Delete
29	COLLECTION_12_SUBMIT	Edit...	Delete
36	COLLECTION_15_SUBMIT	Edit...	Delete
37	COLLECTION_15_WORKFLOW_STEP_1	Edit...	Delete
38	COLLECTION_15_WORKFLOW_STEP_2	Edit...	Delete
44	COLLECTION_16_ADMIN	Edit...	Delete

Image 28 – Group Editor

This page looks very similar to the Eperson's page. You can edit a Group or you can delete Group from this page. To delete a Group, click on the **Delete** button.

To create a new Group, click on the **Create New Group** button.

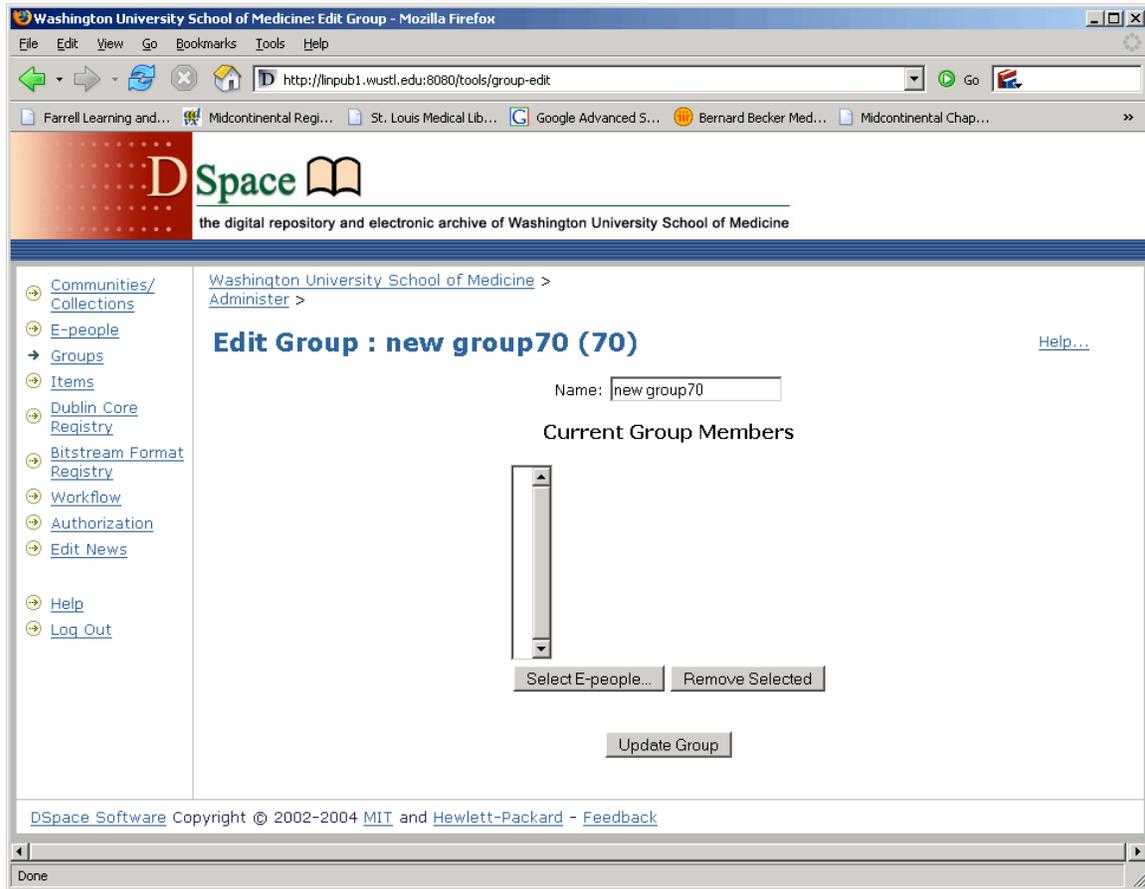


Image 29 – Create New Group

To give the group a name, type in a name for the group in the Name text box. To add people to the group, click on the Select E-people button. In the list in the new window, click on the Add button next to the name or names of the people you wish to add to this group.

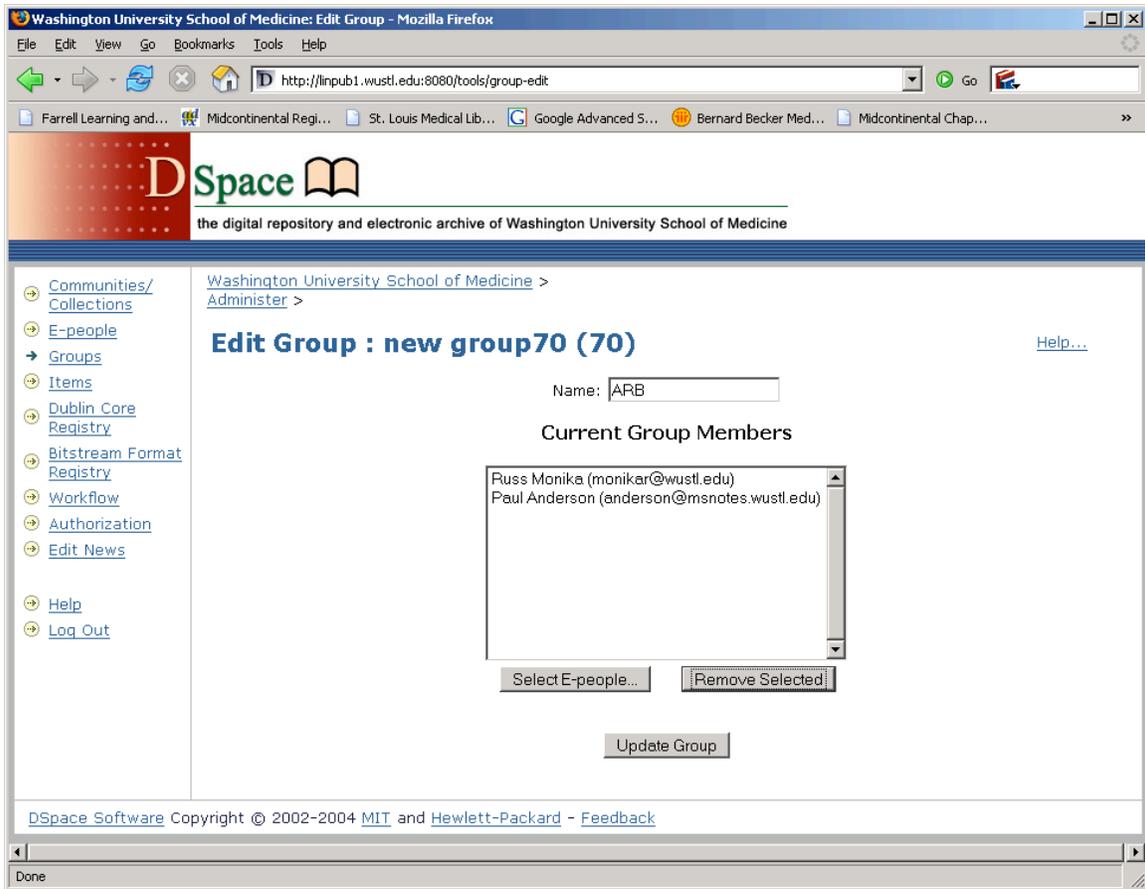


Image 30 – New Group

When finished adding people to the group, close the window and click on the Update Group button. To return to the Groups list, click on the Groups link in the left side Administration menu.

To edit a Group, click on the **Edit** button.

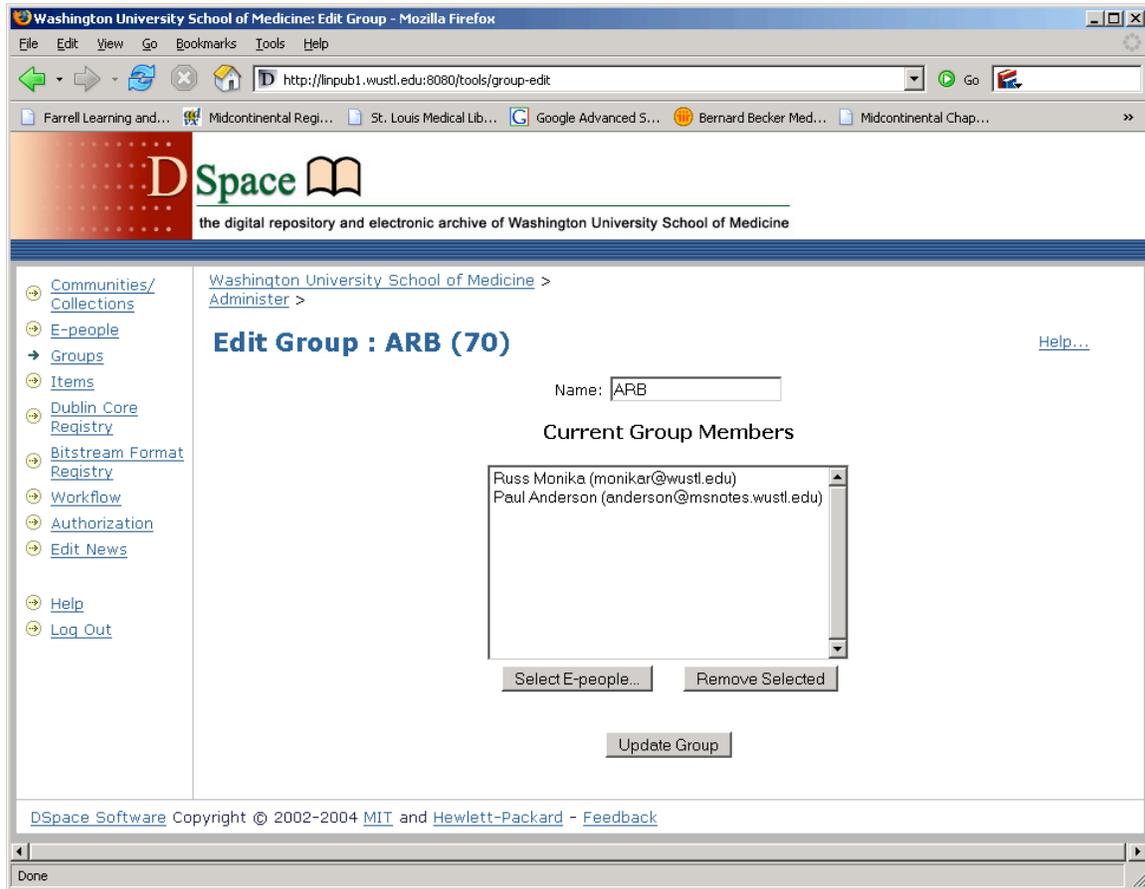


Image 31 – Edit Group

Here you can edit the name of the group and click on the **Update Name** button. You can also **Select EPeople** to add to the group and **Remove Selected** people from the group by clicking on the appropriate buttons.

