

Table of Contents

1. General Information - Mohan
2. Installation -
 - a. System Requirements
 - b. Installation Instructions
3. Workflow - High level
4. Login
5. Admin/Configuration
 - a. Project Management
 - i. Metadata Manager
 1. Media metadata
 - a. Metadata Schema
 - b. Reordering fields
 2. Structural Metadata
 - a. Structural Schema
 - b. Reordering
 3. Page Metadata
 - a. Page metadata Schema
 - ii. File Naming Conventions
 - iii. Folder Naming Conventions
 - iv. Project
 - v. Page Remarks
 - vi. Scanner and OCR Settings - (For windows 7 needs to be run as admin)
 - vii. Language Manager
 - b. Roles and Users
 - i. Roles Creation
 - ii. User creation
6. Workflow
 - a. Metadata Import
 - b. Metadata Entry
 - c. Scanning
 - d. QC0
 - e. Image Processing
 - f. QC1
 - g. OCR
 - h. QC Final
 - i. Packaging
7. Reports
8. Tools
 - a. Merging
 - b. Splitting
 - c. Renaming
 - d. Batch Processing

e.

For each item

1. Introduction
2. Dependencies
3. Example
4. Flow
5. Screen Shot
6. Screen Elements
7. Screen Actions

1. Administration

1.1. Project Management: -

Introduction: Project Management is a structured framework to help a group work productively. Project Management is a method of Organizing Tasks. And it helps in Task Sequencing, Dependency Analysis, Resource Allocation, Scheduling, etc.

Digitization projects are complex, involving multiple roles/people, multiple resources, multiple tasks, multiple decision points and matching of people to several resources and tasks, "Project Management" is helpful in organizing all these tasks and resources.

Digitization of large number of books is also very complex; consider digitization of books as a Project. This digitization also involves lot of manpower, many activities, assigning the manpower to different tasks, etc. Most digitization projects come with specific process and standards which will entail tasks and activities to be executed in a sequential order.

The first part of Project Management is design, define and configure the following

Metadata Schemas,

Media Metadata,

Structural Metadata,

Page Metadata,

Naming conventions

Folder Naming, File Naming

Language manager- to set the User Interface languages

Scanner Settings, OCR Settings

Project settings - Defining a project with the settings designed

1.1.1. Meta data Manager

Introduction: Repositories of documents/ books are stocked and stored with indexes for proper and easy retrieval on demand. Various standards for manual and electronic indexes have been designed and defined for cataloging of different media for storage, retrieval and information with other repositories. With digitization the creation of proper schema for the repositories is of paramount importance.

Metadata that needs to be configured for a project could simplistically be classified as follows

- a. Media metadata: Details of the artifact that is being digitized like the Name of Publication, Author, Publisher etc. This is useful in searching for an artifact.
- b. Structural Metadata: Creation of a table of contents for a book or segmenting articles within a newspaper are examples of structural metadata.
- c. Page Metadata: there a number of rendering and presentation tools available for the digitized content. The metadata of a page is useful in creating a rendered image true to the original giving the details of the fonts, paragraphs, borders and spacing within a page.

The Metadata Manager helps in creating customized or standard schema that can be used across the project. Such standards need to be put in place at the start of the Project and would be helpful for the project if minimal or no changes are made to the schema once the project commences.

1.1.1.1. Media Metadata

1.1.1.1.1. Metadata Schemas

Introduction: Once the metadata schema has been frozen for a project the various elements may be configured in the software. Tabs may be created for various categories of the metadata ex: document dimensions, Publisher details, Copyright details etc, by creating groups within the schema. Each tab is defined by the group definition. Within the groups metadata elements with their attributes can be created.

There is a predefined uneditable tab Basic metadata the elements of which are used for defining the file and the folder naming conventions

Creation of the metadata would be in the following hierarchy

---Metadata Schema Name

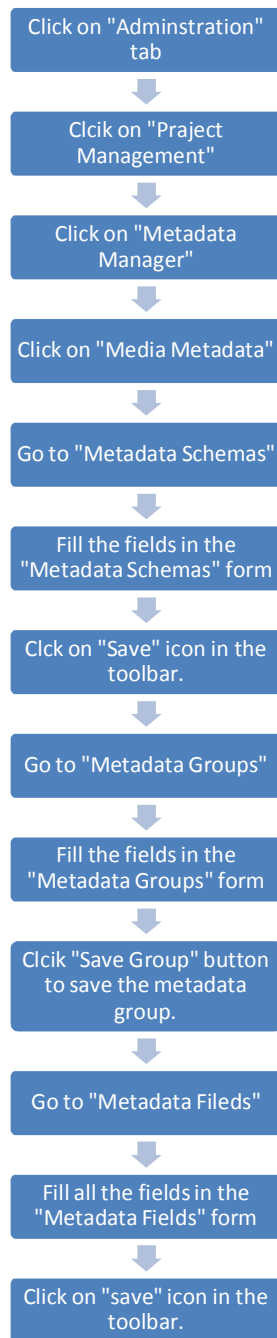
 ---Groups within the schema

 ---Metadata elements with the definition of the type of metadata.

The metadata schema created may be used as the templates for either import of metadata from already created electronic repositories or may be manually entered or both could be done

Dependency: While the creation of these metadata schema does not inherit any dependencies the metadata elements of the Basic metadata will be available for the file naming and folder naming conventions.

Flow for creation of Metadata Schema:



Fields to be filled in the "Metadata Schemas" form: -

- "Metadata Schema Name" - Enter the name of the schema.(Ex: -Dublin Core)
- "Description" - Enter some information regarding the schema.

After saving this form go to "Metadata Group" form.

Fields to be filled in the "Metadata Group" form: -

- "Metadata Group" - Enter the name of the Metadata group (Ex: -Demo Metadata)in the "Dublin Core" Schema.
- "Group Description" - enter the description about the group.

Depends on the requirements groups are created under the Dublin Core Scheme name.

Different operations in this form are: -

- "New Group" - This button is used to create a new group
- "Save Group" - This button is used to save the group
- "Delete Group" - This button is used to delete the group from the list.

The saved groups can be seen in the "Available Goups" filed at the bottom of the field.

When the "Metadata Group" has created, then the "Metadata Field" form will be accessed.

Diiferent fields in the "Metadata Field" form: -

- "Metadata Group" - Select the "Metadata Group" from the drop-down.
- "Field Name" - Enter the field name and this name will be stored in the database and cannot be changed in future.
- "Alias Field Name" - Enter the field name, and this name will be displayed while filling the metadata of the media. And this name can be changed.
- "List Order" - Select the number, which indicates the order of the field in the list.
- "Type of Control": - Select the control from drop-down.(Ex: -Textbox, Combobox,etc)
- "Is Required" - Choose "Yes" or "No". Yes indicates the given field is mandatory while filling the metadata elements for a media. No, indicates the given field is not mandatory.
- "Validation Type" - Select the validation type from the drop-down(Ex: -Phone Numbers, Alphabets, etc).
- "Validation Message" - Enter the message, which will be displayed when there is any error.

