DSpace Basic Tutorial
DSUG Roma 2007

dott. Andrea Bollini – CILEA
dott. Federico Meschini

Agenda

1. Introduction to DSpace System
   - Scenarios & business cases
   - Functionalities
   - Data model

- Metadata & Schema Management

- Submission system configuration (submission.xml, input-forms.xml)

- Understanding the workflow system

- Search & browsing system configuration

- Introduction to development
  6.1 Quick install & run
  6.2 Integration with Eclipse
  6.3 Basic layout modification (JSPUI)

- CRON & utility scripts (batch import, etc.)

Andrea Bollini
Federico Meschini

Introduction to DSpace System

" DSpace **captures** your data in any format – in text, video, audio, and data. It **distributes** it over the web. It **indexes** your work, so users can search and retrieve your items. It **preserves** your digital work over the long term.

DSpace provides a way to manage your research materials and publications in a professionally maintained repository to give them greater visibility and accessibility over time. "

www.dspace.org

- Open-Source software for creating digital library
- Collect and organize an institution intellectual production, collection, describing and disseminating digital objects
- Teachers and researchers: self-archiving, larger research exposure and management support for teaching materials
- Users: full-text retrieval through different search options and network points.

Andrea Bollini
Federico Meschini

1. Creation of Institutional Repositories OA-Compliant

2. Creation of thematic digital libraries

3. Creation of preservation archives

Andrea Bollini
Federico Meschini

Repository: a definition

A repository is an on-line open archive which makes available scientific literature (pre-prints, lectures, thesis, etc.) in digital form. Archive content usually is not subject to peer-review, but every repository has its own internal policies which regulate the archiving.

Andrea Bollini
Federico Meschini

Restraints usually are about:

- Content type: research, teaching

- Content producers: institutional, disciplinary, mixed

And not referring only to open access repositories

- Access model: open, delayed o restricted

Andrea Bollini
Federico Meschini

From policies to functional requirements

Every policy implemented by the archive has consequences on the project functional requirements.

Authentication, Workflow, Metadata &

Authority Files and DRM are only few of the features which are influenced by the repository policies

Andrea Bollini
Federico Meschini

Institutional Repositories - IR

Usually setting up an institutional repository entail the integration of the repository authentication system with the institution centralized one (for example based on LDAP). About the approval in the workflow level usually the IRs do not present particular requirements, and, thanks also to the trust existing between author/employee and institution, it could also be disabled (submitted = archived). When there are restrictive policies for accessing the full-text the authorizations are managed through IP addresses/network.

Andrea Bollini
Federico Meschini

DSUG Roma 2007

Research Repositories

In a research repository the supporting tools for the final user are very important, in particular about the reuse of repository content and the creation of new knowledge. Research material is always quoted by the researchers and therefore metadata export tools to bibliographic format and contextual research and navigation tools would be very useful. In an ideal world in a single repository it would be possible to find different versions of the same paper (pre-print, post-print and published version) all linked together.

Educational Repository

The structure of an educational repository should reflect the organization of the institution courses. This would allow a quick and easy use of the archive by the students. Every teacher should be able to submit and manage his own learning objects.

Andrea Bollini
Federico Meschini

Disciplinary Repositories

The content of a disciplinary repository belongs to a particular scientific sector, sometime even a very specialized one. This implies that particular metadata are used, together with indexes and authority files. The submission workflow process is therefore essential, also to avoid the submission of improper content. Usually submissions are made by authors "unkown" to the supporting institution(s) through independent registration and self-archiving procedurs.

Andrea Bollini
Federico Meschini

Thematic Digital Library

Thematic digital libraries have many issues in common with disciplinary repositories, in particular about access management and submission authorization.

A peculiarity of thematic digital libraries is the particular file format adopted such as images or video.

Therefore the digital library system should be able do manage these formats according to their characteristics, and often the simple download is not sufficient/convenient enough

Andrea Bollini
Federico Meschini

Is DSpace THE solution?

Can DSpace fit all these scenarios?

DSpace can satisfy most of these business case requirements with a good configuration

A complete fulfillment of these requirements is possible but it requires specific developments

Andrea Bollini
Federico Meschini

Functionalities: presentation

Web interface for the activities about submitting, browsing archive management

Out-of-box implementation XHTML 1.0 and WAI compliant

Basic branding features using the web interface: logo upload, customization of presentation texts and CSS

Multilingual – currently 19 languages

Web interface separated from the actual functionalities: starting from 1.5 JSPUI e XMLUI (Manakin) possibility of several presentations

Manakin allows to implement very complex customizations based on XSLT

Andrea Bollini
Federico Meschini

Functionalities: retrieval

- Powerful search-engine based on Lucene, with customizable index and full-text search
- Customizable browsing indexes
- RSS and email notification of new items
- Send to a friend and ask for a copy (patch)
- OpenURL support for accessing SFX based institution services
- SRW/U support (using a patch/add-on module)
- Persistent identifier for every, community and collection: handle -> starting from 1.6 support for other opaque identifiers such as AGR
- OAI-PMH: METS, MODS, DCQualified, MPEG-21 and possibility of creating customized formats

Andrea Bollini
Federico Meschini

Functionalities: metadata

- Management of multiple metadata schema (default DC Qualified)

- Possibility of using a particular metadata schema for each collection

- Possibility of defining the format for every metadata during the submission (name, date, restricted list, etc...)

