# PopulateMetadataFromPubMed

NB: This page is currently under construction

## Populating Metadata in Submission with data from PubMed

Note that currently, this will focus on implementation within the JSP interface. Most of the code is reusable within the XMLUI, however I haven't attempted that implementation yet.

Note also that this is based on making the simplest set of changes to the standard DSpace 1.5 release package. It does not necessarily claim to be a best practise for implementation - particularly with regards to organisation of the Java code - although this may evolve over time.

Finally, note that the implementation is not currently fully tested - the population of the metadata works, but I haven't looked at how the review and workflow processes are affected by this yet.

## Caveats

In order to make the simple set of changes prescribed here in your JSP module, you will need to do a couple of things.

1) In dspace/modules/jspui/pom.xml, you will need to add a compile dependency on servlet-api:

```
        <dependency>
            <groupId>javax.servlet</groupId>
            <artifactId>servlet-api</artifactId>
            <scope>compile</scope>
        </dependency>
```

2) When adding Java code to /dspace/modules/jspui/src/main/java, you will need to create the directory dspace/modules/jspui/src/test/java, otherwise the test plugin will fail (or, you could specify -Dmaven.test.skip=true when running mvn package).

## How To Implement - Java and Crosswalk

Extract the dspace-1.5-release.zip package (downloaded from SourceForge - this will create dspace directory, with the main assembly information, and all the interfaces and applications (jspui, xmlui, lni, sword) inside the dspace/modules directory.

Inside the jpsui module, create a main java source directory - ie. dspace/modules/jspui/src/main/java (you will also need the test directory described above). Within here, create the directories my/dspace/submit/step and my/dspace/app/webui/submit/step.

### dspace/modules/jspui/src/main/java/my/dspace/submit/step/PubmedPrefillStep.java

In dspace/modules/jspui/src/main/java/my/dspace/submit/step, create the file PubmedPrefillStep.java, with the contents from the link below. This contains most of the code for implementing the lookup of metadata in PubMed, and can be reused by an XMLUI implementation.

PubmedPrefillStep.java

### dspace/modules/jspui/src/main/java/my/dspace/app/webui/submit/step/JSPPubmedPrefillStep.java

In dspace/modules/jspui/src/main/java/my/dspace/app/webui/submit/step, create the file JSPPubmedPrefillStep.java, and place the contents below - note that this is almost exactly the JSPSampleStep.java file from DSpace, with minor changes to the DISPLAY_JSP and REVIEW_JSP constants, and an additional call to the JSPStepManager inside the "if (status h1. SampleStep.STATUS_USER_INPUT_ERROR)" test.

This is the code that is specific to JSPUI, and would not be needed for an XMLUI implementation.

JSPPubmedPrefillStep.java

### dspace/config/crosswalks/pmid_dim.xsl

In dspace/config/crosswalks, create the file pmid_dim.xsl. This is the transformtion that will convert the PubMed XML into DIM that can be ingested by DSpace, and should be as below. Note that if you want to customise how PubMed data is mapped into your metadata fields, this transformation file is where you make those changes.

pmid_dim.xsl

## dspace/config/dspace.cfg

You will need to tell the XSLT crosswalk about the new transformation stylesheet for ingesting PubMed data. To do this, edit dspace/config/dspace.cfg, and insert these lines:

```
#Configure XSLT-driven submission crosswalk for PMID
crosswalk.submission.PMID.stylesheet = crosswalks/pmid_dim.xsl
```

Implementing the JSP interface h1. For the submission form, you will need a new file to describe the form to capture the PubMed ID. In dspace/modules /jspui/main/webapp, create submit/pubmed-step.jsp, with the contents of the following link.