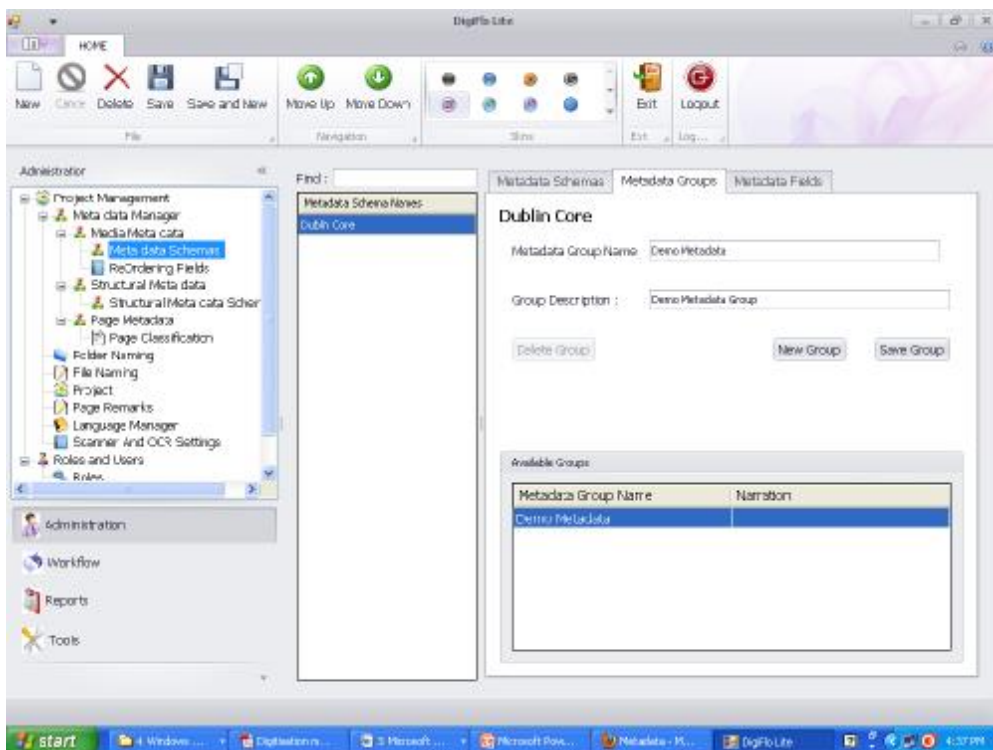
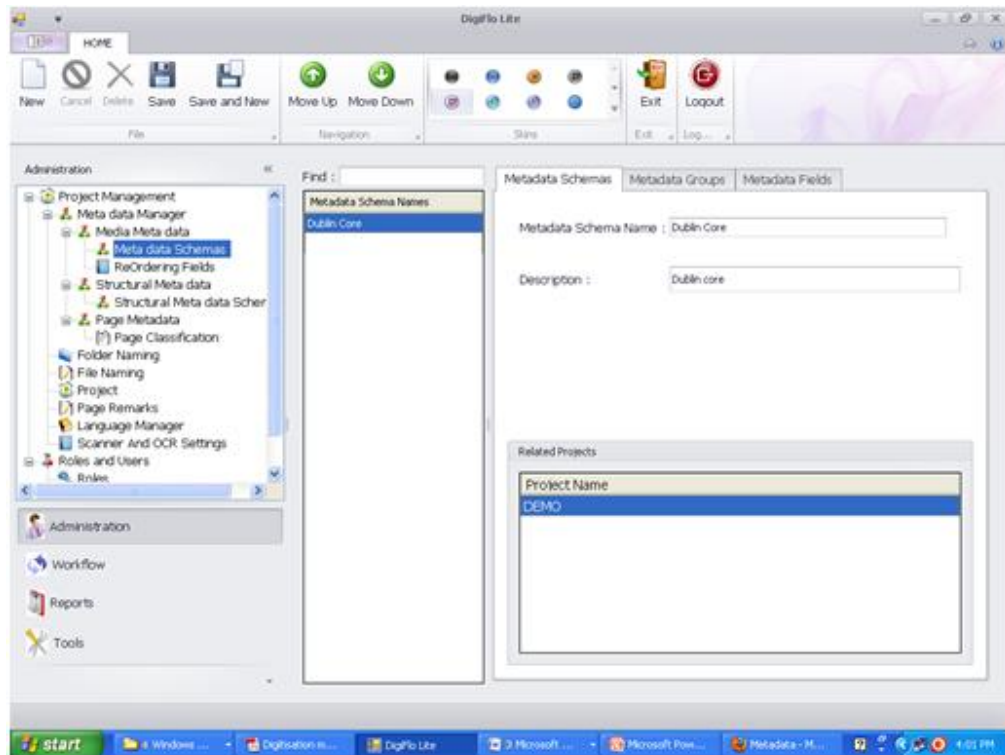
All the fileds that are entered will be displayed in the "Available Fields" box.

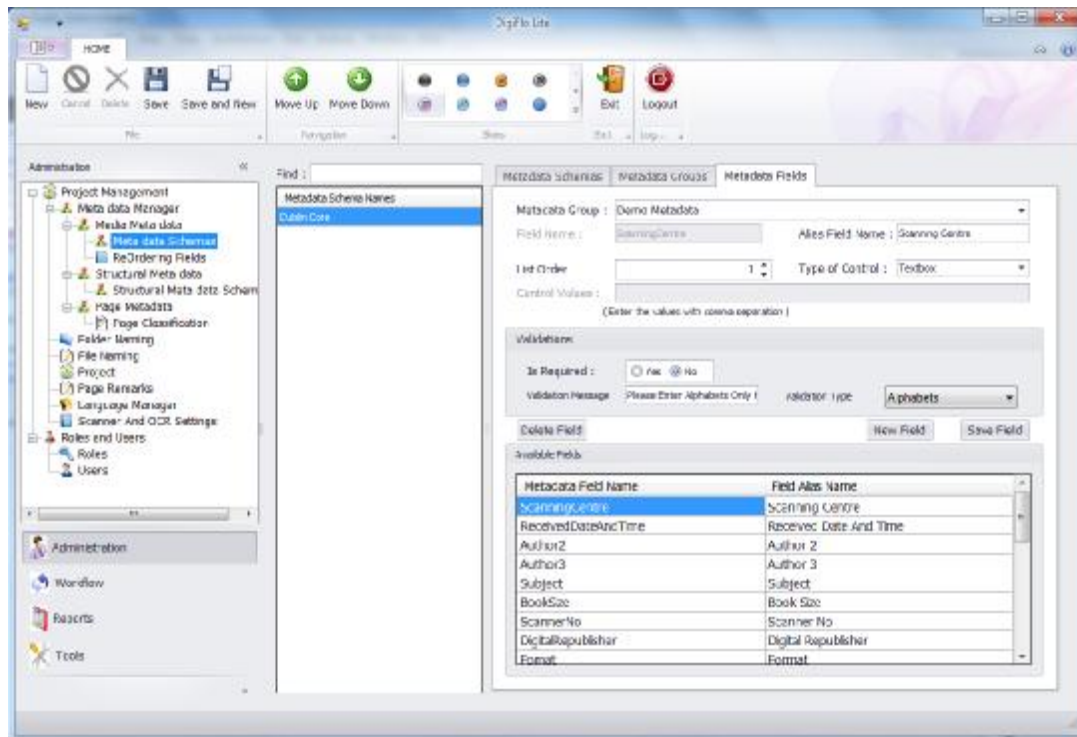
Different operations in this form are: -

- "New Field" - This button is used to create a new field
- "Save Field" - This button is used to save the field
- "Delete Field" - This button is used to delete the filed from the list.

These Created "Metadata fields" will be reflected in the "Metadata Entry" activity of the "Workflow" tab.

The below figures shows the creation of "Metadata Schemas", "Metadata Groups", "Metadata Fields"





Different Fields that are created in the "Metadata Fields" form are as follows,

- "Scanning Centre" -Enter the place where scanning has to be done.
- "Received date and time" - Select the date and time from the drop-down.
- "Author 2" - Enter the second author of the media
- "Author 3" - Enter the third author of the media.
- "Subject" - Enter the subject on which the media has been written.
- "Book Size" - Enter the size of the media.
- "Scanner No" - Enter the scanner number on which scanning of this media is done.
- "Digital Republisher" - Enter the Publisher who is doing this digitization work.
- "Format" - Enter the format of the media
- "Identifier" - Enter the name of the Identifier.
- "Source" - Enter the name of the source from which the information have been taken.
- "Coverage" - Enter the jurisdiction under which the media has been written.
- "Copyrights" - Enter the copyrights
- "Digital Republishing Date" - Enter the date on which the digitization of this media is done.