- UTF-8

- Multilingual metadata support

Andrea Bollini
Federico Meschini

Functionalities: authorization

- LDAP support

- CAS support

- X.509 certificates support

- Grouping based on IP address

- Possibility of using more than a single authentication method at the same time

- Easy development of customized methods

Andrea Bollini
Federico Meschini

Functionalities: import/export

- Import from native XML format, XML+XSL, PDF, customized

- Export to XML with different possibilities, with all the formats used by OAI-PMH (METS, MODS, DCQualified, MPEG-21)
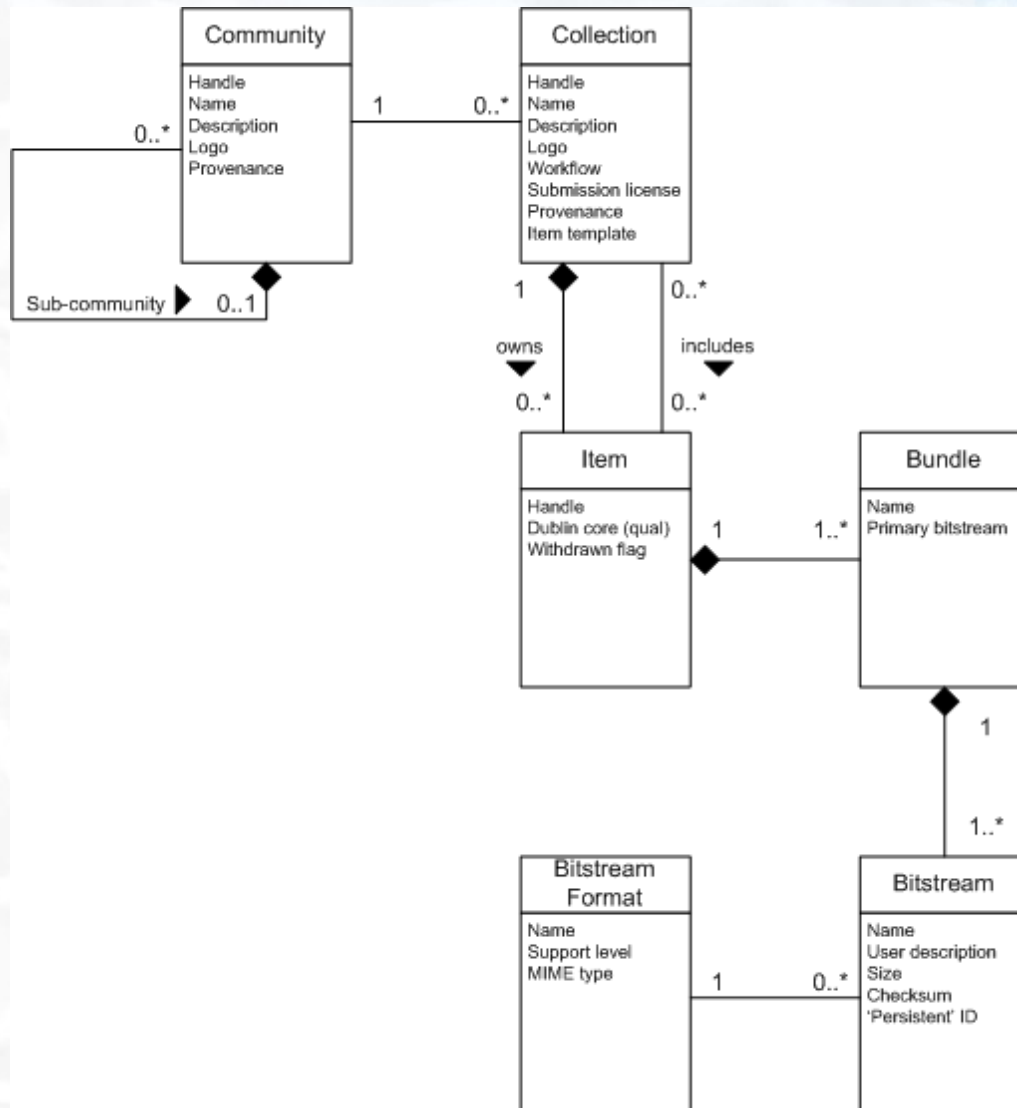
Andrea Bollini
Federico Meschini

DSpace: data models

There are five main entities which are the hub of information structure and aggregation in DSpace

- Communities: an administrative/logic groping of one or more collections (and sub-communities); eg. faculties and departements in a single university, centers for geographically distributed organizations; projects/research area, etc...

- Collections : a grouping of items which are analogous for tipology (metadata) and workflow; currently collections are the fulcrum of archive customizations

- Item: a box which contains both a document metadata and one or more bitstream bundles

- Bundle: a groping of bistreams used to separate the original documents, those obtained from automatic process, (such as full-text extraction), archival and Creative Commons licence

- Bitstream (= digital content, usually a fulltext)

Andrea Bollini
Federico Meschini

DSpace: data models

Andrea Bollini
Federico Meschini

DSUG Roma 2007

Communities

Communities and sub-communities are made up by:

- A set of collections and/or sub-communities

- A customizable home page

- Dedicated Feed RSS, Set OAI-PMH, search and browsing

- A group of users with administrative and managerial role (Community Admin Patch)

They are the logical organization for the collections

Andrea Bollini
Federico Meschini

Collections

Collections are made up by:

- A set of items owned by a particular collection

- Items which are mapped to other collections

- A group of user who are submission-enabled

- A group of user who can access to the full-text

- An approval/revision workflow for the submission in the archive

- Possible delegation of sharing, modification and submission authorizations (Proxy collection administrator)

- Customizable home page

- Dedicated Feed RSS, Set OAI-PMH, search and browsing

In practice a collection is a container of items

Every collection can have a different metadata set and a different workflow process

Andrea Bollini
Federico Meschini

Item

- Persistent Identifier (Handle System – CNRI)

- Metadata

- Sharing amongst collections

- Two groups of visible and not visible bundles (ORIGINAL, TEXT, CC-License, …)

- Full-Text (with authorization management)

- Creative Commons license

Andrea Bollini
Federico Meschini

DSpace: account management

Accounts in DSpace are managed through two entities: eperson and group. Eperson represents a single user, which his/her personal data and is used for accessing the system. A group is an account (or group) aggregator, used to simplify the authorization management (computer science students group, English teachers group, etc …)

Andrea Bollini
Federico Meschini

DSpace: DRM

- Authorization management is very flexible. It is possible to create reading, modification and creation policies for every entity in the data model. Currently webUI and OAI-PMH server don't support restrictions to metadata visibility in a native way.

- At the API level is also possible to define access policies valid only within a defined time range.

Andrea Bollini
Federico Meschini

Institutional Repositories - IR

Centralized authentication integrated with LDAP / X509

CAS / Shibboleth – using a patch

Access to contents based on IP / access network

Andrea Bollini
Federico Meschini

Research repositories

Tools for research support

- Suggest

- Comment (AddOn)

Tools for contextual search and browsing

-> Good metadata

Versioning

-> MIT: descriptive metadata approach

-> DSpace 1.6+: structural metadata approach

Andrea Bollini
Federico Meschini

Educational Repository

Teachers, direct management of their own collections – Collection Admin

Archive structure shaped after the course organization

Course = Collection

Andrea Bollini
Federico Meschini

Community	Faculty	Research Centre	Spin-off and Centre of Excellence
SubCommunity	Faculty of Humanities	English Department	Centre of Excellence in Performance Arts
SubCommunity (level 2)	BA English		
Collection	Course	Type #1..n Articles, conference papers, manuscript, etc.	Type #1..n Articles, conference papers, manuscript, etc.

