National Digital Newspaper Program Digital Viewer and Validator User's Guide

Library of Congress: Office of Strategic Initiatives

Prepared by: Curt Harvey David Brunton

Date of Publication: February 1, 2010

| 1. | Introduction | 4 |
|----|--|----|
| 2. | System Requirements | 5 |
| 3. | Installation | 5 |
| | Download | 6 |
| | Save | 7 |
| | Install | 8 |
| | Choose Destination Folder | |
| | Installation Errors | |
| 4. | Post Install Configuration | |
| | Changing XSLT Files | |
| | Changing Icol I mages | |
| | Default Images | |
| 5 | Starting the DVV | |
| 5. | Open Command Window | |
| | The Command Window | |
| 6 | Batch Processing | |
| 0. | Opening a Batch | |
| | Batch View | |
| | Processing a Batch | |
| | | |
| | Error Reporting The Errors | |
| | | |
| | Clearing Informational Messages | |
| | Validation Progress | |
| | Validation Complete | |
| | LCCN Validation | |
| - | Exemptions | |
| 7. | Viewer Controls | |
| | Issue/Reel Summary | |
| | Load Progress | |
| | Loading Images | |
| | Thumbnail Images | |
| | Sample Info Dialog | |
| | Error Panel | |
| | Viewing Metadata | |
| | Viewing Image Header Data | |
| | Example Header Data | |
| | Viewing Title Essay Data | |
| 8. | Browse Mode | |
| | File Dialog | |
| | The File Tree | 54 |
| | Browsing Images | 55 |
| | Image Loaded | 56 |
| | OCR Data Viewing | |
| | PDF Viewer | |
| 9. | Command Line Usage of the Validation Processor | 59 |
| | Open Command Window | |
| | The Command Window | |
| | Validation Processor Options | 63 |

| PDF filename | 55 |
|---|----|
| TIFF filename | 55 |
| ALTO filename | 55 |
| Validation Examples | 66 |
| BATCH filename | 56 |
| Error Messages | 57 |
| BATCHFILE filename | 57 |
| ISSUEMETSRECORD filename 6 | 59 |
| Validation Signatures7 | 0' |
| File Names | 0' |
| Revalidation7 | 1 |
| Modification of a Validated Resource7 | '2 |
| Capturing Validation Processor Messages7 | '3 |
| Logging Levels | |
| 10. Supplied logging.properties | '5 |
| 11. Changing DOS Window Properties7 | |
| 12. JPEG2000 Plug-ins | 31 |
| 13. Benchmarks | 32 |
| System A | 32 |
| Batch Composition | 32 |
| Batch Processing Statistics | 33 |
| Individual Image Load Times (All load times in seconds) | 33 |
| System A | 33 |
| Load Time | 33 |
| Average Load Time | 33 |
| Average Image Size | 33 |
| Appendix A | 34 |
| Third Party Licensing | 34 |
| JHove | |
| Sun Microsystems Java Runtime Environment 8 | 34 |

1. Introduction

The National Digital Newspaper Program is a partnership between the Library of Congress and the National Endowment for the Humanities. The goal of the project is to create a freely accessible, searchable digital repository of historically significant newspapers published between 1836 and 1922. An overview of the program can be found at <u>http://www.neh.gov/projects/ndnp.html</u>.

One of the guidelines offered to originators of newspaper content (awardees) regarding the validation process is "Validate early, validate once". Awardees are responsible for delivering batches of newspaper issues derived from reels of microfilm. Although there is a non-interactive facility whereby large batches may, and should, be validated, the Digital Viewer and Validator (DVV) offers a visual means to validate any or all issues/reels within a batch. Another, key, aspect of the DVV is its viewing capabilities.

Each issue/reel is composed of three image files (TIFF, JPEG2000, PDF) and an XML file containing the OCR rendition of the microfilmed page. The DVV is capable of displaying each of the three types of images as well as offering viewing capability for the OCR and metadata files. All errors encountered during validation are logged and displayed in the Error Report tab.

2. System Requirements

Operating System – The DVV has been configured to run on Windows.

Memory - The validation of digital images, often many megabytes in size, is a resource intensive operation. A minimum of 512 megabytes of RAM for the Java heap is recommended and the default installation assumes this amount exists.

Disk Storage – A minimum of 176 megabytes of disk space is required to install the DVV. 37 megabytes is released during post-install removal of resources no longer needed, leaving a footprint of about 1395 megabytes. XML and HTML files are generated during normal processing and will add to the space requirement.

Processing power – As with any resource intensive operation, more is better. The DVV has been tested on systems as slow as 500 mhz, with a Java memory specification of 256M for maximum heap size. 64 files in 1 reel with 2 issues validated in just over 2 minutes on this system.

Java – The latest version of the Java Development Kit was used in the creation of the DVV. The setup script installs the proper JRE/JDK, 1.6 as of this release, for the exclusive use of the DVV. Existing instances of the JDK/JRE on a target system are not affected.

Adobe Acrobat Reader 7.0 – This version (or later) of Acrobat Reader must be installed in order to view PDF files. It is free and can be downloaded from <u>http://www.adobe.com</u>.

3. Installation

Installing the DVV requires the following steps.

Download the installation file, dvvInstall_1.4.4_3.7.exe from:

http://www.loc.gov/ndnp/tools/

and save it on your local hard drive. Screenshots of the process are shown below.

Download

When the following web page is displayed, click on dvvInstall_1.4.4_3.7.exe (not shown here) to begin the download.

| 🕹 I | ndex of /ndnp/tools - Mozilla Firefo | ж | | | | | |
|--------------|--|------------------------------|-----------------|-------------|-----|---------------------------|-----|
| <u>F</u> ile | <u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks | <u>T</u> ools <u>H</u> elp | | N | | | |
| | 💽 - C 🗙 🏠 🔳 | http://www.loc.gov/ndnp/tool | sl | hý: | ☆ · | ⊘! • Yahoo! Search | P |
| Iı | ndex of /ndnp/† | tools | | | | | |
| _ | Name | Last modified | Size | Description | | | |
| | Parent Directory dvvInstall 1.4.1 3.1> dvvInstall 1.4.2 3.5> | | - 56M 58M | | | | |
| Don | e | | | | | | //, |

Save

Click 'Save File'.

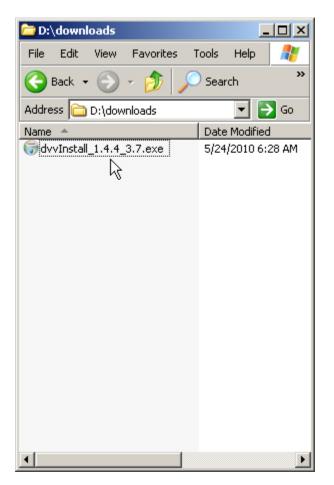
| Opening dvvInstall_1.4.4_3.7.exe | × |
|--|---|
| You have chosen to open | |
| dvvInstall_1.4.4_3.7.exe | |
| which is a: Binary File | |
| from: Y:\lc_rep\NdnpPreingest-Interface\target\install | |
| Would you like to save this file? | |
| Save File Cancel | |

Choose a place to save the install file. In this example, dvvInstall_1.4.4_3.7.exe will be saved as d:\downloads\dvvInstall_1.4.4_3.7.exe.

| Enter name of file | e to save to | | | | <u>? ×</u> |
|-------------------------|---------------|-----------------------|-----|----------|------------|
| Save in: | 🗀 downloads | | • | G 🖻 🖻 🖽 | - |
| My Recent Documents | | | | | |
| Desktop My Documents | | | | | |
| My Computer | | | | | |
| My Network Places | File name: | dvvInstall_1.4.4_3.7. | exe | • | Save |
| | Save as type: | Binary File | | • | Cancel |

Install

Using Windows Explorer or the File Manager facility of My Computer, navigate to the directory in which you saved the install file and double click on dvvInstall_1.4.3_3.6.exe to begin the installation process. The screen shot below shows the saved dvvInstall_1.4.3_3.6.exe, highlighted, at d:\downloads



A security dialog may appear. If so, click Run.

| Open File | - Security \ | Warning | × |
|-----------|-----------------------------|---|---|
| | ıblisher cou s software? | ld not be verified. Are you sure you want to | |
| | Name: | dvvInstall_1.4.4_3.7.exe | |
| | Publisher: | Unknown Publisher | |
| | Type: | Application | |
| | From: | D:\downloads | |
| 🔽 Alwa | ays ask before | e opening this file |] |
| 8 | publisher. Y | is not have a valid digital signature that verifies its You should only run software from publishers you trust. <u>Jecide what software to run?</u> | |

To begin the installation of the DVV double click on the install file name. The following install panel will appear:

| 😽 The Library of Congress - Digital Viewer and Validator (1.4.4_3.7) Setup | - 🗆 🗙 |
|---|-------|
| Choose Install Location Choose the folder in which to install The Library of Congress - Digital Viewer and Validator (1.4.4_3.7). | |
| Setup will install The Library of Congress - Digital Viewer and Validator (1.4.4_3.7) in th following folder. To install in a different folder, click Browse and select another folder. install to start the installation. This install requires 176 MB of free disk space. | |
| Destination Folder c:\dvv Browse | |
| Nullsoft Install System v2.08 | ancel |

Choose Destination Folder

The default directory is c:\dvv. It can be changed by using the 'Browse' button or by overtyping the default. After specifying the Destination Folder, hereinafter referred to as DVV_HOME, click the 'Install' button to continue the installation.

NOTE: If the DVV is already installed in the specified directory, a dialog will appear asking if the previous installation should be replaced, as shown below.

| 🌍 The Li | brary of Congress - Digital Viewer and Validator (1.4.4_3.7) Setup |
|----------|--|
| ? | A previous version of the DVV has been found. Choose OK to replace the existing DVV or Cancel to exit. |
| | OK Cancel |

Click OK to replace the previous version of the DVV with new files. All files in the DVV_HOME\lib directory will be deleted and the new files will be copied to the DVV_HOME\lib directory. All files, with the exception of dvv.properties in the DVV_HOME directory will be overwritten. If dvv.properties exists, it will not be overwritten, but left in place.

Installation Errors

In the event that there are errors during the installation, the following dialog will appear:

| 😚 The Library of Congress - Digital Viewer and Validator (1.4.4_3.7) Setup | - 🗆 🗙 |
|--|--------|
| Installation Aborted Setup was not completed successfully. | |
| Can't write: c:\dvv\dvv.properties | |
| Show details | J |
| | |
| | |
| | |
| | |
| Nullsoft Install System v2.08 | Cancel |

If the DVV is running and setup is attempted, this panel (or one very similar) will appear. Stop the DVV and click "Retry" to continue the installation or "Abort" to exit setup. Do not choose ignore because the installer has not been configured to accommodate this option.

| 🍘 The Li | brary of Congress - Digital Viewer and Validator (1.4.4_3.7) Set 🔀 |
|--------------|--|
| | Error opening file for writing: |
| \mathbf{w} | c:\dvv\dvv.properties |
| | Click Abort to stop the installation, Retry to try again, or Ignore to skip this file. |
| | Abort Retry Ignore |

Install Log

Click 'OK', and then the 'Show details' button to view the install log.

| 😽 The Library of Congress - Digital Viewer and Validator (1.4.4_3.7) Setup | _ 🗆 🗙 |
|---|----------|
| Installation Aborted Setup was not completed successfully. | |
| Can't write: c:\dvv\dvv.properties | |
| Extract: c:\dvv\icons\xml.png Extract: c:\dvv\icons\reel.png Extract: c:\dvv\icons\batch.png Extract: c:\dvv\icons\dir.gf 100% Extract: c:\dvv\lib\NdnpPreingest-Interface-1.4.4_3.7.jar 100% Extract: c:\dvv\lib\NdnpPreingest-Validation-3.7.jar 100% Extract: c:\dvv\lib\bagit-3.1.jar 100% Extract: c:\dvv\lib\jj2000-5.1.jar 100% Extract: c:\dvv\lib\jj2000-5.1.jar 100% Extract: c:\dvv\lib\jj2000-5.1.jar 100% Extract: c:\dvv\publickeys.properties 100% Can't write: c:\dvv\dvv.properties | A |
| Nullsoft Install System v2.08 | Cancel R |

In this case, the DVV was running or the installation directory did not contain enough free space and the installer could not write the necessary files.

Success

A successful installation will end with the following panel:

| 🌍 The Library of Congress - Digital Viewer and Validator (1.4.4_3.7) Setup | _ 🗆 🗙 |
|--|----------|
| Installation Complete Setup was completed successfully. | |
| Completed | |
| Show details | |
| | |
| | |
| | |
| | |
| Nullsoft Install System v2.08 | const. 1 |
| < Back Close | Cancel |

Always click on Show Details to ensure that the installation completed successfully, as shown below.

Notice the error reported above that did not abort the installation.

| 🌍 The Library of Congress - Digital Viewer | and Validator (1.4.4_3.7) Setup | _ 🗆 🗙 |
|--|---------------------------------|--------|
| Installation Complete Setup was completed successfully. | | |
| Completed | | |
| Extract : LICENSE_fr.rtf Extract : LICENSE_it.rtf Extract : LICENSE_ja.rtf Extract : LICENSE_ko.rtf Extract : LICENSE_sv.rtf Extract : LICENSE_sv.rtf Extract : LICENSE_zh_CN.rtf Delete file: c:\dvv\lib\jre.zip Created uninstaller: c:\dvv\Uninstall.exe Uninstaller written Logfile has been closed Completed | | |
| Nullsoft Install System v2.08 | < Back Close | Cancel |

This installation was successful.

