DSpace –

A Sustainable Solution for Institutional Digital Asset Services –

Spanning the Information Asset Value Chain: Ingest, Manage, Preserve, Disseminate

INTERNAL REFERENCE SPECIFICATION

Functionality

Michael J. Bass,
David Stuve,
Robert Tansley

Margret Branschofsky,
Peter Breton [1],
Peter Carmichael [2],
Bill Cattey,
Dan Chudnov,
Joyce Ng

Hewlett-Packard Company
Building 10-500 MIT
77 Massachusetts Avenue, Cambridge MA 02139
+1 617 253 6617

mick_bass@hp.com

Masssachusetts Institute of Technology
Building 10-500 MIT
77 Massachusetts Avenue, Cambridge MA 02139
+1 617 253 xxxx

dspace-dev@mit.edu

---

[1] Under contract to MIT

[2] Under contract to MIT

Version 2002-03-01

CONTENTS

1. AUDIENCE

This document is intended to be a reference source for the functionality of DSpace. It is aimed at individuals who have some familiarity with digital library services and technology and who require more detail than other collateral provides (for example: new DSpace team members, project managers and team members at DSpace early federator institutions, research collaborators, and operations staff). The document is intended to be useful to provide source material for other collateral (e.g. user's manuals, functional overviews, etc.). The document does not assume a naïve reader. However, since the intended audience spans multiple disciplines (library science, computer science, operations, product/service management) we will attempt to define all terms inline before initial use.

2. FUNCTIONALITY

This section describes functionality of the DSpace release that supports early adopters at MIT, February 2002. Targeted future functionality, and research directions are described in the "DSpace Futures" whitepaper.

2.1 Supported Clients

DSpace is offered as a web-accessible service. Users access the service via a web browser. DSpace supports virtually any browser for dissemination. To access submit and manage/administer functionality, web-browsers must support web forms and file upload. To fully access all DSpace functionality, we recommend:

- Netscape Navigator v4.0 or later
- Internet Explorer v5.0 or later

2.2 Core Concepts and Relationships

Familiarity with these concepts and the relationships among them will aid understanding of DSpace functionality.

2.2.1 Site

A DSpace Site is a specific installation of the DSpace software, upon which services are offered that are backed by the committement of a host institution.

2.2.2 Host Institution

A DSpace Host Institution is that institution to whom use rights for submitted material are granted, and which stands behind the commitments that are made during the course of offering services atop the DSpace software platform. In the first instance of the DSpace Site at MIT, the host institution is MIT. For example, submitters offer distribution rights for submitted content to the host institution (MIT); The host institution (MIT) makes some commitment regarding storage and preservation of the submitted materials, and that it will use such materials in accordance with the rights granted.

2.2.3 User

A DSpace User is an individual who uses the DSpace system, by visiting a DSpace site with their web browser.

DSpace users can be at any given time either unknown, or credentialed to the system to some degree, for example via username/password or IP-address-based network presence.

DSpace keeps some basic information for *registered users* (email address, name, credential information), so that they can take advantage of all of the systems functionality (For example: Submission, My DSpace).

2.2.4 Group of Users

DSpace administrators can organize DSpace users into groups, which may be used to define participants in a role within a collection's submission process (e.g. "approver"). Policy statements can also refer to groups of users (for example, allow users in group "thesis-administrators" to edit the collection metadata for the thesis collection).

While it is true that there may be some organizational or socio-political group of people that correspond to an overall social "community", DSpace functionality is concerned with defined groups for specific roles within the community. For example: Who can edit the community's home page? Who can add collections to the community? Who can submit items to a collection's submission process?

2.2.5 Community

A DSpace Community is a convenient entry point or "portal" into the corpus of material in the repository.

A Community consists of a configurable home page for the community, a set of collections referred to by the community, and a group of users with management and administrative responsibility for the community.

Because communities must be administered, DSpace communities typically correspond – at least initially – to an organizational entity, for example, a school, department, laboratory, or research center.

2.2.6 Collection

A DSpace Collection groups together a set of DSpace items that are related in some way.

A DSpace Collection consists of a configurable home page for the collection, a set of items referred to by the collection, a configurable submission process for content entry into the collection, and a group of users with management and administrative responsibility for the collection.

Users who submit items to DSpace can choose a collection to submit to. Further, DSpace administrators can re-organize items into another collection – or even multiple collections – after their initial submission. Collections typically contain items that are similar in some dimension (for example: source, purpose, existing series or audience, subject matter, research topic). Administrators can also use collections to organize a submission process for consistency of submitted content (for example, with respect to scope of content, metadata requirements, required bitstream formats, etc.)