- "Copyrights Owner" - Enter the name of the person or institution who owns the copyrights of this media.
- "Copyrights Expiry date" - Enter the date on which the copyrights will be expired.
- "ISBN No" - Enter the "International Standard Book Number" for the given media
- "Contributor" - Enter the name of the Contributor, who contributed information to the preparation of the media.
- "Total No of pages" - Enter the number of pages of the media.
- "Book Bank" - Enter the name of the Book Bank from where this media has been taken
- "Received From" - Enter the name of the person or the institution who have given the book for digitization.
- "Received by" - Enter the name of the person who has received the media.
- "Publisher Place" - Enter the place of the publisher.
- "Publishing Date" - Enter the date on which the media has been published.
- "Document type" - Enter the type of document (Ex: -book)
- "Language" - Enter the name of the language in which the media has been published

Different operations that can be performed in the "Metadata Scheme": -

- "New" - This operation is used to get an empty form to create a new Metadata Schema.
- "Cancel" - To cancel the Metadata Schema before it has saved.
- "Delete" - To delete the Metadata Schema even after saving it.
- "Save" - To save the Metadata Schemas into the database after filling all the fields in the form.
- "Save and New" - To get an empty form for creating a new Metadata Schema after successfully saving the previous Schema.
- "Move Up" - To move the created Metadata Schema up the order
- "Move Down" - To move the created Metadata Schema down the order

1.1.1.1.2. ReOrdering Fields: -

Introduction: The elements of the metadata schema that have been created would appear in the workflow section metadata entry/ import. If there is an order in which the users would desire to enter the data this could be configured using the reordering option. The elements may be moved up or down and will appear in respective sequence in the Metadata data entry form.

Different Fields available in the "ReOrdering Fields" form are as follows

- "Metadata Schemas" - Select the Metadata Schema from the drop-down.
- "Metadata Group" - Select the Metadata Group from the drop-down.
- "Metadata Fields" - When the "Metadata Schemas" and "Metadata Group" fields are selected, then all the Metadata fields in the "Metadata Group" will be displayed here.

Reordering can be done by using the "Move UP" and "Move Down" icons in the toolbar.

After the Re-Ordering, click on "save" icon in the toolbar to save the Order of Metadata Fields.

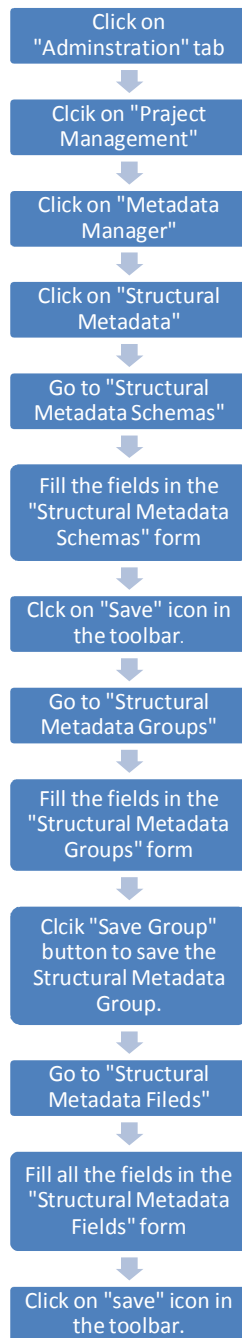
1.1.1.2. Structural Metadata

Introduction: Structural metadata simplistically stated is the data which defines the internal segmentation of a document. For structured documents like books the fields that need to be defined are normally related to the table of contents. While further complex data can be extracted and embedded in the data it will need to be programmatically extracted. The elements of the internal segmentation of a document may be defined and during the process of OCR or Eyeball checking the elements may be populated.

1.1.1.2.1. Structural Metadata Schemas: -

Introduction: The predefined Structural metadata may be configured using the interface and these could be used at f

The process of creating the Metadata Fields(Metadata elements): -



Fields to be filled in the "Structural Metadata Schemas" tab: -

- "Structural Metadata Schema Name" - Enter the name of the schema. (Ex: -ALTO)
- "Description" - Enter some information regarding the schema.

After saving this form go to "Structural Metadata Group" form.

Fields to be filled in the "Metadata Group" tab: -

- "Structural Metadata Group" - Enter the name of the Structural Metadata group in the "ALTO" Schema.
- "Group Description" - Enter the description about the group.

Depends on the requirements, groups are created under the ALTO Scheme name.

Different operations in this form are: -

- "New Group" - This button is used to create a new group
- "Save Group" - This button is used to save the group
- "Delete Group" - This button is used to delete the group from the list.

The saved groups can be seen in the "Available Groups" filed at the bottom of the field.

When the "Structural Metadata Group" tab has created, then the "Structural Metadata Fields" form will be accessed.

Diiferent fields in the "Structural Metadata Fields" tab -

- "Metadata Group" - Select the "Metadata Group" from the drop-down.
- "Structural Field Name" - Enter the field name and this name will be stored in the database and cannot be changed in future.
- "Structural Alias Field Name" - Enter the field name, and this name will be displayed while filling the metadata of the media. And this name can be changed.
- "List Order" - Select the number, which indicates the order of the field in the list.
- "Type of Control": - Select the control from drop-down. (Ex: -Textbox, Combobox, etc)
- "Is Required" - Choose "Yes" or "No". Yes indicates the given field is mandatory while filling the metadata elements for a media. No, indicates the given field is not mandatory.
- "Validation Type" - Select the validation type from the drop-down (Ex: -Phone Numbers, Alphabets, etc).
- "Validation Message" - Enter the message, which will be displayed when there is any error.

All the fields that are entered will be displayed in the "Available Fields" box.

Different operations in this form are: -

- "New Field" - This button is used to create a new field
- "Save Field" - This button is used to save the field
- "Delete Field" - This button is used to delete the field from the list.

These Created "Structural Metadata fields" will be reflected in the "Metadata Entry" activity of the "Workflow" tab.

1.1.2. Folder Naming: -

Introduction: One of the principal requirements for a making a rendering engine or a presentation layer effective is to have data with consistent nomenclature. Sometimes when media have nominal metadata the practice could be to embed the entire metadata in the folder/file naming mechanism. Also that various digital formats of data may need to be generated there needs to be parity in the files and the mechanisms that they are stored.

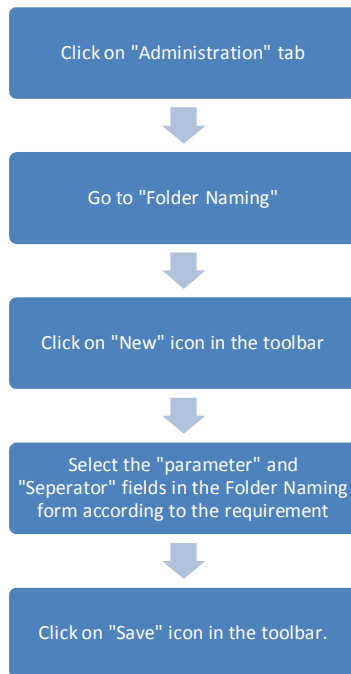
For many archiving projects multiple copies of the image, archive copy(image as scanned), master copy(processed image) and render copies (various low to medium res versions for distribution) need to be created. These would have the same information but in different image formats/resolutions depending on the use or purpose of the image. The folder naming convention for the various packets of images can be configured based on how the project has been planned.

The folder naming convention gets the data elements from the metadata and hence avoids errors while creation of the folders as the person scanning may or may not know the language the document he is scanning is in. The rigorous and error prone activity of manually creating the folder names is eliminated.

The segregation of the images at the time of finally packaging the end delivery is also made easy by directing the images from folders to the respective storage/servers.

Dependencies: The folder naming conventions take data elements for naming from the fields in the Basic metadata.

Process of naming a folder: -



The parameters in the Field Naming form are the “Basic Metadata” of the form, i.e. Media Title, Barcode, OCLCNumber, Author, Publisher, and Region.

The “Basic Metadata” for the Marketing Management book.

- Media Title: - Marketing Management
- Barcode: -15234
- OCLCNumber: - OCM23568521
- Author: -Kotler
- Publisher: - Pearson
- Region: -London

And “Separators” are used to separate the parameters.

Different Separators used here are: -

- –
- ~
- @

So after selecting the “New” icon in the tool bar, a new form with “Parameters” and “Separators” as fields will be opened.