pubmed-step.jsp

You will also need to add the required messages - in dspace/modules/jspui/src/resources/Messages.properties, add the following:

```
jsp.submit.progressbar.pubmed-prefill = Prefill
jsp.submit.pubmed-prefill.title = Prefill Metadata
jsp.submit.pubmed-prefill.heading = Submit: Describe your item
jsp.submit.pubmed-prefill.elem1 = There are two methods of submitting items to this repository: you can either
enter the item's descriptive information (metadata) manually or pre-populate some of the fields using a PubMed
ID. You will then be able to add a file or files.  You will be able to review and edit your submission before
it is archived. You will also be required to accept a standard licence agreement.
jsp.submit.pubmed-prefill.elem2 = If you have a PubMed ID you can enter it below. The information available in
PubMed will be used to pre-populate the submission form. You will still be able to update any of these fields
or add additional metadata.
jsp.submit.pubmed-prefill.elem3 = To manually enter the item metadata click the 'Next&nbsp;&gt;' button below
to go straight to the submission form.
jsp.submit.pubmed-prefill.elem4 = NOTE: Some publishers have certain conditions about you self-archiving work
they have already published. You can look up their policies on <a href="http://www.sherpa.ac.uk/romeo.php"
target="blank">SHERPA's Romeo database</a>.
jsp.submit.pubmed-prefill.error.identifier = The identifier entered is not recognized
jsp.submit.pubmed-prefill.info.identifier = We are unable locate the published article. Please check the ID or
continue the submission manually.
jsp.submit.pubmed-prefill.pubmid.label = PubMed ID
jsp.submit.pubmed-prefill.forcerefresh.label = Refresh item with information retrieved via the above IDs (note:
this will erase and replace all existing information)
```

Final Configuration h1. h2. Configurable Submission

Now that the step has been created, with it's appropriate forms and text messages, you can replace the standard initial questions step, with the Pubmed population one. To do so, edit dspace/config/item-submission.xml, and replace the code:

```
          <step>
            <heading>submit.progressbar.initial-questions</heading>
                <processing-class>org.dspace.submit.step.InitialQuestionsStep</processing-class>
                <jspui-binding>org.dspace.app.webui.submit.step.JSPInitialQuestionsStep</jspui-
binding>
                <xmlui-binding>org.dspace.app.xmlui.aspect.submission.submit.InitialQuestionsStep</xmlui-
binding>
            <workflow-editable>true</workflow-editable>
          </step>
```

with

```
          <step>
            <heading>submit.progressbar.pubmed-prefill</heading>
                <processing-class>my.dspace.submit.step.PubmedPrefillStep</processing-class>
                <jspui-binding>my.dspace.app.webui.submit.step.JSPPubmedPrefillStep</jspui-
binding>
                <xmlui-binding>my.dspace.app.xmlui.aspect.submission.submit.PubmedPrefillStep</xmlui-binding>
            <workflow-editable>false</workflow-editable>
          </step>
```

## Displaying the identifiers

We are going to add additional identifier fields to the DSpace metadata schema - for these to show up correctly in the submission interface, we need to adjust the input-forms config. Open dspace/config/input-forms.xml, and in the 'form-value-pairs' section, inside the value pairs for the common_identifiers, add the following configuration:

```

     <pair>
         <displayed-value>PubMed ID</displayed-value>
         <stored-value>pmid</stored-value>
     </pair>
     <pair>
         <displayed-value>DOI</displayed-value>
         <stored-value>doi</stored-value>
     </pair>
```

Installation ==

Now you can build and install DSpace as you normally would - from the 'dspace' directory, run

```
mvn package
```

, and then in the target directory, either

```
ant fresh_install
```

or

```
ant update
```

(note that if you are updating an existing installation, you will need to

```
ant init_configs
```

to install the new configuration elements).

## DC metadata changes

Start Tomcat with the newly built application. But before attempting to use the new submission step, you need to add the two new identifiers to the Dublin Core schema. Go to the Administrator interface, and from metadata registry, select the Dublin Core metadata.

Now add a metadata element 'identifier' with the qualifier 'pmid', and another element 'identifier' with the qualifier 'doi'.

## Prefilling metatdata

Congratulations - you should now be able to use the PubMed prefill step. As an authorised user, start a new submission process. Once you have identified the collection that you wish to submit to, instead of being presented with the normal initial questions, you will see a page requesting a PubMed ID. You can skip this step simply by leaving the form blank, but if you insert a PubMed ID (ie. 17381832) and then click next, you will find that the next page is already filled out with the title, authors, etc. - all retrieved from PubMed for the ID that you specified.