2.2.7 Approval Process

In institutional settings, decisions about what enters the archive often must be distributed, with decisions made close to communities. In DSpace, each collection can have its own approval process, and specify individuals who will participate in it. Such approval processes can range from very simple, to multi-stage review with multiple individuals participating at each stage.

2.2.8 Item

A DSpace Item is a logical grouping of a useful set of content and metadata that are related in some way. Items correspond to

"Archival Atoms" in DSpace. Examples of DSpace items include: a working paper, a conference presentation, a monograph (book), an annotated series of images, a video clip, materials for a course lecture, a research paper with auxiliary material (e.g. dataset, extended bibliography, rich media images).

2.2.9 Bitstream

In DSpace, a bitstream is simply a reproducible sequence of bits, with a corresponding bitstream format. Bitstreams typically correspond to content or metadata files that are submitted to DSpace.

2.2.10 Bitstream Format

Because preservation services are an important aspect of the DSpace service offering for Libraries, it is important to capture the specific formats of files that users submit. In DSpace, a bitstream format is a unique and consistent way to refer to a particular file format.

An integral part of a bitstream format is an either implicit or explicit notion of how material in that format can be interpreted. For example, the interpretation for bitstreams encoded in the JPEG standard for still image compression is defined explicitly in the Standard ISO/IEC 10918-1. The interpretation of bitstreams in Microsoft Word 2000 format is defined implicitly, through reference to the Microsoft Word 2000 application.

Bitstream formats can be more specific than mime-types or file suffixes. For example, "application/ms-word" and ".doc" span multiple versions of the Microsoft Word application, each of which produces bitstreams with presumably different semantics.

2.3 Submit

2.3.1.1 Anonymous vs.Credentialed Access

Any user can use DSpace to search and browse items and collections that are globally accessible. Only users who are registered with DSpace can submit items, administer items, collections, or communities, or view items that are not globally accessible.

DSpace will ask that the user authenticate themselves to the system whenever the system's current policy configuration indicates that some credential is required to perform the action.

2.3.1.2 Register with DSpace

Users that have been "bulk identified" (See "Bulk Identify Users", section 2.6.7.1) into the system may register with DSpace, by providing a valid email address to uniquely identify them, and a corresponding password. (X.509 certificates will be supported shortly). The user registration process:

- Gathers the user's email address

- Verifies that individuals are registering using an email address that they can access. In other words, it is not possible to register for DSpace using someone else's email address.

- Allows the user to initially set their user profile: password and basic personal information (name, etc.).

- Stores the password securely within the system

2.3.1.3 Forgotten Passwords

DSpace provides a secure process for users who have forgotten their password to select a new password without human administrative intervention. Users provide their email address, and the system mails them a special hyperlink which, when clicked, allows them to update their password.

2.3.1.4 Edit User Profile

Registered DSpace users can edit the basic personal information that the system keeps for them:

- Lastname, Firstname

- Contact Telephone (optional)

- Password

The system tracks users by the single email address with which they initially registered. A user's mail address cannot currently be changed via the end-user interface, but can be changed using the administrative user interface. Thus, end users can update their email address through a call to the help desk.

2.3.2 Submit to Collection

DSpace provides an easy way for materials to enter the archive in a distributed fashion across the host institution. The DSpace submission process gates entry of submitted material into the DSpace archive. Materials within DSpace are always in one of three stages:

- Items Being Assembled – the submitting user is still in the process of entering the Item's Metadata, or uploading the file(s) to be included with the item.

- Submissions Pending Archive – the submitting user has submitted the item to the initial Collection's submission approval process. But this approval process is not yet complete, and the item is not yet archived.

- Archived Items – the item is archived, having been approved for entry into the collection to which it was submitted.
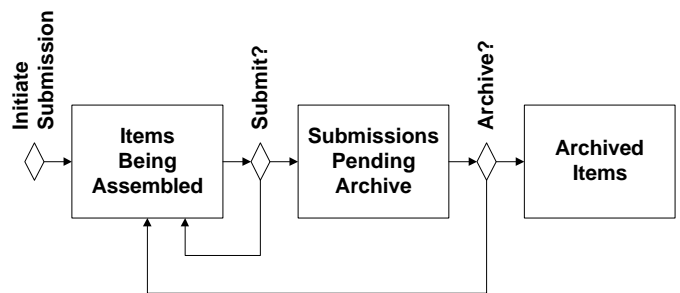
Authorized users can initiate submissions to a DSpace collection that they have chosen. Users can choose the collection to submit to in one of two ways:

- Explicitly, from a list of all collections in DSpace (accessible from DSpace home)

- Contextually, by first navigating to the collection of interest (accessible from each collection's home page)

2.3.3 Submit Baseline Metadata

With their submission, users can specify baseline metadata. Users fill in the same form for all submitted items. This baseline metadata will enable users to search across all of DSpace, and to easily find submitted items in the future. The baseline metadata requested for each submitted item is based upon the qualified Dublin Core Metadata Schema, adapted to DSpace requirements by MIT Libraries.