Andrea Bollini
Federico Meschini

Disciplinary Repositories

Specific metadata -> it is possible to add new metadata schema

Authority files -> limited support for controlled lists and basic classification

Approval workflow

Self archiving

Andrea Bollini
Federico Meschini

Thematic Digital Libraries

XMLUI and JSPUI are the areas where most of the community development efforts are focused on

Base out-of-box support for the images

- Thumbnail

- Preview

What about Video? There is much interest about this in the community! Tomorrow go to see the presentation: **DSpace Integration with a Video Indexing and Summariz**

Andrea Bollini
Federico Meschini

Metadata & Schema Management

- In DSpace a metadata is characterized by
- Schema
- Element
- Qualifier
- Language
- Value
- schema and element.qualifier are bind in a metadata registry

Andrea Bollini
Federico Meschini

Metadata & Schema Management

- WebUI allows to create/record new schema in the registry or to modify the existing ones, adding new metadata types

- Metadata management screencast

- It is also possible to move a metadata from a schema to another (this allows a data reorganization in DSpace installations before version 1.4)

Andrea Bollini
Federico Meschini

Metadata & Schema Management

- New metadata can be used by the administrator to modify the description of already submitted data, using the webUI

- New metadata are automatically showed in the 'show full item record'

Andrea Bollini
Federico Meschini

Metadata & Schema Management

- It is possible to modify the 'show simple item record' using the file config/dspace.cfg

- In particular is possible to define:
  - The metadata set and its default display sorting
  - Different sets to use for one or more collections

Andrea Bollini
Federico Meschini

Metadata & Schema Management

```
webui.itemdisplay.default = dc.title, dc.title.alternative, dc.contributor.*, \
                            dc.subject, dc.date.issued(date), dc.publisher, \
                            dc.identifier.citation, dc.relation.ispartofseries, \
                            dc.description.abstract, dc.description, \
                            dc.identifier.govdoc, dc.identifier.uri(link), \
                            dc.identifier.isbn, dc.identifier.issn, \
                            dc.identifier.ismn, dc.identifier
```

(link) -> metadata content will be rendered as an hypertextual link

(date) -> metadata content will be rendered according to the standard used by the selected language (locale)

The metadata label in the 'show simple item record' will the the value of the key **i18n** `metadata.schema.element.qualifier` (e.g. `metadata.dc.title, metadata.dc.identifier.citation`)

Andrea Bollini
Federico Meschini

Metadata & Schema Management

```
webui.itemdisplay.<style>.collections = <collection handle>, ...
webui.itemdisplay.presentation.collections = 123456789/2, 123456789/..
webui.itemdisplay.tutorial.collections = 123456789/3


webui.itemdisplay.presentation = dc.contributor.*, dc.title, \
                        dc.relation.ispartof, dc.identifier.citation, \
                        dc.identifier.uri(link), my.custom.metadata, \
                        dc.description.abstract



webui.itemdisplay.tutorial = dc.title, dc.contributor.*, \
                        dc.relation.ispartof, dc.identifier.citation, \
                        dc.description.tableofcontents, my.custom.metadata, \
                        dc.identifier.uri(link)
```

Remember to add the related i18n keys in the
`Messages.properties` file, if necessary

Andrea Bollini
Federico Meschini

Metadata & Schema Management

`webui.itemdisplay.presentation`

Please use this identifier to cite or link to this item: `http://hdl.handle.net/123456789/4`

Authors:	Mornati, Susanna Bollini, Andrea
Title:	New opportunities for Institutional Repositories: the evaluation challange. A case study
Conference/Journal name:	DSUG Roma 2007
Citation:	Mornati S.; Bollini, A. - New opportunities for Institutional Repositories: the evaluation challange. A case study. Oral presentation at the DSUG Rome 2007
URI:	http://hdl.handle.net/123456789/4
Abstract:	The population of Institutional Repositories (IRs) is one of the most controversial issues within the OAI community. Ideological approaches, based exclusively on the authors. awareness that scholarly communication should become controlled by the scholarly community, have so far proved to be largely unsatisfactory. Relying on open-access deposit mandates from institutions, research funders, governments, looks not entirely appropriate if imposition is the only reason to deposit. To overcome the limits of the present communication system, present stakeholders that get advantages from the their dominant position (e.g. the big publishers) are replaced by stakeholders that get advantages from a reversed situation (e.g. the authors, from a wider circulation of knowledge). But this lies partly beyond the capabilities of any single institution. An easier strategy for change is from inside the system, accepting its rules but acting on its weak points. Change management analysis unfolds two key elements that we may affect: levers and constraints. Both elements are well known in the Open Access (OA) environment: levers have been extensively explored, such as author awareness, spiralling journal costs, etc., while constraints are harder to remove. A main constraint lies in the relation between research evaluation and commercial publishing. In many countries, including Italy, evaluation is based on the ISI Impact Factor (IF), biasing the communication system towards a restricted number of channels that are the most cited journals, mainly belonging to the commercial circuit. Even though in some disciplinary fields the IF cannot be calculated, in others it constitutes the only evaluation parameter. In order to substitute it with more reliable and article-based metrics, these have to be more deeply investigated, particularly as far as their relevance, appropriateness, and usability are concerned. IRs and the software packages on which they are based may become building blocks of a new infrastructure for metrics, a network for the production and analysis of data about usage, cross citations, and any other parameters and indicators that can contribute to assess the quality of a research work. The IR has to be the centre for institutional internal evaluation activities, and consequently the tool to collect evaluation data at the national level, guaranteeing both the governance and control on the institutional research activities and their output, and the capability to undergo external evaluation by exposing institutional records.

Appears in Collections: Presentation

Handle: 123456789/2

Files in This Item:

File	Description	Size	Format	
evaluation-challange.ppt		988 kB	Microsoft Powerpoint	View/Open

[Show full item record]

Andrea Bollini
Federico Meschini

Metadata & Schema Management
webui.itemdisplay.tutorial

Please use this identifier to cite or link to this item: `http://hdl.handle.net/123456789/5`

Title: Basic tutorial

Authors: Bollini, Andrea
Meschini, Federico

Title&Authors in reverse order

Conference/Journal name: DSUG Roma 2007

Citation: Bollini, A.; Meschini, F. - Basic Tutorial at DSUG Roma 2007

Table of Contents: 1. Introduction to DSpace System - Scenarios & business cases - Functionalities - Data model 2. Metadata & Schema Management 3. Submission system configuration (input-forms.xml, submission.xml) 4. Understanding the workflow system 5. Search & browsing system configuration 6. Introduction to development - Quick install & run - Integration with Eclipse - Basic layout modification (JSPUI) - Patch and module installation 7. CRON & utility scripts (batch import, etc.)