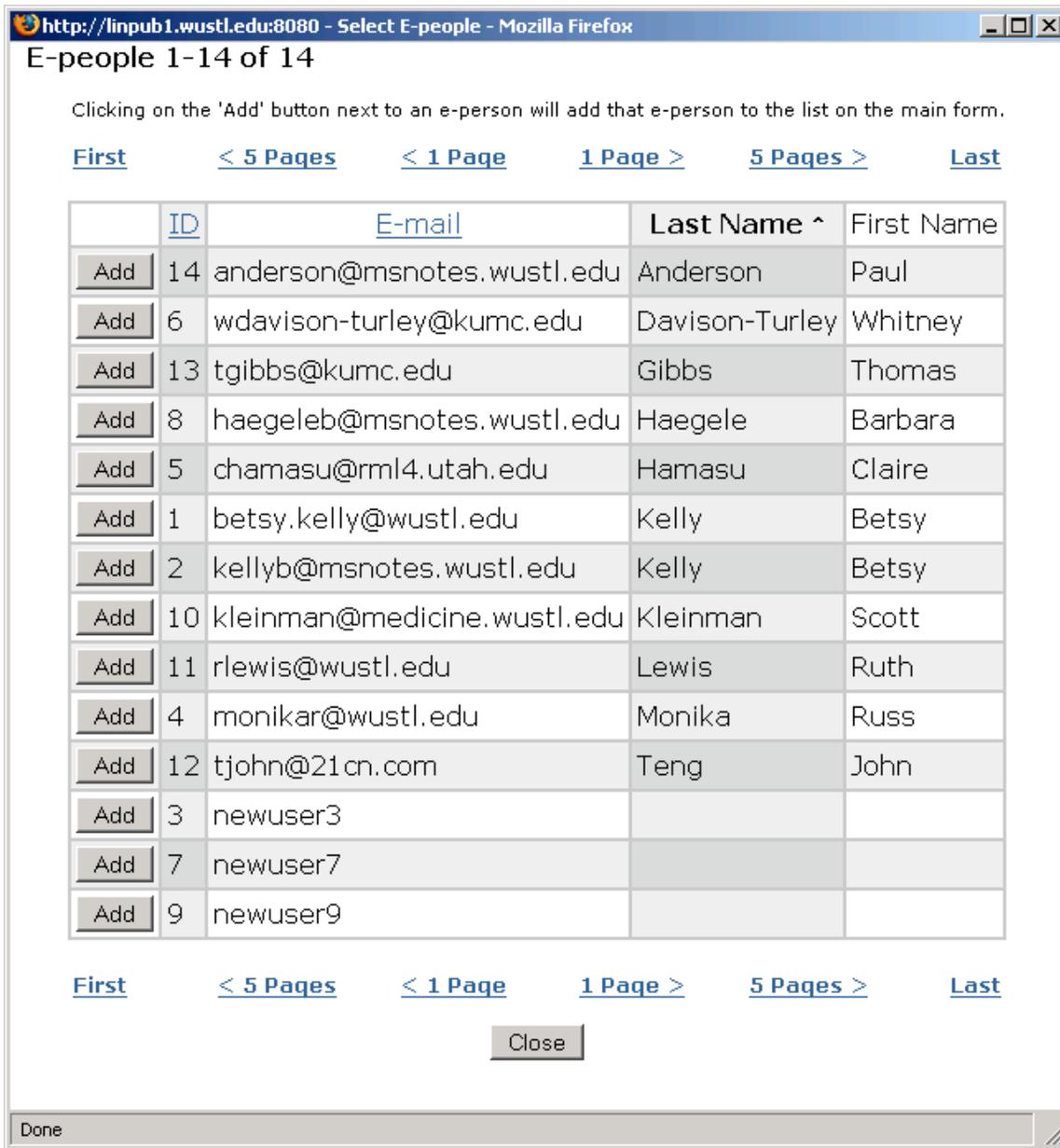


Image 32 – Adding an EPerson

Click on the Add button next to a person's name to add them to the group. When finished, click on the Close button to close the window and return to the Edit Group page.

Groups Notes

- *Reasons for creating Groups*
 - *Restrict users who can submit to a collection. Create a submitter group. When a new user is added to DSpace who can submit to a collection, add that person to the submitter group.*

- *Restrict users who can read a collection.*
- *Create a group for a specific workflow. Create a workflow for a collection and leave the users blank. Then assign the workflow group to the workflow.*

Authorization Policies

Each Eperson and Group must have the correct Authorizations in order to function in DSpace. The Anonymous Group must have Read authorization in order to view the content of DSpace. Other users and groups performing other tasks must have the correct authorizations to perform those tasks.

The four basic authorizations in DSpace are:

READ – The ability to read records.

WRITE – The ability to make edits to records.

ADD – The ability to add a record to the archive.

REMOVE – The ability to delete records.

Assigning Authorizations

Here is the Authorizations Page.

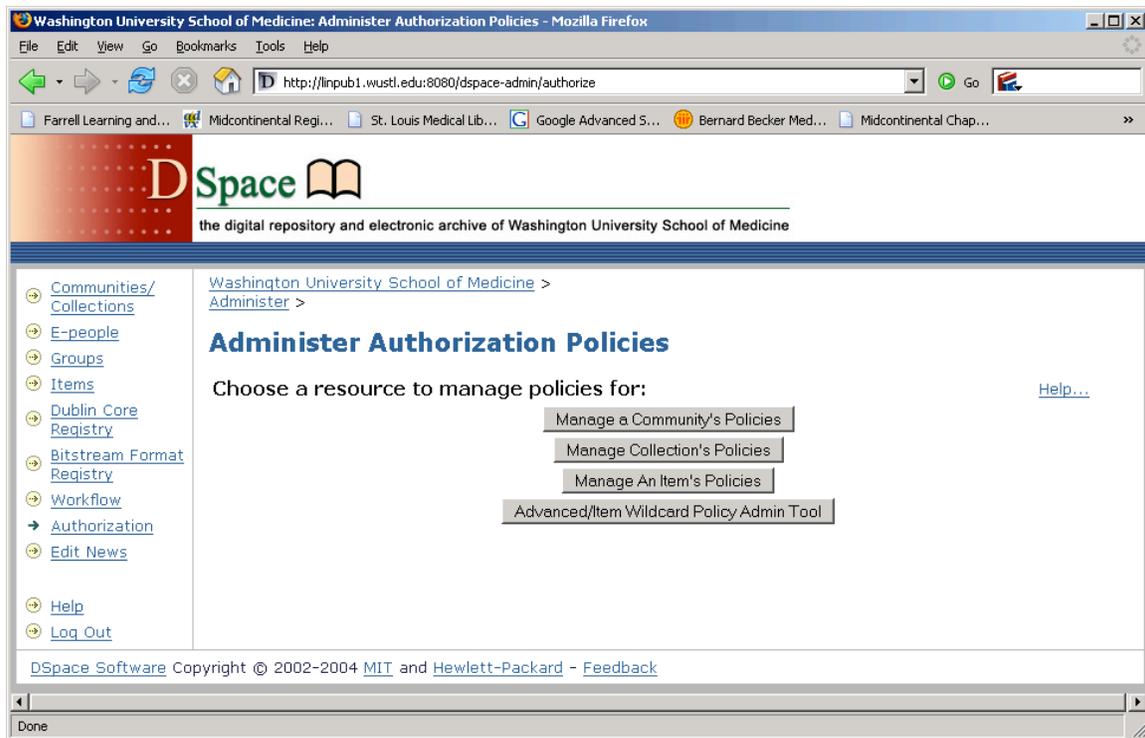


Image 33 – Administer Authorization Policies

On this page, you can administer the following authorization policies:

Manage a Community's Policies

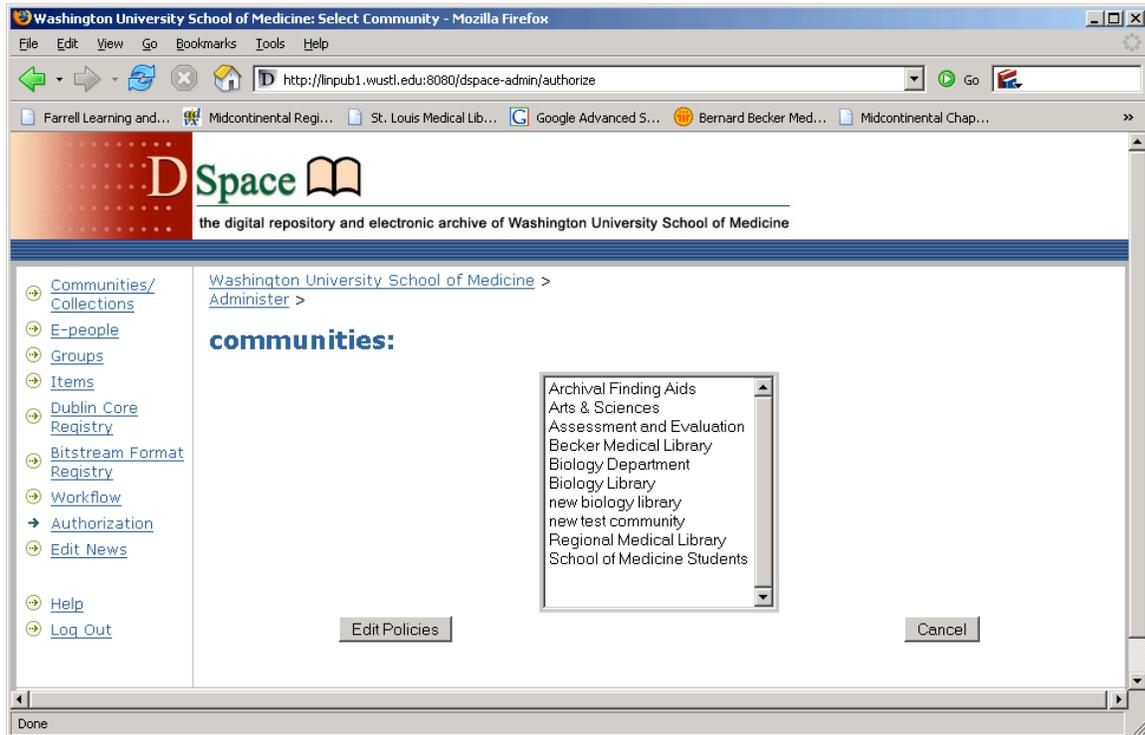


Image 34 – Community's Policies

On this page, you select a community.

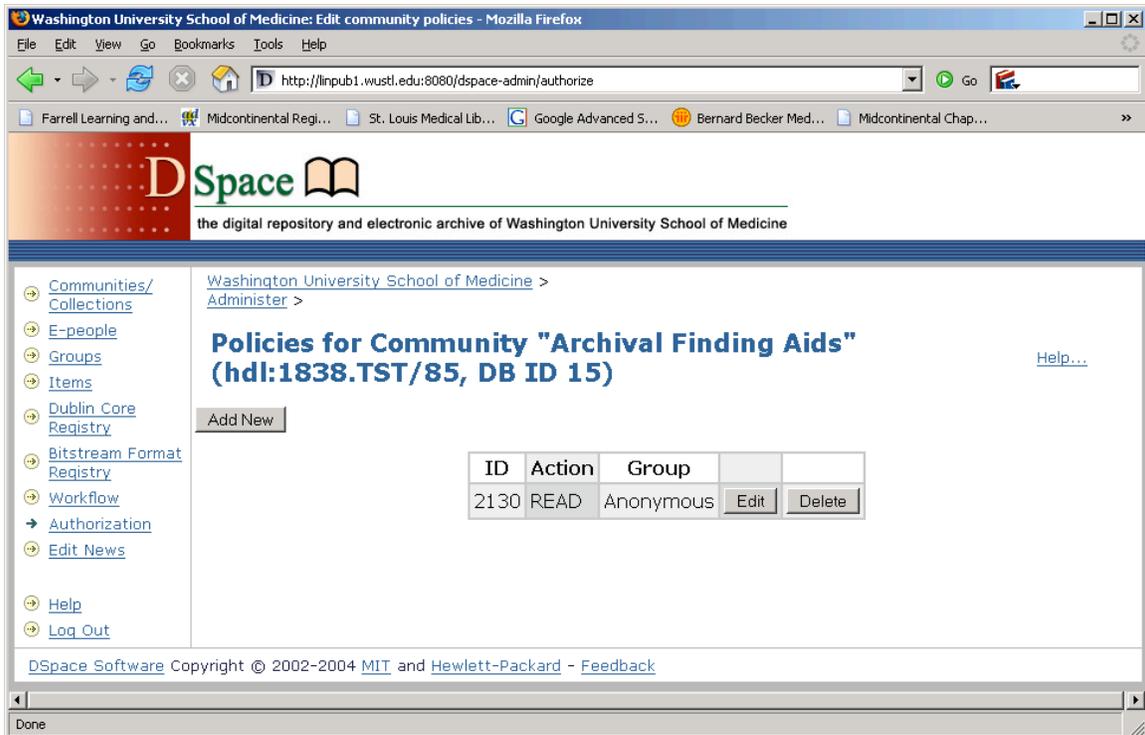


Image 35 –Policies for a Community

Here you can edit or delete existing policies or add new policies. Since allowing anonymous users to read your collection is a necessary policy, we won't touch that one.

To create a new policy, click on the **Add New** button.