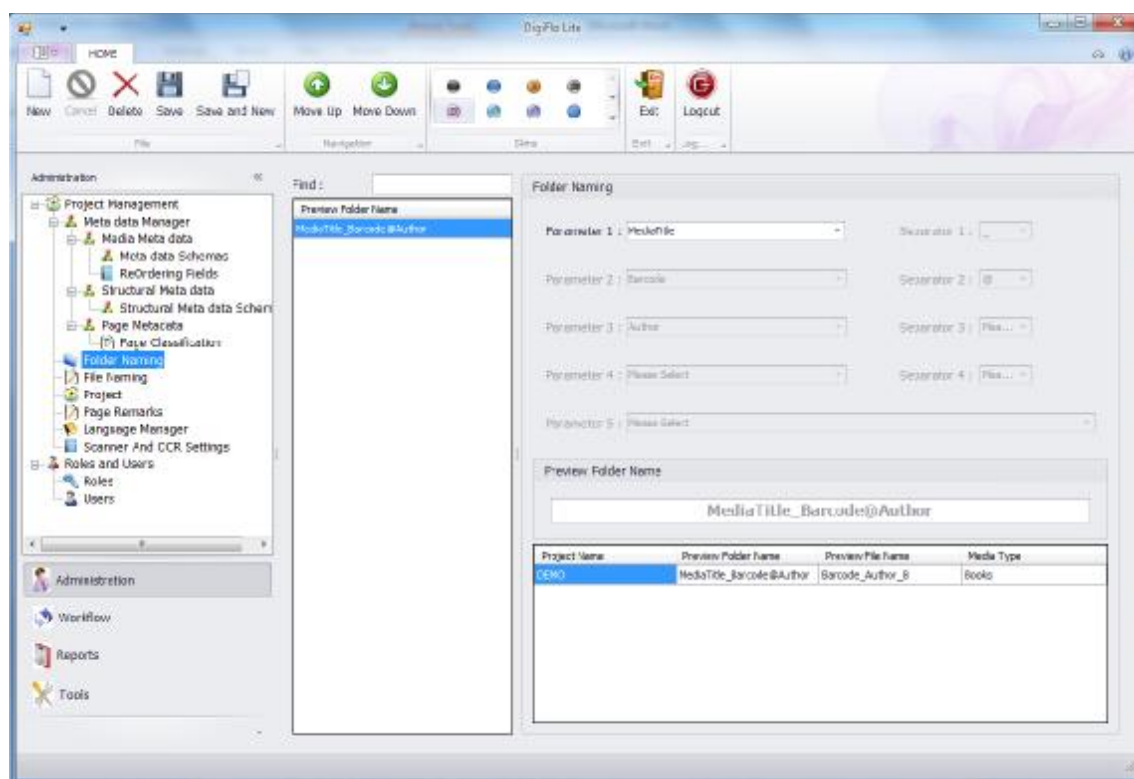
And now select

- Parameter 1 - "Media Title"
- Separator 1 - "_"
- Parameter 2 - "Barcode"
- Separator 2 - "@"
- Parameter - "Author"

And now the selected Folder name will be displayed in the "Previous Folder Name" field as "Media Title_Barcode@Author"

And in the digitization process of the "Marketing Management" book, the files will be stored in the "Marketing Management_15234@Kotler".

The below figure shows the "Folder Naming" form.



And whenever the folder name is created, automatically the sub folders will be created. They are as follows: -

- Marketing Management_15234@Kotler_HTML
- Marketing Management_15234@Kotler_OTIFF
- Marketing Management_15234@Kotler_PDF
- Marketing Management_15234@Kotler_PTIF
- Marketing Management_15234@Kotler_RTF

Different operations that can be performed in this "Folder Naming" are: -

- "New" - This icon is used to get an empty form to create a new "Folder Naming".
- "Cancel" - To cancel the "folder name" before it has saved
- "Delete" - To delete the "folder name" even after saving the role
- "Save" - To save the "Folder name".
- "Save and New" - To get an empty form for creating a new "Folder Naming" form after successfully saving the previous folder name.
- "Move Up" - To move the created "folder names" up the order
- "Move Down" - To move the created "folder names" down the order

1.1.3. File Naming: -

Introduction: Like the naming convention for a folder the file naming convention also may be configured. While the 8.3 mechanism of a number with 8 digits and a file extension of 3 characters was in vogue with the older operating systems, the windows environment has allowed for file naming upto 256 characters.

The naming convention could use the elements defined in the Basic metadata and the user may also define the number of leading zeros in the file naming convention with separators between elements.

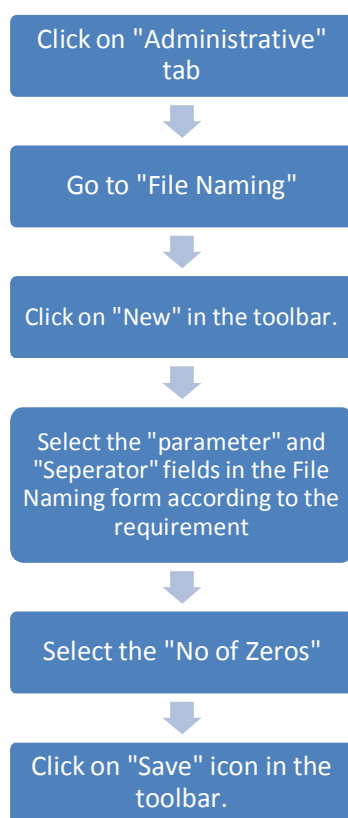
Dependencies: Metadata fields in the Basic Metadatatab .

The output images will be stored in the folder with the proper naming. And also the images which are stored have to be properly named to get an unique identity and for easy searching of the images. The images stored as files with a certain name. Proper file Naming will make the process easy while submitting the output to the customers.

Certain parameters are used to name the files. These parameters are the "Basic Metadata" of the media. (Ex: - Media Title, Barcode, OCLCNumber, Author, Publisher, Region). And the separators are used to separate the parameters. (Ex: - _, ~, @). And entering of "Number of zeros" will differentiate the files from one another.

Ex: - If a book "Marketing Management" has to be digitized, and in the digitization process this book has to undergo different steps. And the output images are stored with proper file names. And the File name is "Marketing Management@OCM23568521_000000 (image number)". The below illustration shows you the process of naming the file as "Marketing Management@OCM23568521_000000 (image number)".

Process of File Naming: -



The parameters in the File Naming form are the "Basic Metadata" of the book, i.e. Media Title, Barcode, OCLCNumber, Author, Publisher, and Region.

The "Basic Metadata" for the Marketing Management book.

- Media Title: - Marketing Management
- Barcode: -15234

- OCLCNumber: - OCM23568521
- Author: -Kotler
- Publisher: - Pearson
- Region: -London

And "Separators" are used to separate the parameters.

Different Separators used here are: -

- -
- ~
- @

So after selecting the "New" icon in the tool bar, a new form with "Parameters" and "Separators" as fields will be opened.