Custom Metadata: this is a default value that can be replaced by the submitter *custom metadata*

URI: http://hdl.handle.net/123456789/5

Handle: 123456789/3

Appears in Collections: tutorial

Files in This Item:

File	Description	Size	Format	
tutorial-basic.ppt	slides	1,24 MB	Microsoft Powerpoint	View/Open
create-administrator.swf	screencast - createadministrator account	2,39 MB	application/x-shockwave-flash	View/Open
create-administrator.htm	wrapper html for create administrator screencast	659 B	HTML	View/Open

Show full item record

This item is licensed under a Creative Commons License

SOME RIGHTS RESERVED

Items in DSpace are protected by copyright, with all rights reserved, unless otherwise indicated.

Andrea Bollini
Federico Meschini

Metadata & Schema Management
webui.itemdisplay.tutorial

This happens if you forget to define the metadata.shema.element.qualifier i18n key!

Please use this identifier to cite or link to this item: http://hdl.handle.net/123456789/5

Title:	Basic tutorial
Authors:	Bollini, Andrea Meschini, Federico
???metadata.dc.relation.ispartof???:	DSUG Roma 2007
Citation:	Bollini, A.; Meschini, F. - Basic Tutorial at DSUG Roma 2007
???metadata.dc.description.tableofcontents???:	1. Introduction to DSpace System - Scenarios & business cases - Functionalities - Data model 2. Metadata & Schema Management 3. Submission system configuration (input-forms.xml, submission.xml) 4. Understanding the workflow system 5. Search & browsing system configuration 6. Introduction to development - Quick install & run - Integration with Eclipse - Basic layout modification (JSPUI) - Patch and module installation 7. CRON & utility scripts (batch import, etc.)
???metadata.my.custom.metadata???:	this is a default value that can be replaced by the submitter
URI:	http://hdl.handle.net/123456789/5
Appears in Collections:	tutorial

Files in This Item:

File	Description	Size	Format	
tutorial-basic.ppt	slides	1,24 MB	Microsoft Powerpoint	View/Open
create-administrator.swf	screencast - createadministrator account	2,39 MB	application/x-shockwave-flash	View/Open
create-administrator.htm	wrapper html for create administrator screencast	659 B	HTML	View/Open

Show full item record

This item is licensed under a Creative Commons License

SOME RIGHTS RESERVED
creative commons

Items in DSpace are protected by copyright, with all rights reserved, unless otherwise indicated.

Andrea Bollini
Federico Meschini

Submission system configuration

- How does the submission process work in DSpace?

- Default submission process screencast

- Two main characteristics:
  - Submission workflow steps
  - Required metadata

> The submission process is influenced by the collection where it takes place

Andrea Bollini
Federico Meschini

Submission system configuration

- Required metadata: config/input-form.xml

- Three sections:
  - form-map
  - form-definition
  - form-value-pairs

Andrea Bollini
Federico Meschini

Submission system configuration: form mapping

- The form-map section `<form-map>…</form-map>`
- It contains nodes such as:

  `<name-map collection-handle="xxxxx/yyyy" form-name="myform" />`

- The collection-handle attribute has the form handle_prefix/collection_handle
- The form-name attribute refers to a node defined in the `<form-definition>…</form-definition>` section
- This section should contain at least the default mapping, which is defined by the `collection-handle="default"` attribute

e.g. `<name-map collection-handle="default" form-name="traditional" />`

Andrea Bollini
Federico Meschini

Submission system configuration

- The form-definitions section: <form-definitions>... </form-definitions>

- This section contains one or more <form name="XXX" /> elements which define what metadata have to be displayed and their order during the submission / workflow process

- The name attribute is the element unique identifier and it is used in the form-map section to associate it with one or more collections

- Each form is described by one or more <page /> elements, which allow to distribute the metadata on one or more pages

- The page elements contain the field elements, and each field element describes a single metadata

Andrea Bollini
Federico Meschini

Submission system configuration: metadata defininition

`<field>`

`<dc-schema>dc</dc-schema>`	The schema name, as defined in the registry
`<dc-element>contributor</dc-element>`	The metadata element
`<dc-qualifier>author</dc-qualifier>`	The metadata qualifier. When the element does not have a qualifier the element should be empty e.g. `<dc-qualifier></ dc-qualifier>` or `<dc-qualifier />`
`<repeatable>true</repeatable>`	Whether the element is repeatable or not
`<label>Authors</label>`	The label to be used during hte submission process
`<input-type>name</input-type>`	The data entry modality. More details later...
`<hint>Enter the names of the authors of this item below.</hint>`	The help text to be showed to the submitter
`<required></required>`	The error message to be showed when the metadata is not submitted. If this element is left empty the metadata would be not considered mandatory

`</field>`

Andrea Bollini
Federico Meschini

Submission system configuration: metadata defininition

- **Free Data Entry:**
    - onebox
    - twobox
    - textarea
- **Specific Data Entry**
    - series
    - name
    - date
- **Controlled Data Entry:**
    - dropdown
    - list
    - qualdrop_value

Andrea Bollini
Federico Meschini

- **onebox** – The most simple type. A single text-entry box.

Enter the main title of the item.

Title al Repositories: the evaluation challange. A case study

Andrea Bollini
Federico Meschini

- **twobox** -- A pair of simple text-entry boxes, used for *repeatable* values such as the DC subject item

Enter appropriate subject keywords or phrases below.

Subject Keywords

Institutional Reposi	DSpace
Remove	Remove
evaluation	change manageme
Remove	Remove
open access		Add More

- **twobox** -- A pair of simple text-entry boxes, used for *repeatable* values such as the DC subject item

Current Hack in JSPUI: When the metadata is relation.ispartofseries a specific hint is displayed for the two input, using **i18n** keys
`jsp.submit.edit-metadata.seriesname - jsp.submit.edit-metadata.paperno`

Enter the series and number assigned to this item by your community.

Series Name	Report or Paper No.	
Series/Report No. [ ]	[ ]	Add More

- **textarea** -- Large block of text that can be entered on multiple lines, e.g. for an abstract.

Enter the abstract of the item below.