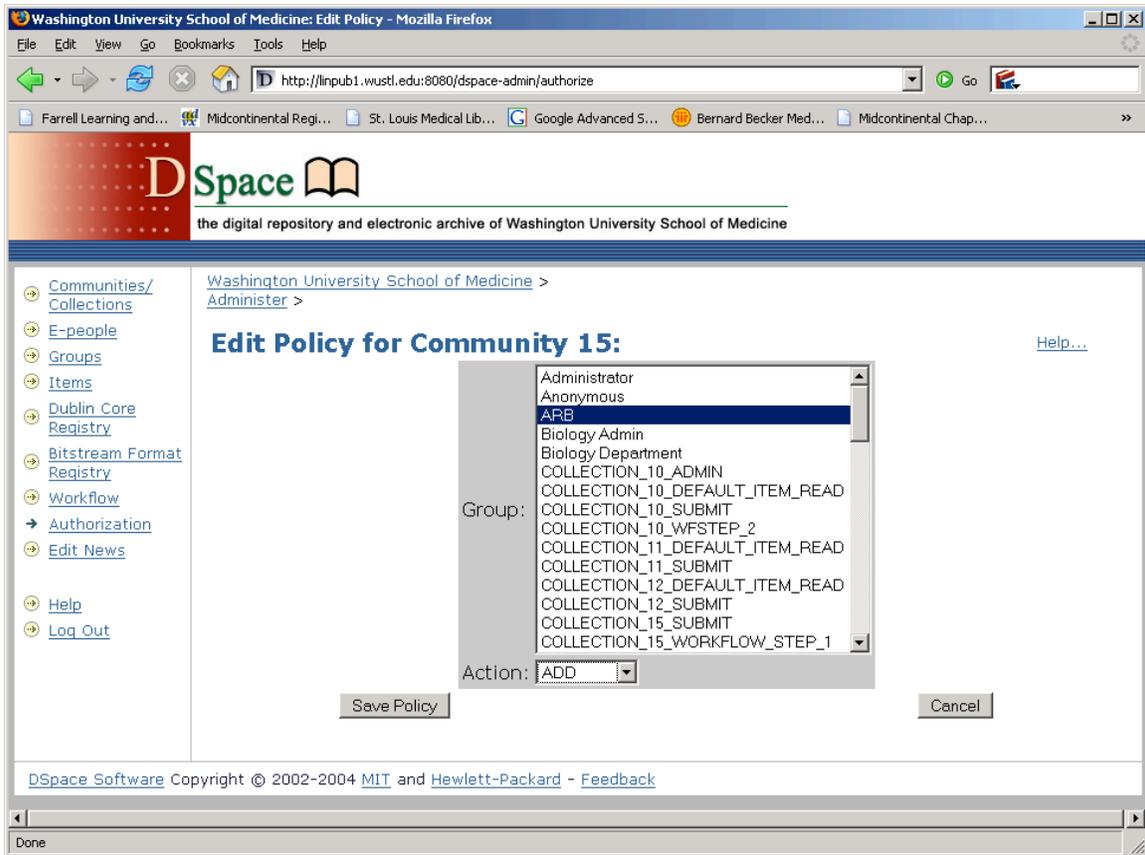


Image 36 – New Policy for a Community

Here you can assign a task to a group for the Community. **Select the Group** and **select the action**. Then **click on the Save Policy** and it will be added to the Community's Policies.

Manage a Collection's Policies

Editing a Collection's Policies is the same as for a Community. When you click on the **Manage a Collections Policies** button, you get a page similar to the Edit Policy for Community page (Image 22). You **select a group** and **select a task** and then **click on the Save Policy** button.

An Item

You can edit or assign an authorization to an individual item in DSpace. Click on the **Manage an Item's Policies** button to see the Select an Item Page.

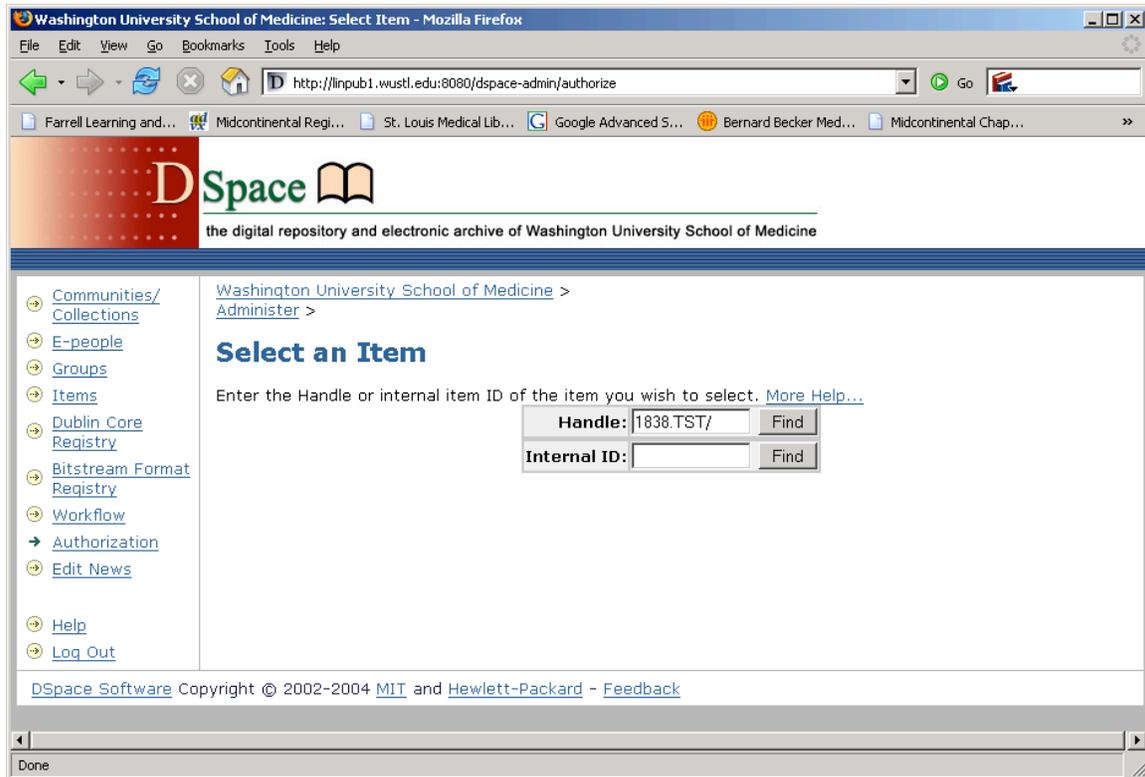


Image 37 – Select an Item

You need to search for the items handle to finish the handle number to find the item you want to assign an authorization. Once you have the handle number, type it into the top box in the search form.

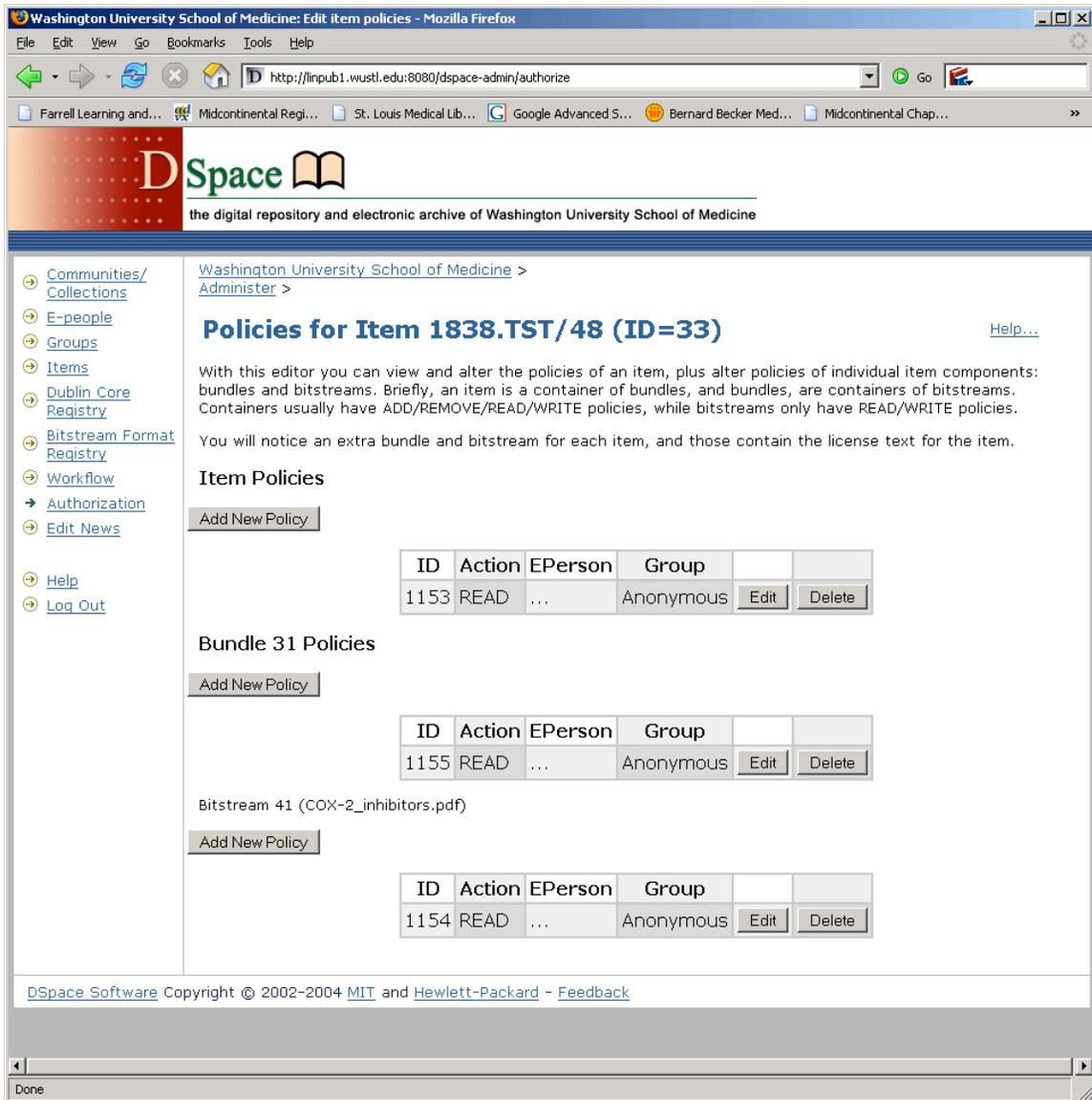


Image 38 – Edit an Item’s Policies

Here is the text from the Edit an Item’s Policies Page. This is a page that we at WUSTL don’t use often.

“With this editor you can view and alter the policies of an item, plus alter policies of individual item components: bundles and bitstreams. Briefly, an item is a container of bundles, and bundles, are containers of bitstreams. Containers usually have ADD/REMOVE/READ/WRITE policies, while bitstreams only have READ/WRITE policies.

You will notice an extra bundle and bitstream for each item, and those contain the license text for the item.”