4. Post Install Configuration

There is a file named dvv.properties in the DVV_HOME directory which contains the following runtime options for the DVV:

| Property Name | Default Value | Possible | Purpose |
|-----------------------|--------------------------|----------------------|--|
| | | Values | |
| paths.check.always | false | true false | If true, perform Batch Match when batch is opened. If false, Batch Match must be started manually. |
| batch.validate.always | false | true false | If true, validation starts automatically when batch is opened. If false, validation must be started manually. |
| show.hidden.files | false | true false | If true, hidden files will be shown and considered to be present during Batch Match. If false, hidden files will not be shown and will not be considered present during Batch Match. |
| dvv.version | Set by installer | Current | Part of title, assists in |
| icon.validated.yes | icons/greenCheckmark.gif | version File name | troubleshooting. Image file for icon. |
| icon.validated.no | icons/redX.gif | File name | Image file for icon. |
| icon.batch | icons/batch.png | File name | Image file for icon. |
| icon.folder | icons/dir.gif | File name | Image file for icon. |
| icon.folder.open | Icons/dir_open.gif | | initige file for feori. |
| icon.parent.folder | icons/dirup.gif | File name | Image file for icon. |
| icon.issue | icons/issue.png | File name | Image file for icon. |
| icon.reel | icons/reel.png | File name | Image file for icon. |
| icon.jp2 | icons/jp2.png | File name | Image file for icon. |
| icon.pdf | icons/pdf.png | File name | Image file for icon. |
| icon.tif | icons/tif.png | File name | Image file for icon. |
| icon.xml | icons/xml.png | File name | Image file for icon. |
| icon.unknown | icons/unknown.png | File name | Image file for icon. |
| dvvbrowser.html.path | DVV_HOME/html | Directory name | Temporary storage of generated HTML. |
| mets.reel.xsl | metsReel.xsl | File name | XSLT for reel data. |
| mets.issue.xsl | metsIssue.xsl | File name | XSLT for issue data. |
| mets.ocr.xsl | metsOcr.xsl | File name | XSLT for OCR data |

| pdf.header.xsl | pdfHeader.xsl | File name | XSLT for Derivative |
|--------------------|-------------------|-----------|---------------------|
| | | | Header |
| master.header.xsl | masterHeader.xsl | File name | XSLT for Master |
| | | | Header |
| service.header.xsl | serviceHeader.xsl | File name | XSLT for Service |
| | | | Header |

Properties that are newly introduced will not appear in the dvv.properties file until the DVV has been started at least once subsequent to the installation of the new version.

Changing XSLT Files

Metadata files for Issues and Reels, as well as OCR files, are stored as XML. To make the analysis of these files less tedious, each has been converted from XML into an HTML page that presents the data contained therein, via an embedded web browser, in a much more readable fashion. The process of conversion is accomplished by applying Xml Stylesheet Language Transformation (XSLT) to the XML files. Each of the file types mentioned here has a specific XSLT rules file associated with it in the dvv.properties file.

The mets.reel.xsl property assigns the file name metsReel.xsl as the default XSLT rules file for Mets Reel data transformation. Change the value of this property to a customized XSLT file of your choice and restart the DVV to make it the new rules file for Mets Reel data transformation. After the DVV has been started, changes can be made to the rules without restarting the DVV.

The same methodology applies to changing XSLT rules files for Mets Issue and OCR files, as well as Master, Derivative and Service Header data files.

Changing Icon Images

There are two ways to change the images used by the DVV as icons. The first is to replace the image files mentioned in the table above with different files of the same name. The second way is to change the value of the property which refers to the icon to be changed.

Default Images

| : | |
|--------------------------|--------------|
| icons/greenCheckmark.gif | Y |
| icons/redX.gif | X |
| icons/batch.png | |
| | Batch |
| icons/dir.gif | |
| icons/dir_open.gif | |
| icons/dirup.gif | t. , |
| icons/issue.png | 23) |
| icons/reel.png | * |
| icons/jp2.png | JPEG 2000 |
| icons/pdf.png | PDF |
| icons/tif.png | TIFF |
| icons/xml.png | OCR XML |
| icons/unknown.png | |

5. Starting the DVV

To start the Digital Viewer and Validator, open a command (DOS) window and change to the directory in which the DVV is installed.

Open Command Window

First, click on 'Start' in the Windows menu bar.



Then click 'Run' to initiate the following dialog:

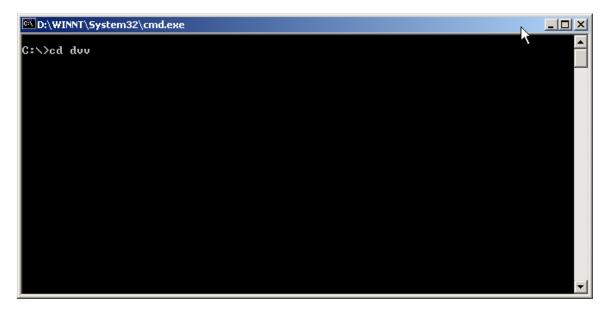


Make sure the value contained in the text field is 'cmd' (without the quotes) and then click 'OK'.

The Command Window

| D:\WINNT\System32\cmd.exe | |
|---------------------------|---------|
| C:\> | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

Change to the directory in which the DVV is installed as shown below:



Then issue the following command:

| D:\WINNT\System32\cmd.exe | |
|---------------------------|---|
| C:/>cd dvv | |
| C:\dvv>locdvv | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | - |

to start the Digital Viewer and Validator.

6. Batch Processing

Opening a Batch

To open a batch, click on 'File' and then select 'Open'. This action will cause a file manager dialog to open. Navigate to the batch file for the batch to be viewed/validated, click the batch file name to select it and then click on 'Open' to load the batch file.

Note – if accessing a batch located on a remote drive, ensure that the remote drive is mapped locally (using Tools, Map Network Drive from the Windows File Manager interface) and use the facilities of the DVV file manager interface to navigate to the remote batch file.

File paths containing spaces will not be processed properly. To ensure proper processing of batches with spaces in the path, specify the short name for the non-8dot3 file name. To determine the short name, list the directory with the /X switch, as shown below.

| M D05 | |
|--|--------------------------|
| C:∖Documents and Settings>dir /X Volume in drive C has no label. Volume Serial Number is F066-FEF4 | <u> </u> |
| Directory of C:\Documents and Settings | |
| | dministrator 11 Users |
| C:\Documents and Settings>_ | |
| | |
| | |
| | |
| | - |
| <u>۱</u> | |

Do not specify the Universal Naming Convention name, such as:

 $\mbox{wachine99}d\$

Shown below is an example of the File, Open step.

| 🕌 LOC Dig | | | |)r | | | | | |
|-------------------|-----|-----|-----|---------|--------|-------------|-----------------|------|--|
| <u>File</u> Batch | | | | | | | | | |
| Open D | JP2 | PDF | OCR | Headers | Thumbs | Title Essay | Metadata Viewer | Info | |
| Browse | v | | | | | | | | |
| E <u>x</u> it | | | | | | | | | |
| | | | | | | | | | |

The following illustration shows the file manager dialog being used to open a file named demoBatch.xml.

| 🕌 Open | | | |
|------------------------|---------------|----------------------|---------------------|
| Look <u>i</u> n: 📑 s | sn85066387 | | |
| 00100 | 1480426 | BATCH.xml | oneMonthBatch.xml |
| 00100 | 480438 | BATCH_1.xml | oneMonthBatch_1.xml |
| 00100 | 480451 | demoBatch.xml | 🗋 twoDayBatch.xml |
| 00100 | 1480463 | 🗋 demoBatch_1.xnរំ | twoDayBatch_1.xml |
| <u> </u> | 480487 | 🗋 fiveMonthBatch.xml | |
| <u> </u> |)480529 | fiveMonthBatch_1.xm | il 👘 |
| • | | | |
| File <u>N</u> ame: | demoBatch.xml | | |
| Files of <u>T</u> ype: | All Files | | - |
| | | | Open Cancel |

Batch View

| 🛎 L(| DC Digital Viewer and Validat | 0F | | | | | | | |
|------|----------------------------------|-----|-----|---------|--------|-------------|-----------------|------|--|
| | Batch Exemptions Help | | | | | | | | |
| 🗋 di | | PDF | OCR | Headers | Thumbs | Title Essay | Metadata Viewer | Info | |
| | Verify Batch | | | | | | | | |
| | V <u>a</u> lidate All | | | | | | | | |
| | Validate All <u>U</u> nvalidated | | | | | | | | |
| | Validate Selected | | | | | | | | |
| | Vie <u>w</u> | | | | | | | | |
| | Cancel Current Processing | | | | | | | | |
| | <u>C</u> lear Info Messages | | | | | | | | |
| 1 | | | | | | | | | |

The Batch View option causes the DVV to load and parse the batch, issue and reel files associated with the just opened (but not loaded) batch.xml file. After a batch has been processed by the View initiation process, images can be viewed as described below in Viewer Controls.

Another method of viewing images, with much less overhead, is subsequently described in the Browse Mode section.

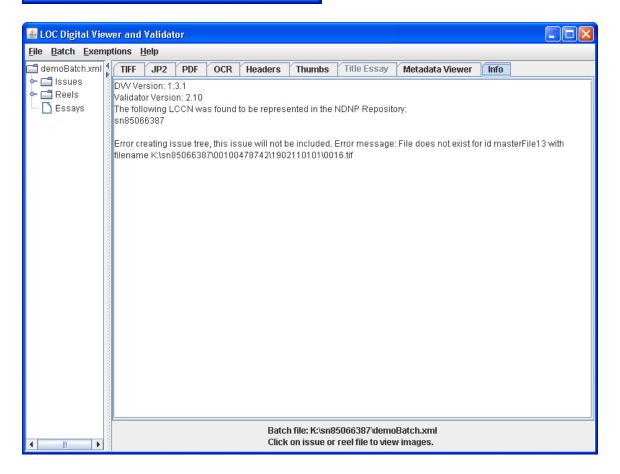
Only after invoking Batch View to load a batch file will the Exemptions menu choice be made available.

Note also that a report of valid LCCNs is included. If any of the LCCNs referenced in the batch had not been present in the NDNP Repository, those LCCNs would have been included in a list of LCCNs not represented in the NDNP Repository.

Batch Errors

When the batch file is opened with the View option, Validation Library routines automatically verify that all files referenced from within the batch metadata files are present. If a file is missing, an alert dialog is presented and appropriate entries are added to the Error Report.

| 🖆 DVV Additional Info Report 🛛 🛛 🔀 |
|---|
| Error creating issue tree, this issue will not be included. Error message: File does not exist for id masterFile13 with filename K:\sn85066387\00100478742\1902110101\0016.tif |
| Continue Exit DVV |



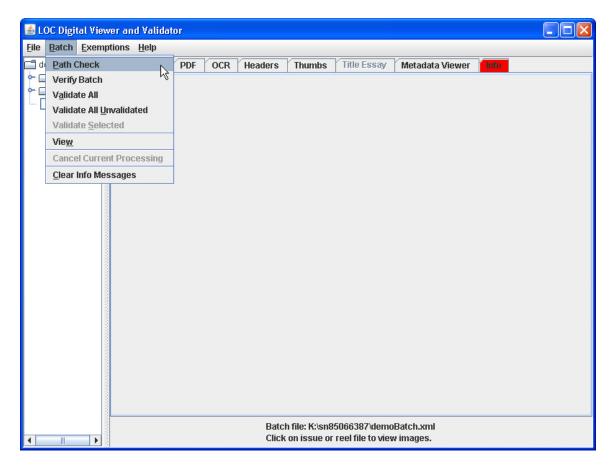
Processing a Batch

There are four validation operations that may be performed on a batch, Batch Match, Validate All, Validate All Unvalidated and Validate Selected, plus Verify which checks the digital signatures to ensure that they are valid and have not been modified. Note that the Batch menu will not be available unless a batch has been opened, as in the previous step.

Batch Match examines all files in the batch directory, and all subordinate directories, to ensure that every file referenced from within the batch metadata files is present. If the batch specifies a file that is not present, the Batch Match process will fail. If there are files present that are not referenced from within the batch, this will also cause the Batch Match process to fail.

All files within a batch must reside in the same directory as the batch XML file or a subordinate directory.

To begin the Batch Match operation, click 'Batch' on the menu bar and then select 'Path Check', as shown below. The Path Check option is available only after the View option has been selected and the batch loaded.



During Batch Match processing, a progress panel at the bottom of the DVV user interface displays the current step and the elapsed time.

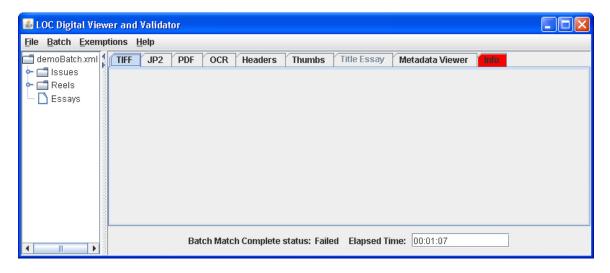
| 🛎 LOC Digital Viewer an | d Validator | | | k | | | |
|--|----------------|-------------------------|----------------|-------------|------------------|--------|---|
| <u>File</u> <u>Batch</u> <u>Exemptions</u> | | | | | | | |
| 📑 demoBatch.xml 👔 🔳 | JP2 PDF | OCR Headers | Thumbs | Title Essay | Metadata Viewer | Info | |
| 🕈 🗂 Issues 🛛 👔 | | | | | | | |
| ← 🚍 Reels - 🗋 Essays | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | Batch Match St | t atus Step: Loa | ding Files Pre | sent List | Elapsed Time: 00 | :00:02 |] |

If the Batch Match process fails, an error dialog will report the failure.

| 📓 DVV Additional Info Report 🛛 🛛 🔀 | | | | | | | | |
|---|--|--|--|--|--|--|--|--|
| DVV Batch Match failed. See Info/Log for details. | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| Continue Exit DVV | | | | | | | | |

Error Reporting

If errors are encountered during the Batch Match or Validation process, they are sent to a log file and also to the Info tab of the DVV. If there are any unviewed errors, the 'Info' tab background is set to red until it is clicked on, at which time it reverts to the default background color. Below is an example of a Batch Match that failed.