Abstract

Submission system configuration: input-form.xml

`<input-type>`...`</input-type>` `<vocabulary />`

- When one of the three possible free data entry is used:
  - onebox
  - twobox
  - textarea

  It is possible to define a taxonomy, through the vocabulary element. The use of the taxonomy elements can be made mandatory using the closed attribute

  `<vocabulary [closed="false"]>vocabularyname</vocabulary>`

- The taxonomy definition is made in the vocabularyname.xml file, in the config/controlled-vocabularies folder

- The rendering mode uses javascript not compatible with WAI 2.0 standard, therefore the function is turned off by default. To turn it on

  `webui.controlledvocabulary.enable = true`

Andrea Bollini
Federico Meschini

Submission system configuration: input-form.xml
`<input-type>`…`</input-type>` `<vocabulary />`

The taxonomies are described using XML with this (very simple) structure:

```xml
<node id="acmccs98" label="ACMCCS98">
  <isComposedBy>
   <node id="A." label="General Literature">
    <isComposedBy>
     <node id="A.0" label="GENERAL"/>
     <node id="A.1" label="INTRODUCTORY AND SURVEY"/>

     …
    </isComposedBy>
   </node>

   …
  </isComposedBy>
</node>
```

Andrea Bollini
Federico Meschini

- **name** -- Personal name, with different fields for last name name and first name. The whole metadata will be recorded in the db as last name, first name

A specific hint is displayed for the two input fields, using the **i18n** keys
`jsp.submit.edit-metadata.lastname - jsp.submit.edit-metadata.firstname`

Enter the names of the authors of this item below.

	Last name e.g. **Smith**	First name(s) + "Jr" e.g. **Donald Jr**	
Authors	Mornati	Susanna	Remove This Entry
	Bollini	Andrea	
			Add More

Andrea Bollini
Federico Meschini

- **date** -- Calendar date. When the date is required at least the year should be present.

A specific label for the three input fields is displayed, using the **i18n** keys
```
jsp.submit.edit-metadata.year - jsp.submit.edit-metadata.month -
                    jsp.submit.edit-metadata.day

The month names use the java standard i18n  with java.text and the key
                  jsp.submit.edit-metadata.no_month
```

Please give the date of previous publication or public distribution below. You can leave out the day and/or month if they aren't applicable.
Date of Issue *Month:* (No Month) ⌄ *Day:* [   ] *Year:* [   ]

Andrea Bollini
Federico Meschini

- **series** -- A couple of simple text-entry boxes which allow the input of two strings which will be recorded as STRING1;STRING2

Beware: this function is not documented and could be changed or removed!

Andrea Bollini
Federico Meschini

DSUG Roma 2007

Submission system configuration: input-form.xml
`<input-type>...</input-type>`

- **dropdown** – value(s) can be selected from a "drop-down" menu list. **Note:** You must also include a value for the value-pairs-name attribute which specify a list of menu entries, which will be used for this item (the list is defined in the last section of input-form.xml: value-pairs). Use it to make a choice from a restricted set of options, such as for the language item.

Select the type(s) of content of the item. To select more than one value in the list, you may have to hold down the "CTRL" or "Shift" key.

Type
- Animation
- Article
- Book
- Book chapter
- Dataset
- Learning Object

Rendering when `<repeatable>true</repeatable>`

Select the language of the main content of the item. If the language does not appear in the list below, please select 'Other'. If the content does not really have a language (for example, if it is a dataset or an image) please select 'N/A'.

Language N/A

Rendering when `<repeatable>false</repeatable>`

Andrea Bollini
Federico Meschini

- **list** -- value(s) can be selected using a checkbox or radio button list. If the repeatable attribute is set to true, a list of checkboxes is displayed. If the repeatable attribute is set to false, a list of radio buttons is displayed. **Note:** You must also include a value for the value-pairs-name attribute which specify the list of values to be used (the list is defined in the last section of input-form.xml: value-pairs), from which to choose for this item.

Target audience
☐ option 1   ☐ option 4
☐ option 2   ☐ option 5
☐ option 3   Rendering when `<repeatable>true</repeatable>`

Has your article been peer-reviewed?

Peer-reviewed ○ yes
○ no   Rendering when `<repeatable>false</repeatable>`

Andrea Bollini
Federico Meschini

- **qualdrop_value** – It is possible to enter a "qualified value", which includes *both* a qualifier from a drop-down menu and a free-text value. This is used to enter items like alternate identifers and codes for a submitted item, e.g. the DC identifier field. **Note:** As in the case of the dropdown&list type, you must include the value-pairs-name attribute to specify a menu choice list.

If the item has any identification numbers or codes associated with it, please enter the types and the actual numbers or codes below.

Identifiers | ISSN ▾

Add More

Andrea Bollini
Federico Meschini

- This section contain elements such as:

```
<value-pairs value-pairs-name="common_identifiers" dc-
    term="identifier" />
```

- The `value-pairs-name` attribute is the list unique identifier and must correspond to the `value-pairs-name` attribute of the `input-type` elements for dropdown, lists and qualdrop_value. `fields` (metadata) have to use it to implement that particular list

- Every list item is defined by this kind of element:

```
<pair>
        <displayed-value>ISSN</displayed-value>
        <stored-value>issn</stored-value>
</pair>
```

Andrea Bollini
Federico Meschini

Submission system configuration: proof-of-concept

- Assuming that we have a particular collection about the DSUG tutorial and presentations, we could use the following configuration: input-form.xml

- `dc.type, dc.relation.ispartof` and `dc.language.iso` elements will be defined by default and based on the collection

- Custom submission screencast

Andrea Bollini
Federico Meschini

Submission system configuration: step definition

- Starting from version 1.5 it will be possible to change the order of the submission steps, delete them or add customized ones
- The configuration will be made using an xml file in the config folder
  - item-submission-JSPUI.xml
  - item-submission-XMLUI.xml
- Structured in three parts
  - submission-map
  - step-definitions
  - submission-definitions

Andrea Bollini
Federico Meschini

- Item-submission-[UI].xml is organized in three main sections
  - submission-map
  - step-definitions
  - submission-definitions

Andrea Bollini
Federico Meschini

- Similar to the `form-map` section
- It contains elements such as:

  `<name-map collection-handle="xxxxx/yyyy" submission-name="myform" />`

- The collection-handle attribute has the form handle_prefix/collection_handle
- The `submission-name` attribute refers to an element defined in the section

  `<submission-definition>`...`</submission-definition>`

- This section has to contain at least the default mapping defined by the attribute `collection-handle="default"`

  e.g. `<name-map collection-handle="default" submission-name="traditional" />`

Andrea Bollini
Federico Meschini

- Contains elements such as `<step id="step_name" />` `where the id` attribute is the step unique identifier
- A step is a single phase of the user submission process and can be composed by more than one page.
- This section must contain at least the definition of two fundamental steps:
  - `collection`
  - `complete`
- Collection is ALWAYS the first step and it links the items in the submission process to a particular collection – The default implementation is able to automatically detect the collection, when the process has been started form a collection home page using the "submit to this collection" function
- complete is ALWAYS the last step and it starts the workflow process
- The steps defined in this section can be shared by different collection processes