The elements captured through the end-user interface are:

- Author(s);
  Zero or more supported. Currently no authority control for authors (i.e. DSpace does not currently know that "Samuel Clemens" and "Mark Twain" are the same author, nor does it distinguish well between two authors that share the same name).

- Title(s);
  Including alternative titles, if applicable.

- Date of Issue

- Series Name and Report Number; For example "Sloan School of Management Working Papers, Number 2002-128"

  If applicable. Zero or more supported.

- Identifiers;
  Including ISBN, ISSN, ISMN, URI, and other. Zero or more supported.

- Language (in which submitted material is written).

- Subject Keywords;

  Zero or more supported. Currently no thesauri or authority control for subject keywords.

- Abstract

- Sponsors / Funding Codes

- Other Description

Additional elements are modeled and stored internally. Some of these are generated automatically by the system. Others may be managed by DSpace administrators using the admin user interface.

2.3.4 Submit Domain-Specific Metadata

In addition to DSpace baseline metadata, Users can submit domain-specific metadata that is specific to their item, or that is required by the curators of the collection to which they are submitting, by uploading a file that contains the relevant metadata.

For example, a collection of images might require that each image be submitted with GIS metadata indicating the location of the corresponding image.

Users may annotate their item using as many metadata formats as desired.

DSpace calculates and maintains an MD5 checksum of each file uploaded with the item, that can be used by DSpace administrators and by users to verify the integrity of the content and metadata within the system.

DSpace will maintain each of these metadata bitstreams with the item, and make them available with the item. DSpace does not yet support discovery based upon domain-specific metadata submitted in this fashion.

2.3.5 Upload File(s) to Item

Submitted DSpace Items are destined to become "archival atoms", that is some amount or boundary of material that make sense together. As such, each item can include multiple pieces of content. Each piece of content might comprise several files.

Users must submit one or more files to be included with their item. For example, a user might submit a conference paper, along with presentation materials actually used at the conference. A researcher might submit a pre-print of an article, along with dataset(s) that would enable other researchers to independently reproduce the research results.

DSpace calculates and retains a checksum of each file uploaded with the item that can be used by DSpace administrators and by users to verify the integrity of the content and metadata within the system. This checksum can be obtained through the end-user interface.

DSpace attempts to recognize the bitstream format of each uploaded file. If it cannot match the bitstream format to a known format from the system's bitstream format registry, it asks the user to describe the format so that library administrators can track and support important emerging formats over time. See "Manage Bitstream Format Registry", Section 2.6.6.

2.3.6 Grant Distribution License

To enable the host institution to administer, preserve, and distribute the submitted material, DSpace asks the user to grant to the institution a non-exclusive license to distribute the material, and to translate it for the purposes of preservation.

License agreements can vary by Collection, and are specified by the collection administrator(s) for the Collection (see "Specify Collection's Required License", section 2.6.1.4).

Because license terms are likely to change over time based on the needs of submitters and the host institution, DSpace stores a copy of the license that was granted at the time of submittal as a bitstream within the item, so that the specific terms agreed upon are always available.

2.3.7 Augment & Approve Submissions

When the user submits an item to a DSpace collection, the system routes the submission through the approval process previously configured for that collection (see "Configure Item Approval Process", section 2.6.1.6). The approval process can vary by collection, and can include any combination of the approval roles defined below. Further, groups of users can be used to associate any number of users with each of the roles that are configured for the process.

The system routes the submission to individuals (if any) who have been chosen to assume each of the following roles for the target collection:

- *Reviewers:* review the content of the submission for appropriateness to the collection. Reviewers act as the gatekeepers for the collection. Reviewers are empowered to return a submission back to the submitter because it is deemed inappropriate for the

collection. Reviewers do not edit submission metadata.

- *Approvers:* check the submission for completeness and/or obvious errors (e.g. wrong file uploaded). Approvers can edit the submission's metadata to fix obvious errors, and are empowered to return a submission back to the submitter because it is incomplete or in error.