And now select

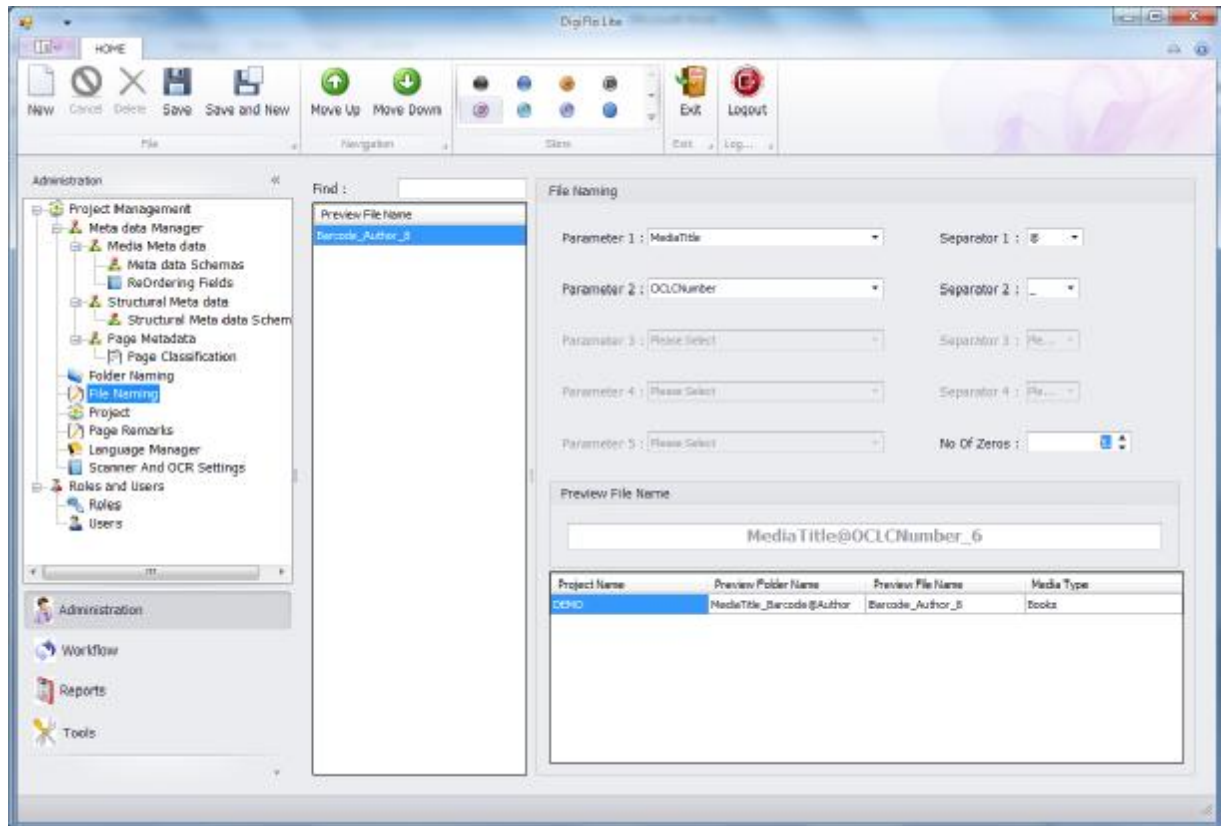
- Parameter 1 - "Media Title"
- Separator 1 - "@"
- Parameter 2 - "OCLCNumber"
- Separator 2 - "_"

Select "No of Zeros" as "6".

And now the selected File name will be displayed in the "Previous File Name" field as "Media Title_Barcode@Author"

And in the digitization process of the "Marketing Management" book, the files will be stored as the "Marketing Management@ OCM23568521_000000 (image number)".

The below figure shows the File Naming form: -



Different operations that can be performed in this “File Naming” are: -

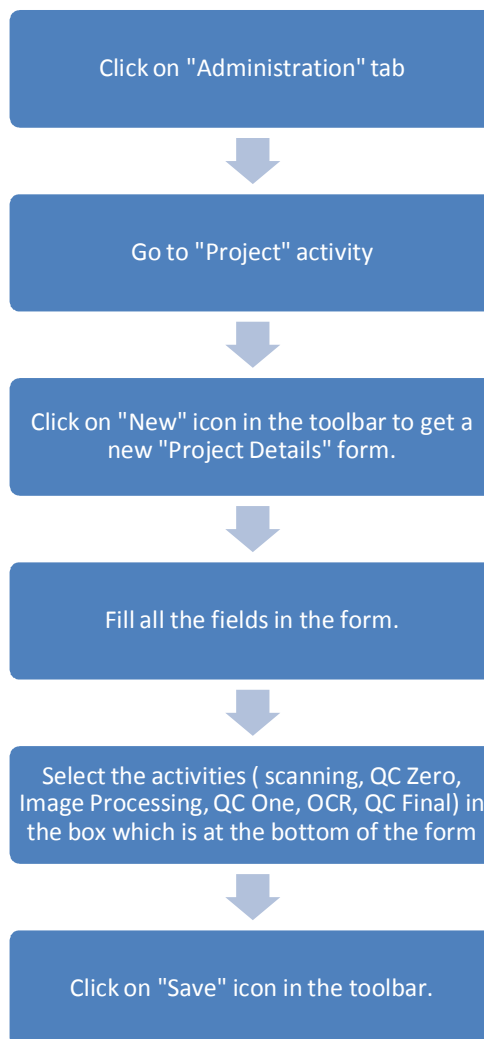
- “New” - This icon is used to get an empty form to create a new “File Naming”.
- “Cancel” - To cancel the “file name” before it has saved
- “Delete” - To delete the “file name” even after saving the role
- “Save” - To save the “File name”.
- “Save and New” - To get an empty form for creating a new “File Naming” form after successfully saving the previous file name.
- “Move Up” - To move the created “file names” up the order
- “Move Down” - To move the created “file names” down the order

1.1.4. Project: -

Introduction: The creation of a project is a culmination of the standards procedures and processes that have been created. The project interface, helps in bringing together the various metadata, file and folder naming standards and defining the workflow for the task ahead. The various tasks and activities within the workflow may be selected/deselected and the defined workflow and the parameters get reflected in the workflow section of the software.

Dependencies: Metadata Schema, File/Folder naming conventions need to be created and then selected

The process of selecting the project: -



Fields that are to be filled in the "Project Details" form: -

- "Project Name" - Enter the name of the project, which is to be executed.
- "Preview File Name" - Select the "File Naming" format from the drop-down.
- "Preview Folder Name" - Select the "Folder Naming" format from the drop-down.
- "Media type" - Select the Media (Ex: - Books, Newspapers, etc)
- "Media Meta data" - Select the Media Metadata from the drop-down.
- "Structural Metadata" - Select the structural Metadata from the drop-down.
- "Page Metadata" - Select the page metadata from the drop-down.
- "Description" - Enter the Description about the project, if necessary.

There are six predefined workflow stages in the product these stages can be toggled as desired during the design stage. They can be flagged to be on or off and would appear or be hidden in the workflow tabs. This is to enable the users to remove any portion of the workflow that they would not need or would seem redundant. The stages with their description are described below.

- Scanning : The books/documents whose metadata has been entered is then available to the scanning operators in their queue, they select the book and scan them. The file and folder naming schema that have been configured initially will be created.
- Quality Control Zero (QC0): The scanned books are checked for the quality of scan both parametric through the image EXIF data for the standard parameters of the project like the resolution, file format, compression, file sizes etc. and there is a viewer incorporated to eyeball check the images. Missing pages are checked for at this stage. A QC report of the images is generated and can be accessed at any later point in time. Normally the output of the QC0 is the Archive Copy of the Images.
- Image Processing: The scanned books/documents are then image processed and corrections made for skew, despeckling, color correction, artifact and noise band removal then resized and padded. The output of this process is a clean and corrected image.
- Quality Control One (QC1): Processed images are subject to the same checks that happen after scanning but at this stage, additionally the page sizes are also checked for. It is from this stage that the structural metadata of the documents are created. Documents/books that clear this stage would be stored as Master copies which can be used for further conversion of documents for the presentation Layer.
- OCR and creation of Structural metadata: The master images created would be OCRed and the structural metadata would be created from these. Different processes would be used for structured documents like books and for unstructured documents like Newspapers.
 - Structured Documents(like books): Pages would be ocred, Table of Contents created and images etc would be tagged. The Alto files of the pages would be created and depending on the requirement a complete OCR Correction may be done. The images are then stored in the respective presentation layer formats like HTML/PDF/Djvu.
 - Unstructured Documents(Like newspapers): The master images created would be zoned and the requisite METS data would be extracted from the images. All