Advanced/Wildcard Policy Admin Tool

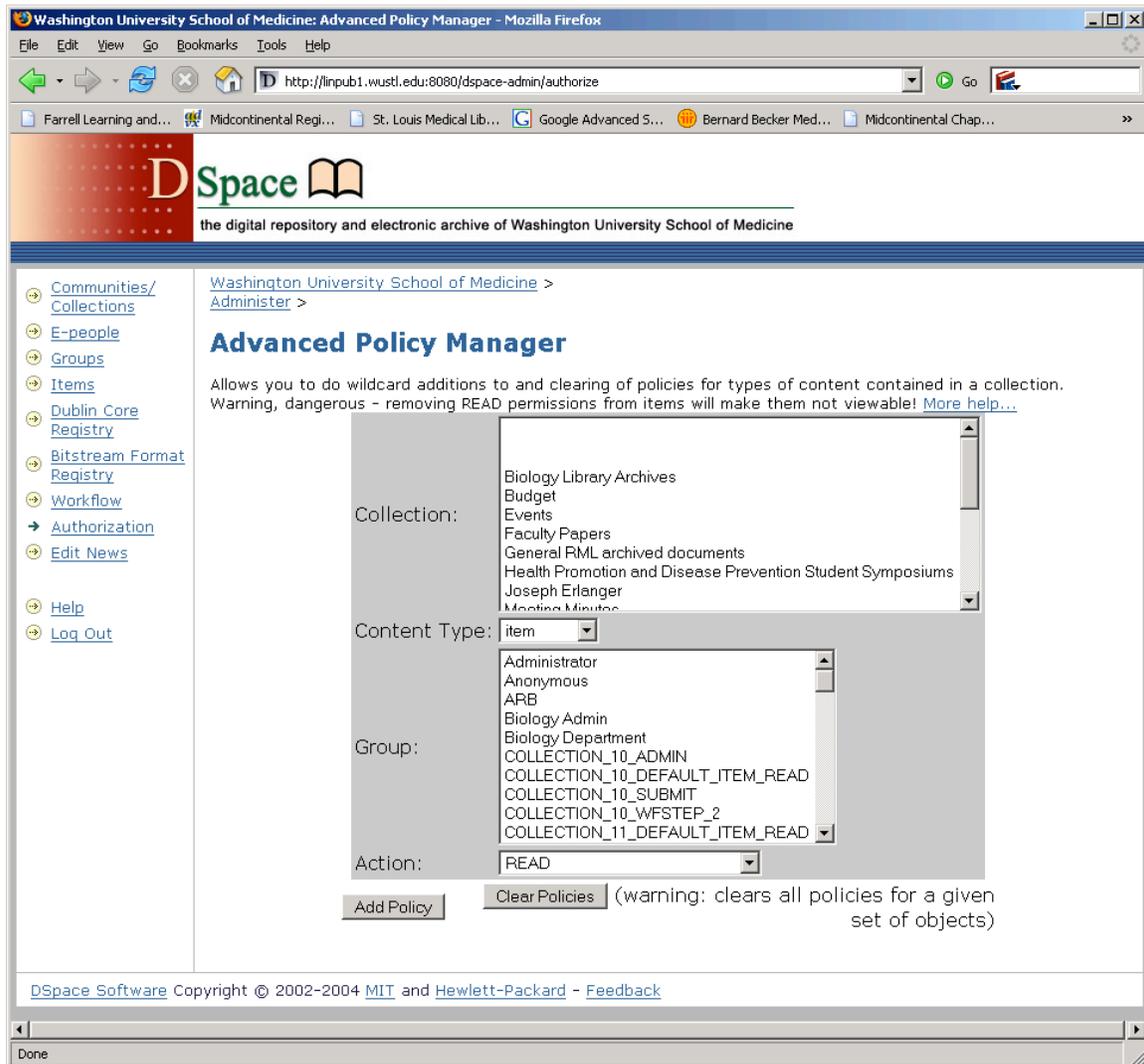


Image 39 – Advanced Policy Manager

Here is the text from the Advanced Policy Administration page:

“Allows you to do wildcard additions to and clearing of policies for types of content contained in a collection. Warning, dangerous - removing READ permissions from items will make them not viewable!”

This is not a page that we at WUSTL use very much.

Web Interface

All of the files for the Web Interface are **Java Server Pages (JSP)** and are written in that format. It sounds daunting, but only a strong knowledge of HTML is really needed to edit the basic pages. **We found no need to edit the JSP programming in DSpace.** All we did was edit the HTML coded embedded in the JSP pages.

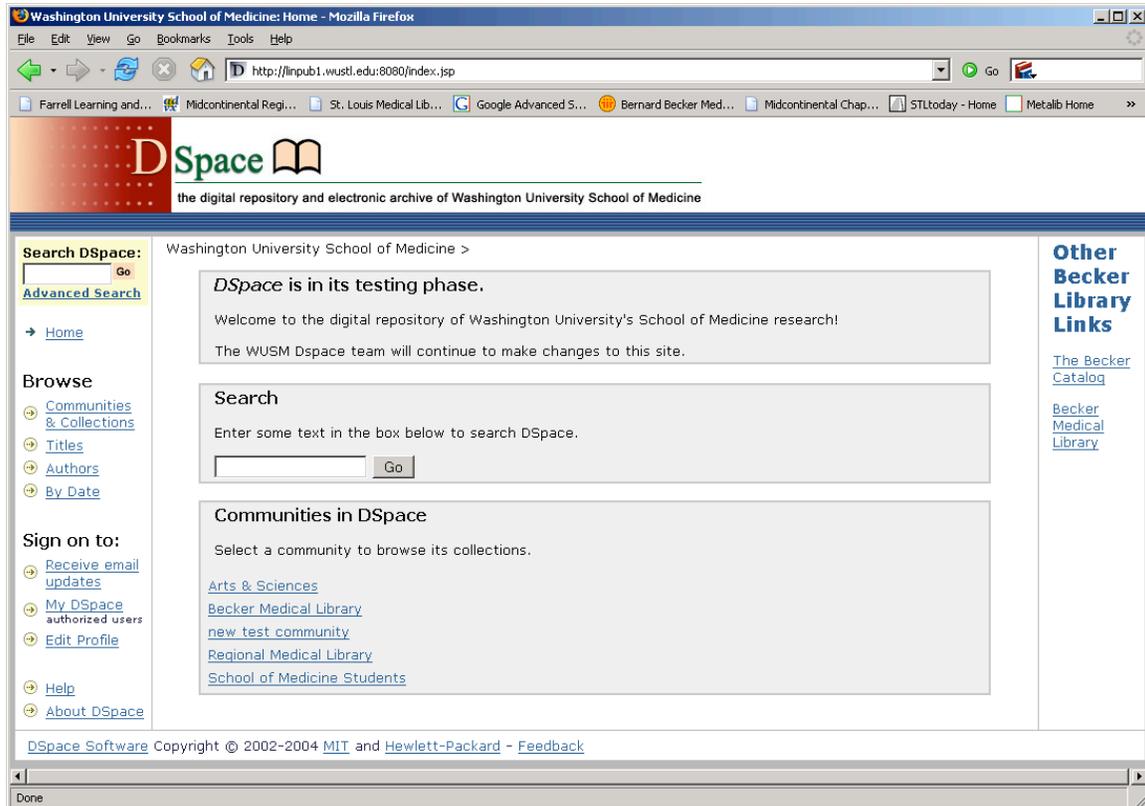


Image 40 – DSpace Main Page

Brief Overview

Here is a brief over of the step we at WUSM took to get DSpace running. Detailed instructions of how we accomplished these steps will follow.

1. Identify the files that make up the Main Page.
Several individual files make up the main page. They are combined by the home.jsp file.
2. Copy files that are to be edited and used into the **/home/dspace_source/dspace-1.2-source/jsp/local** directory.
All of the files that you edit and use in your local installation of DSpace should be saved into the /local directory of the /jsp directory. Copy the **/images** directory there as well. Then you can add and delete images whenever you want. The **/local** directory should have the identical directory structure as the **/jsp** directory. DSpace will look for files in the **/jsp/local** directory first, then look for the file in the **/jsp** directory. The files that are saved in the **/local** directory are not touched when a new version is installed.
3. Edit those files to make DSpace our own.
You can edit the files. You can add your own images. You will need to change the path in all of the tags to the /local directory for the images in your files. You don't have to refer to MIT or HP. You will want to add links to your own sites and email addresses. Again, you shouldn't have to change any of the JSP programming.
4. Edit the styles.css.jsp stylesheet to use the files in the /local directory.
You will have to change a statement at the beginning of the stylesheet so that it will use the local stylesheet. Instructions are later in this document. The stylesheet is very involved and most styles are best left alone, but this is the file that governs all styles in all files in the site.

The follow sections will give you specific instructions about accomplishing these steps.

Identify Files that Make Up Main Page

The first step is to determine which files make up the main page of the Web Interface. The main page is made up of several component files that are then put together to make up the main page.

Those files are:

Home.jsp – This file is the one that puts together most of the components for the main page. It calls for the components/news.jsp file for the top box in the middle column. It creates the Search section in the middle box of the middle column. And it brings up the list of Communities in the bottom box of the middle column. It also is where you can add text and link in the right column of the main page if you would like.

Index.jsp – This is the main page that browser will load first when the go to <http://linpub1.wustl.edu:8080/>. I launches sever Java Applets that allow DSpace to operate properly. It also launches home.jsp which loads the rest of the main page components.

Styles.css.jsp – This is the style sheet for the site. Even though it has a .jsp file extension, it is really a Cascading Style Sheet. Most of the styles will be left at their default settings, but you will want to change some of them. We will discuss which ones later.

Components/news.jsp – This is the file that fills out the top box in the middle column. This is where you would put News items and other announcements about your DSpace installation. Other news information can be put here as well.

Layout/footer-default.jsp – This file is what makes up the bottom part of the main page below the three columns. By default, there are some MIT and HP icons there. This page is set up like the header-default.jsp page in that it has a three column table in it but with only one row.

Layout/header-default.jsp – This file makes up the top part of the page. It includes the banner and the blue stripe. It is a table with two rows and three columns. The default file has the DSpace image in the leftmost cell of the first row and an MIT image in the rightmost row. In the center cell of the top row is text and a link to the About DSpace page at MIT. As you can see, we merged the first two cells and put in our own banner and left aligned it. We put in the Medical School logo in the rightmost cell and right aligned it. This way, as the browser windows widens, it will separate from the banner and align with the third column of the body of the rest of the page. The second row contains the blue stripe image.

Layout/location-bar.jsp – This file brings up the name of the DSpace installation which shows up above the top box in the middle column. This is the name of the institution sponsoring the DSpace installation.

Layout/navbar-default.jsp – This file is what makes of the contents of the left column navigation from below the blue bar from the header to right above the footer. It starts with the Search text box and ends with the About DSpace link. Most of the time, you won't be editing this file very much.

Copy Files to Edit to /Local Directory

All files that you are editing and using for your local installation of DSpace **must** be copied into the **dspace/jsp/local** directory in the DSpace directory structure. The files should follow the same directory structure as they do in the **/jsp** directory. That is where you can edit and save the files. They should not be edited in their default locations.

Style Sheet – styles.css.jsp

This is the file that set the display styles for all of the pages in DSpace. It is basically a CSS page with some minimal JSP code to fit in with the other JSP files. There are several edits to make to the file and procedures to do after the file is uploaded to make it work.

- The file has a typo. There is a JPS comment between the declaration that change the tag and the declarations that create the classes. The first line of the comment ends with --%., when it should end with --%>. The end bracket is crucial as it throws off the rest of the style sheet.
- You must delete all references to looking for the local style sheet. This file is the local style sheet. In the first JSP statement, leave in the text/css line, the imageURL line, and the Netscape 4.x stuff. Delete the if/then stuff about the local style sheet stuff. It causes a loop that shuts down DSpace. See Appendix A, Editing the Styles.css.jsp Page.
- After you upload the styles.css.jsp to the dspace/jsp/local directory, you must restart the server for it to recognize the local style sheet. Otherwise, it will stop using the old style sheet but the styles in the new sheet won't take effect, so no styles will show at all.
- Once the new style sheet is in use, in order to avoid restarting the server many times, I recommend editing the styles.css.jsp on the server using Pico. There is no real benefit editing a CSS page is a program like Dreamweaver since you have to edit it in Code View anyway. You'll probably want to FTP the file after editing it for backup purposes.

Item Importer and Exporter in DSpace

DSpace provides command line tools to import and export in batches into collections in the system. They launch Java applets to do the importing and exporting. I have spent some time working with these tools, applets, and the documentation and have successfully imported and exported items. These are my notes about both processes. I have enlisted the help of both our staff and the DSpace-Tech listserv in making these applets work on our DSpace.

DSpace Simple Archive Format

The Simple Archive Format is what DSpace used to create an archive for the items. The archive is simply a directory full of items, with a subdirectory for each item. Each subdirectory contains a file for the item's descriptive metadata, and the files that make up the item.

Here is what the [archive](#) should look like:

```
archive_directory/
  item_000/
    dublin_core.xml  -- qualified Dublin Core metadata
    contents         -- text file containing one line per filename
    file1.doc        -- files to be added as bitstreams to the item
    file2.pdf
    ...
  item_001/
    dublin_core.xml  -- qualified Dublin Core metadata
    contents         -- text file containing one line per filename
    file1.doc        -- files to be added as bitstreams to the item
    file2.pdf
    ...
etc.
```

Notes:

- *The contents file contains the name of the files that are part of the item, one filename per line.*
- *You must include the path to Java in the PATH environment variable.*
- *I put the archive_directory in the dspace/bin directory. Several messages in the listserv mentioned that their importing efforts didn't work until they did this.*

Dublin_core.xml

The dublin_core.xml file has a strict format. The root element is <dublin_core> and the only element allowed in the document is <dcvalue>. Only three attributes are allowed in the <dc_value> element. They are:

<element> -- the Dublin Core element
<qualifier> -- the element's qualifier
<language> -- (optional) language code for the element

Here is an example of a [dublin_core.xml](#) document used in importing an item.

```
<dcvalue element="contributor" qualifier="author">Walter, Ed</dcvalue>
<dcvalue element="date" qualifier="issued">2001-12</dcvalue>
<dcvalue element="identifier" qualifier="citation">News@Becker,1(1),Nov-Dec
2001</dcvalue>
<dcvalue element="description">Quarterly newsletter of the Becker Medical Library. Also
available on the library website at
http://becker.wustl.edu/newsletter/v1i1/index.htm</dcvalue>
```

If there is no qualifier, then the qualifier="none" is used.