- *Metadata Editors:* check and/or augment the submission's metadata. For example, a metadata editor may be assigned to add the appropriate series name and number to each submission in a collection. Metadata Editors can only edit the submission's metadata.

Each collection may specify zero or more individuals to assume each of these roles. If zero individuals are chosen, the system skips that stage of the approval process. For example, a collection might have a simple process for a series of working papers where all submissions are routed to a single staff metadata editor. This metadata editor would assign the series number to each submission. All submitted items would be approved and enter the archive – in this configuration it would be the responsibility of each submitter to crosscheck their own submissions and decide that they were suitable for the archive. Another collection might additionally define a reviewer role to ensure that submitted material is of appropriate nature to be included in the collection. Yet another collection might use all three roles. The intent is to allow communities flexibility in meeting their collection management needs while avoiding the institutional paralysis that often results from too much flexibility or too many choices.

DSpace users in each role are notified via email when they have pending submissions that require their attention. These tasks are also available to them from the "My DSpace" section of the DSpace system (See "View Pending Approval Tasks", section 2.3.8.3).

As each task is completed, the submission is routed to the individual(s) responsible for the next stage of review, if any. The system sends email(s) to each of these Users, with links to their pending task in the DSpace system.

When the submission either enters the DSpace archive or is returned to the submitter, the system notifies the submitter by email about the status of their submission.

2.3.8 My DSpace

DSpace offers each user personalized access to information within the system through their *My DSpace* page. As appropriate given their role(s) in the system, users can view their:

- Items being assembled

- submissions pending archive

- archived Items that they submitted

- review tasks for Items pending archive

DSpace filters these sections so that only the sections relevant to each user are presented.

2.3.8.1 View Archived Items Submitted by User

Users can easily view all of the archived items that they submitted from their personalized *My DSpace* page.

2.3.8.2 View Items being Assembled

Users can access submissions that were interrupted while partially complete (because key information was not available, because all files were not available for upload, or simply because the user was called away for an extended coffee break) from the "Items being Assembled" section of their personalized *My DSpace* page.

2.3.8.3 View Pending Approval Tasks

Users who are taking part in the approval process for one or more collections within DSpace can view, select, and perform tasks that require their attention from the "Pending Approval Tasks" section of their personalized *My DSpace* page.

2.4 Archive

2.4.1 Assign/Resolve Citeable Persistent Name

DSpace creates a persistent name for submitted material that enters the archive. The name that DSpace assigns is intended to be valid and resolvable in perpetuity; the institution hosting the DSpace system stands behind that commitment, even if the DSpace system in place at the time of submission is removed, revised, or replaced. The original host institution may at some point assign that commitment to another institution (conveying responsibility for maintenance/administration of the submitted material to the other institution).

DSpace currently creates and assigns persistent names for:

- archived items

The names assigned are free of semantics about the material that they refer to. Further, they are decoupled from both logical and physical storage currently in use for associated material. Assigned names are registered using the CNRI Handle System. CNRI provides an http resolver at:

http://hdl.handle.net/<handlename>

CNRI also makes plug-ins available for most popular browsers, that allow Handles to be resolved without need for an http resolver. DSpace does not require that DSpace users install the handle plug-in, and is agnostic about its use.

The name assigned by DSpace is appropriate for citing the submitted material from other digital materials, or from print. For example, a handle assigned by the DSpace system, embedded with an http-based resolution request from CNRI, is:

http://hdl.handle.net/1721.2/27

This work might be cited digitally or in print as:

> Bass, Mick et. Al.. "DSpace – A Sustainable Solution for Digital Asset Services". *MIT DSpace,* http://hdl.handle.net/1721.2/27.

2.4.2 Store Bitstreams

DSpace provides storage for the individual bitstream(s) associated with submitted, archived, items.

Users can access the stored bitstreams for any item through that item's overview (see "View Item Overview", Section 2.5.4).

DSpace allows additional physical storage for bitstreams to be added, and for stored bitstreams to be easily moved from one physical store to another. That is, if one disk array fills up, DSpace makes it easy to add another, and to distribute content appropriately across storage.

DSpace will *at a minimum* remember the bit sequences in the bitstreams associated with a submission and be able to return those bitstreams to future users. Achieving at least this service level is of critical importance.

Additional preservation information and/or services may also be offered using the DSpace platform. That is, additional preservation services may include additional information that enhances the probability that bitstreams returned in the future will actually be understandable and useable by community designated by the host institution.