Andrea Bollini
Federico Meschini

- ## The definition of a single step is composed by 4 sub-elements

  - `processing-class` (REQUIRED): the JAVA class to use to process the step (need to `implement` `org.dspace.app.webui.submit.JSPStep` AND `extends` `org.dspace.submit.AbstractProcessingStep`)

  - `heading`: the i18n key to show in the navigation bar (if it is missing the step will not appear in the navigation bar)

  - `review-jsp`: the jsp fragment to use in the prospectic verification step

  - `workflow-editable` (default `true`): boolean which indicates whether a step is available in the workflow process

Submission system configuration: item-submission-[UI].xml
`<step-definitions>`...`</step-definitions>`

```xml
<step id="collection">
  <heading></heading> <!-- can specify heading, if you want it to
                          appear in Progress Bar -->
  <processing-class>
      org.dspace.app.webui.submit.step.JSPSelectCollectionStep
  </processing-class>
  <workflow-editable>false</workflow-editable>
</step>
```

These special steps are NEVER editable in workflow

```xml
<step id="complete">
  <heading>jsp.submit.progressbar.complete</heading>
  <processing-class>
      org.dspace.app.webui.submit.step.JSPCompleteStep
  </processing-class>
  <workflow-editable>false</workflow-editable>
</step>
```

Andrea Bollini
Federico Meschini

- Similar to `form-definition` section
- Contains elements such as:

    `<submission-process name="myprocess" />`

- The `name` attribute is the process unique identifiers and is used in the `submission-map` section to link collections to processes
- This section should contain at least the default process, usually named "default"

    `<submission-process name="default" />`

Andrea Bollini
Federico Meschini

- Every process is defined by a succession of steps which are defined locally (using the same rules of the step-definitions section) or it refers to some general steps using the id attribute

```xml
<!--Step 2 will be to Describe the item.-->
<step>
    <heading>jsp.submit.progressbar.describe</heading>
    <processing-class>
            org.dspace.app.webui.submit.step.JSPDescribeStep
    </processing-class>
    <review-jsp>/submit/review-metadata.jsp</review-jsp>
    <workflow-editable>true</workflow-editable>
</step>
<step id="general-available-step"/>
```

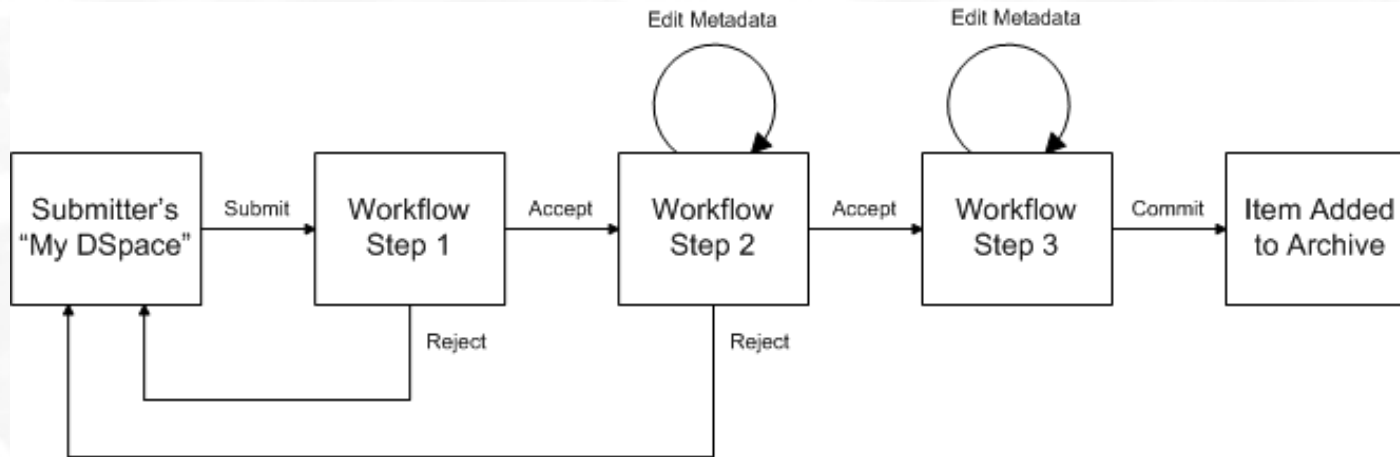- The step order in the xml file determines their execution order in the submission process

Andrea Bollini
Federico Meschini

Understanding the workflow system

- The workflow system currently implemented is very simple. During the Architecture Review phase (DSpace 2.0) it has been decided to expand it using an external framework

- It is composed by 3 steps at most, which are executed in a sequence. For every collection is it possible to decide which step(s) should be activated

Andrea Bollini
Federico Meschini

Understanding the workflow system

Edit Metadata

Edit Metadata

| Submitter's "My DSpace" | Submit → | Workflow Step 1 | Accept → | Workflow Step 2 | Accept → | Workflow Step 3 | Commit → | Item Added to Archive |

Reject

Reject

- **step1** - accept submission for inclusion, or reject submission.
- **step2** - can edit metadata provided by the user, but cannot change the submitted files. Can accept submission for inclusion, or reject submission.
- **step3** - can edit metadata provided by the user, but cannot change the submitted files. Must then commit to archive; may not reject submission.

Andrea Bollini
Federico Meschini

Understanding the workflow system

- Every collection can organize its own workflow independently from the other collections

- Every step is associated to a reviewer group: standard DSpace group which could be composed by one or more eperson and/or sub-groups

- The system administrator can change the group composition using the web interface

Andrea Bollini
Federico Meschini

Understanding the workflow system

- If your archive is organized following the structure departments = community and type of materials = collection, probably you would always have the same revision group for a step (step1 department director, step3 librarians). It is useful to create a specific group and insert it as a subgroup in the workflow groups of every collection. In this way it would be possible to change this subgroup only once.

Andrea Bollini
Federico Meschini

DSUG Roma 2007

Understanding the workflow system

KNOW ISSUES: BE CAREFUL TO EDIT WORKFLOW OF LIVE COLLECTIONS!