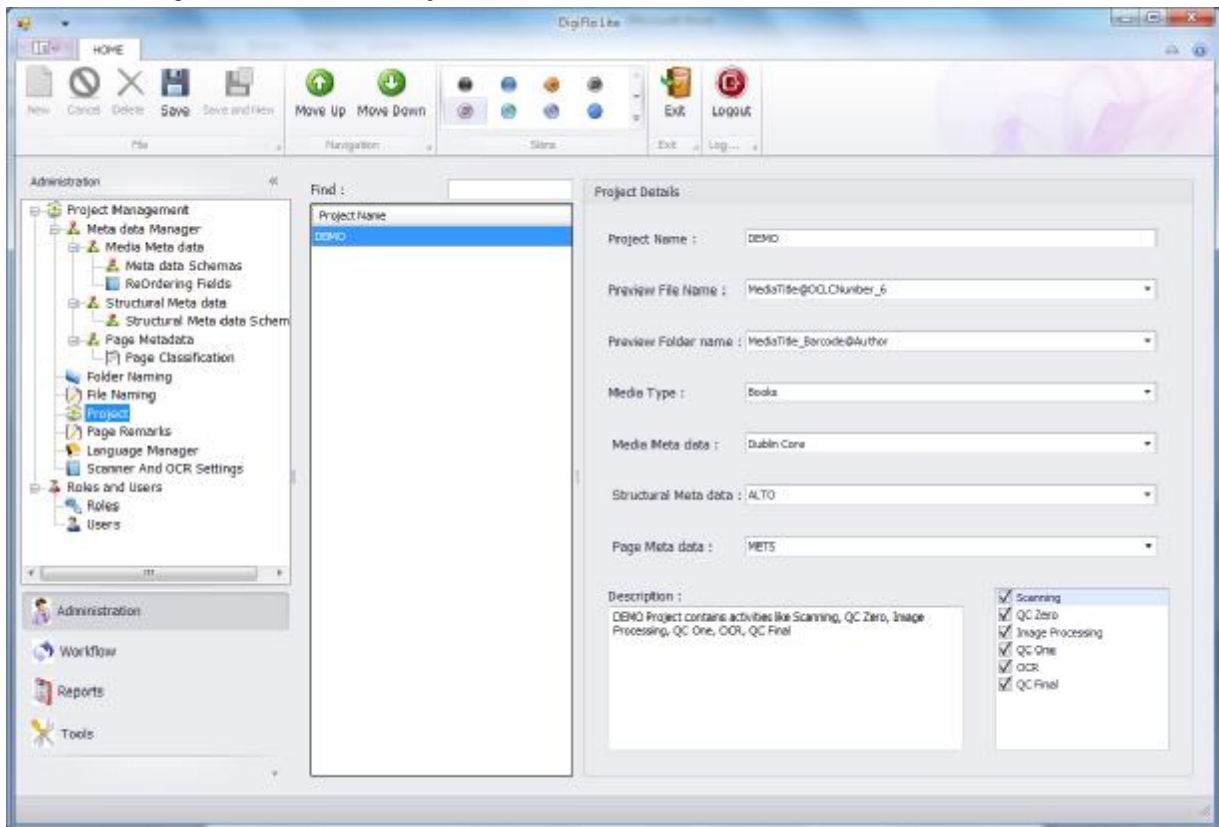
articles will be tagged with the metadata defined and Alto of the pages also need to be created.

- Quality Control Final (QCF): At this level the entire packaging , the file/folder file naming and the document EXIF data is validated and once the QCF is passed the render documents and the metadata are ready for the packaging.

At the bottom of the “Project Details” form, a grid with all the activities is displayed. According the requirement of the customer, the activities have to be selected. The above example shows the project contains activities like Scanning, QC Zero, Image Processing, QC One, OCR, QC Final. So, all the activities in the grid have to be selected.

After filling the fields and selecting the activities, click on “save” icon in the toolbar to save the “Project Details” form.

The Below figure shows the “Project Details” form: -



Different Operations that are performed in the “Users” are as follows: -

- “New” - This operation is used to get an empty form to create a new Project.
- “Cancel” - To cancel the Project before it is saved
- “Delete” - To delete the project even after saving the Project

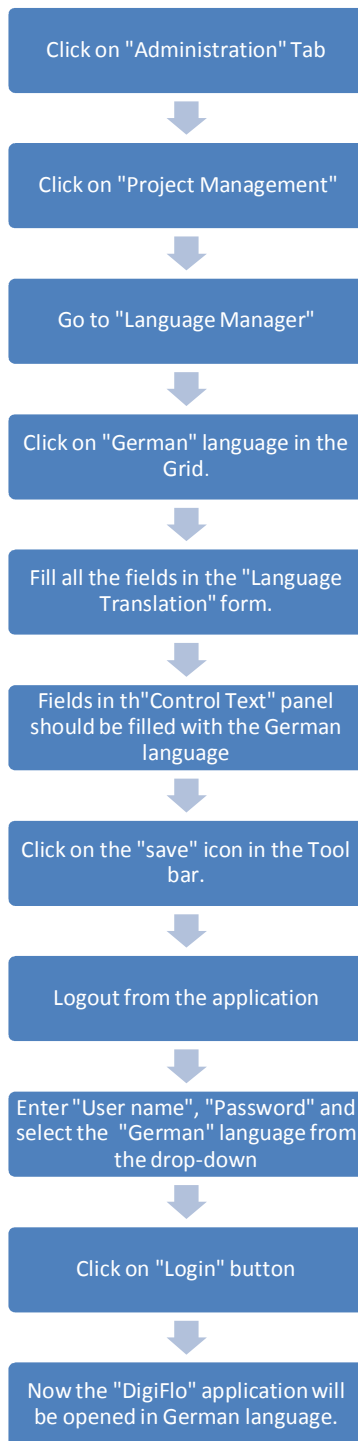
- "Save" - To save the Project into the database after filling all the fields in the form.
- "Save and New" - To get an empty form for creating a new Project after successfully saving a Project record into the database.
- "Move Up" - To move the created Projects up the order
- "Move Down" - To move the created Projects down the order

1.1.5. Language Manager: -

Introduction: In an attempt to make Digiflo Global, it comes with a Language Manger to configure the Languagesin the User Interfaces. The base language of the software is English, the labels and the actions across all the interfaces can be translated either manually or using an auto translation feature. Any text that could not be translated automatically can be edited in the text boxes. In case the text boxes are left empty the labels would be displayed in English. Language setting for data entry can be set from the Language button of Windows.

Once the translation to the any particular language is completed the user needs to logout and select the display language of choice at the time of relogin.

Process of Translating English Language into Germany Language using "Language Manager": -



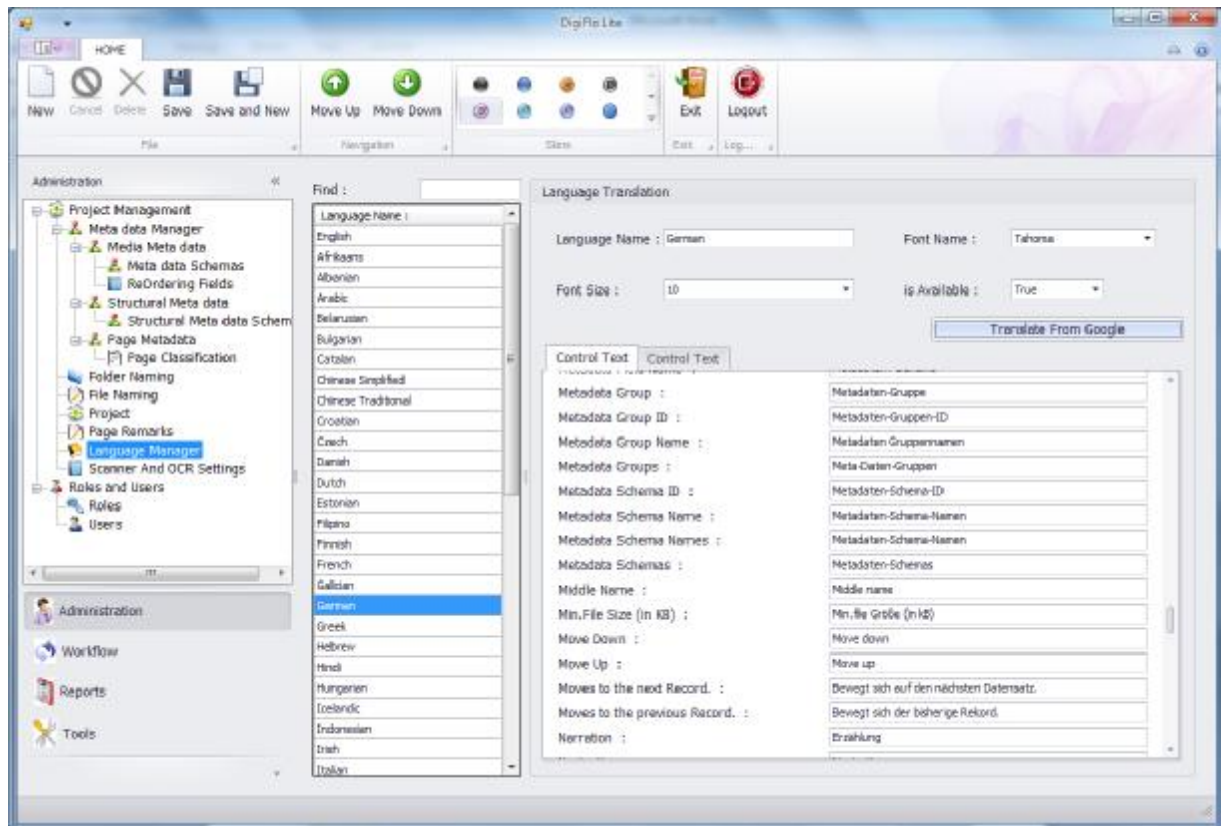
The above figure clearly shows that the DigiFlo application which is in "English" language has been translated into "German" language. Now all the text interfaces in the application are in the german language.