2.4.3 Generate Bitstream Checksums

For each bitstream maintained within the system, DSpace generates and stores an MD5 checksum that can be used either by users or by downstream preservation services to verify the integrity of the stored bitstreams over time. Checksums for each of an Item's bitstreams are accessible from the Item Overview.

2.4.4 Identify Bitstream Format and Preservation Service Level

DSpace maintains a bitstream format for each bitstream stored within the system, in addition to maintaining the sequence of bits associated with that bitstream.

Bitstream formats are chosen from a registry of known formats, the Bitstream Format Registry, which is maintained by DSpace administrators. See "Manage Bitstream Format Registry", Section 2.6.6. Note that bitstream formats may be more specific than mime-types.

DSpace attempts to identify the bitstream format automatically. If it is unable to do so, it requests that the submitting user identify the bitstream format from those within the bitstream format registry.

If the bitstream format of a submitted bitstream has not yet been entered into the bitstream format registry, the system requests a prose description of the format, so that DSpace administrators can consider whether or not this format should be incorporated into the registry.

DSpace administrators can identify a preservation service level for each entry in the bitstream format registry. The possible preservation service levels are depicted in the following table:

	unsupported	**supported**
Unknown	0, not known	
Known	1, forming guidance	2, supported

In the table above, "known" means that:

> DSpace administrators have named the file format, its version, mimetype, and any other relevant information (for example, relationships with other formats) in the system's bitstream format registry;

and

a process is in place to actively identify incoming submissions using the named format. The de facto process for format identification is a heuristic-based attempt by the system to identify the format, followed by the opportunity for the submitting user and/or metadata editor to override from the list of formats in the bitstream format registry.

"supported" means either:

> DSpace administrators have procured and stored sufficient specifications documenting the format – either within DSpace or in analog format within a trusted library;

or

> DSpace administrators have procured and stored within DSpace representative code for processing the format.

This implies the following preservation service levels:

- *Level 0, not known:* The bitstream will be maintained and returned upon request in the future. However, because the format of the submitted bitstream is not known to the host institution, no further preservation statements can be made. Members of the archive's designated community may or may not be able to successfully use or understand the (identically returned) bitstream in the future.

- *Level 1, forming guidance:* The bitstream will be maintained and returned upon request in the future. The host institution is aware of and has named the format of the bitstream. However, the host institution has not committed to attempt to preserve the utility of bitstreams of this format into the future.

  The format may not be supported because the host institution has not yet had the opportunity to consider it in any depth, or the host institution may have considered it and decided that the format is inherently difficult or impossible to preserve.

  The host institution will attempt to offer guidance to users about the status of investigations into formats at this level. For some formats, especially those that are widely used or supported by partners (e.g. Information Systems or Academic Computing Departments) the host institution may choose to offer further guidance suggesting additional formats, alternative formats, or techniques that users can use to reduce the risk that their submission will be difficult to use or understand in the future.

  Bitstream Format Preservation Guidance offered by the host institution will be captured in Items that they maintain in a DSpace collection created for this purpose. These items will be referred to from the bitstream format registry.

- *Level 2, supported:* The bitstream will be maintained and returned upon request in the future. Additionally, the host institution believes that the ability to use and understand the submitted material can be reasonably

preserved into the future, and commits to exercise its best effort to do so, in a manner appropriate to the material's context. Preservation techniques may include emulation, migration, transformation services, or other strategies. Such strategies will be outlined and made available to users over time as appropriate.

In addition to capturing the bitstream format, the system also captures prose text capturing the purpose, use, or description of the bitstream (for example, "preprint of journal article", "additional images referred to in article", "research dataset").

2.4.5 Store Item Provenance Information

DSpace maintains provenance information for each item in the archive, in qualified Dublin Core <description.provenance> values.

In addition to the system-generated provenance values described below, administrators can annotate items with additional human-generated provenance values.

The system automatically creates and stores the indicated human-readable provenance information upon each of the following system events:

- When accessioning (gathering of metadata and uploaded files) is completed:

  "Submitted by [name] and accessioned in DSpace on [date]."

- When an item's first issue is in DSpace:

  "Made available in DSpace on [date]."

- When an item is approved and enters the archive:

  "Accepted by [name] on [date]."

2.4.6 Store Object History, Serializations

DSpace offers history functionality to provide an audit trail of the administration of the archive, to provide data supporting root-cause analysis, and to support human-moderated rollback capabilities.

Whenever selected events of archival interest occur within the DSpace system, the system creates and stores a snapshot of the objects involved, and the relationships among them.