The Errors

| 📓 LOC Digital Viewer and Validator | | | | | | | |
|---|--|---|--|--|--|--|--|
| <u>F</u> ile <u>B</u> atch <u>Exemptions</u> <u>H</u> elp | | | | | | | |
| demoBatch.xml | | | | | | | |
| | k:\sn85066387\00100480529\0001.pdf k:\sn85066387\00100480529\0001.tif k:\sn85066387\00100480529\0001.xml k:\sn85066387\00100480529\0002.jp2 | | | | | | |
| • | k/sn85066387/00100480529/0002.pdf k/sn85066387/00100480529/0002.pdf k/sn85066387/00100480529/0002 if | - | | | | | |

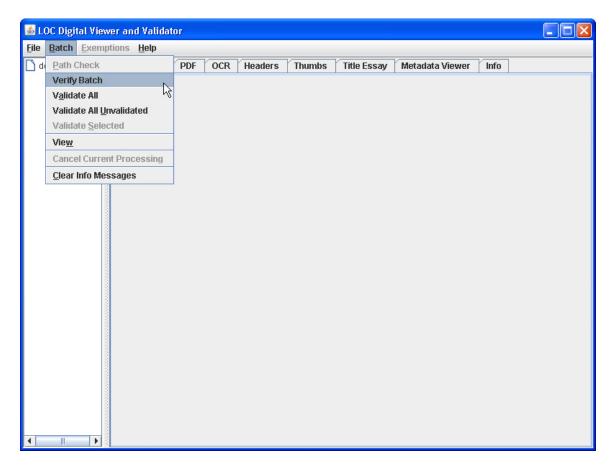
Clearing Informational Messages

If the contents of the 'Info' tab become unnecessary, they can be cleared. Clearing the messages also closes the current message log and reinitializes a new log. All logs are stored in the DVV_HOME/logs directory. The message log is written to disk only after Batch/Clear Errors or File/Exit has been invoked. Terminating the DVV via other methods will result in the loss of the current message log contents.

| 🛎 L0 | OC Digital Viewer and Validat | tor 📃 🗖 🔀 | | | | | | | |
|---|------------------------------------|--|--|--|--|--|--|--|--|
| <u>F</u> ile | Batch Exemptions Help | | | | | | | | |
| 🗖 di | Path Check | PDF OCR Headers Thumbs Title Essay Metadata Viewer Info | | | | | | | |
| ٩¢ | Verify Batch | 3.1 | | | | | | | |
| l ° ⊓ | V <u>a</u> lidate All | n: 2.10 | | | | | | | |
| - 4 | Validate All <u>U</u> nvalidated | CCN was found to be represented in the NDNP Repository: | | | | | | | |
| | Validate Selected | | | | | | | | |
| | Vie <u>w</u> | sue tree, this issue will not be included. Error message: File does not exist for id masterFile13 with 5066387\00100478742\1902110101\0016.tif n failed, the following extraneous files (not represented anywhere in the batch) were found on the local | | | | | | | |
| | Cancel Current Processing | | | | | | | | |
| | Clear Info Messages N | | | | | | | | |
| | 12 | 100100480529\0001.jp2 \00100480529\0001.pdf | | | | | | | |
| | 2 | \00100480529\0001.tif | | | | | | | |
| k:\sn85066387\00100480529\0001.xml | | | | | | | | | |
| | k:\sn85066387\00100480529\0002.jp2 | | | | | | | | |
| k\sn85066387\00100480529\0002.pdf k\sn85066387\00100480529\0002.tif k\sn85066387\00100480529\0002 tif | | | | | | | | | |
| | m r stantinn sat | | | | | | | | |

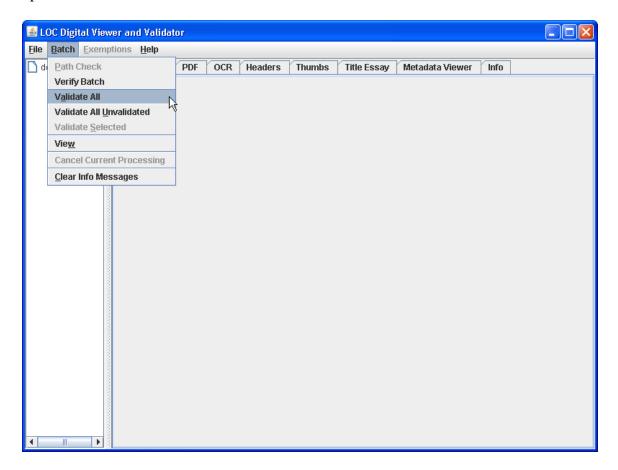
Batch Verification

Batch verification, which analyzes the digital signatures of all files in the batch, is initiated by selecting the "Verify Batch" menu item as shown below. Note that Verify Batch does not load the batch file tree. Use Batch/View to make the batch file tree available.



Batch Validation

Validation of a batch begins by clicking 'Batch' on the menu bar and then selecting one of 'Validate All', 'Validate All Unvalidated' or 'Validate Selected'. Validate All will perform validation on all files within the batch, regardless of whether they have been validated before. Validate All Unvalidated will perform validation only on those files in an issue/reel that has not been validated. Validate Selected will perform validation on the selected issue or reel. The Validate Selected option is available only after the View option has been selected and the batch loaded.



Validation Progress

As with Batch Match, a progress panel, updated approximately every second, is displayed during batch validation. In order to provide a visual cue that the Validation process is making steady progress, the name of the file currently being visited is displayed in the progress panel. Other pertintent statistics shown on the the panel are:

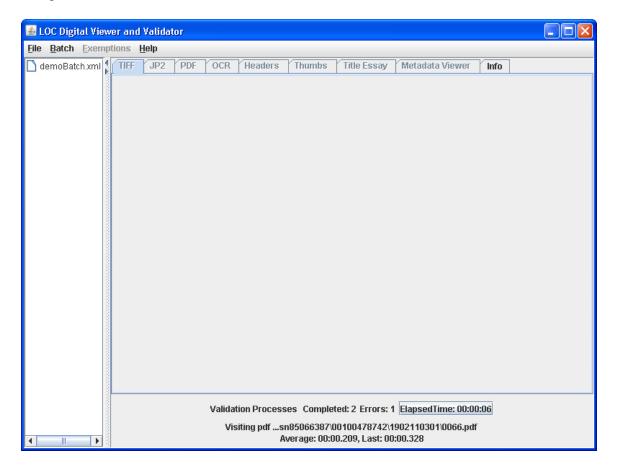
Validation Processes Completed: The number of files that have been validated.

Errors: The number of validation errors that have occurred. An error describing the nature of the failure will be found in the message log and in the Info tab.

Elapsed Time: How long this validation has been running.

Average: The average amount of time (for this batch) to process the type of file currently being visited.

Last: The amount of time it took to process that most recent file of the type currently being visited.



Validation Complete

At the completion of the Validation process, a dialog is initiated to alert the user that validation is complete and the progress panel is updated to show the results of the validation. A sample dialog:

| 🕌 DVV Additional Info Report | × | | | | | | |
|------------------------------|---|--|--|--|--|--|--|
| Validation completed. | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| Continue Exit DVV | | | | | | | |

Values are:

Total Tasks: The number of files validated. Note: this number will be less than the highest number seen in Validation Processes Started. This is because ancillary tasks that are part of the validation process are shown in this count. These tasks are subtracted from the total.

Good Issues/Reels: The number of issues/reels which validated successfully.

Failed Issues/Reels: The number of issues/reels that failed validation.

Good Derivatives/Masters/Ocrs/Services: The number of successful validations for each of the indicated file types.

Failed Derivatives/Masters/Ocrs/Services: The number of failed validations for each of the indicated file types.

Elapsed Time: The amount of time it took to validate all the files in the batch.

Unresolved Errors: the number of good issues, reels, derivatives, masters, ocrs and services plus the number of failed issues, reels, derivatives, masters, ocrs and services should equal the number of processes completed. There may be instances where the number of processes started does not equal the number of processes completed. This is due to the possibility that an unresolved error has occurred. Unresolved errors are reported on and detailed in the log.

| LOC Digital Viewer and Validator Image: Second Se | | | | | | | |
|---|--|---|--|--|--|--|--|
| 🗖 demoBatch.xml [| TIFF JP2 PDF OCR Headers Thumbs Title Essay Metadata Viewer Info | | | | | | |
| ← 급 Issues ← 급 Reels └ D Essays | The following LCCN was found to be represented in the NDNP Repository: sn85066387 validation successes: 56 pdf validation successes: 14 tiff validation successes: 14 ocr validation successes: 14 ip 2 validation successes: 0 pdf verification successes: 0 ocr verification successes: 0 integrity check successes: 0 integrity successes: 0 or integrity successes: 0 iff integrity successes: 0 issue validation success: 0 issue validation success: 0 issue verification success: 0 issue validation success: 0 issue validation success: 0 issue verification success: 0 issue validation failures: 1 | | | | | | |
| | vaalualidatian failuvaa: 0 | • | | | | | |
| ElapsedTime: 00:36:20 | | | | | | | |
| | Good Issues: 0 Reels: 0 Failed Issues: 1 Reels: 0 Unresolved Errors: 0 | | | | | | |
| < III > | Good PDF: 14 TIF: 14 OCR: 14 JP2: 14 Failed PDF: 0 TIF: 0 OCR: 0 JP2: 0 | | | | | | |

After a batch, issue or reel has been successfully validated, a digital signature will be created for the respective batch, issue or reel. Once a signature has been applied, modification of any existing file within the batch would invalidate all signatures. To prevent this situation, a new batch, issue or reel metadata file is created each time validation is successful, thereby ensuring that the original contents remain unchanged. A new metadata file will have _n appended to the file name just before the .xml suffix, where n is a number, starting at 1, that is incremented every time successful validation, by a new signer, occurs. For this reason, the batch XML file with the highest "validation count" value should be opened when viewing a batch. The batch XML file will only be updated with the names of the new issue/reel files and have its validation count incremented if Validate All has been selected.

LCCN Validation

As each newspaper title within a batch, identified by its LCCN, is encountered a runtime query is dispatched to the live NDNP application to determine the validity of the LCCN. A message, reporting the absence or presence of that LCCN in the NDNP newspaper title database, will be written to the log as shown below.

| 📓 LOC Digital Viewer and Validator | | | | | | | | |
|------------------------------------|---|---------|--|--|--|--|--|--|
| <u>File B</u> atch Exem | ptions <u>H</u> elp | | | | | | | |
| 📑 demoBatch.xml | TIFF JP2 PDF OCR Headers Thumbs Title Essay Metadata Viewer Info | | | | | | | |
| 🗣 🚍 Issues | The following LCCN was found to be represented in the NDNP Repository: | | | | | | | |
| ► 📑 Reels | sn85066387 | | | | | | | |
| 🗆 🗋 Essays | | | | | | | | |
| | validation successes: 56 | | | | | | | |
| | pdf validation successes: 14 tiff validation successes: 14 | | | | | | | |
| | orrvalidation successes: 14 | | | | | | | |
| | ip2 validation successes: 14 | | | | | | | |
| | verification successes: 0 | | | | | | | |
| | pdf verification successes: 0 | | | | | | | |
| | tiff verification successes: 0 | | | | | | | |
| | ocr verification successes: 0 | | | | | | | |
| | jp2 verification successes: 0 | | | | | | | |
| | integrity check successes: 10 | | | | | | | |
| | pdf integrity successes: 0 | | | | | | | |
| | tiff integrity successes: 0 | = | | | | | | |
| | ocr integrity successes: 0 | | | | | | | |
| | ip2 integrity successes: 0 | | | | | | | |
| | reel validation success: 0 | | | | | | | |
| | issue verification success: 0 | | | | | | | |
| | reel verification success: 0 | | | | | | | |
| | issue integrity success: 7 | | | | | | | |
| | reel integrity success: 3 | | | | | | | |
| | issue validation failures: 1 | - | | | | | | |
| | real validation failuraa: 0 | · · | | | | | | |
| | ElapsedTime: 00:36:20 | | | | | | | |
| | | | | | | | | |
| | Good Issues: 0 Reels: 0 Failed Issues: 1 Reels: 0 Unresolved Errors: 0 | | | | | | | |
| <hr/> | Good PDF: 14 TIF: 14 OCR: 14 JP2: 14 Failed PDF: 0 TIF: 0 OCR: 0 JP2: 0 | | | | | | | |

Canceling Long Running Tasks

In the event that a long running task, such as validation or verification, needs to be canceled, select Batch/Cancel Current Processing, as shown below.

| 🛃 L(| DC Digital Viewer and Validat | 01 | | | | | | | |
|--------------|----------------------------------|-----|--------|-------------|------------|-----------------------------------|------------------------------|------|--|
| <u>F</u> ile | Batch Exemptions Help | | | | | | | | |
| 🗋 di | Path Check | PDF | OCR | Headers | Thumbs | Title Essay | Metadata Viewer | Info | |
| | Verify Batch | | | | | | | | |
| | V <u>a</u> lidate All | | | | | | | | |
| | Validate All <u>U</u> nvalidated | | | | | | | | |
| | Validate Selected | | | | | | | | |
| | Vie <u>w</u> | | | | | | | | |
| | Cancel Current Processing | | | | | | | | |
| | Clear Info Messages | | | | | | | | |
| | | | | | | | | | |
| | | | Valida | tion Proces | ses Comple | ted: 1 Errors: 0 | ElapsedTime: 00:0 | 0:06 | |
| • | | | Vi | | | 00100478742\1)0.200, Last: 00 | 902110101\0017.pd :00.312 | f | |

An Additional Info dialog, specific to the task that was cancelled, will be presented. Example:

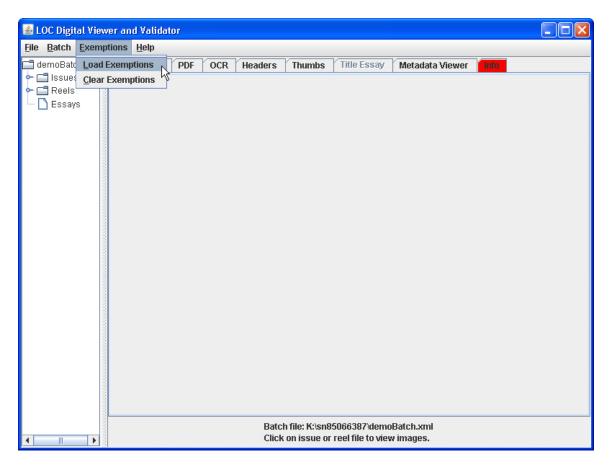
| 🕌 DVV Additional Info Report | × |
|--|---|
| Batch validation processing cancelled. | |
| | |
| | |
| | |
| | |
| | |
| [] | |
| Continue Exit DVV | |

Exemptions

Exemptions support batches that deviate from NDNP requirements by relaxing rules that are checked during validation. Exemptions are only allowed for legitimate purposes and must be approved by NDNP. As such, only the mechanics of introducing a preconfigured exemption set are presented here.