Notes:

- *I was able to import items with dcvalues that didn't have qualifiers. You are supposed to use a qualifier="none" but I had problems with it. I took out the qualifier attribute and those items imported fine.*
- *The file that determines what elements can be used is the **/dspace/config/registries/dublin-core-types.xml** file. It contains the metadata for the acceptable metadata for the items in DSpace. I used this file, along with the records of existing, similar items in DSpace to determine what metadata to put in the Dublin_core.xml file.*
- *Also in the same directory is the **bitstream-formats.xml** file. This file holds all of the bitstream formats that are acceptable in DSpace. When you add a bitstream type, this file is edited by the system.*

Importing Items

The item importer is in the `[org.dspace.app.itemimport.ItemImport]`, and is run with the `dsrun` utility in the `dspace/bin` directory.

To import a group of items with an Eperson as submitter, the command syntax is:

```
dsrun org.dspace.app.itemimporter.ItemImporter --add --eperson=joe@wustl.edu --collection=collectionID --source=source_dir --mapfile=mapfile
```

- **eperson** – a valid eperson registered in DSpace
- **collectionID** – the database number found when you click on the Collections link
- **source_directory** – the archive_directory for the items
- **mapfile** – a name you give to a file that DSpace create when it imports the items. This file can use this file to unimport files.

When all elements are structured correctly, the [import utility works well and quickly](#). If there is an error in an element in the metadata file, the output of the utility tells you which element made the program stop.

Here is an example. One time, the program stopped because of the citation element.

```
Element: publisher Qualifier: none Value: Becker Medical Library
Element: language Qualifier: iso Value: en_US
Processing contents file: archive_directory/item_000/contents
Bitstream: news@beckerv1i1.pdf
Processing handle file: handle
It appears there is no handle file -- generating one
java.sql.SQLException: bad_dublin_core indentifier citation
at org.dspace.content.Item.update(Item.java:1085)
at org.dspace.content.InstallItem.installItem(InstallItem.java:181)
at org.dspace.content.InstallItem.installItem(InstallItem.java:90)
at org.dspace.app.itemimport.ItemImport.addItem(ItemImport.java:475)
at org.dspace.app.itemimport.ItemImport.addItems(ItemImport.java:333)
at org.dspace.app.itemimport.ItemImport.main(ItemImport.java:281)
java.sql.SQLException: bad_dublin_core indentifier citation
[dspace@linpub1 bin]$
```

Here the program tells us that the Dublin Core is bad. This gives you a guide to what needs fixing in the `dublin_core.xml` file before you run the importer utility again.

Notes:

- *All the files must follow all of the above guidelines. Any error and the process will stop and show an error message.*

Exporting Items

The item exporter is in the `|org.dspace.app.itemexport|ItemExport|`, and is run with the **dsrun** utility in the **dSPACE/bin** directory. You can export several items or a single item using the exporter utility.

To export all items in a collection, use the following syntax:

```
dsrun org.dspace.app.itemexport.ItemExport --type=COLLECTION --id=collID --dest=dest_dir --number=seq_num
```

Notes:

- *The collection ID can either be the database ID or the handle.*
- *The dest_dir is a directory that you create to hold the files that you export. The system will create subdirectories for each item in the collection that will hold all of the files for that item. I created a directory in the dspace/bin directory just to be consistent with the imported files. I gave the directory the name of the collection.*
- *The seq_num is a number you assign that will begin the sequence of items that are exported. There can be no alpha characters or dashes in the number.*

To export a single item in a collection, use the following syntax:

```
dsrun org.dspace.app.itemexport.ItemExport --type=ITEM --id=itemID --dest=dest_dir --number=seq_num
```

Notes:

- *The item ID is the item's handle. A handle file is exported and can be used when importing into another DSpace machine so that it can keep its same handle.*
- *The dest_dir is a directory that you create to hold the files that you export. The system will create subdirectories for each item in the collection that will hold all of the files for that item. I created a directory in the dspace/bin directory just to be consistent with the imported files. I gave the directory the name of the collection.*
- *The seq_num is a number you assign that will begin the sequence of items that are exported. There can be no alpha characters or dashes in the number.*

Here is what the directory structure of the archive_directory looks like for imported items.

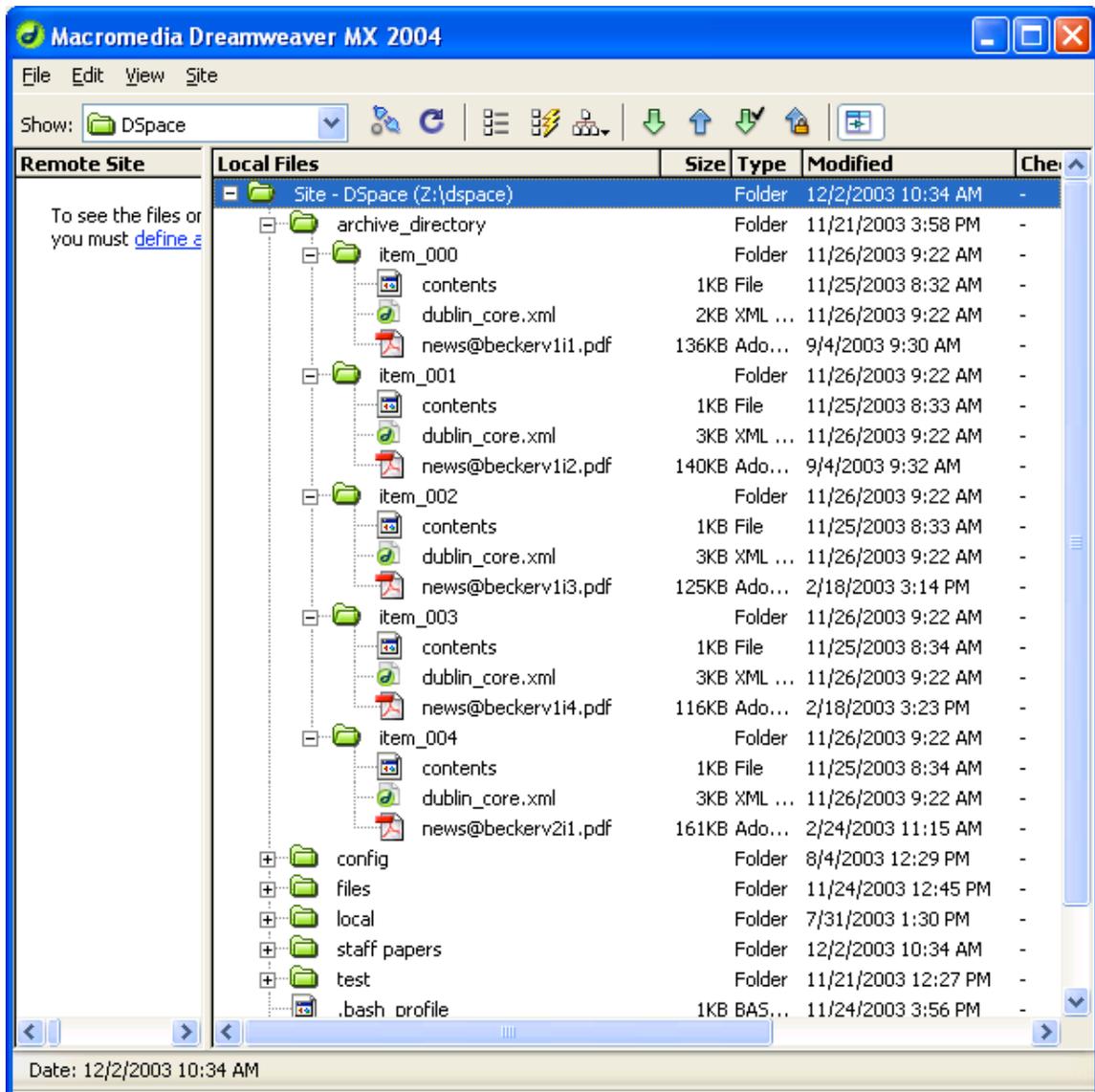


Image 41 – Import files

Here is an example of a complete **dublin_core.xml** document used in importing an item.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE dublin_core [
  <!ELEMENT dublin_core (dcvalue+)>
  <!ELEMENT dcvalue (#PCDATA)>
  <!ATTLIST dcvalue element CDATA "none">
  <!ATTLIST dcvalue qualifier CDATA "none">
]>

<dublin_core>
  <dcvalue element="contributor" qualifier="author">Schoening, Paul</dcvalue>
  <dcvalue element="contributor" qualifier="author">Murray, Carol</dcvalue>
  <dcvalue element="contributor" qualifier="author">Prendergast, Neville</dcvalue>
  <dcvalue element="contributor" qualifier="author">Kelly, Betsy</dcvalue>
  <dcvalue element="contributor" qualifier="author">Halbrook, Barbara</dcvalue>
  <dcvalue element="contributor" qualifier="author">Ferguson, Erin</dcvalue>
  <dcvalue element="contributor" qualifier="author">Walter, Ed</dcvalue>
  <dcvalue element="date" qualifier="issued">2001-12</dcvalue>
  <dcvalue element="identifier" qualifier="citation">News@Becker,1(1),Nov-Dec 2001</dcvalue>
  <dcvalue element="description">Quarterly newsletter of the Becker Medical Library. Also available on the
library website at http://becker.wustl.edu/newsletter/v1i1/index.htm</dcvalue>
  <dcvalue element="description" qualifier="abstract">Newsletter of the Becker Medical Library, Washington
University School of Medicine</dcvalue>
  <dcvalue element="format" qualifier="mimetype">application/pdf</dcvalue>
  <dcvalue element="title">News@Becker</dcvalue>
  <dcvalue element="type">Newsletter</dcvalue>
  <dcvalue element="subject">Web of Science</dcvalue>
  <dcvalue element="subject">proxy</dcvalue>
  <dcvalue element="subject">electronic resources</dcvalue>
  <dcvalue element="subject">Archives and Rare Books</dcvalue>
  <dcvalue element="subject">Digital Gallery</dcvalue>
  <dcvalue element="subject">MSCNS</dcvalue>
  <dcvalue element="publisher">Becker Medical Library</dcvalue>
  <dcvalue element="language" qualifier="iso">en_US</dcvalue>
</dublin_core>
```

Here is a log of a successful import of five items into the News@Becker collection.

