Validator Overview

Compare the Validator editions

**Validator Enterprise**
- For ambitious large-scale projects.
- 5 million book pages or
- 1,25 million newspaper pages per year
- Validator client: 20
- Validator server: 8

**Validator Pro**
- For ambitious midsize projects.
- 1 million book pages or
- 250,000 newspaper pages per year
- Validator client: 4
- Validator server: 2

**Validator Basic**
- For ambitious small projects.
- 400,000 book pages or
- 100,000 newspaper pages per year
- Validator client: 2
- Validator server: 1

View the interactive presentation [here](#).
1.1 Start

Validator is a tools package used to automatically validate digital objects, such as image files, PDF files or XML data (METS, ALTO).

The validation is performed in two steps: Automated and Manual validation.
Automated or Manual validation

Validator is a tools package used to automatically validate digital objects, such as image files, PDF files or XML data (METS, ALTO).

The validation is performed in two steps: Automated and Manual validation.

The **Automated Validation** is checking all files for specific parameters and integrity and ensures all data is of consistent quality.

The **Manual Validation** is performed by operators using the Validator user interface. They check samples for image quality, meta data, text quality or other parameters that cannot be validated by a machine (subjective criteria).

**Benefits**
Check the Validator's benefits for your production environment.

**Basic workflow**
The standard workflow for the validation process has four main stages.

**Contact**
Get in contact with CCS for more information and for a demo session.
1.2 Interface

With Validator you can view and check the validation results and statistics, inspect the validated files and perform manual tests.

Validator allows to arrange, resize, close or hide windows for each file on screen, or even leave them floating on a second screen.
1.3 Tools

The Toolbox on right hand side of the interface shows a set of tools to enhance the dWValidator visualizations.

Tool sets are available for:
- Images
- ALTO text and image view
- ALTO articles view
- CSV files
- METS output
1.4 Options

Easy to configure:

- **Users**
  Add or remove users and configure access levels.

- **View managers**
  Tell the application how to display files.

- **E-mail notifications**
  Send e-mail if a batch validation is completed.
1.5 Results

See the validation results for the entire batch, a document, a docblock or a file.

Validation report formats:
- HTML (easy to browse)
- XML (machine readable for further processing)
- PDF
- Excel and CSV
- HTML

Statistics:
- View a summary report of the batches
- View pie chart representing the status of the batches
1.6 Views

Manual QA can be done in a view especially configured for the selected file type.
<?xml version="1.0" encoding="UTF-8" ?>
<alto xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xmlns="http://www.loc.gov/standards/alto/v3#
     xsi:schemaLocation="http://www.loc.gov/standards/alto/v3/alto-3-1.xsd"
     SCHEMAMVERSION="3.1"
     xmlns:xlink="http://www.w3.org/1999/xlink">
  <Description>
    <MeasurementUnit>mm10</MeasurementUnit>
    <sourceImageInformation>
      <fileNmae>../MASTER/00002.tif</fileName>
    </sourceImageInformation>
    <OCRProcessing ID="OCRPROCESSING_1">
      <preProcessingStep>
        <processingSoftware>
          <softwareCreator>CCS Content Conversion Specialists GmbH, Germany</softwareCreator>
          <softwareName>CCS_docWorks</softwareName>
          <softwareVersion>6.9.0.114</softwareVersion>
          <processingSoftware>
        </preProcessingStep>
      </OCRProcessing>
    </OCRProcessing>
  </Description>
  <Stories>
    <TextStyle ID="TXT_0" FONTSIZE="21"
      FONTFAMILY="Verdana_bold" FONTSTYLE="bold"/>
    <TextStyle ID="TXT_1" FONTSIZE="31"
      FONTFAMILY="Verdana_bold" FONTSTYLE="bold"/>
    <TextStyle ID="TXT_2" FONTSIZE="9"
      FONTFAMILY="Times New Roman_Bold" FONTSTYLE="bold"/>
  </Stories>
</alto>
Articles view
Blocks view
<?xml version="1.0" encoding="UTF-8"?>
<metis xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
http://schema.ccs-gmbh.com/docworks/version20/mix_jp2.xsd"
  <metsHdr CREATEDATE="2016-03-10T18:06:03" LASTMODDATE="2016-03-10T18:06:03">
    <mets:agent ROLE="CREATOR" TYPE="OTHER" OTHERTYPE="SOFTWARE">
      <name>CCS docWorks/METAe Version 6.9.1</name>
      <note>docWorks ID: 515</note>
    </mets:agent>
  </metsHdr>
  <mdWrap MIMETYPE="text/xml" MDTYPE="MODS" LABEL="Bibliographic meta-data of the printed version">
    <xmlData>
      <MODS xmlns="http://www.loc.gov/mods/3.2">
        <mods:titleInfo id="MODSMD_PRINT_TT1" lang="en">19440310_01</mods:titleInfo>
        <mods:title>19440310_01</mods:title>
      </MODS>
      <MODS xmlns="http://www.loc.gov/mods/3.2">
        <mods:languageTerm type="code" authority="rfc3066">en</mods:languageTerm>
      </MODS>
      <MODS xmlns="http://www.loc.gov/mods/3.2">
        <mods:identifier type="local">19440310</mods:identifier>
      </MODS>
    </xmlData>
  </mdWrap>
</mets>
KORAMAEN JOULUKIRKKO