- Changes in one step will take place only for the submissions which have still to reach it and not for the ones which are already there!
- If a reviewer is added he will NOT see the submissions which are already in his assigned step
- If a reviewer is removed he WILL still see the current submissions, until they are moved in a preceding or following step
- If a reviewer is removed, he would still able to accept/reject the submissions in progress, unless it is the final step. In this case he will need to have an ADD collection permission to approve/reject the submission
- The system administrator can reject all the submissions, sending them back in the user workspaces, solving therefore possible problems

Andrea Bollini
Federico Meschini

Supervision and Collaboration

- A supervision order system exists to bind groups of other users (thesis supervisors) to an item in someone's pre-submission workspace.
- The bound group can have system policies associated with it that allow different levels of interaction with the student's item; a small set of default policy groups are provided with:
    - Full editorial control (allows edit metadata & add new files)
    - View item contents
    - No policies
- Once the default set has been applied, a system administrator may modify them as they would any other policy set in DSpace
- This functionality could also be used in situations where researchers wish to collaborate on a particular submission, although there is no particular collaborative workspace functionality.

Andrea Bollini
Federico Meschini

Search & browsing system

- DSpace utilizes Lucene as a search engine
- Indexes are kept in customizable filesystem directory (usually `[dspace]/search`)
- The browsing system instead uses a specific logic of its own(package org.dspace.browse)
- The information are recorded in the database
- In version 1.5 the performance have been really improved, together with the possibilities of configuration

Andrea Bollini
Federico Meschini

Search configuration

Working on the DSpace configuration file config/dspace.cfg it is possible

- How many and which indexes should be made available

- Which metadata should be included in every index

- The default boolean operator to apply when there are more than one keywords in the same input text

- How to normalize and extract the keywords putting them in an index

Andrea Bollini
Federico Meschini

Search configuration: index definition

- Indexes are defined in dspace.cfg using keys with the form

`search.index.<n> = indexname:schema.element.qualifier`

- <n> is a progressive number which is only used to used properly process the configuration file

- It is possible to use more than one metadata for the same index, using the same index name for different search.index (1 and 2 for example)

- It is also possible to insert all the element qualifiers using the jolly character *

- Basic search utilizes an index called default, which contains all the other indexes + all the extracted full-texts + all the collection and community names

Andrea Bollini
Federico Meschini

Search configuration: index definition default configuration

```
search.index.1 = author:dc.contributor.*
search.index.2 = author:dc.creator.*
search.index.3 = title:dc.title.*
search.index.4 = keyword:dc.subject.*
search.index.5 = abstract:dc.description.abstract
search.index.6 = author:dc.description.statementofresponsibility
search.index.7 = series:dc.relation.ispartofseries
search.index.8 = abstract:dc.description.tableofcontents
search.index.9 = mime:dc.format.mimetype
search.index.10 = sponsor:dc.description.sponsorship
search.index.11 = identifier:dc.identifier.*
search.index.12 = language:dc.language.iso
```

Currently the web interface is not automatically generated from this configuration

Therefore it will be necessary a manual change

Andrea Bollini
Federico Meschini

Search configuration

- The default boolean operator used when there are more keywords in the same input text (advanced search) or in the basic search form is `OR` (more results), but it can be set up using the key

```
search.operator = OR
```

Andrea Bollini
Federico Meschini

Search configuration

- The normalization and extraction mode used to insert the metadata in an index is defined by the key `search.analyzer` and it must have the full name of the java class which will be used

  e.g. `search.analyzer = org.dspace.search.DSAnalyzer`

  Where the defined class MUST

  `extends org.apache.lucene.analysis.Analyzer`

- The default implementation utilizes the following lucene filters:

StandardFilter	Split-up of the single tokens using <space> comma etc.
LowerCaseFilter	Case insensitive behaviour
StopFilter	Stopwords filter: a, am, and, are, as, at, be, but, by, for, if, in, into, is, it, no, not, of, on, or, the, to, was
PorterStemFilter	Removal of the common English end words (e.g. s for plural nouns)

Andrea Bollini
Federico Meschini

Browse configuration

Working on the DSpace configuration file config/dspace.cfg it will be possible to define

- The indexes which should be available

- Which metadata should be included in an index

- The sorting mode of every index

- The display mode of the results and the inter-navigation mode between the item page and the browsing system.

Andrea Bollini
Federico Meschini

Browse configuration: index definition

Is it possible to define two index types:

- Full indexes, which show an item list (e.g. title, date) – full indexes must use the available metadata also for sorting option
- Single value indexes, which show a metadata list (e.g. author, subject)

```
webui.browse.index.<n> = indexname:item:sortname
```

Full index

```
webui.browse.sort-option.<n> = sortname:schema.element[.qualifier|.*]:datatype
```

Sorting option

Single Value index

```
webui.browse.index.<n> = indexname:metadata:schema.element[.qualifier|.*]:datatype
```

The datatype determines the index rendering mode and the sorting algorithm to be used

Andrea Bollini
Federico Meschini

Browse configuration: index definition - data type

The following datatypes are defined by default:

`title` – this metadata is used to link the item page

`date` – this metadata is a date and will use a chronological sorting

`text` – this metadata will use by default an alphabetical sorting

It is also possible to define new datatype which will be treated as text for rendering issues, but which could redefine the sorting alghoritm using the key

```
plugin.named.org.dspace.browse.BrowseOrderDelegate= \
org.dspace.browse.BrowseOrderTitleMarc21=title
```

Andrea Bollini
Federico Meschini

Browse configuration: display behaviour

- The definition of the metadata to be used to inter-navigate is defined using the keys of the type

```
webui.browse.link.<n> = author:dc.contributor.*
```

- The definition of the metadata to be used in the summary tables (browsing or research results) is defined using the key

```
webui.itemlist.columns = dc.date.issued(date), \
    dc.title, dc.contributor.*
```

Screencast about the browsing functions

Many other options are available and is possible to discover the exploring the dspace.cfg file and reading the comments ☺

Andrea Bollini
Federico Meschini

Search & Browse System: conclusion

- To Remember! Any change made in the research indexes has to be manually done also in the web interface

- Any change in the search configuration implies the recreation of the indexes (you can use the script

  `[dspace]/bin/index-all`

- Any change to the browsing configuration can imply the creation/update of db tables, and it is possible to use the command

  `[dspace]/bin/dsrun org.dspace.browse.IndexBrowse –f –r`

Andrea Bollini
Federico Meschini

Introduction to Development

- DSpace is a complex application made up by several modules

- API

- Web Interface (Two: one based on Servlet/JSP – the other on Coocon/XML/XSL)

- Command line scripts

- WebServices

- ... put your module here ...