Fields that are to be entered in the Language Translation form: -

- "Language Name" - Enter the name of the language
- "Font Name" - Select the font name from drop-down.
- "Font Size" - Select the font size from drop-down.
- "is Available" - Select "True" to make the language appear on the "Language" field in Login Form. Select "False" to make the language not appear on the "Language" field in Login form.
- "Control Text" - It contains all the text interfaces fields in the application.

All the fields in the "Control Text" should be filled with the german language manually or by using the "Translate From Google" button. This german language filled in these fields are displayed in the DigiFlo application, when it is opened with the german language.

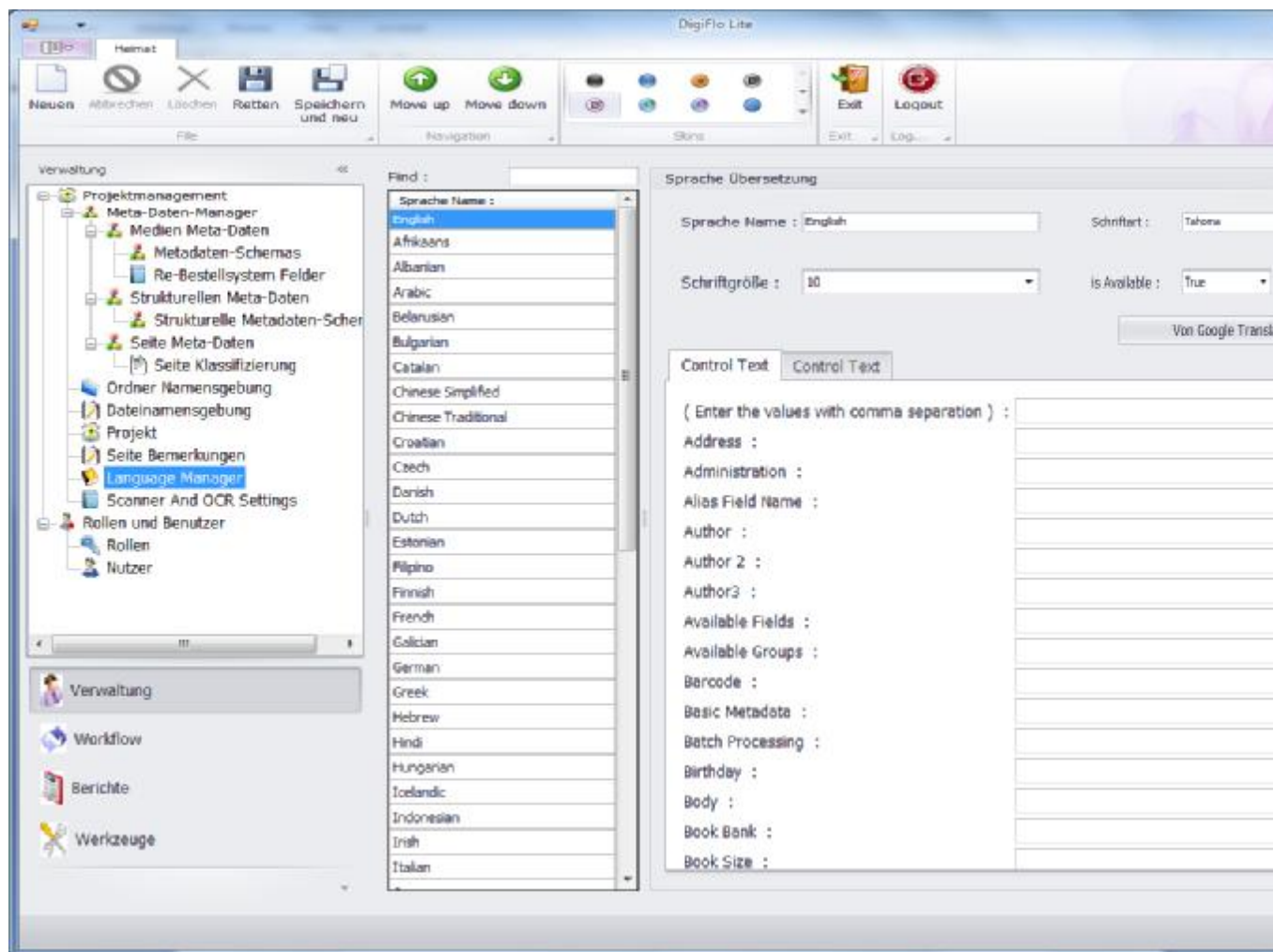
The below figure shows that German language has been selected and all the fields in the "Language Translation" form are filled and all the fields in the "Control Text" panel are filled with the german words.



The below figure shows the "Login Form" of the "DigiFlo" application, and username and password have to be entered and select the language as "German" to get the german language in the application.



After pressing the "Login" button in the "Login Form", theDigiFlo application will be opened in German language. The below figure shows this



Different Operations that are performed in the "Language Manager" are as follows: -

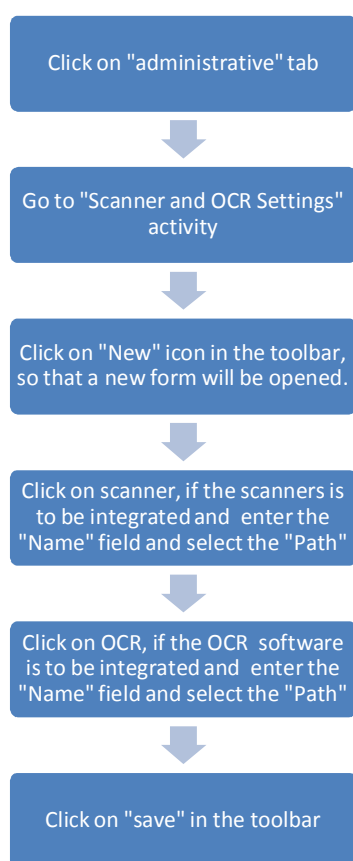
- "New" - This operation is used to get an empty form to create a new language.
- "Cancel" - To cancel the user before it has saved
- "Delete" - To delete the user even after saving the language
- "Save" - To save the language into the database after filling all the fields in the form.
- "Save and New" - To get an empty form for creating a new language after successfully saving a language record into the database.
- "Move Up" - To move the created language up the order
- "Move Down" - To move the created language down the order

1.1.6. Scanner and OCR Settings: -

Scanning of papers and OCR of images are the two different steps in the digitization process. Scanning of papers is done by the book scanners and OCR of the images is done by the OCR software.