The information in DSpace history snapshots is recorded using open standards. The generated history snapshots are graph-oriented, and are usable outside the DSpace system by emerging, standard, semi-structured data manipulation toolkits. Specifically, the history snapshots adopt the "ABC" data model from the Harmony project

> <http://www.ilrt.bris.ac.uk/discovery/harmony/docs/abc/abc_draft.html>,

and implement this data model using RDF, see

> <http://www.w3.org/RDF/ >.

Taken together, the history snapshots provide a time-based record of significant changes to the DSpace corpus. The history data does not provide current information about the archive; it provides a record of significant changes that occurred in the past. History snapshots are read-only. Once created, they are never changed. Further, the intent from the outset is that they never be deleted.

DSpace offers a basic lookup capability for history snapshots, that provides an interested administrator with a list of all history snapshots that pertain to a specified item.

Dspace does not currently offer query functionality that interprets the contents of history snapshots. Doing so is a goal of the DSpace research agenda.

2.4.6.1 History Events

The following events within DSpace create history snapshots:

- Communities
  - *create / modify / delete Community*
  - *add collection to community*
- Collections
  - *create / modify / delete Collection*
  - *add item to collection*
- Items
  - *create/modify/delete Item*
  - *assign handle to item*
  - *modify item contents (bitstreams, metadata fields, etc.)*
- Users
  - *create / modify / delete User*
- Submission Process
  - *submission approval process completed*

2.4.6.2 Example

An item is submitted to a collection via bulk upload. The Item is eventually added to the collection. At this time, DSpace creates a history snapshot that records information about the submittal.

The history snapshot includes the following new resources (all with unique ids):

- an event. This event will be annotated with the time that the addition occurred, and be used to relate the addition to the resulting state of the archive (see next bullet).
- a state. This resource provides a way to refer to the state of some subset of the archive. Events within the archive cause some prior state to transition to a subsequent state. Examining the relationships between states and events can allow administrators to understand specifically what has occurred within the archive, and how it got to be the way that it currently is.
- an action. This models the addition itself. Other actions might also be modeled, if they happen atomically with the event. Actions are typically performed by agents or tools. These tools may have specific versions. The action was initiated by a particular DSpace user. All of these assertions can then be models as relationships with the action resource.

For example, the system might includes the following relationships in the snapshot:

event —atTime→ time

event —hasOutput→ state

Item —inState→ state

state —contains→ Item

action —creates→ Item

event —hasAction→ action

action —usesTool→ DSpace Upload

action —hasAgent→ User

The system further includes serializations that capture the state of archival objects participating in these relationships (in this case, the Item, the User, and the DSpace Upload).

2.5 Disseminate

2.5.1 Search
DSpace offers users the capability to search DSpace for items of interest.

2.5.1.1 Query Features
DSpace offers the following search features through its web-based user interface:

- Search all of DSpace

- Bounded search, within all of a specified community's collections.

- Bounded search, within a specified collection

- Simple search

  Searches fields: author, title, keywords

- Case insensitive search

  All searches are case insensitive.

- Truncation, Constraints

  program*
  ® program, programme, programming

  ?i?er
  ® diver, wiper

  +gone +wind
  ® "gone with the wind"

  but not "here today, gone tomorrow"
  or "the cold north wind"

- Word stemming

  searches for "processed" match any of [process, processing, processor]

- Stop words

  common words (e.g. "a", "an", "the") are omitted from the search

These query features are available for DSpace baseline metadata. Other metadata submitted with the item (i.e. in bitstreams within the item) is currently neither indexed nor searchable by DSpace.

2.5.1.2 Finding Newly Submitted Items
When new items are submitted to DSpace, they can be discovered immediately via search.

2.5.1.3 Search Results
Once the user specifies a search, DSpace performs the search and produces a result set. DSpace displays the result set, including a terse description for each item in the results.

From the terse item description displayed in the search results, the user may select a desired item to view its *Item Overview* (see Section 2.5.3).

2.5.2 Browse
DSpace users may browse the contents of DSpace in the following ways.

2.5.2.1 Browse Communities and Collections
DSpace administrators organize DSpace items into collections, and include collections in communities. Users can browse the structure of communities and collections, in an outline view. Each outline entry includes text describing the contents and/or purpose of the corresponding community or collection. Users can select an entry from the outline view to access the home page of the selected community or collection, from which further bounded browse or search can be performed.

2.5.2.2 Browse Items
From the DSpace home page, users can browse all items in DSpace by title, author, or issue date.

From a community or collection home page, users can initiate a bounded browse within that community or collection. DSpace supports bounded browse by title, author, or issue date.