A default exemption set file, with a single default exemption is supplied, but is not loaded when the DVV is initialized. Rather, exemptions must be explicitly loaded as shown below. Located in DVV_HOME, the name of the file is exemptionset.xml. To specify a different filename of the exemption set, select Load Exemptions from the Exemptions menu and follow the steps typical for a file dialog.

Exemptions are recognized only during validation and are not considered during the verify process.



The default exemption is:

mets_original_relateditem_physicaldescription_child

Which exempts issue and reel files from the rule that states the original related item must have a physicalDescription child.

Another exemption that is often useful:

reel_target_child_dmdid_mods

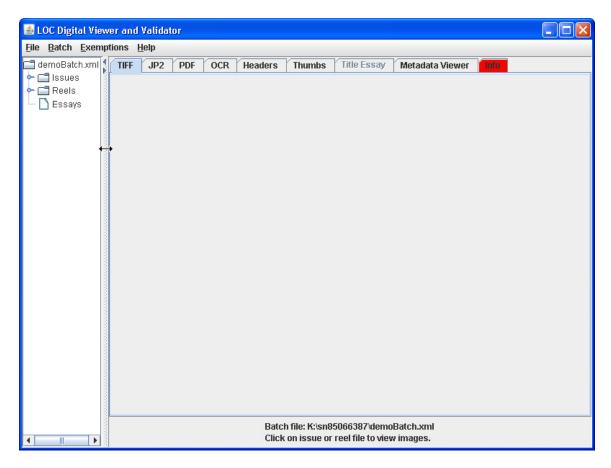
exempts a reel file from the rule that states that the target div child of techtargetreel div must have a DMDID for a Mods dmdSec.

A sample exemptions file with all both of the aforementioned exemptions is shown below:

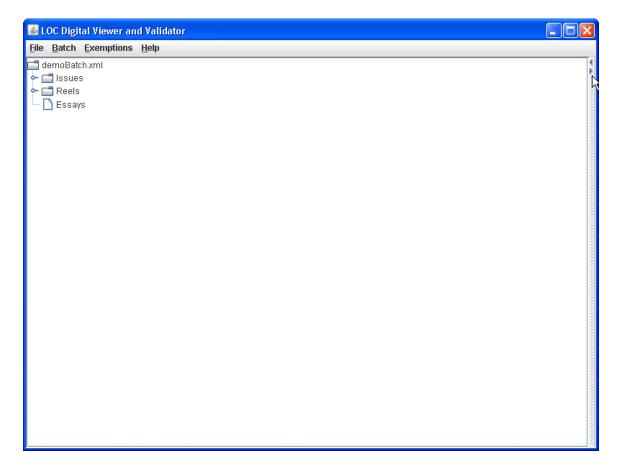
<?xml version="1.0" encoding="UTF-8"?> <exemptionSet xmlns="http://www.loc.gov/ndnp"> <code>mets_original_relateditem_physicaldescription_child</code> <code>reel_target_child_dmdid_mods</code> </exemptionSet>

7. Viewer Controls

The viewer is separated into two areas. The right portion of the viewer contains tabs which are used to view error information, the contents of issue and reel files, title essays, and the images which make up a batch. The left portion of the viewer contains a tree structure which shows the issues, reels and title essays contained in the batch. The relative amount of space allocated to each side may be adjusted by positioning the cursor over the dotted vertical bar at which time the cursor's appearance will change, as shown below. Once the cursor has changed, click and hold down the left mouse button and move the mouse to the right or left to adjust the position of the divider. Release the left mouse button to end this adjustment.



Another method of changing the allocation of space between the two sections is to click on the small arrows at the top of the divider. Shown below is the result of clicking on the right pointing arrow to minimize the viewer side and maximize the tree side.



To expand the list of issues, click on the toggle widget next to the folder labeled "Issues".

The result:

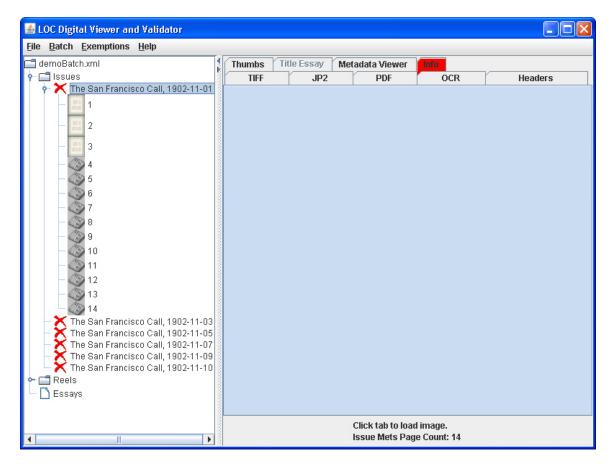
| LOC Digital Viewer and Validator | | | | | |
|--|--------|-------------|---|------|---------|
| <u>File Batch Exemptions H</u> elp | | | | | |
| 📑 demoBatch.xml 🚦 | Thumbs | Title Essay | Metadata Viewer | Info | |
| | TIFF | JP2 | PDF | OCR | Headers |
| — X The San Francisco Call, 1902-11-03 — X The San Francisco Call, 1902-11-05 | | | | | |
| - X The San Francisco Call, 1902-11-07 X The San Francisco Call, 1902-11-09 | | | | | |
| The San Francisco Call, 1902-11-10 | | | | | |
| Essays | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | tch file: K:\sn850663 ck on issue or reel fi | | |

The divider between the batch tree and the viewing panel is adjusted automatically.

To return the viewing panel to its optimum size, click the toggle widget for the "Issues" folder and the collapse of the batch tree will cause the viewing panel to be automatically resized.

Issue/Reel Summary

When an issue or reel is selected, a summary of that issue/reel is displayed at the bottom of the viewer panel which indicates the number of Mets pages within the issue/reel. To view an image of a given type, click on the viewing tab corresponding the type to be viewed, as shown below for Master, and the image will be loaded. Only one image can be loaded at a time.



The red X next to the issue above indicates that the corresponding issue/reel has not been validated. A green checkmark next to an issue/reel means that validation upon that issue/reel has been successfully performed. A page with light toned zeroes and ones represents a page that has not been digitized, but may have been published. A green checkmark overlaying the icon for a non-digitized page is used to indicate an issue that has not been digitized, but may have been digitized and has not been digitized page will be found next to an issue that is not digitized and has not been validated. The method to change these icons is discussed on page 14.

Clicking on an issue that has not been validated will not expand the page tree for that issue. The only tab, other than the info tab, that will work is the Metadata Viewer tab.

Clicking on a page that has not been digitized will result in a page info display with a filename of consisting of one of the three Issue Present indicators describing a non-

digitized issue/page. If the Issue Present Comment is available it will be appended to the end of the Issue Present Indicator.

Load Progress

While an image is being loaded a progress panel will be displayed at the bottom of the viewer panel, as shown below. For Master files the progress bar indicates the percentage of completion. Due to the loading mechanisms available for Service, Derivative and OCR files, it is not possible to determine how much progress has been made relative to completion, so an "indeterminate" progress bar is used.

| 🕌 LOC Digital Viewer and Validator | | | | | |
|---|--------|------------------|---------------------|-----------------|-------------|
| <u>File Batch Exemptions H</u> elp | | 6 | | | |
| 🗖 demoBatch.xml 🚦 | Thumbs | Title Essay | Metadata Viewer | Info | |
| 🕈 🚍 Issues | TIFF | JP2 | PDF | OCR | Headers |
| Issues The San Francisco Call, 1902-11-03 The San Francisco Call, 1902-11-03 The San Francisco Call, 1902-11-05 1 2 3 4 5 6 7 8 9 10 11 22 3 4 5 6 7 8 9 10 11 22 8 9 10 11 22 33 4 5 6 7 8 9 10 11 22 33 44 5 10 11 24 55 16 The San Francisco Call, 1902-11-07 The San Francisco Call, 1902-11-09 | | le to load image | Filename: Not avail | | |
| | | Loading ser | vice image 1 of 16 | _ | |
| | | Filename: K: | \sn85066387\001004 | 178742\19021105 | 01\0082.jp2 |

Loading Images

After the image has been loaded, the progress panel is updated with information about the image. To load the next image of a given type, click on the "Next" button just below the horizontal scroll bar. Another way to display a different image is to click on the page number under issue being viewed in the Batch Tree. The Page Number dropdown box, between the Next and Previous buttons may also be used.

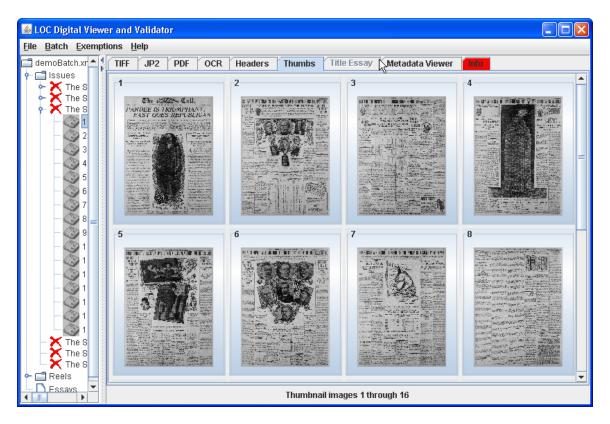
The page number icons displayed under an issue also serve as the Master image control. Whenever the page number icon is used to change the page being displayed, the corresponding Master image will be loaded and the Master tab will be selected. Subsequent selection of any other image type tab (Service, Derivative or OCR) will cause the corresponding page image of that type to be loaded. While viewing other image types, the Master image control remains on the most recent Master image numbe viewed. This provides the ability to skip around and view images "at random".

To zoom in our out, click the appropriate 'Zoom' button just above the image. The zoom amount, which defaults to 25 percent, may be adjusted by overtyping the value in the text field between 'By' and '%'.



Thumbnail Images

Thumbnail size images, derived from the list of master files, are available on the Thumbs tab.



Each thumbnail image has a border containing the image number, which corresponds to the page number shown in the Batch Tree. Clicking on a thumbnail image will cause the Master image of the same page number to be loaded and the Master tab will be selected.

Load Errors

If an error is encountered that precludes the viewing of an image, an information panel, often accompanied by an informational dialog, is displayed.

Sample Info Dialog

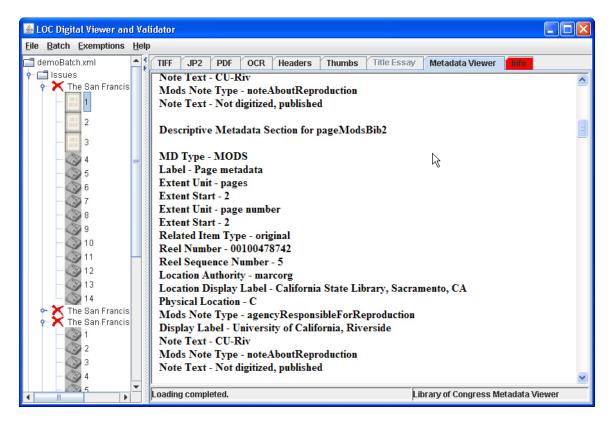


Error Panel

| 🕌 LOC Digital Viewer and Validator | |
|--|---|
| <u>File B</u> atch <u>E</u> xemptions <u>H</u> elp | |
| 🗂 demoBatch.xml 📃 | Thumbs Title Essay Metadata Viewer Thio |
| •- 🚍 Issues | TIFF JP2 PDF OCR Headers |
| The San Francisco Call, 1902-11-01 | Zoom In Zoom Out By 25 % |
| - 101 2 | Unable to load image Filename: Not available Reason: Not digitized, published |
| 2 3 4 5 6 7 8 9 10 11 12 13 14 The San Francisco Call, 1902-11-03 The San Francisco Call, 1902-11-03 1 2 3 4 4 5 5 6 7 8 9 9 10 11 12 13 14 14 14 15 14 15 14 15 16 17 18 19 10 11 10 11 12 13 14 14 14 14 14 14 14 14 14 14 | |
| - ∞ 4 = - ∞ 5 | |
| - 🐼 6 | |
| - 🕢 7 | |
| 8 | |
| 9 | |
| - 310 | |
| - 2 12 | |
| - 🐼 13 | |
| - 🐼 14 | |
| The San Francisco Call, 1902-11-03 | |
| • 🗙 The San Francisco Call, 1902-11-05 | |
| - 2 | |
| - 3 | < Previous 1 Vext> |
| ▲ ⁴ | Unable to load image Not available |

Viewing Metadata

When an issue or reel is selected, the xml metadata contained within the Mets File is converted, via XML Stylesheet Language Transformation (XSLT), to a formatted HTML page and displayed in a browser embedded in the Metadata Viewer tab. An example of the page is shown below.



See the Post Install Configuration section for details on changing the XSL files that control formatting and the location of generated HTML files.

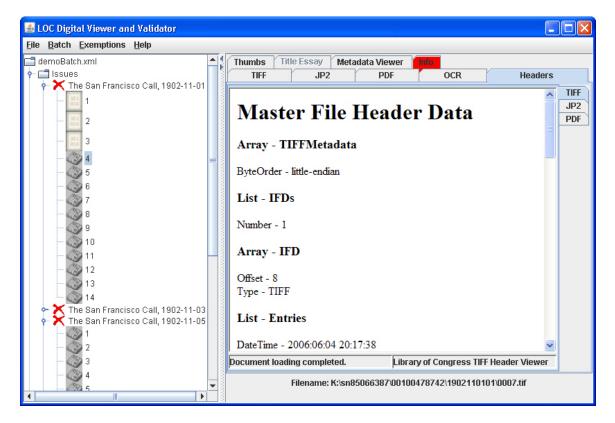
Viewing Image Header Data

Header data is available for Master, Derivative and Service files and can be viewed in the Headers tab. As shown below, there is a tab for each of the image types mentioned.

| 🕌 LOC Digital Viewer and | d Validator | |
|---|--|------|
| <u>F</u> ile <u>B</u> atch <u>E</u> xemptions ! | Help | |
| 📑 demoBatch.xml 🚽 | TIFF JP2 PDF OCR Headers Thumbs Title Essay Metadata Viewer Info | |
| | | TIFF |
| 🔶 🗙 The San Francis | | JP2 |
| 850 | | PDF |
| - 2 | | |
| 3 | | |
| - (730) A | | |
| - 23 5 | | |
| - 6 | | |
| - 23 7 | | |
| - 🖉 8 | | |
| - 🔊 9 | | |
| - 🐼 10 | | |
| - 🐼 11 - | | |
| - 🐼 12 | | |
| 13 | | |
| → 🥥 14 ► 🗙 The San Francis | | |
| - The San Francis | | |
| - 🖉 1 | | |
| - 🐼 2 | | |
| - 3 | | |
| 4 | | |
| I ■ 5 | | |

Header data can only be viewed after a batch has been loaded or, when in Browse mode, after an image has been loaded for viewing. To view header data, click the appropriate tab on the right side of the viewing panel.