```
[dSPACE@linpub1 bin]$ dsrun org.dSPACE.app.itemimport.ItemImport --add --
eperson=monikar@msnotes.wustl.edu --collection=12 --source=archive_directory --mapfile=map7
Adding items from directory: archive_directory
Generating mapfile: map7
Adding item from directory item_000
  Loading dublin core from archive_directory/item_000/dublin_core.xml
  Element: contributor Qualifier: author Value: Schoening, Paul
  Element: contributor Qualifier: author Value: Murray, Carol
  Element: contributor Qualifier: author Value: Prendergast, Neville
  Element: contributor Qualifier: author Value: Kelly, Betsy
  Element: contributor Qualifier: author Value: Halbrook, Barbara
  Element: contributor Qualifier: author Value: Ferguson, Erin
  Element: contributor Qualifier: author Value: Walter, Ed
  Element: date Qualifier: issued Value: 2001-12
  Element: identifier Qualifier: citation Value: News@Becker,1(1),Nov-Dec 2001
  Element: description Qualifier: none Value: Quarterly newsletter of the Becker Medical Library. Also
available on the library website at http://becker.wustl.edu/newsletter/v1i1/index.htm
  Element: description Qualifier: abstract Value: Newsletter of the Becker Medical Library, Washington
University School of Medicine
  Element: format Qualifier: mimetype Value: application/pdf
  Element: title Qualifier: none Value: News@Becker
  Element: type Qualifier: none Value: Newsletter
  Element: subject Qualifier: none Value: Web of Science
  Element: subject Qualifier: none Value: proxy
  Element: subject Qualifier: none Value: electronic resources
  Element: subject Qualifier: none Value: Archives and Rare Books
  Element: subject Qualifier: none Value: Digital Gallery
  Element: subject Qualifier: none Value: MSCNS
  Element: publisher Qualifier: none Value: Becker Medical Library
  Element: language Qualifier: iso Value: en_US
  Processing contents file: archive_directory/item_000/contents
  Bitstream: news@beckerv1i1.pdf
Processing handle file: handle
It appears there is no handle file -- generating one
0 item_000
Adding item from directory item_001
  Loading dublin core from archive_directory/item_001/dublin_core.xml
  Element: contributor Qualifier: author Value: Schoening, Paul
  Element: contributor Qualifier: author Value: Murray, Carol
  Element: contributor Qualifier: author Value: Prendergast, Neville
  Element: contributor Qualifier: author Value: Sullivan, Christina
  Element: contributor Qualifier: author Value: Lipsey, Kim
  Element: contributor Qualifier: author Value: Vekerdy, Lilla
  Element: contributor Qualifier: author Value: Kelly, Betsy
  Element: contributor Qualifier: author Value: Halbrook, Barbara
  Element: contributor Qualifier: author Value: Ferguson, Erin
  Element: contributor Qualifier: author Value: Walter, Ed
  Element: date Qualifier: issued Value: 2002-04
  Element: identifier Qualifier: citation Value: News@Becker,1(2),Mar-Apr 2002
  Element: description Qualifier: none Value: Quarterly newsletter of the Becker Medical Library. Also
available on the library website at http://becker.wustl.edu/newsletter/v1i2/index.htm
  Element: description Qualifier: abstract Value: Newsletter of the Becker Medical Library, Washington
University School of Medicine
  Element: format Qualifier: mimetype Value: application/pdf
  Element: title Qualifier: none Value: News@Becker
  Element: type Qualifier: none Value: Newsletter
  Element: subject Qualifier: none Value: Web of Science
```

Element: subject Qualifier: none Value: Paul Schoening
 Element: subject Qualifier: none Value: Library Director
 Element: subject Qualifier: none Value: Archives and Rare Books
 Element: subject Qualifier: none Value: Digital Gallery
 Element: subject Qualifier: none Value: evidence-based medicine
 Element: subject Qualifier: none Value: Journal Citation Reports
 Element: subject Qualifier: none Value: journal impact factor
 Element: subject Qualifier: none Value: anthrax
 Element: subject Qualifier: none Value: smallpox
 Element: subject Qualifier: none Value: rare infections
 Element: subject Qualifier: none Value: bioterrorism
 Element: subject Qualifier: none Value: Tyler Collection
 Element: subject Qualifier: none Value: The Practice of Medicine
 Element: subject Qualifier: none Value: bioterrorism
 Element: subject Qualifier: none Value: Centers for Disease Control
 Element: subject Qualifier: none Value: CDC
 Element: publisher Qualifier: none Value: Becker Medical Library
 Element: language Qualifier: iso Value: en_US
 Processing contents file: archive_directory/item_001/contents
 Bitstream: news@beckerv1i2.pdf
 Processing handle file: handle
 It appears there is no handle file -- generating one
 1 item_001
 Adding item from directory item_002
 Loading dublin core from archive_directory/item_002/dublin_core.xml
 Element: contributor Qualifier: author Value: Schoening, Paul
 Element: contributor Qualifier: author Value: Grant, Georgiana L.
 Element: contributor Qualifier: author Value: Kelly, Betsy
 Element: contributor Qualifier: author Value: Halbrook, Barbara
 Element: contributor Qualifier: author Value: Ferguson, Erin
 Element: contributor Qualifier: author Value: Walter, Ed
 Element: date Qualifier: issued Value: 2002-07
 Element: identifier Qualifier: citation Value: News@Becker,1(3),Jun-Jul 2002
 Element: description Qualifier: none Value: Quarterly newsletter of the Becker Medical Library. Also available on the library website at <http://becker.wustl.edu/newsletter/v1i3/index.htm>
 Element: description Qualifier: abstract Value: Newsletter of the Becker Medical Library, Washington University School of Medicine
 Element: format Qualifier: mimetype Value: application/pdf
 Element: title Qualifier: none Value: News@Becker
 Element: type Qualifier: none Value: Newsletter
 Element: subject Qualifier: none Value: Exhibits
 Element: subject Qualifier: none Value: Deafness in Disguise
 Element: subject Qualifier: none Value: Instructional Technologies and Library Systems
 Element: subject Qualifier: none Value: Archives and Rare Books
 Element: subject Qualifier: none Value: presentations
 Element: subject Qualifier: none Value: Interlibrary Loan
 Element: subject Qualifier: none Value: evidence-based medicine
 Element: subject Qualifier: none Value: Cochrane Library
 Element: subject Qualifier: none Value: Family Resource Center
 Element: subject Qualifier: none Value: St. Louis Children's Hospital
 Element: subject Qualifier: none Value: concealed hearing devices
 Element: subject Qualifier: none Value: electronic journal usage
 Element: publisher Qualifier: none Value: Becker Medical Library
 Element: language Qualifier: iso Value: en_US
 Processing contents file: archive_directory/item_002/contents
 Bitstream: news@beckerv1i3.pdf
 Processing handle file: handle
 It appears there is no handle file -- generating one
 2 item_002
 Adding item from directory item_003
 Loading dublin core from archive_directory/item_003/dublin_core.xml
 Element: contributor Qualifier: author Value: Schoening, Paul

Element: contributor Qualifier: author Value: Lipsey, Kim
Element: contributor Qualifier: author Value: Anderson, Paul
Element: contributor Qualifier: author Value: Kelly, Betsy
Element: contributor Qualifier: author Value: Halbrook, Barbara
Element: contributor Qualifier: author Value: Ferguson, Erin
Element: contributor Qualifier: author Value: Walter, Ed
Element: date Qualifier: issued Value: 2002-10
Element: identifier Qualifier: citation Value: News@Becker,1(4),Sep-Oct 2002
Element: description Qualifier: none Value: Quarterly newsletter of the Becker Medical Library. Also available on the library website at <http://becker.wustl.edu/newsletter/v1i4/index.htm>.
Element: description Qualifier: abstract Value: Newsletter of the Becker Medical Library, Washington University School of Medicine
Element: format Qualifier: mimetype Value: application/pdf
Element: title Qualifier: none Value: News@Becker
Element: type Qualifier: none Value: Newsletter
Element: subject Qualifier: none Value: Cowdry, E. V.
Element: subject Qualifier: none Value: Proteome BioKnowledge Library
Element: subject Qualifier: none Value: WUSM's electronic repository
Element: subject Qualifier: none Value: informatics
Element: subject Qualifier: none Value: Wang, Lili
Element: subject Qualifier: none Value: EndNote
Element: subject Qualifier: none Value: citations
Element: subject Qualifier: none Value: finding aids
Element: subject Qualifier: none Value: archives
Element: subject Qualifier: none Value: web resources
Element: subject Qualifier: none Value: wireless networking
Element: subject Qualifier: none Value: Medical School Computing and Networking Services
Element: subject Qualifier: none Value: Instructional Technologies and Library Systems
Element: subject Qualifier: none Value: Instructions to Authors in the Health Sciences
Element: subject Qualifier: none Value: Karolinska Institutet: Diseases, Disorders and Related Topics
Element: subject Qualifier: none Value: computer classes
Element: publisher Qualifier: none Value: Becker Medical Library
Element: language Qualifier: iso Value: en_US
Processing contents file: archive_directory/item_003/contents
Bitstream: news@beckerv1i4.pdf

Processing handle file: handle

It appears there is no handle file -- generating one

3 item_003

Adding item from directory item_004

Loading dublin core from archive_directory/item_004/dublin_core.xml

Element: contributor Qualifier: author Value: Schoening, Paul
Element: contributor Qualifier: author Value: Lipsey, Kim
Element: contributor Qualifier: author Value: Kelly, Betsy
Element: contributor Qualifier: author Value: Halbrook, Barbara
Element: contributor Qualifier: author Value: Ferguson, Erin
Element: contributor Qualifier: author Value: Walter, Ed
Element: date Qualifier: issued Value: 2002-01
Element: identifier Qualifier: citation Value: News@Becker,2(1),Dec 2002-Jan 2003
Element: description Qualifier: none Value: Quarterly newsletter of the Becker Medical Library. Also available on the library website at <http://becker.wustl.edu/newsletter/v2i1/index.htm>
Element: description Qualifier: abstract Value: Newsletter of the Becker Medical Library, Washington University School of Medicine

University School of Medicine

Element: format Qualifier: mimetype Value: application/pdf
Element: title Qualifier: none Value: News@Becker
Element: type Qualifier: none Value: Newsletter
Element: subject Qualifier: none Value: MDConsult
Element: subject Qualifier: none Value: Wang, Lili
Element: subject Qualifier: none Value: electronic resources
Element: subject Qualifier: none Value: Archives and Rare Books
Element: subject Qualifier: none Value: HIPAA
Element: subject Qualifier: none Value: bioinformatics
Element: subject Qualifier: none Value: USA PATRIOT Act

Element: subject Qualifier: none Value: grant
Element: subject Qualifier: none Value: Institute of Museum and Library Services
Element: subject Qualifier: none Value: Missouri State Library
Element: subject Qualifier: none Value: 21 General Hospital
Element: subject Qualifier: none Value: WWI
Element: subject Qualifier: none Value: WWII
Element: subject Qualifier: none Value: Create Change
Element: subject Qualifier: none Value: scholarly communication
Element: subject Qualifier: none Value: medical students
Element: subject Qualifier: none Value: Pharmacia
Element: subject Qualifier: none Value: Pfizer
Element: subject Qualifier: none Value: Norman Rockwell
Element: subject Qualifier: none Value: art exhibit, paintings
Element: subject Qualifier: none Value: LibQual+ survey
Element: subject Qualifier: none Value: wireless networking
Element: publisher Qualifier: none Value: Becker Medical Library
Element: language Qualifier: iso Value: en_US
Processing contents file: archive_directory/item_004/contents
Bitstream: news@beckerv2i1.pdf
Processing handle file: handle
It appears there is no handle file -- generating one
4 item_004
[dspace@linpub1 bin]\$

Here is the log for a successful export from the News@Becker collection.