KOIRAMÄEN JOULUKIRKKO

Mauri Kunnas

»Kile ja Elsa herätkin. Aamupuro odottaa«, sanoo äiti. »On aika lähteä joulukirkoon!«
### CSV/Excel view

**Table View**

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>vogNr</td>
<td>master</td>
<td>access</td>
<td>mets</td>
</tr>
<tr>
<td>MMKB15_000055001 00001</td>
<td>MMKB15_000055001_00001_master.p1</td>
<td>MMKB15_000055001_00001_access.p1</td>
<td>MMKB15_000055001_00001_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00002</td>
<td>MMKB15_000055001_00002_master.p1</td>
<td>MMKB15_000055001_00002_access.p1</td>
<td>MMKB15_000055001_00002_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00003</td>
<td>MMKB15_000055001_00003_master.p1</td>
<td>MMKB15_000055001_00003_access.p1</td>
<td>MMKB15_000055001_00003_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00004</td>
<td>MMKB15_000055001_00004_master.p1</td>
<td>MMKB15_000055001_00004_access.p1</td>
<td>MMKB15_000055001_00004_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00005</td>
<td>MMKB15_000055001_00005_master.p1</td>
<td>MMKB15_000055001_00005_access.p1</td>
<td>MMKB15_000055001_00005_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00006</td>
<td>MMKB15_000055001_00006_master.p1</td>
<td>MMKB15_000055001_00006_access.p1</td>
<td>MMKB15_000055001_00006_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00007</td>
<td>MMKB15_000055001_00007_master.p1</td>
<td>MMKB15_000055001_00007_access.p1</td>
<td>MMKB15_000055001_00007_mets.p1</td>
<td></td>
</tr>
<tr>
<td>MMKB15_000055001 00008</td>
<td>MMKB15_000055001_00008_master.p1</td>
<td>MMKB15_000055001_00008_access.p1</td>
<td>MMKB15_000055001_00008_mets.p1</td>
<td></td>
</tr>
</tbody>
</table>

**Toolbox**

Set delimiter for CSV:
- comma
- semicolon
General description

What is documentValidator?
Getting started
The validation process
The validation results
Checking the files
Generating reports
Manual QA

Projects
Use one of the links below to view the tests specific to a project.
- no projects available -

What is documentValidator?
documentValidator is a tool used to automatically validate digital objects, such as image files, PDF files, XML data (METS, ALTO) and other derivatives. Validation is the process of checking if something satisfies certain criterion. Examples would include checking if a statement is true (valid), or if data is compliant with a certain standard. For digital objects there are many variations for output data. Consequently, a validation tool has to be configured to validate each component of a digital object according to the respective requirements. The validation is typically performed in two steps: Automated validation and Manual validation.

The Automated Validation is checking all files for specific parameters and integrity and ensures all data is of consistent quality. The Manual Validation is performed by people using the Validator user interface, checking samples for image quality, meta data, text quality or other parameters. The validator software provides reports of the results which are essential as manifest files.

Getting started

Download video
What is dWValidator?

dWValidator is a tool package used to automatically validate digital objects, such as image files, PDF files, XML data (METS, ALTO) and other derivatives. Validation is the process of checking if something satisfies certain criteria. Examples would include checking if a statement is true (valid), or if data is compliant with a certain standard. For digital objects there are many variations for output data.

The validation is typically performed in two steps: Automated validation and Manual validation.

- The Automated Validation is checking all files for specific parameters and integrity and ensures all data is of consistent quality.
- The Manual Validation is performed by operators using the Validator user interface, checking samples for image quality, meta data, text quality or other parameters that cannot be validated by a machine (subjective criteria).

The dWValidator software provides reports and statistics of the results.

The dWValidator’s benefits:

- Independent quality assurance tool
- Detailed reporting as XML, HTML, PDF, custom format
- Integrates customizable automated tests
- Integrates customizable selection for manual QA data set
- Specific tools for manual QA (combined views, color picker, measurement tools, histograms, etc.)
- Multithread testing for higher performance

The dWValidator is developed as a QA tool for digitization services in order to:

- Reduce QA costs
- Avoid human errors from manual QA process
- Merge different tools we use for QA
- Ensure 100% accuracy for measurable properties
- Create reports to be submitted with deliveries
1.7 Reports

Generate validation reports in various formats to view outside of dWValidator.

Show all batches of a selected project, with details about their status, number of documents, and number of files.
1.8 Benefits

Validator Benefits

- Independent quality assurance (QA) tool
- Avoid human errors from manual QA process
- Reduce QA costs
- Ensure 100% accuracy for measurable properties
- Customizable automated tests
- Customizable tools for manual QA (combined views, color picker, measurement tools, histograms, etc.)
- Multithread testing for higher performance
- Detailed reporting as XML, HTML, PDF, custom format
- Reports can be generated as Excel or Csv
1.9 Workflow

Example basic workflow for the validation of docWorks documents

- Use dWConfigurationManager to create/edit project configurations that will be used to validate documents.
- Use dWControlCenter to start the import and validation of the batches.
- Check the progress of the automatic validation.
- Use Validator to check the results and get reports and statistics.
1.10 Contact

Get in Contact with CCS

CCS Content Conversion Specialists GmbH
Weidestraße 134
22083 Hamburg, Germany
Phone: +49-(0)40-227 130 0
Fax: +49-(0)40-227 130 11
E-Mail: dWSupport@content-conversion.com
Website: www.content-conversion.com