Example Header Data

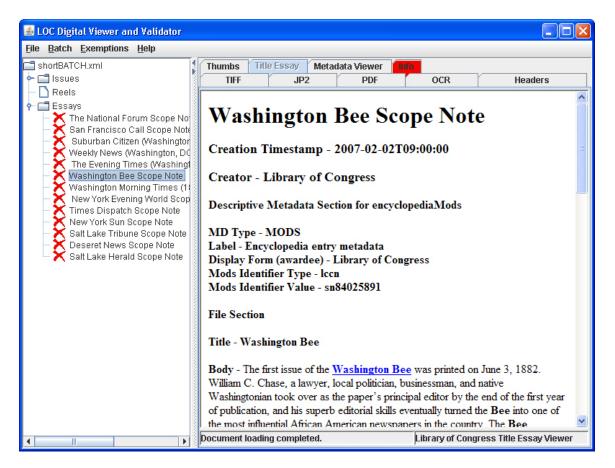


The data presented in the Header view panel is an HTML representation of the header data extracted from the image file. To afford ease of reformatting, the header data is extracted as XML and then converted to HTML via XSLT stylesheet processing. The properties used to specify the names of the XSLT rules file for each image type are discussed in the Post Install Configuration section.

Intermediate XML files are stored in the directory specified in the dvvbrowser.html.path property.

Viewing Title Essay Data

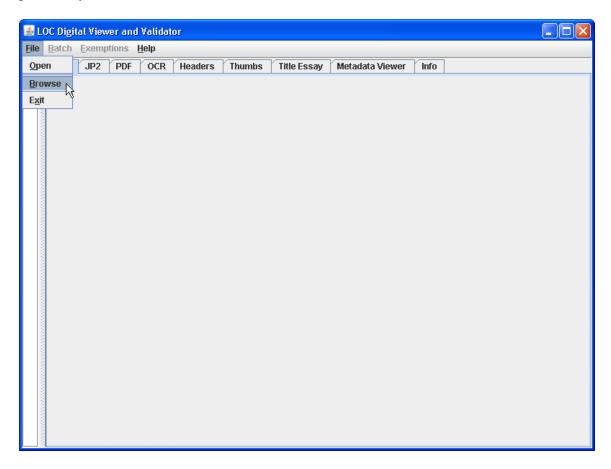
To expand the list of title essays, click on the toggle widget next to the folder labeled "Essays". The essay will be displayed in the panel on the right when the essay name is clicked.



8. Browse Mode

There may be times when it is desirable to browse images without performing the Batch Match or Validation functions. To use the DVV in Browse mode, select Browse from the File menu. A file dialog will be presented from which the starting point of the browse operation may be selected. Only directories can be selected when initiating Browse mode.

Browse mode does not use the metadata found in the batch, issue and reel files. As a result, loading the file structure of a large batch is much faster in Browse mode because those files don't need to be parsed. One side effect of this gain in performance is the loss of correlation between files that comprise a batch, issue or reel. Viewing a batch that observes the relationships established through metadata is accomplished with the previously described Batch View function.



An example of the file dialog is shown below. **File Dialog**

| 🕌 Open | | |
|------------------------|-------------------|-------------|
| Look <u>i</u> n: 🗐 I | on 'marine9' (K:) | |
| autorun | | |
| C Mac | | |
| sn8506638 | | |
| T WD_Windo |)ws_tools | |
| | | |
| | | |
| File <u>N</u> ame: | K:1 | |
| Files of <u>T</u> ype: | All Files | • |
| | | Open Cancel |

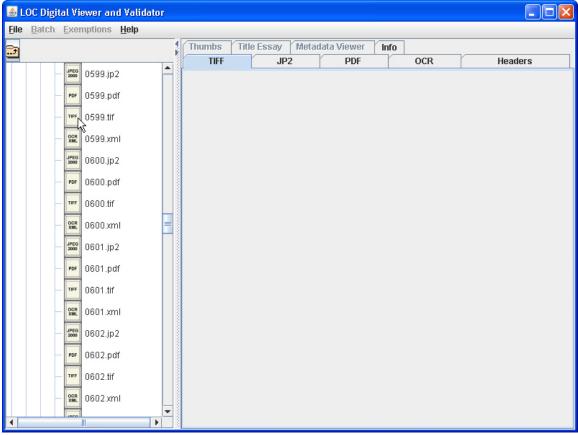
The File Tree

| 🕌 LOC Digital Viewer and V | /alidato | ır. | | | | | | | [| |
|---------------------------------|----------|-----|-----|-----|---------|--------|-------------|-----------------|------|--|
| <u>File Batch Exemptions He</u> | _ | | | | | | | | | |
| a | TIFF | JP2 | PDF | OCR | Headers | Thumbs | Title Essay | Metadata Viewer | Info | |
| | | | | | | | | | | |
| Mac | | | | | | | | | | |
| 🗢 🦳 autorun | | | | | | | | | | |
| ← 🛅 WD_Windows_Tools | | | | | | | | | | |
| - autorun.inf | | | | | | | | | | |
| ⊷ 🛅 sn85066387 | | | | | | | | | | |
| 🕶 🧰 Recycled | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |

The screen shot above shows the top level directories and files on the K drive. The icon above the drive letter is the Up Folder button, which will change the tree to display the parent folder of the one selected. When the selected folder has no parent, as shown here, the File Dialog will be presented when the Up Folder button is clicked.

Browsing Images

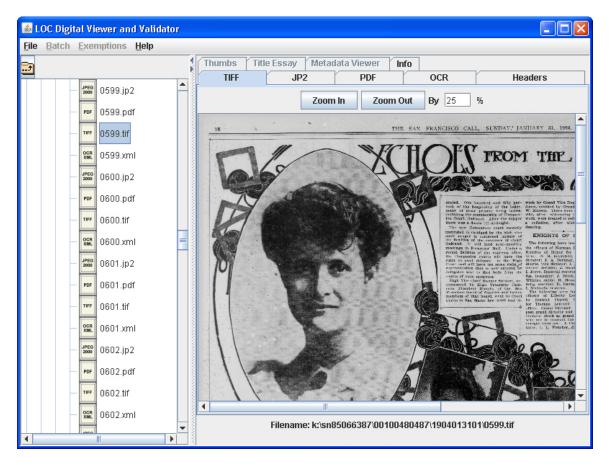
When in browse mode, selection of any of the four file types typically associated with a batch image will cause the clicked image to be loaded into the appropriate tab. In the example below, the icon for the image file 0599.tif is about to be clicked.



This will cause 0599.tif to be loaded and the Master tab will be selected. Another action resulting from this click is the creation of an image set, comprised of all image files typically associated with a batch that begin with the same filename (in this case 0599). When any of the tabs Master, Service, Derivative or OCR are clicked, the appropriate image set member will be loaded, if necessary, and displayed.

Clicking on a tab while another image is loading will display the current image, if there is one. A concurrent image load operation will not be initiated.

Image Loaded

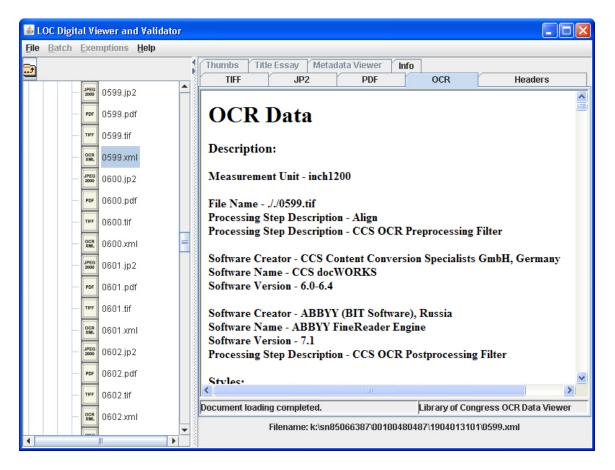


The relationship established between image files by metadata is not available when browsing, hence the absence of Next and Previous buttons. Zoom functionality is still available.

To view the images of another image set, click on an image file of a different name.

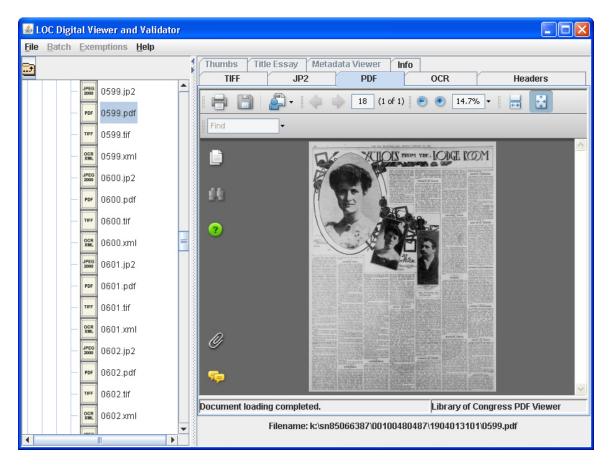
OCR Data Viewing

The output of OCR data interpretation is an XML file that specifies the makeup of an entire page, or series of pages. Each word on a page is represented as an XML element, with attributes that describe font styles, location, confidence level of successful word recognition, among other things. As might be expected, manually viewing an XML file of this breadth and detail is time consuming and prone to error. To partially alleviate these concerns, the OCR XML file is transformed, via XSLT, into a formatted web page to make viewing simpler. The OCR tab contains an embedded web browser in which the web page is displayed. As with Issue and Reel files, the XLST used to effect the transformation from XML to HTML may be customized as described in the section regarding Post Install Configuration. Shown below is an example of the OCR data as a web page.



PDF Viewer

Adobe Acrobat Reader 7.0 installs a PDF viewer plug-in for Internet Explorer. The DVV's embedded browser loads PDF files utilizing the Acrobat plug-in. The plug-in includes many features, explained in the Acrobat Reader documentation. An example of the PDF viewer panel is shown below:



Note that on systems that cannot accommodate a JVM size of 512MB, the load operation will be noticeably slow. The loading process is controlled entirely by the Acrobat Reader software and load times should match those experienced with a standalone instance of Acrobat Reader.

9. Command Line Usage of the Validation Processor

The Validation Processor is the command line component of the DVV and can be used to validate the contents of an entire batch, the contents of a single issue or reel within a batch, or a single image file. To run the Validation Processor, open a command (DOS) window and change to the directory in which the DVV was installed.

Open Command Window

First, click on 'Start' in the Windows menu bar.

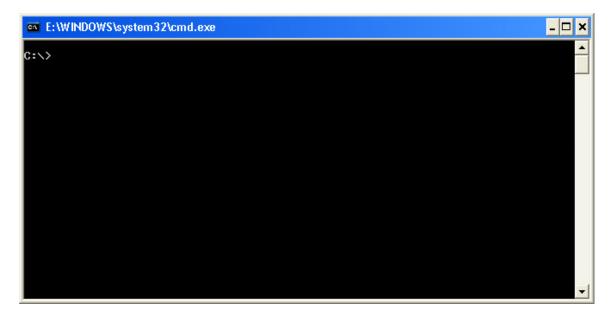


Then click 'Run' to initiate the following dialog:

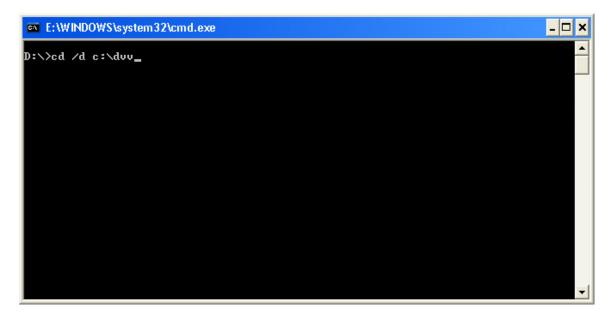
| Run | ? 🔀 |
|-------|--|
| - | Type the name of a program, folder, document, or Internet resource, and Windows will open it for you. |
| Open: | cmd |
| | OK Cancel Browse |

Make sure the value contained in the text field is 'cmd' (without the quotes) and then click 'OK'.

The Command Window

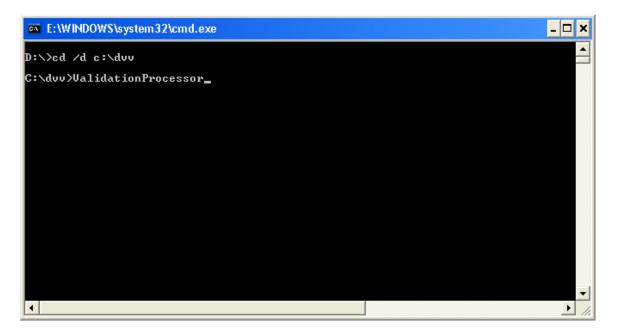


Change to the directory in which the DVV is installed as shown below:



The /d in the command above instructs the operating system to change drives, if necessary. In this case the command originates from the D drive and is switching to the C drive.