Andrea Bollini
Federico Meschini

Introduction to Development: requirements

- DBRMs: PostgreSQL 7.3+ or Oracle 9+
- JDK5
- Ant 1.7
- Maven 2
- Servlet Container like Tomcat

Introduction to Development

ATTENTION! Currently the new Dspace Maven2 build process is still a work-in-progress so check that this info are updated

- DSpace source files can be downloaded from sourceforge.net as a zipped archive (eg .tar.gz) or directly from the SVN repository. In the next slide we will refer to the unzipped or checked out directory using [dspace-source]

- The DSpace installation process alwasy produces at least the command line interface (only maintenance scripts not for browsing or other uses ☺ !) which detects the installation directory [dspace]

- The web modules (as JSPUI) create a subdirectory into [dspace]/webapp which has to be copied in the proper servelet container directory (for example for tomcat this usually is the [tomcat-home]/webapps directory)

Andrea Bollini
Federico Meschini

Introduction to Development

- Starting from version 1.5 DSpace will use Maven2 as project management tool and this will also allow a physical separation of the several DSpace "modules"
- The source directory [dspace-source] will contain the following subdirectories by defaul

dspace	There are the configuration file templates and the shell scripts. There is also the main pom.xml file with the instruction to maven2 for creating the binary distribution
dspace-pom	There are the maven2 configuration files for the dspace modules
language-packs	There are the i18n property files to be used for the interface translation
dspace-api	There are the java classes which implement and allow the access to the data model, together with the common business logic
dspace-jspui	Web application which implements the user interface and is based on Servlet/JSP
dspace-xmlui	Integration of Manakin project: web application which implements the user interface based on XML/XSL
dspace-lni	WebServices for the remote integration of DSpace with other applications: WebDav standard

Andrea Bollini
Federico Meschini

Introduction to Development

USE:

- Eclipse (with J2EE plugins - m2eclipse – subclipse) http://www.eclipse.org
- IDEA (the outcoming version 7 appears to be the simplest one to integrate with DSpace) http://www.jetbrains.com/idea/
- or your favourite IDE (and contribuite with an how-to for integration on the wiki!)
- Firefox Web Developer plugin
- https://addons.mozilla.org/it/firefox/addon/60

BROWSE (from "DSpace How-To Guide – Other useful resources"):

- W3Schools Tutorials/Guides (http://www.w3schools.com/) – introductory tutorials, references, and examples for XHTML, CSS, XML, XSLT, SQL, among many others.
- HTML Dog (http://www.htmldog.com/) – HTML and CSS References and tutorials for all levels of knowledge
- Holy CSS Zeldman! (http://www.dezwozhere.com/links.html) – all things CSS, with intermixed links to good HTML, Javascript, and web design sites.
- MaxDesign (http://css.maxdesign.com.au/) - CSS tutorials for web developers.
- Covers customizing page layout using CSS-based floats, lists and selectors.

Andrea Bollini
Federico Meschini

Introduction to Development

- Before modifying DSpace code or start developing a new feature, it is useful to check what is already available and what "is being cooked"
- Sourceforge.net http://sourceforge.net/projects/dspace
  - feature request http://sourceforge.net/tracker/?group_id=19984&atid=369984
  - patch code http://sourceforge.net/tracker/?group_id=19984&atid=319984
- DSpace How-To Guide (by Tim Donohue, Scott Phillips, Dorothea Salo)
- DSpace Wiki http://wiki.dspace.org
- DSpace documentation - (which is also available in the folder [dspace]/docs)
- Mailing lists http://sourceforge.net/search/?group_id=19984&type_of_search=mlists
  - dspace-devel (new developments)
  - dspace-tech (issues about a released version)

Andrea Bollini
Federico Meschini

DSUG Roma 2007

- DSpace uses an i18n property file for all the interface messages so if you want to modify some text you need to edit the file:

  `[dspace-source]/dspace-api/src/main/resource/Messages.properties`

  or your specific locale which is in

  `[dspace-source]/language-packs/yourlocale/Messages_ISO.properties`

- From version 1.5 DSpace has a built-in support for multilingual interface which can be configured from dspace.cfg using the following option

  ```
  default.locale = en
  webui.supported.locales = en, de
  ```

Andrea Bollini
Federico Meschini

DSUG Roma 2007

Introduction to Development: simple layout modification

Files:

- [dspace-source]/dspace-webui/src/main/webapp/jsp/layout/*.jsp
- [dspace-source]/dspace-webui/src/main/webapp/jsp/styles.css.jsp

Andrea Bollini
Federico Meschini

from "DSpace How-To Guide - Change overall layout"

DSUG Roma 2007

Introduction to Development: cmd line script

- All the scripts are in the [dspace]/bin directory
- For most of the java scripts there is a bash wrapper which allows a quick running (index-all, create-administrator, etc.)
- There is also script, both in bash (linux) and a .bat (windows) versions, which simplifies the execution of the java scripts setting the proper VM options (memory and classpath): `dsrun`

  `dsrun org.dspace.administrer.CommunityFiliator`

- It is also possible to use the --help option to have more information about every script!
- Eclipse configuration for executing java "scripts" (screencast create-admin)

Andrea Bollini
Federico Meschini

Introduction to Development: cmd line script

index-all	org.dspace.search.DSIndexer org.dspace.browse.IndexBrowse	(Re)creation of search and browsing indexes (options–r –f to fully recreate the browsing table)
create-administrator	org.dspace.administer.CreateAdministrator	Creation of administrator(s)
structure-builder	org.dspace.administer.StructBuilder	Batch creation of communities and collections
import	org.dspace.app.itemimport.ItemImport	Item and bitstreams import using specific folders and files described in an xml file
packager	org.dspace.app.packager.Packager	A very flexible import script, which allows to import items inside .pdf and .zip files

Andrea Bollini
Federico Meschini

Introduction to Development: CRON Job

filter-media	org.dspace.app.mediafilter.MediaFilterManager	Format conversion for full-text extraction, thumbnail creation and preview It is configured using the Plug In Manager
generate-sitemaps	org.dspace.app.sitemap.GenerateSitemaps	Google sitemap creation
sub-daily	org.dspace.eperson.Subscribe	Invio delle mail di notifica
checker	org.dspace.app.checker.ChecksumChecker	Check of assetstore integrity
stat-general stat-monthly stat-report-general stat-report-monthly	Scripts Perl	DSpace log parsing and processing of the use statistics (mostly administrative information)

Andrea Bollini
Federico Meschini

How to solve a problem with DSpace

- Check the content of dspace log file: usually [dspace]/log/dspace.log

- Set [dspace]/config/log4j.properties on an higher level (like WARN or DEBUG) and try again to see more details about the error

- Search the dspace-tech mailing list for similar problems http://sourceforge.net/mailarchive/forum.php?forum_name=dspace-te

- If you don't find a solution send an email to the list dspace-tech@lists.sourceforge.net including the full error java stack trace and a description of the operation

Andrea Bollini
Federico Meschini