All browses support paginated results displays, with links for previous and next page, as well as shortcuts that allow the user to jump to a particular location within the browse results (for example, browse authors beginning with 'S').

Browse by date and title display one entry for each item. The displayed entry is linked to the corresponding item overview (see Section 2.5.3).

Browse by author collapses all items by same author into a single entry, which links to a list of all items having authors with that name. The user may sort this list either by title or by date. Entries from the browsed author's item list link to the corresponding item overview. Future authority control services for authors may allow browse to refrain from collapsing works created by different authors having the same name.

2.5.2.3 New Collections
New collections are displayed in a sidebar on the DSpace home page, and linked to the corresponding collection home page, from where they can be browsed.

2.5.2.4 Recent Submissions
Recently submitted items to each collection are displayed in a sidebar on each collection's home page.

2.5.3 Search, Browse, and Authorization
Search and browse currently display brief descriptions for all items, whether or not the user is authorized to view the item. If the user clicks on the brief description of an item that she is not

authorized to view, the system will refuse access to the item overview.

2.5.4 View Item Overview

Terse item descriptions in either the search view or browse view are linked to the corresponding item overview. Authorized users can view an item's overview, which displays the item's DSpace core metadata, lists the collection(s) that include it, and provides links to each of the bitstreams that the item comprises.

2.5.5 Download Item Bitstream(s)

Users may download each of the bitstream(s) in the item by selecting the bitstream from those listed in the item's item overview.

Bitstreams that:

(1) have a known bitstream format, that format which

(2) itself has a corresponding mime-type

will be reasonably rendered in-browser by the user's web browser.

Alternatively, the user may download the file using the standard "Save Target As" functionality within their web browser.

2.6 Manage and Administer

In addition to the end-user interface available to all DSpace users, DSpace also offers an administrative user interface, from which administrative functionality can be conveniently accessed. Like the end-user interface, the administrative user interface is web-client based.

Note that this user interface offers access to the *system functionality* corresponding to DSpace administrative tasks. This section describes that functionality, which is typically orthogonal to the *instituitional policy* regarding appropriate circumstances for invocation of each task.

2.6.1 Administer Communities and Collections

2.6.1.1 Create/Delete Community/Collection

Authorized DSpace users can create a new community, or create a new collection and associate it with an existing community. Admin users can also delete an existing community or collection.

2.6.1.2 Edit Community/Collection Home Page

Authorized users can edit key information about a community or collection. The information is used primarily within the home page of the community or collection, and includes:

- Name
- Short Description
  (used in community/collection browse)
- Intro Text
- Logo (upload bitstream from local filesystem)
- Copyright Text
- Sidebar Text

Intro Text, Copyright Text, and Sidebar Text may include arbitrary HTML. This allows some degree of user formatting control, and allows links to pages outside DSpace to be easily embedded.

2.6.1.3 Edit Collection Provenance Description

The system maintains a provenance description for each collection. Authorized users can use the administrative user interface to edit this description.

2.6.1.4 Specify Collection's Required License

When users submit an item to a DSpace collection, the system asks them to grant an appropriate license to the host institution (See "Grant Distribution License", Section 2.3.6).

Authorized users can specify the license text that is presented upon submission to the collection, a copy of which is stored with the item as a consequence of submission.

DSpace administrators can also specify authorization policies regarding distribution of items within each collection so that they are consistent with the license requested and granted for items submitted to the collection. See "Administer Authorization Policies", Section 2.6.4.1.

2.6.1.5 Specify Collection Metadata Defaults

Appropriate default values for DSpace core metadata may vary from collection to collection. DSpace maintains an optional "template item" for each collection, from which the system copies default value(s) for each DSpace core metadata field.

Authorized users can create such a template item for a collection, and edit the metadata values within it, which will then be used to initially populate each item subsequently submitted to the collection.

For example, this functionality could be used to specify a series name that corresponds to a collection.

2.6.1.6 Configure Item Approval Process

Each collection may optionally associate a group of users with each of the "reviewer", "approver", and "editor" roles described in Section 2.3.7.

If in a particular collection, any of these roles has no associated group of users, then that role is omitted from that collection's submission process. When an administrator associates a group with a particular role for a collection, DSpace will notify all users in that group when tasks are queued for that role. Authorized users can therefore configure the submission process by:

- Associating group(s) corresponding to each submission process role (or "action"). See "Administer Authorization Policies", Section 2.6.4.1.
- Maintaining the appropriate users within each such group. See "Administer Users and Groups", Section 2.6.3.