Then issue the following command:



to start the Validation Processor. Starting the Validation Processor with no arguments will cause the list of possible arguments to be displayed:

| C:\duv>validationprocessor |
|--|
| USAGE: |
| gov.loc.ndnp.preingest.validation.ValidationProcessor BATCHFILE filename [exemptionset_filename] gov.loc.ndnp.preingest.validation.ValidationProcessor BATCH filename no_updatelupdate_ignore_existing!no_update_ignore_ex |
| gov. loc.ndnp.preingest.validation.validationProcessor ENCVCLOPEDIAENTRYMEISECORD filename no_update_ignore_exist |
| yov.loc.ndnp.preingest.validation.ValidationProcessor NEWSPAPERTILLEMEISRECORDFILE filename [exemptionset filename] |
| gov.loc.ndnp.preingest.validation.UalidationProcessor NEWSPAPERTITLEMETSRECORD filename no_updatelupdate_ignore_existing} |
| gov.loc.ndnp.preingest.validation.ValidationProcessor ISSUEMETSRECORDFILE filename [exemptionset_filename] |
| gov.loc.ndnp.preingest.validation.ValidationProcessor <u>ISSUEMETSRECORD</u> filename no_updatelupdate_ignore_existing!no_update |
| gov.loc.ndnp.preingest.validation.ValidationProcessor REELMETSRECORDFLLE filename |
| gov.loc.ndnp.preingest.validation.ValidationProcessor REELMETSRECORD filename no_update lupdate_ignore_existingino_update_ gov.loc.ndnp.preingest.validation.ValidationProcessor PDF filename textixulinone [exemptionset_filename] |
| gov.loc.ndnp.preingest.validation.validationProcessor TIFF filename textixminione texemptionset_filename] gov.loc.ndnp.preingest.validation.ValidationProcessor TIFF filename textixminione texemptionset_filename] |
| gov.loc.ndnp.preingest.validation.ValidationProcessor MARCXMLBIBRECOR filename |
| gov.loc.ndnp.preingest.validation.ValidationProcessor MARCXMLHOLDINGRECORD filename |
| gov.loc.ndnp.preingest.validation.ValidationProcessor MODSBIBRECORD filename |
| gov.loc.ndnp.preingest.validation.ValidationProcessor ALTO filename [exemptionset_filename] |
| gov.loc.ndnp.preingest.validation.ValidationProcessor JPEC2000 filename textixmlinone [exemptionset_filename] gov.loc.ndnp.preingest.validation.ValidationProcessor BAG filename textixnlinone |
| gov. loc.indnp.preingest.validation.validationProcessor CREATEBAG batch filename |
| goving in the second second second second second second second in the second second in the second second second |
| Exemptionset_filename is optional. |
| Filename for BAG is the bag directory (not the data directory). |
| To create a bag, the batch must be in a directory named data. The bag will be create in place. |
| Exit of 0 indicates validation was successful |
| Exit of 1 indicates validation was unsurcessful Exit of 2 indicates that an error occurred |
| Exit of 2 indicates that an error occurred C. C. August and C. August an |
| |
| |
| |
| |
| |
| |
| |

When using the default configuration for a command window, the options will sometimes need to be wrapped onto the next line, making it difficult to read. Instructions for changing the properties of the DOS window are contained in a later chapter. The default logging options for the Validation Processor will produce many lines of output, indicating that particular attention should be paid to the Screen Buffer Size height option, which should be set to 9999.

Validation Processor Options

As shown earlier, the list of possible arguments for the Validation Processor is as follows.

Notes about argument format. Arguments underlined below, such as <u>filename</u>, indicate that the argument must be replaced by the user with an appropriate value. Arguments enclosed in braces, such as {no_update | update | update_ignore_existing }, indicate that one of the enclosed values, must be supplied as shown. A vertical bar separates each of the possible values. An argument enclosed with brackets is optional.

If accessing a batch located on a remote drive, ensure that the remote drive is mapped locally (using Tools, Map Network Drive from the Windows File Manager interface). Do not specify the Universal Naming Convention name, such as:

\\machine99\d\$\batch.xml

See the earlier section, **Opening a Batch**, for a warning about spaces in file paths.

In this example, the "All Users" directory short name is "ALLUSE~1".

BATCHFILE <u>filename</u> [exemptionset_filename]

Invocation of the Validation Processor with the BATCHFILE and filename arguments causes the file of the supplied filename to be validated. This produces an analysis of the batch file only. No other files in the batch are validated and no validation signature is generated for the batch file. Specification of an exemption set filename will cause the exemptions described therein to be considered during the validation process.

BATCH <u>filename</u> {no_update [exemptionset_filename] |

update_ignore_existing [exemptionset_filename] |
no_update_ignore_existing [exemptionset_filename] |
update(validation signatures) [exemptionset_filename] |
verify [exemptionset_filename] |
verify_plan_for_restart output_log_file exemptionset_filename] |
verify perform restart input log file [exemptionset filename] }

Running the Validation Processor with the BATCH, filename and update type arguments is used to perform validation on an entire batch. The filename argument specifies the name of the batch file. Specification of an exemption set filename will cause the exemptions described therein to be considered during validation process.

no_update also causes all issues and reels within a batch to undergo validation, but no digital signatures will be generated. As with the update option, verification of the signatures of previously validated files is performed.

update_ignore_existing causes all issues/reels within a batch to be validated regardless of the value of an existing signature, which will be updated if necessary.

no_update_ignore_existing a combination of no_update and update_ignore_existing, causes all issues/reels within a batch to be validated regardless of the presence of an existing signature.

If update is supplied as the update type argument, validation is attempted only on batch entities that do not have existing, valid digital signatures. By ignoring issues/reels with a valid digital signature, this option avoids redundant processing. Verification, however, will be performed to ensure that the digital signatures are valid.

verify issued to cause verification of all digital signatures. No validation is performed and no signatures are updated. Exemption sets are not used during verify processing and will be ignored if included as a parameter.

verify_plan_for_restart output_log_filename is used to start the verification process, but also log the progress of the verification to the named output log file. If the verification fails, the following command can be used to restart the verification at the point of failure.

verify_perform_restart input_log_filename starts the verification process at the point of failure. Before a file in the batch is verified, a check is made to see if it has already been verified (its presence in the log signifies that it has). If so, the next file in the batch is processed.

NEWSPAPERTITLEMETSRECORDFILE <u>filename</u> [exemptionset_filename] This processing option is used to validate a Mets Record File for the title represented by the filename argument. Newspaper Title Mets Record Files are created as part of the repository ingest process and, as such, not available to awardees. This option has no value in the awardee environment.

NEWSPAPERTITLEMETSRECORD <u>filename</u> {no_update |

update_ignore_existing | no_update_ignore_existing | update (validation signatures) } [exemptionset filename]

The NEWSPAPERTITLEMETSRECORD argument is used with the filename and update type options to validate the contents of a title, to include the Mets Record File and all issue and reel files contained therein. Update type options are as described previously in the discussion of the BATCH option. Newspaper Title Mets Record Files are created as part of the repository ingest process and, as such, not available to awardees. This option has no value in the awardee environment.

ISSUEMETSRECORDFILE <u>filename</u> [exemptionset_filename]

This processing option is used to validate a Mets Record File for the issue represented by the filename argument. It does not validate any other files in the issue, just the Mets Record File ("issuename".xml), and no validation signature is generated.

ISSUEMETSRECORD <u>filename</u> {no_update | update_ignore_existing | no_update_ignore_existing update(validation signatures) | verify} exemptionset_filename]

The ISSUEMETSRECORD argument is used with the filename and update type options to validate the contents of an issue, to include the Mets Record File and all issue files contained therein. Update type options are as described previously in the discussion of the BATCH option.

REELMETSRECORDFILE <u>filename</u>

This processing option is used to validate a Mets Record File for the reel represented by the filename argument. It does not validate any other files in the reel, just the Mets Record File ("reelname".xml), and no validation signature is generated.

REELMETSRECORD <u>filename</u> {no_update | update_ignore_existing | no_update_ignore_existing update(validation signatures) | verify} exemptionset_filename]

The REELMETSRECORD argument is used with the filename and update type options to validate the contents of a reel, to include the Mets Record File and all reel files contained therein. Update type options are as described previously in the discussion of the BATCH option.

PDF filename

This processing option is used to validate the PDF (derivative) image file named by the filename argument. No validation signature is generated when this processing option is used.

TIFF filename

This processing option is used to validate the TIFF (master) image file named by the filename argument. No validation signature is generated when this processing option is used.

ALTO filename

This processing option is used to validate the ALTO (OCR) image file named by the filename argument. No validation signature is generated when this processing option is used.

JPEG2000 <u>filename</u>

This processing option is used to validate the JPEG2000 (service) image file named by the filename argument. No validation signature is generated when this processing option is used.

MARCXMLBIBRECORD filename

This processing option is used to validate the MARC XML Bibliographic file named by the filename argument. No validation signature is generated when this processing option is used.

MARCXMLHOLDINGRECORD filename

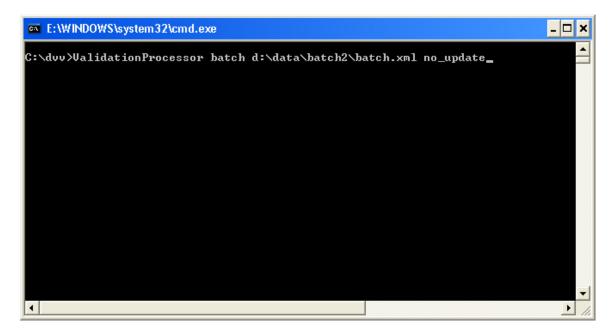
This processing option is used to validate the MARC XML Holding Record file named by the filename argument. No validation signature is generated when this processing option is used.

MODSBIBRECORD filename

This function has not been implemented.

Validation Examples

BATCH filename



In the example above, the batch identified by d:\batch2\batch.xml will be validated with the no_update option.

| E:\WINDOWS\system32\cmd.exe |
|---|
| Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEUERE: UALIDATION: Failure: Original related item must have a physicalDescripti 1/mods:mods[1]/mods:relatedItem[1] (D:\data\batch2\sn84026749\100492350\10049235 Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEUERE: UALIDATION: Failure: Original related item must have a physicalDescripti 1/mods:mods[1]/mods:relatedItem[1] (D:\data\batch2\sn84026749\100492350\10049235 Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEUERE: UALIDATION: Failure: Target div child of techtargetreel div must have a iv[1]/div[1]/div[1] (D:\data\batch2\sn84026749\100492350\100492350\xml, 2005-11- Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEVERE: UALIDATION: Failure: Target div child of techtargetreel div must have a iv[1]/div[1]/div[1] (D:\data\batch2\sn84026749\100492350\100492350\xml, 2005-11- Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEVERE: UALIDATION: Failure: Target div child of techtargetreel div must have a iv[1]/div[1]/div[2] (D:\data\batch2\sn84026749\100492350\100492350\xml, 2005-11- Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEVERE: UALIDATION: Failure: Target div child of techtargetreel div must have a iv[1]/div[1]/div[3] (D:\data\batch2\sn84026749\100492350\100492350\xml, 2005-11- Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEVERE: UALIDATION: Failure: Target div child of techtargetreel div must have a iv[1]/div[1]/div[4] (D:\data\batch2\sn84026749\100492350\100492350\xml, 2005-11- Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEVERE: UALIDATION: Failure: Target div child of techtargetreel div must have a iv[1]/div[1]/div[4] (D:\data\batch2\sn84026749\100492350\100492350\xml, 2005-11- Aug 2, 2008 3:54:00 PM gov.loc.ndnp.preingest.utilities.logging.JavaUtilPreinges SEVERE: UALIDATION: Failure |
| iv[1]/div[1]/div[5] (D:\data\batch2\sn84026749\100492350\100492350.xml, 2005-11- Completed in 9 seconds false C:\dvv> |
| |

A successful validation will produce many lines of output ending with the number of seconds taken to validate the batch and the last line will contain the single word "true".

A batch that has failed validation, shown above, is indicated by the single word "false" as the last line in the output.

Error Messages

Error messages reported by the Validation Processor typically take the following form: (this message split into three separate lines for the sake of clarity)

SEVERE: VALIDATIONERROR: Stream places multiple text strings in same location (1 0 0 1 482 1170.17)

f:\sn82015056\100492945\1910091701\0050.pdf, 2005-07-08T13:05:06.000-0400)

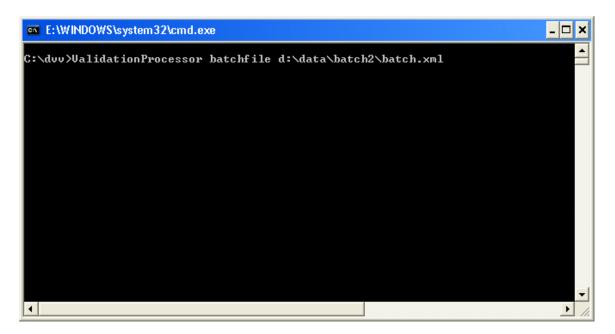
at gov.loc.ndnp.ingest.validation.NdnpLoggerErrorHandler

The first line describes the error.

The second line is the filename, plus last date/time modified.

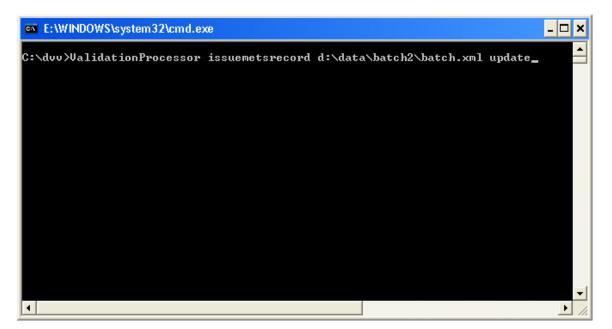
The third line is the location within the Validation Library from which the error message was reported.

BATCHFILE <u>filename</u>



An example of the BATCHFILE option. Note that the update type argument is not present.

ISSUEMETSRECORD <u>filename</u>



The command prompt has been shifted in the above screenshot to ensure that all ValidationProcessor arguments are visible.

Not all possible argument possibilities have been shown due to the inherent similarities in all of them.

Validation Signatures

After a batch, issue or reel has been validated, a digital signature will be created for the respective batch if any, not necessarily all, content within the issue/reel is valid. This avoids the need to revalidate the entire issue/reel during subsequent validation runs. As an example, consider a batch with five metadata files:

| 🔁 E:\tmp\batchA | | |
|--------------------------|----------------|-------------------|
| File Edit View Favorites | Tools Help | h5 🔢 |
| 🗢 Back 🔹 🤿 👻 🔯 Se | arch 🔓 Folders | 🎯 🖻 🗣 🔹 |
| Address 🚞 E:\tmp\batchA | | ▼ @Go |
| Name 🛆 | Size | Modified |
| 🔮 batchA.xml | 1 KB | 9/23/2005 1:59 PM |
| 🔮 100492945.xml | 30 KB | 9/24/2005 7:59 AM |
| 😫 1910091001.×ml | 13 KB | 7/8/2005 12:50 PM |
| 🔮 1910091701.×ml | 13 KB | 7/8/2005 12:59 PM |
| 1910092401.×ml | 13 KB | 7/8/2005 1:07 PM |
| | | |
| | | |
| 5 object(s) 66.9 |) КВ 📃 Му | Computer // |

The files in the above illustration were copied into a common directory for the purpose of displaying the attributes of all five at the same time. The files used in the actual validation process are physically located in other directories.