```
[dspace@linpub1 bin]$ dsrun org.dspace.app.itemexport.ItemExport --type=COLLECTI
ON --id=12 --dest=news@becker --number=001
Exporting from collection: 12
Beginning export
Exporting item to 1
Exporting Item 49 to news@becker/1
Attempting to create file news@becker/1/dublin_core.xml
Exporting item to 2
Exporting Item 63 to news@becker/2
Attempting to create file news@becker/2/dublin_core.xml
Exporting item to 3
Exporting Item 70 to news@becker/3
Attempting to create file news@becker/3/dublin_core.xml
Exporting item to 4
Exporting Item 71 to news@becker/4
Attempting to create file news@becker/4/dublin_core.xml
Exporting item to 5
Exporting Item 72 to news@becker/5
Attempting to create file news@becker/5/dublin_core.xml
Exporting item to 6
Exporting Item 73 to news@becker/6
Attempting to create file news@becker/6/dublin_core.xml
Exporting item to 7
Exporting Item 74 to news@becker/7
Attempting to create file news@becker/7/dublin_core.xml
[dspace@linpub1 bin]$
```

Here is a screen shot of the directory structure of the exported items from the News@Becker collection.

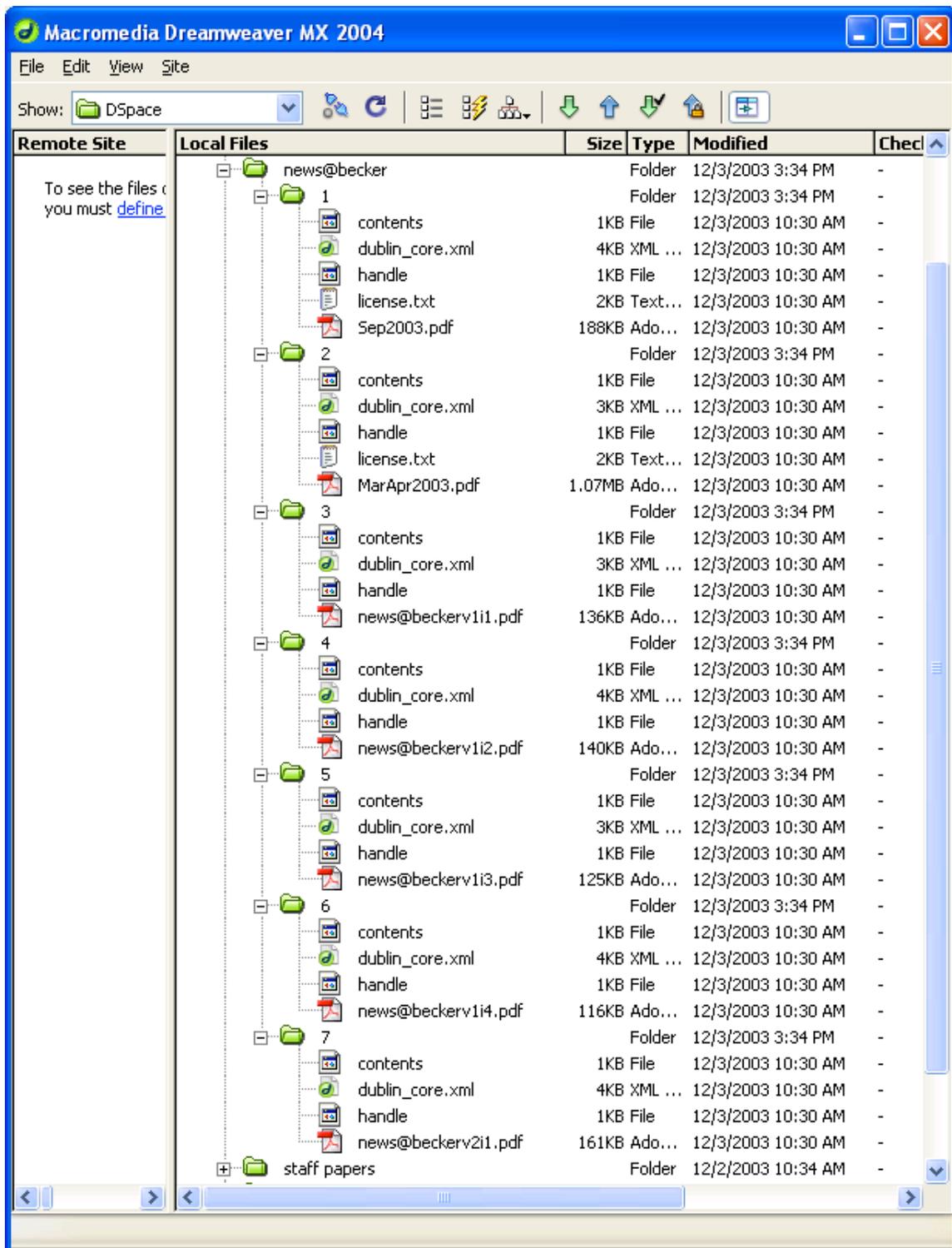


Image 42 – Exported files

Using Exported Items to Import Items into DSpace

You can use items that you have exported from DSpace to import into another DSpace server or to import into the same server after deleting the items.

Reasons for doing this would include:

- Transferring the items to a production server
- Re-importing the items to have new handles assigned to them if your handle server has been changed

You can just turn around the files without making some edits first.

Step 1 – Delete the **licence.txt** and **handle** files – These files are the byproducts of the exporting process and are not part of the importing process. They should be deleted as they are not needed and will throw off the import command in DSpace.

Step 2 – Edit the **contents** file. – This file lists the files you are importing into DSpace. During the export command, it adds the licence.txt name to this file. You need to delete this name because you are not importing the licence.txt file. It will again throw off the import command in DSpace.

Step 3 – Edit the **dublin_core.xml** file. – This is the file that provides the metadata for the items record in DSpace. The export process adds the handle at the time of exporting to the file as a URI metadata field. If you don't delete this line, the new record will have two URIs and both will look like handles. Delete this line and DSpace will generate a new handle for the new record.

Step 4 – Rename the folders. – The folders that hold each collections items that you exported have a simple numbering system. To import the items, the first folder should be named **item_000** and the subsequent folders should be numbered sequentially. The folders in the **bin/archive_directory** should be emptied and then overwritten with the new files. Be sure to delete all of the contents of the folders on the server. The item_000, item_001, etc. folder shouldn't have any old files in them from previous imports.

Appendix A

Editing the Syles.css.jsp File

In order for browsers to use the files in the /local directory first, before they use the default files, you must edit the styles.css.jsp file. At the beginning of the file, there is an if/then statement that you must remove. You must use just the then part of the statement.

The statement is this one:

```
// Make sure the browser knows we're a stylesheet
response.setContentType("text/css");

// Check for locally modified version
String localVersion = JSPManager.getLocalJSP("/styles.css.jsp");

    if (!localVersion.equals("/styles.css.jsp"))
    {
        // Local version is different
        JSPManager.showJSP(request, response, "/styles.css.jsp");
    }
else
{
    // Use this version (site default)

String imageUrl = request.getContextPath() + "/image/";

// Netscape 4.x?
boolean usingNetscape4 = false;
String userAgent = request.getHeader( "User-Agent" );
if( userAgent != null && userAgent.startsWith( "Mozilla/4" ) )
{
    usingNetscape4 = true;
}
```

You should change it to read:

```
String imageUrl = request.getContextPath() + "/image/";

// Netscape 4.x?
boolean usingNetscape4 = false;
String userAgent = request.getHeader( "User-Agent" );
if( userAgent != null && userAgent.startsWith( "Mozilla/4" ) )
{
    usingNetscape4 = true;
}
```

Appendix B

DSpace Administration Instructions

Path for Dublin Core XML file

`/home/dspace_source/dspace-1.2-source/config/registries/Dublin-core-types.xml`

Path to Tomcat /work directory – if changes to the jsp or stylesheet don't register, this is one place to check. You can delete all of the files in this directory, stop Tomcat and then re-start Tomcat to clear the server's cache.

`/usr/jakarta-tomcat-4.1.30/work/Standalone/localhost` – This should be cleared before stopping and restarting Tomcat

Commands to stop and re-start Tomcat:

Stop Tomcat:

```
JAVA_HOME=/usr/java/j2sdk1.4.1_02
export JAVA_HOME
/root/Jakarta-tomcat-4.1.18/bin/shutdown.sh
```

Re-Start Tomcat:

```
JAVA_HOME=/usr/java/j2sdk1.4.1_02
export JAVA_HOME
/root/Jakarta-tomcat-4.1.18/bin/startup.sh
```

DSpace Flowchart

1. Create a Community. (Administrator)
2. Create a Collection. (Administrator)
 - Create Workflows
3. Create groups (Administrator) – The ability to add, edit, and perform other tasks in a collection is assigned to groups of users in a collection. Users are assigned to these groups by the administrator.
4. Create authorizations by groups (Administrator) – The different authorizations (ADD, WRITE, READ) are assigned by the Administrator.

Anonymous must have READ so the collection can be read by all.

Workflow 1 and 2 don't need any special authorization, unless you are going to use Workflow 1 only. Then it must have the ADD authorization.

If you use only Workflow 1, you cannot edit the metadata that was provided by the submitter.

Workflow 3 must have ADD authorization.

If only one Workflow is needed, Workflow 2 is the best. It can reject a submission as well as edit the metadata.

Appendix C - DSpace Files

The files that make up the Web Interface for DSpace are stored in the **jsp** directory off of the **dspace-1.2-source directory**. Here is the path to the **jsp** directory:

/home/dspace_source/dspace-1.2-source/jsp.

Here is a screen capture of the directory structure.

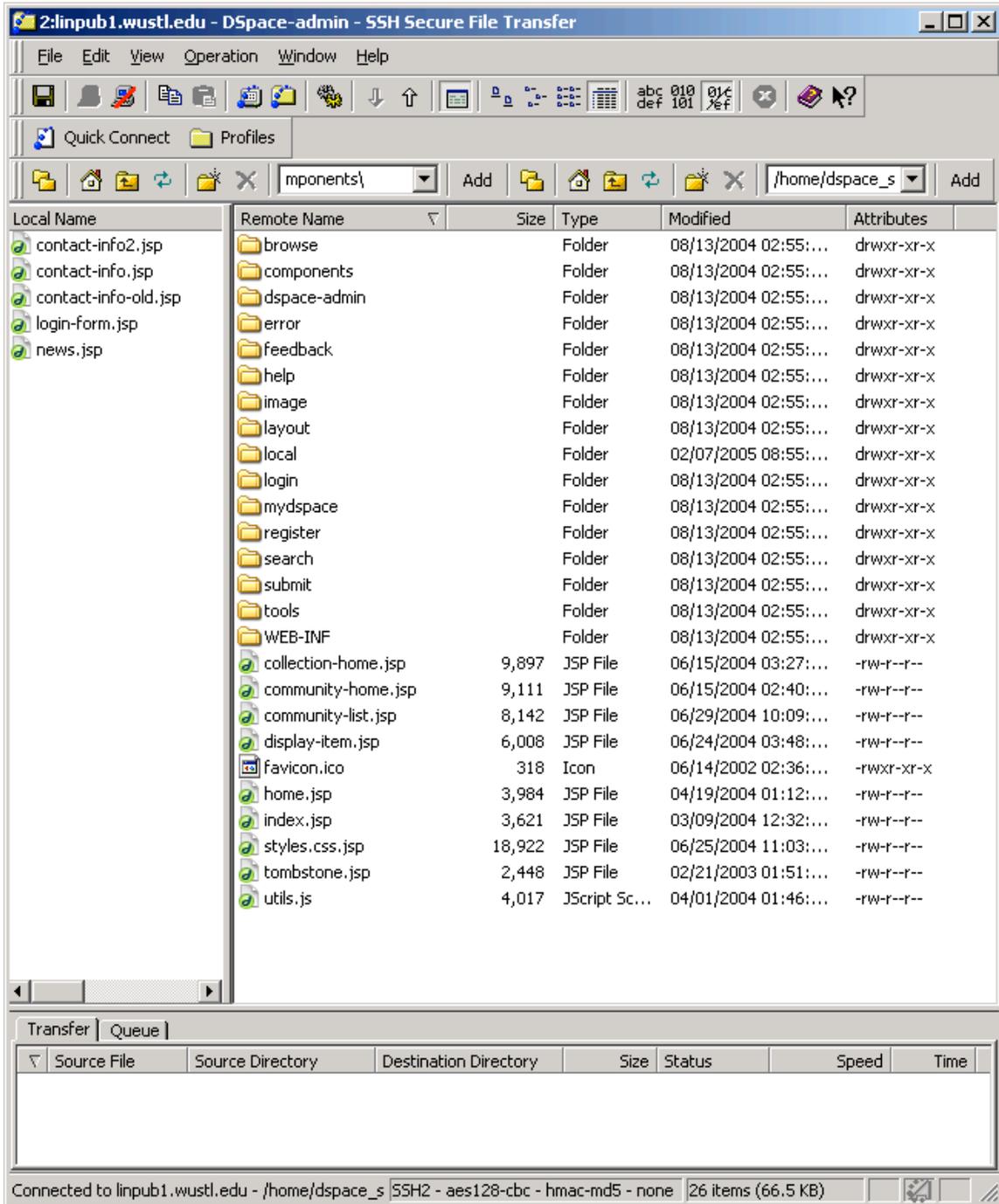


Image 43 – /jsp directory

/jsp/local – where the locally edited files are stored and used

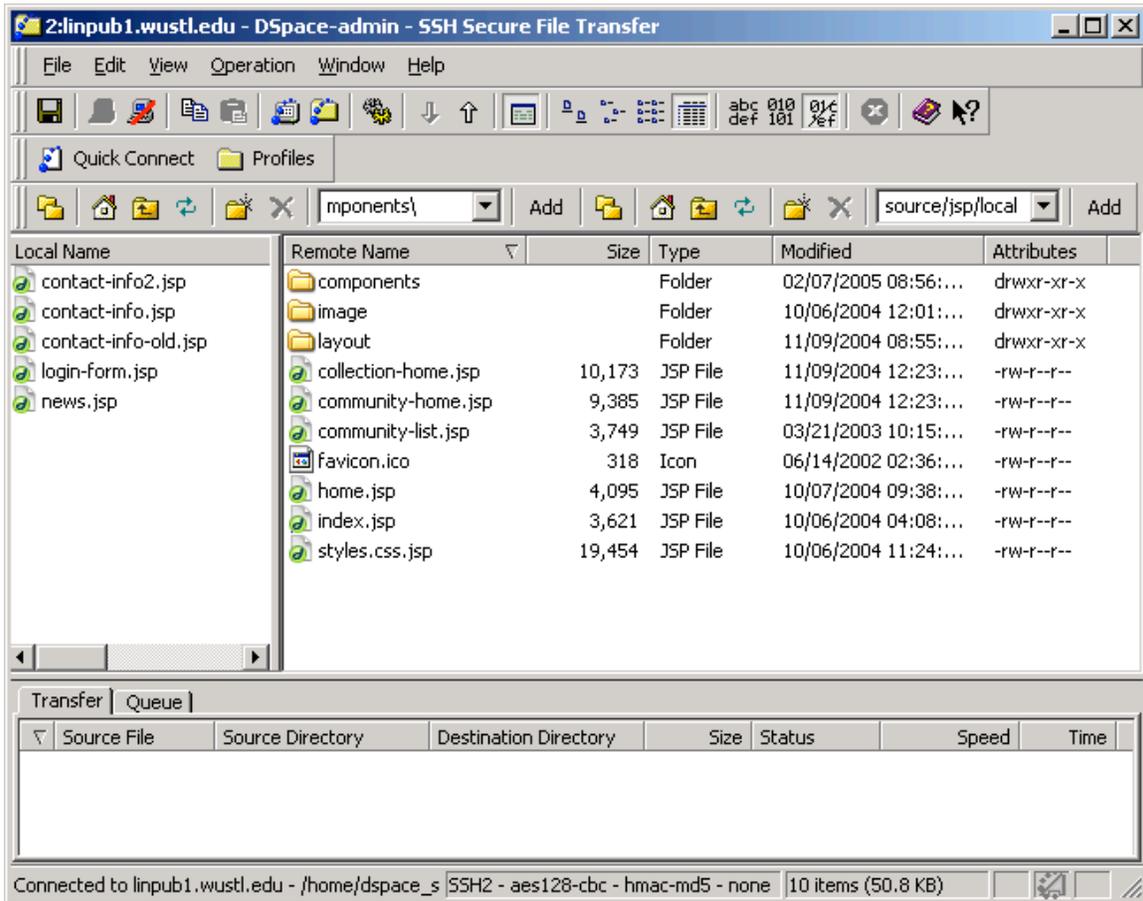


Image 44 - /jsp/local directory

/jsp/image – Image files for DSpace

Below is the list of the contents of DSpace/jsp/image directory. They are the original DSpace images that come with the installation. If you are going to use them in your local configuration, you should copy them into the /jsp/local/image directory as well as any local images you create.

- Arrow.gif – right facing arrow in left side frame
- Arrow-highlight.gif – right-facing arrow with transparent background
- Banner-large.gif – Dspace banner with star background – wider and taller
- Banner-small.gif – Dspace banner with star background – narrower and shorter
- Dome-bluewhite-smaller.gif – MIT libraries logo in top right-hand corner of main page
- Dspace-blue.gif – Dspace logo in top left corner of all pages
- Flying-man.jpg – man flying from Earth
- Hp.gif – HP logo in footer of pages
- Hp-mit.gif – HP + MIT logo in footer of pages
- Search-go.gif – Go button with yellow background
- Sfx-link.gif – MIT-SFX links logo
- Star-background.jpg – star field background image
- Stripe.gif – blue stripe image under the header on each page.

Appendix D - DSpace Flowchart

1. Create a Community. (Administrator)
2. Create a Collection. (Administrator)

Create Workflows

3. Create groups (Administrator) – The ability to ADD, EDIT, READ, and perform other tasks in a collection is assigned to groups of users in a collection. Users are assigned to these groups by the administrator.
4. Create authorizations by groups (Administrator) – The different authorizations (ADD, WRITE, and READ) are assigned by the Administrator.
 - Anonymous must have READ so the collection can be read by all.
 - Workflow 1 and 2 don't need any special authorization, unless you are going to use Workflow 1 only. Then it must have the ADD authorization.
 - If you use only Workflow 1, you cannot edit the metadata that was provided by the submitter.
 - Workflow 3 must have ADD authorization.
 - If only one Workflow is needed, Workflow 2 is the best. It can reject a submission as well as edit the metadata.

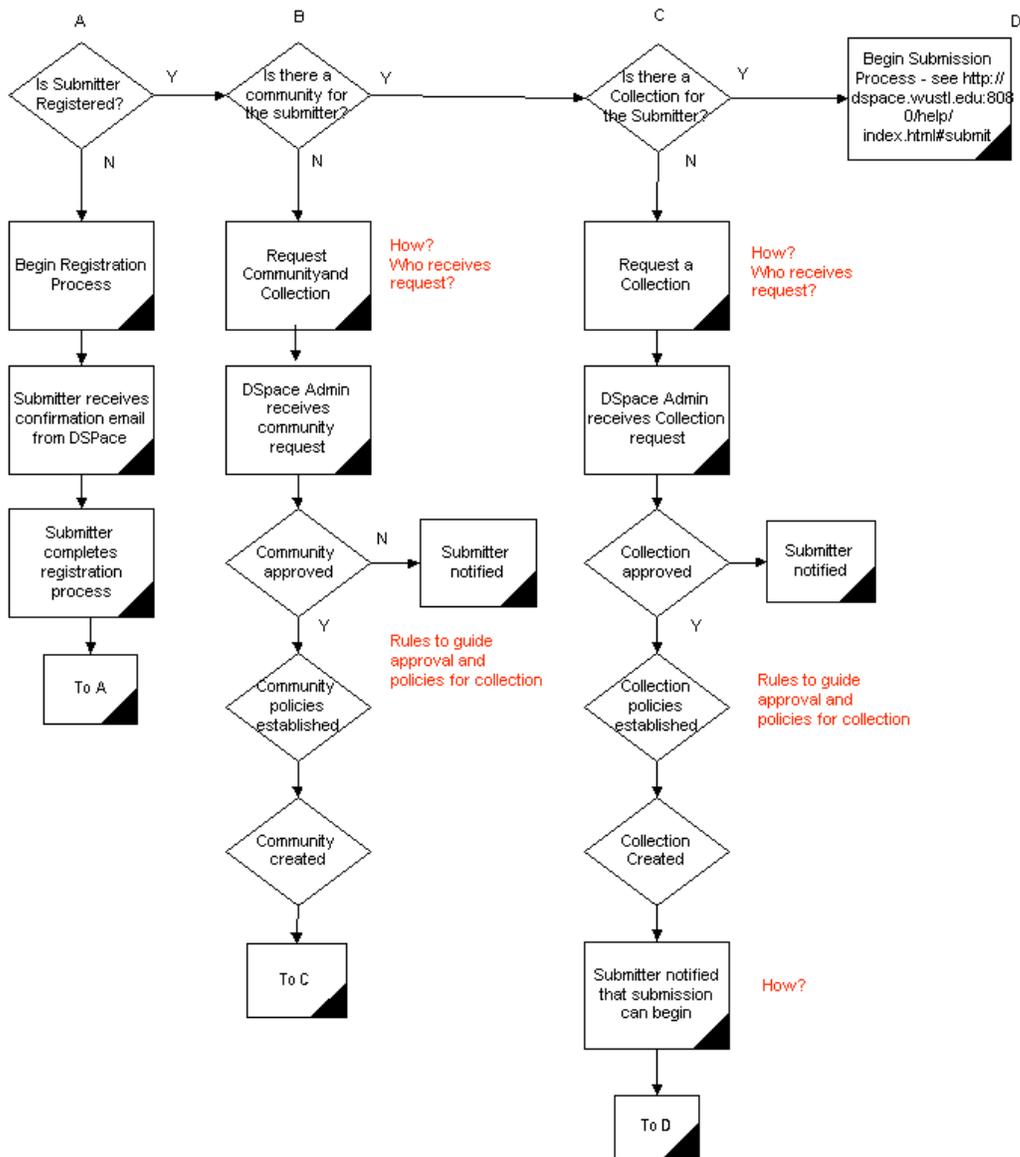


Image 45 – DSpace Flowchart

Appendix E - Customizing the Web User Interface

The Web UI is implemented using Java Servlets which handle the business logic, and JavaServer Pages (JSPs) which produce the HTML pages sent to an end-user. Since the JSPs are much closer to HTML than Java code, altering the look and feel of DSpace is relatively easy.

To make it even easier, DSpace allows you to 'override' the JSPs included in the source distribution with modified versions that are stored in a separate place, so when it comes to updating your site with a new DSpace release, your modified versions will not be overwritten.

However, note that the data (attributes) passed from an underlying Servlet to the JSP may change between versions, so you may have to modify your customized Servlet to deal with the new data.

The JSPs are stored in *[dspace-source]/jsp*. Place your edited version of a JSP in the *[dspace-source]/jsp/local* directory, with the same path as the original. If they exist, these will be used in preference to the distributed versions in *[dspace-source]/jsp*. For example:

DSpace default

Locally-modified version

[dspace-source]/jsp/community-list.jsp *[dspace-source]/jsp/local/community-list.jsp*
[dspace-source]/jsp/mydspace/main.jsp *[dspace-source]/jsp/local/mydspace/main.jsp*

Heavy use is made of a style sheet, in *[dspace-source]/jsp/styles.css.jsp*. If you make edits, call the local version *[dspace-source]/jsp/local/styles.css.jsp*, and it will be used automatically in preference to the default, as described above.

Fonts and colors can be easily changed using the stylesheet. The stylesheet is a JSP so that the user's browser version can be detected and the stylesheet tweaked accordingly.

The 'layout' of each page, that is, the top and bottom banners and the navigation bar, are determined by the JSPs *[dspace-source]/jsp/layout/header-*.jsp* and *[dspace-source]/jsp/layout/footer-*.jsp*. You can provide modified versions of these (in *[dspace-source]/jsp/local/layout*, or define more styles and apply them to pages by using the "style" attribute of the `dspace:layout` tag.

After you've customized your JSPs, **you must rebuild the DSpace Web application**. If you haven't already built and installed it, follow the install directions. Otherwise, follow the steps below:

1. Rebuild the `dspace.war` file by running the following command from your `[dspace-source]` directory:

```
ant -Dconfig=[dspace]/config/dspace.cfg build_wars
```

Building .WAR files

```
-bash-2.05b$ cd /home/dspace_source/dspace-1.2-source
-bash-2.05b$ ant -Dconfig=/dspace/config/dspace.cfg build_wars
Buildfile: build.xml
```

compile:

build_wars:

```
[copy] Copying 1 file to /home/dspace_source/dspace-1.2-source/build/jsp
[copy] Copying 28 files to /home/dspace_source/dspace-1.2-source/build/jsp
[war] Building war: /home/dspace_source/dspace-1.2-source/build/dspace.war
```

```
BUILD SUCCESSFUL
Total time: 17 seconds
-bash-2.05b$
```

2. Shut down Tomcat, and **delete** the existing `[tomcat]/webapps/dspace` directory.

Shutdown Tomcat.

```
-bash-2.05b$ /usr/jakarta-tomcat-4.1.30/bin/shutdown.sh
Using CATALINA_BASE: /usr/jakarta-tomcat-4.1.30
Using CATALINA_HOME: /usr/jakarta-tomcat-4.1.30
Using CATALINA_TMPDIR: /usr/jakarta-tomcat-4.1.30/temp
Using JAVA_HOME: /usr/java/j2sdk1.4.2_05
```

Delete existing `dspace.war` file.

3. Copy the new `.war` file to the Tomcat `webapps` directory:

```
cp [dspace-source]/build/dspace.war [tomcat]/webapps
```

Copy `dspace.war` file.

```
-bash-2.05b$ cp /home/dspace_source/dspace-1.2-source/build/dspace.war /usr/jakarta-tomcat-4.1.30/webapps/dspace.war
```

Customizing Web Files (JSP Pages)

4. Restart Tomcat.

```
-bash-2.05b$ /usr/jakarta-tomcat-4.1.30/bin/startup.sh
Using CATALINA_BASE: /usr/jakarta-tomcat-4.1.30
Using CATALINA_HOME: /usr/jakarta-tomcat-4.1.30
Using CATALINA_TMPDIR: /usr/jakarta-tomcat-4.1.30/temp
Using JAVA_HOME: /usr/java/j2sdk1.4.2_05
-bash-2.05b$cd ..
```

When you restart the web server you should see your customized JSPs.