2.6.2 Administer Items

Authorized users can change the content and metadata of an item, or expunge the item.

Note that the history subsystem retains information about all changes, including the user who performed the change.

2.6.2.1 Add/Change/Remove MetadataValues

Authorized users can use the admin user interface to:

- Add baseline metadata: that is, add to an item any number of values for any qualifer in the system's registry of Dublin Core Qualifiers.

- Change baseline metadata: that is, edit any existing value(s) for any qualifier

- Remove baseline metadata: that is, remove any/all existing value(s) for any qualifier

2.6.2.2 Add/Delete Bitstreams

Authorized users can use the admin user interface to:

- add to an item any number of additional bitstreams

- remove any existing bitstream(s) from the item

2.6.2.3 Expunge Item

- Authorized users can use the admin UI to expunge an item from the system. This deletes the item completely

2.6.3 Administer Users and Groups

2.6.3.1 Users

Authorized users can use the admin user interface to:

- Identify users who may register with DSpace. The email address of valid users must be entered in this way before users will be able to register for the system.

  The requirement to identify users before registering with DSpace stems from our early adopter requirement of limiting use of DSpace to the MIT environment. We wish individuals outside MIT to be able to register for DSpace, but must pre-qualify their registration. This restriction will be removed after the early adopter trials.

- Edit the personal information stored for a specified user, including their email, first name, last name, phone, password, and whether or not they are active in DSpace.

- Delete a specified user from the system.

2.6.3.2 Groups

Authorized users can create and maintain groups of users. DSpace uses these groups and authorization policies to determine which users are authorized to perform particular actions within the system (See next section).

Authorized users can use the admin user interface to:

- Create a group of users

- Name the group
  [where is group name used?]

- Add users to the group

- Remove users from the group

2.6.4 Authenticate / Authorize Users

DSpace uses a set of machine-interpretable policy statements to determine whether or not a given user is authorized to perform a given action upon a given resource.

These policy statements are configurable.

2.6.4.1 Administer Authorization Policies

DSpace keeps authorization policies that allow it to understand what credentials are required (if any) to undertake particular actions upon particular resources. Authorized Users can administer these policies.

DSpace policies can be added that:

- Set a default distribution policy for all items within a collection. For example all items within the collection might be made visible globally. Alternatively, they might be made visible only within MIT.

- Specify groups of users permitted to submit to a collection

- Specify reviewers, approvers, and metadata editors for a collection's submission process.

- Configure who may modify the items within a collection

- Configure who may administer communities and collections themselves.

Because DSpace's mechanisms to assert and interpret policy statements are quite flexible, the above list is not all-inclusive. Multiple policy statements can apply to a particular resource, enabling many situations to be handled (for example, provide view access within MIT, and also allow access to a specified group of DSpace users outside MIT).

2.6.5 List / Abort Submission Proccesses

Authorized users can view the state of open submission processes in the system, and can abort a selected submission process.

When a submission process is aborted, any pending tasks are removed from the "My DSpace" tasklists of approvers, reviewers, and/or metadata editors. The submitted item remains accessible to the submitter from their "My DSpace" page.

2.6.6 Manage Bitstream Format Registry

Authorized users can add, edit, or delete a format in the system's registry of bitstream formats. DSpace uses this registry to allow users to identify the format of bitstreams they submit to DSpace. Typically this registry will be maintained DSpace Administrators and library staff. In the future, management of the registry may come to be shared across several institutions.

For each bitstream format, the administering user can specify the mime-type, short description, description, and the preservation support level determined by the host institution for bitstreams in this format.

Formats already in the registry can be deleted only if they are not used by any bitstreams in the system.

2.6.7 Bulk Import

DSpace system administrators can bulk import items into the system, and bulk identify users to the system.

Unlike all of the preceding administrative functionality, the bulk import tools are *not* available in the web-client-based administrative user interface. Rather, bulk import tools are driven from a command-line interface.

2.6.7.1 Bulk Identify Users

DSpace administrators can bulk identify users to the system.

For example, users could be extracted from a legacy system such as the MIT roles database and made available within DSpace.

2.6.7.2 Bulk Import Items

DSpace administrators can bulk import items into the system.

For example, an entire series of scanned material such as the MIT Press out-of-print corpus can be made available within DSpace.

Current Bulk Import tools implement a crosswalk from MARC to DSpace baseline metadata (in qualified Dublin Core). When content is imported that has already been MARC catalogued, the original MARC is serialized as a Bitstream and retained with each imported Item.