File Names

The name of the first signed batch, issue or reel file to be generated will be the original name of the batch, issue or reel file with _1 appended to the file name, preceding the file suffix. Shown below are the batch, issue and reel files after a validation run in which an error was reported. Even though a validation error was reported, all files received a validation signature.

| 🔁 E:\tmp\batchA | | | 、 | | × |
|---------------------------------|-------|-----------|--------------|-------------|-----|
| File Edit View Favorites | Tools | Help I | 4 | | ŀ |
| 🗘 Back 🔹 🤿 👻 🔯 🥘 Se | earch | 🔁 Folders | 3 B | Ì. | » |
| Address 🔂 E:\tmp\batchA | | | • | <i>∂</i> Go | 1 |
| Name 🛆 | | Size | Modified | | |
| 😫 batcha_1.xml | | 2 KB | 9/25/2005 9: | 02 AM | |
| 🔮 100492945_1.×ml | | 61 KB | 9/25/2005 9: | 02 AM | |
| 1910091001_1.xml | | 40 KB | 9/25/2005 8: | 58 AM | |
| 🔮 1910091701_1.xml | | 40 KB | 9/25/2005 8: | 59 AM | |
| 🕋 1910092401_1.xml | | 37 KB | 9/25/2005 9: | DO AM 🖥 | ₹I |
| • | | | | ₽ | |
| Type: XML Document Size: 36, 36 | .6 KB | 🖳 My | Computer | | //, |

Of note is the larger size of each metadata file. This is because of the digital signature that is contained within the generated file. A subsequent validation run against this batch, using batchA.xml as the input file, would cause all of these signature files to be updated.

To illustrate this point, the error message reported for this validation run was:

SEVERE: VALIDATIONERROR:

Scanner model name (Tag 272) does not contain scanner serial number (NDNP Profile) (f:\sn82015056\100492945\1910092401\0054.tif, 2005-06-22T10:10:26.000-0400) at gov.loc.ndnp.ingest.validation.NdnpLoggerErrorHandler

Revalidation

Correcting the problem, by replacing the TIFF file with a valid image and revalidating the batch, using batchA.xml as the input file, causes updates to the same five files, as shown in the following screen shot.

| 🔁 E:\tmp\batchA | | | N | |
|--------------------------|-------|---------|-------------|------------|
| File Edit View Favorites | Tools | ; Help | 43 | - |
| 🔃 Back 🔹 🔿 👻 🔯 Se | earch | Folders | 🕲 Pir 🖣 | e » |
| Address 🔂 E:\tmp\batchA | | | • |] @Go |
| Name 🗸 | | Size | Modified | |
| 1910092401_1.xml | | 40 KB | 9/25/2005 1 | 0:03 AM |
| 🔮 1910091701_1.xml | | 40 KB | 9/25/2005 1 | 0:02 AM |
| 1910091001_1.xml | | 40 KB | 9/25/2005 1 | 0:01 AM |
| 100492945_1.xml | | 61 KB | 9/25/2005 1 | 0:05 AM |
| 🔮 batcha_1.xml | | 2 KB | 9/25/2005 1 | 0:05 Alv 🖵 |
| • | | | | F |
| 15 object(s) 42 | 2 KB | 🖳 My | Computer | |

Because "update" was supplied as the update type argument, the existing signed metadata files were updated. Of note is the modification time on each of the files. Each of these validation runs took roughly the same amount of time, indicating that all resources in the batch were validated. This is because the "original" metadata files were used to determine if validation had been previously performed on any batch resources. Since the validation results were recorded within the "_1.xml" metadata files, and batchA.xml was used as input, the Validation Processor was unable to determine that validation had previously occurred. Had this been a large batch, with perhaps thousands of files, the amount of time spent revalidating files would have been costly.

To avoid this problem, always use the batch metadata file with the highest generated number. To continue this example, validation was performed using batchA_1.xml as the input file.

The result of revalidation when using batchA_1.xml as input was that the entire run took significantly less time – 1 minute as opposed to about 4 minutes, and a new file, batchA_2.xml was generated.

The creation of the new file, batchA_2.xml, prompts a question as to why the Validation Processor did not just update file batchA_1.xml rather than creating a new file. The answer is that a signature file reflects the state of *another* resource within the batch and cannot logically reflect its own state. The convention is that the Validation Processor knows that it generated the _1 and increments that value by one to create the new file name.

Thusly, a new batch, issue or reel metadata file is created each time the Validation Processor must record the state of a resource, within the scope of that metadata file, that has not previously been recorded. In this case, batchA_1.xml was the input to the Validation Processor and it cannot record its own state. As such, the Validation Processor creates a new file, based on the name of the input file.

Modification of a Validated Resource

The modification of a resource that has been previously validated will cause the related signature file to become invalid. As an example, let it be stated that master file 0009.tif, whose validated state is recorded in file 1910092401_1.xml, has been replaced subsequent to the validation which created 1910092401_1.xml. If a validation is performed, using batchA_1.xml as input, a message very close in text to the following will be generated:

SEVERE: VALIDATIONERROR: Master or service file has wrong number of ADMID references at mets[1]/fileSec[1]/fileGrp[3]/file[1] for f:\sn82015056\100492945\1910092401\1910092401_1.xml (after modification by validator) at gov.loc.ndnp.ingest.validation.NdnpLoggerErrorHandler

The key to this message is the phrase "(after modification by validator)". This is most often an indication that something within the batch has been modified after it has been validated and signed. When this happens the entire batch must be revalidated to ensure that everything is properly signed.

Capturing Validation Processor Messages

When processing a large batch, it is quite possible that the number of messages generated by the Validation Processor will exceed the number allowed by the DOS window. To ensure that no messages are lost, the Validation Processor also logs all messages to a file in the VP_HOME directory named ValidationLibrary.log. If the installation defaults are chosen, VP_HOME will be c:\dvv.

The log file is reused to conserve space, but the logging configuration can be customized to create a new file each time the Validation Processor is executed. Changing the defaults is accomplished by editing the file named logging.properties in the VP_HOME directory.

To create a unique log filename each time the Validation Processor runs, add %u to the filename, just before the suffix in the java.util.logging.FileHandler.pattern property. For example:

This is the default: java.util.logging.FileHandler.pattern = ../validationprocessor.log

Change it to this: java.util.logging.FileHandler.pattern = ../validationprocessor%u.log

To save the logs in a location other than VP_HOME, change the relative directory and filename supplied to any other relative or absolute directory and filename. If, for instance, the objective is to save all log files, with unique names, in a directory named D:\LOC\logs, the java.util.logging.FileHandler.pattern property would be set to:

java.util.logging.FileHandler.pattern = D:/LOC/logs/validationprocessor%u.log

Notice the use of the forward slashes.

Logging Levels

The amount of detail captured in the log, and displayed in the DOS window, is controlled by the **level** properties. Possible values for each of the three level properties, from lowest level of detail to highest level of detail are OFF, SEVERE, WARNING, INFO, FINE, FINER, FINEST, ALL.

As might be expected, a specification of OFF turns off logging for the facility specified.

INFO is the default level for the Validation Library. It provides information in an amount that can be useful, but not overwhelming. The effect on performance is deemed acceptable. Specification of levels "above" INFO are not recommended.

SEVERE is used when only error messages are needed.

The java.util.logging.FileHandler.level property controls the level of detail going into the log file.

The java.util.logging.ConsoleHandler.level property controls the granularity of information displayed in the DOS window.

The .level property specifies the level for all logging facilities. Notice the . before level. The value of this property is set to OFF for the Validation Processor and overridden to a higher level for each validation processor facility as needed.

Note that the verification setting defaults to off. To receive messages generated by the Validation Processor pertaining to verification, perform the following edit:

Before:

#VERIFICATION.level = OFF

After:

VERIFICATION.level = INFO

10. Supplied logging.properties

Shown below is a copy of the logging.properties file as supplied. Properties mentioned are in bold type.

"handlers" specifies a comma separated list of log Handler
classes. These handlers will be installed during VM startup.
Note that these classes must be on the system classpath.
By default we only configure a ConsoleHandler, which will only
show messages at the INFO and above levels.
#handlers= java.util.logging.ConsoleHandler

To also add the FileHandler, use the following line instead. handlers= java.util.logging.FileHandler, java.util.logging.ConsoleHandler

Default global logging level.

This specifies which kinds of events are logged across

all loggers. For any given facility this global level

can be overriden by a facility specific level

Note that the ConsoleHandler also has a separate level

setting to limit messages printed to the console.

.level= OFF

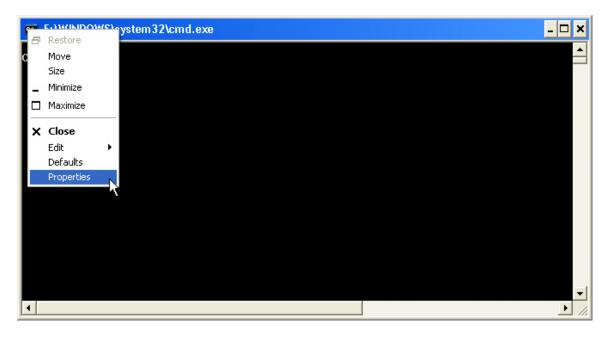
default file output is in DVV install directory. java.util.logging.FileHandler.pattern = validationprocessor.log java.util.logging.FileHandler.limit = 0 java.util.logging.FileHandler.count = 1 java.util.logging.FileHandler.level = INFO # java.util.logging.FileHandler.formatter = java.util.logging.XMLFormatter java.util.logging.FileHandler.formatter = java.util.logging.SimpleFormatter

Limit the message that are printed on the console to INFO and above. java.util.logging.ConsoleHandler.level = INFO java.util.logging.ConsoleHandler.formatter = java.util.logging.SimpleFormatter

For example, set the com.xyz.foo logger to only log SEVERE
messages:
gov.loc.ndnp.level = SEVERE
#edu.harvard.hul.ois.jhove = OFF
VALIDATION.level = INFO
INTEGRITY.level = SEVERE
#VERIFICATION.level = OFF
#INGEST.level = OFF

11. Changing DOS Window Properties

To change the properties of the DOS window, click in the upper left corner of the window itself:



Then click on the Properties option:

| E:\WINDOWS\system32\cmd.e | exe" Properties | ?× |
|--|------------------------------------|---------------------------------------|
| Image: Second constraints Options Font Layout Colors Cursor Size Small Medium Large Command History Buffer Size: So Mumber of Buffers: Discard Old Duplicates | Exe" Properties Display Options | · · · · · · · · · · · · · · · · · · · |
| | OK Ca | ancel |

To change the layout of the command window, click on the Layout tab:

| "E:\WINDOWS\system32\ci | md.exe" Properti | es 🕜 🔀 |
|--|--|---|
| Options Font Layout Colo | urs | |
| Options Font Layout Cold Window Preview | Screen Buffer Size <u>W</u> idth: <u>H</u> eight: Window Size Wjdth: H <u>e</u> ight: Window Position <u>L</u> eft: <u>I</u> op: ▼ <u>Let system positi</u> | 140 • 300 • 120 • 27 • 13 • 52 • 52 • |
| | OK | Cancel |

and then change the value of Screen Buffer Size Width to 140 and the value of the Window Size Width to 120. This will make the window wide enough to show all of the possible arguments on a single line each. Since each line is now wider than the screen width, a scroll bar is added, by the system, to enable viewing the entire line.

The value of the Screen Buffer Height value should be set to 9999.

The result:

| C:\>cd duw C:\>cd duw C:\>cd duw SGGE: gou lcc.ndnp.preingest.ualidation.UalidationProcessor BATCHFILE filename [exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor BATCHFILE filename no_updatelupdate_ignore_exist gou.lcc.ndnp.preingest.ualidation.UalidationProcessor ENCYCLOPEDIGENTRYMETSRECORD filename no_updatelupdate_ignore_exist gou.lcc.ndnp.preingest.ualidation.UalidationProcessor NEUSPOPERTITLEMETSRECORD filename no_updatelupdate_ignore_exist gou.lcc.ndnp.preingest.ualidation.UalidationProcessor NEUSPOPERTITLEMETSRECORD filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor ISSUEMETSRECORD file filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor ISSUEMETSRECORD file filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor ISSUEMETSRECORD file filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor ISSUEMETSRECORD file filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor REELMETSRECORD file filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor REELMETSRECORD file filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor REELMETSRECORD file (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMMETSRECORD filename (exemptionset_filename] gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ualidation.UalidationProcessor MERCMLINEDENCORD filename gou.lcc.ndnp.preingest.ua | E:\WINDOWS\system32\cmd.exe |
|--|--|
| | <pre>rdow.WalidationProcessor SAGE: bu. loc.ndmp.preingest.walidation.UalidationProcessor bu. loc.ndmp.preingest.validation.UalidationProcessor bu. loc.ndmp</pre> |

Scrolled to show most arguments, keep scrolling right to see more:

| E:\WINDOWS\system32\cmd.exe | × |
|---|----------|
| | |
| BATCHFILE filename [exemptionset_filename] | |
| BATCH filename no_update/update_ignore_existing¦no_update_ignore_existing¦update(validation_signatures)¦verify [exempti ENCYCLOPEDIAENTRYMETSRECORD filename no_update!update_ignore_existing¦no_update_ignore_existing update(validation_signa NEWSPAPERTITLEMETSRECORDFILE filename [exemptionset_filename] | .o it |
| NEWSPAPERTITLEMETSRECORD filename no_updatelupdate_ignore_existing no_update_ignore_existing update(validation signatur ISSUEMETSRECORDFILE filename [exemptionset_filename] | |
| ISSUEMETSRECORD filename no_updatelupdate_ignore_existing!no_update_ignore_existing!update(validation signatures) [exem REELMETSRECORDFILE filename REELMETSRECORD filename no_update!update_ignore_existing!no_update_ignore_existing!update(validation signatures) [exemp | |
| PDF filename textixmlinone [exemptionset_filename] TIFF filename textixmlinone [exemptionset_filename] | |
| MARCXHLBIBRECORD filename MARCXMLHOLDINGRECORD filename MODSBIBRECORD filename | |
| ALTO filename [exemptionset_filename] JPEG2000 filename text;xml;none [exemptionset_filename] | |
| xt. Select xml to output representation information as xml. Select none to not output representation information. | |
| | |
| | - |
| | |

12. JPEG2000 Plug-ins

There are a number of JPEG2000 plug-ins available. Documentation for each can be obtained from their respective websites:

http://www.lizardtech.com

http://www.luratech.com

http://www.ermapper.com

13. Benchmarks

Benchmark testing has been performed to provide an estimate of the time necessary to validate batches of various sizes. The computers used in the benchmarks had the following configurations.

Caveat – The statistics regarding PDF images have been superceded by the implementation of a new PDF viewer. Load times for PDF files in the current implmentation are very similar to times mentioned for TIFF files. Load times mentioned here for JP2 files are reflective of default functionality and have no meaning when a third party JPEG2000 plug-in has been selected as the JP2 viewer.

System A

Processor – AMD Athlon XP 2700+ (2166mhz) RAM – 1.5 GB PC3200 DDR SDRAM (200mhz) Input Hard Drive – LaCie FW800, firewire connection JVM Memory Specification 512M

System B

Processor – AMD Duron (800mhz) RAM – 1.5 GB PC100 ECC SDRAM (100mhz) Input Hard Drive – LaCie FW800, 100mbps LAN connection JVM Memory Specification 512M

System C

Processor – AMD Athlon (600mhz) RAM – 384 MB PC100 SDRAM (100mhz) Input Hard Drive – LaCie FW800, 100mbps LAN connection JVM Memory Specification 512M

Batch Composition

17 Issues

320 Files 2,776,850,364 Bytes

1 Reel

73 Files 696,800,728 Bytes

393 Files 3,473,651,092 Bytes

Batch Processing Statistics

| | Elapsed Time | Seconds/File | Bytes/Second |
|----------|------------------------------|--------------|--------------|
| System A | 32 Min 50 Sec (1970 sec) | 5.012 | 1,763,274 |
| System B | 1 Hr 59 Min 2 Sec (7142 sec) | 18.173 | 486,369 |
| System C | 2 Hr 39 Min 8 Sec | 24.295 | 363,809 |

Individual Image Load Times (All load times in seconds)

| | System A | System B | System C | File Size |
|--------|-----------|-----------|-----------|------------|
| | Load Time | Load Time | Load Time | |
| TIFF 1 | 5.219 | 8.812 | 9.544 | 33,187,284 |
| TIFF 2 | 3.656 | 6.078 | 5.949 | 33,386,837 |
| TIFF 3 | 3.578 | 5.829 | 5.639 | 34,748,209 |
| TIFF 4 | 4.078 | 5.127 | 6.369 | 33,973,832 |
| OCR 1 | 1.235 | 4.346 | 6.079 | 1,113,867 |
| OCR 2 | 1.812 | 6.520 | 7.380 | 1,874,634 |
| OCR 3 | 1.438 | 4.636 | 4.857 | 1,492,965 |
| OCR 4 | 1.250 | 4.006 | 4.627 | 1,288,908 |
| PDF 1 | 16.734 | 51.248 | 97.931 | 1,660,272 |
| PDF 2 | 21.953 | 75.338 | 138.710 | 1,823,497 |
| PDF 3 | 20.156 | 66.365 | 234.107 | 1,755,839 |
| PDF 4 | 18.766 | 61.849 | 225.574 | 1,824,611 |
| JP2 1 | 28.672 | 76.620 | 84.942 | 4,149,444 |
| JP2 2 | 26.937 | 88.547 | 83.641 | 4,174,393 |
| JP2 3 | 29.078 | 89.969 | 87.696 | 4,344,565 |
| JP2 4 | 28.672 | 88.217 | 85.833 | 4,247,739 |

Average Load Time

| | System A | System B | System C |
|------|----------|----------|----------|
| TIFF | 4.132 | 6.461 | 6.875 |
| OCR | 1.434 | 4.877 | 5.736 |
| PDF | 19.402 | 63.7 | 174.08 |
| JP2 | 28.34 | 85.838 | 85.528 |

Average Image Size

| TIFF | 33,824,040 |
|------|------------|
| OCR | 1,442,593 |
| PDF | 1,766,054 |
| JP2 | 4,229,035 |

Appendix A

Third Party Licensing

This appendix contains the licenses of third party software included as part of the distribution of the Validation Library.

JHove

The most current version of this license is available at:

http://hul.harvard.edu/jhove/index.html (This is not a direct link. Scroll to the License section.)

and

http://www.gnu.org/licenses/licenses.html#LGPL

JHOVE - JSTOR/Harvard Object Validation Environment Copyright 2003-2005 by JSTOR and the President and Fellows of Harvard College

This program is free software; you can redistribute it and/or modify it under the terms of the GNU Lessor General Public License as published by the Free Software Foundation; either version 2.1 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Sun Microsystems Java Runtime Environment

The most current version of this license is available at:

http://java.sun.com/j2se/1.5.0/jre-1_5_0_05-license.txt

Sun Microsystems, Inc. Binary Code License Agreement

for the JAVA 2 PLATFORM STANDARD EDITION RUNTIME ENVIRONMENT $5.0\,$

SUN MICROSYSTEMS, INC. ("SUN") IS WILLING TO LICENSE THE

SOFTWARE IDENTIFIED BELOW TO YOU ONLY UPON THE CONDITION THAT YOU ACCEPT ALL OF THE TERMS CONTAINED IN THIS BINARY CODE LICENSE AGREEMENT AND SUPPLEMENTAL LICENSE TERMS (COLLECTIVELY "AGREEMENT"). PLEASE READ THE AGREEMENT CAREFULLY. BY DOWNLOADING OR INSTALLING THIS SOFTWARE, YOU ACCEPT THE TERMS OF THE AGREEMENT. INDICATE ACCEPTANCE BY SELECTING THE "ACCEPT" BUTTON AT THE BOTTOM OF THE AGREEMENT. IF YOU ARE NOT WILLING TO BE BOUND BY ALL THE TERMS, SELECT THE "DECLINE" BUTTON AT THE BOTTOM OF THE AGREEMENT AND THE DOWNLOAD OR INSTALL PROCESS WILL NOT CONTINUE.

1. DEFINITIONS. "Software" means the identified above in binary form, any other machine readable materials (including, but not limited to, libraries, source files, header files, and data files), any updates or error corrections provided by Sun, and any user manuals, programming guides and other documentation provided to you by Sun under this Agreement. "Programs" mean Java applets and applications intended to run on the Java 2 Platform Standard Edition (J2SE platform) platform on Java-enabled general purpose desktop computers and servers.

2. LICENSE TO USE. Subject to the terms and conditions of this Agreement, including, but not limited to the Java Technology Restrictions of the Supplemental License Terms, Sun grants you a non-exclusive, non-transferable, limited license without license fees to reproduce and use internally Software complete and unmodified for the sole purpose of running Programs. Additional licenses for developers and/or publishers are granted in the Supplemental License Terms.

3. RESTRICTIONS. Software is confidential and copyrighted. Title to Software and all associated intellectual property rights is retained by Sun and/or its licensors. Unless enforcement is prohibited by applicable law, you may not modify, decompile, or reverse engineer Software. You acknowledge that Licensed Software is not designed or intended for use in the design, construction, operation or maintenance of any nuclear facility. Sun Microsystems, Inc. disclaims any express or implied warranty of fitness for such uses. No right, title or interest in or to any trademark, service mark, logo or trade name of Sun or its licensors is granted under this Agreement. Additional restrictions for developers and/or publishers licenses are set forth in the Supplemental License Terms.

4. LIMITED WARRANTY. Sun warrants to you that for a period of ninety (90) days from the date of purchase, as evidenced by a copy of the receipt, the media on which Software is furnished (if any) will be free of defects in materials and workmanship under normal use. Except for the foregoing, Software is provided "AS IS". Your exclusive remedy and Sun's entire liability under this limited warranty will be at Sun's option to replace Software media or refund the fee paid for Software. Any implied warranties on the Software

are limited to 90 days. Some states do not allow limitations on duration of an implied warranty, so the above may not apply to you. This limited warranty gives you specific legal rights. You may have others, which vary from state to state.

5. DISCLAIMER OF WARRANTY. UNLESS SPECIFIED IN THIS AGREEMENT, ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT THESE DISCLAIMERS ARE HELD TO BE LEGALLY INVALID.

6. LIMITATION OF LIABILITY. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL SUN OR ITS LICENSORS BE LIABLE FOR ANY LOST REVENUE, PROFIT OR DATA, OR FOR SPECIAL, INDIRECT, CONSEQUENTIAL, INCIDENTAL OR PUNITIVE DAMAGES, HOWEVER CAUSED REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF OR RELATED TO THE USE OF OR INABILITY TO USE SOFTWARE, EVEN IF SUN HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In no event will Sun's liability to you, whether in contract, tort (including negligence), or otherwise, exceed the amount paid by you for Software under this Agreement. The foregoing limitations will apply even if the above stated warranty fails of its essential purpose. Some states do not allow the exclusion of incidental or consequential damages, so some of the terms above may not be applicable to you.

7. TERMINATION. This Agreement is effective until terminated. You may terminate this Agreement at any time by destroying all copies of Software. This Agreement will terminate immediately without notice from Sun if you fail to comply with any provision of this Agreement. Either party may terminate this Agreement immediately should any Software become, or in either party's opinion be likely to become, the subject of a claim of infringement of any intellectual property right. Upon Termination, you must destroy all copies of Software.

8. EXPORT REGULATIONS. All Software and technical data delivered under this Agreement are subject to US export control laws and may be subject to export or import regulations in other countries. You agree to comply strictly with all such laws and regulations and acknowledge that you have the responsibility to obtain such licenses to export, re-export, or import as may be required after delivery to you.

9. TRADEMARKS AND LOGOS. You acknowledge and agree as between you and Sun that Sun owns the SUN, SOLARIS, JAVA, JINI, FORTE, and iPLANET trademarks and all SUN, SOLARIS, JAVA, JINI, FORTE, and iPLANET-related trademarks, service marks, logos and other brand designations ("Sun Marks"), and you agree to comply with the Sun Trademark and Logo Usage Requirements currently located at

http://www.sun.com/policies/trademarks. Any use you make of the Sun Marks inures to Sun's benefit.

10. U.S. GOVERNMENT RESTRICTED RIGHTS. If Software is being acquired by or on behalf of the U.S. Government or by a U.S. Government prime contractor or subcontractor (at any tier), then the Government's rights in Software and accompanying documentation will be only as set forth in this Agreement; this is in accordance with 48 CFR 227.7201 through 227.7202-4 (for Department of Defense (DOD) acquisitions) and with 48 CFR 2.101 and 12.212 (for non-DOD acquisitions).

11. GOVERNING LAW. Any action related to this Agreement will be governed by California law and controlling U.S. federal law. No choice of law rules of any jurisdiction will apply.

12. SEVERABILITY. If any provision of this Agreement is held to be unenforceable, this Agreement will remain in effect with the provision omitted, unless omission would frustrate the intent of the parties, in which case this Agreement will immediately terminate.

13. INTEGRATION. This Agreement is the entire agreement between you and Sun relating to its subject matter. It supersedes all prior or contemporaneous oral or written communications, proposals, representations and warranties and prevails over any conflicting or additional terms of any quote, order, acknowledgment, or other communication between the parties relating to its subject matter during the term of this Agreement. No modification of this Agreement will be binding, unless in writing and signed by an authorized representative of each party.

SUPPLEMENTAL LICENSE TERMS

These Supplemental License Terms add to or modify the terms of the Binary Code License Agreement. Capitalized terms not defined in these Supplemental Terms shall have the same meanings ascribed to them in the Binary Code License Agreement . These Supplemental Terms shall supersede any inconsistent or conflicting terms in the Binary Code License Agreement, or in any license contained within the Software.

A. Software Internal Use and Development License Grant. Subject to the terms and conditions of this Agreement and restrictions and exceptions set forth in the Software "README" file, including, but not limited to the Java Technology Restrictions of these Supplemental Terms, Sun grants you a non-exclusive, non-transferable, limited license without fees to reproduce internally and use internally the Software complete and unmodified for the purpose of designing, developing, and testing your Programs.

B. License to Distribute Software. Subject to the terms and conditions of this Agreement and restrictions and

exceptions set forth in the Software README file, including, but not limited to the Java Technology Restrictions of these Supplemental Terms, Sun grants you a non-exclusive, non-transferable, limited license without fees to reproduce and distribute the Software, provided that (i) you distribute the Software complete and unmodified and only bundled as part of, and for the sole purpose of running, your Programs, (ii) the Programs add significant and primary functionality to the Software, (iii) you do not distribute additional software intended to replace any component(s) of the Software, (iv) you do not remove or alter any proprietary legends or notices contained in the Software, (v) you only distribute the Software subject to a license agreement that protects Sun's interests consistent with the terms contained in this Agreement, and (vi) you agree to defend and indemnify Sun and its licensors from and against any damages, costs, liabilities, settlement amounts and/or expenses (including attorneys' fees) incurred in connection with any claim, lawsuit or action by any third party that arises or results from the use or distribution of any and all Programs and/or Software.

C. Java Technology Restrictions. You may not create, modify, or change the behavior of, or authorize your licensees to create, modify, or change the behavior of, classes, interfaces, or subpackages that are in any way identified as "java", "javax", "sun" or similar convention as specified by Sun in any naming convention designation.

D. Source Code. Software may contain source code that, unless expressly licensed for other purposes, is provided solely for reference purposes pursuant to the terms of this Agreement. Source code may not be redistributed unless expressly provided for in this Agreement.

E. Third Party Code. Additional copyright notices and license terms applicable to portions of the Software are set forth in the THIRDPARTYLICENSEREADME.txt file. In addition to any terms and conditions of any third party opensource/freeware license identified in the THIRDPARTYLICENSEREADME.txt file, the disclaimer of warranty and limitation of liability provisions in paragraphs 5 and 6 of the Binary Code License Agreement shall apply to all Software in this distribution.

For inquiries please contact: Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, California 95054, U.S.A. (LFI#141623/Form ID#011801)