So the scanner and the OCR software have to be integrated with the DigiFlo application. The DigiFlo application has to fire the Scanner software and OCR software. So to fire those two softwares, the path of those two software has to be set in the DigiFlo Application. So this "Scanner and OCR Settings" activity allows the DigiFlo users to set the path to the scanner and the OCR software.

Process of setting the path to the Scanner and the OCR software: -



Fields in the "Scanner and OCR Settings" form

Form - 1: -Scanner: -

- Name: - Enter the name of the scanner
- Path: - Select the path of the scanner exe file, so that the scanner will be integrated with the DigiFlo application.

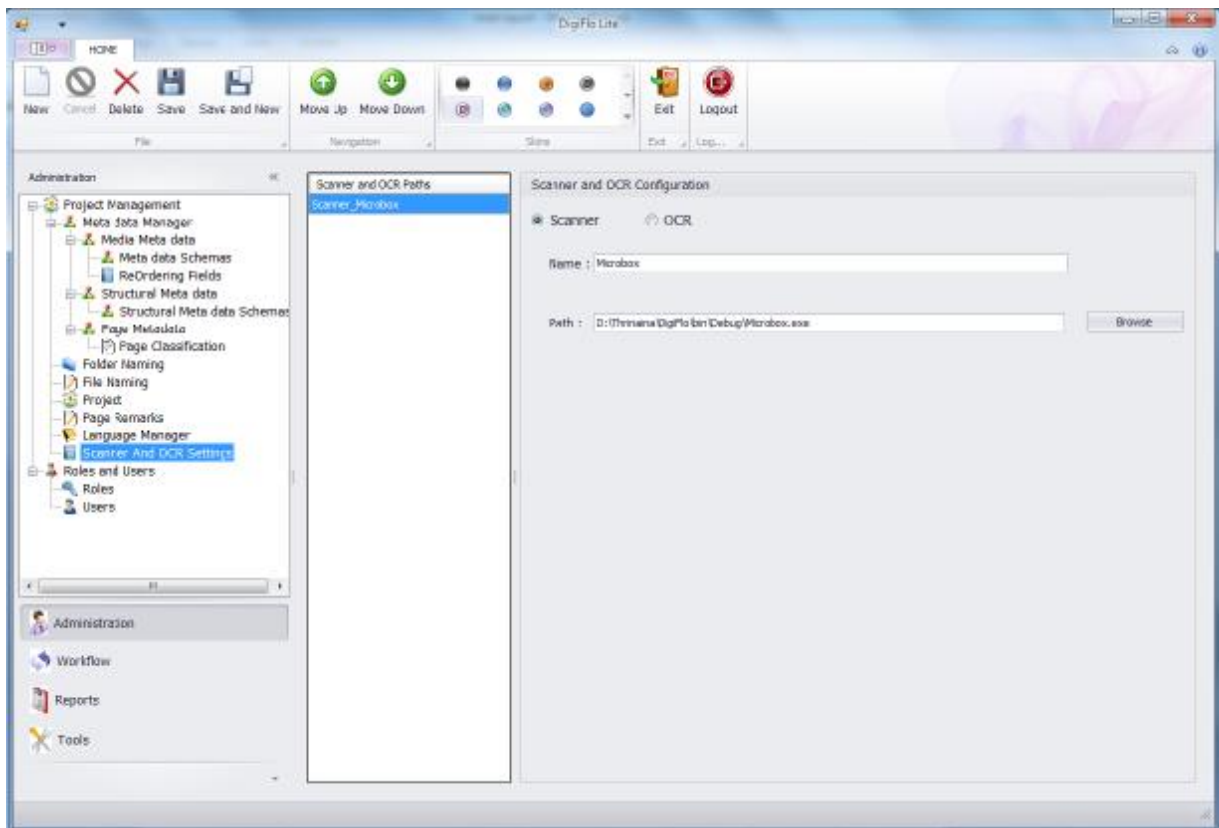
Form - 2: - OCR: -

- Name: - Enter the name of the OCR Software, which is used.
- Path: - Select the path of the OCR software exe file, so that the OCR software will be integrated with the DigiFlo application.

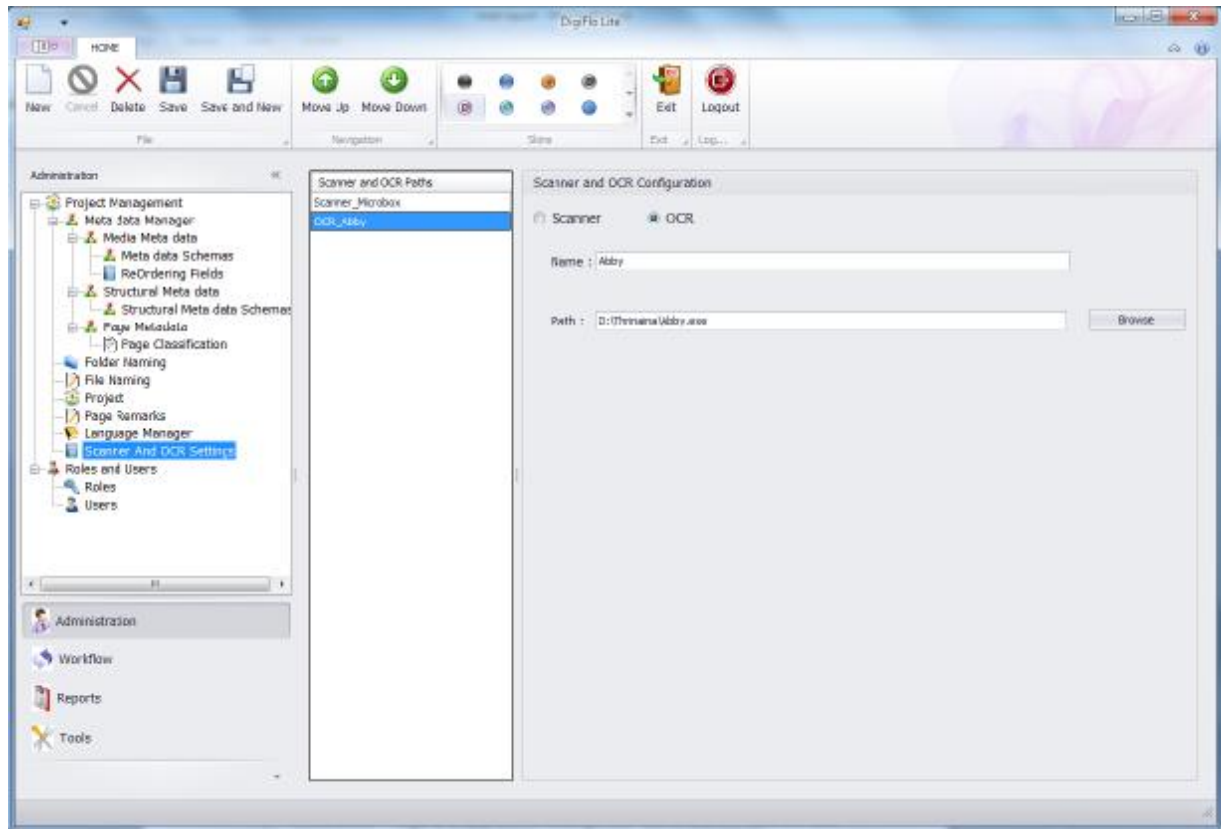
After the filling the required path, click on “save” icon in the toolbar to save the form.

The below figure shows the “Scanner and OCR settings” form” : -

Scanner Setting Form: -



OCR Setting Form: -



Different operations that can be performed in this "Scanner and OCR Settings" are: -

- "New" - This icon is used to get an empty "Scanner and OCR Configuration" form.
- "Cancel" - To cancel the "Scanner and OCR Paths" before it has saved
- "Delete" - To delete the "Scanner and OCR Paths" even after saving the role
- "Save" - To save the "Scanner and OCR" Path.
- "Save and New" - To get an empty form for creating a new "Scanner and OCR" path form after successfully saving the previous path.
- "Move Up" - To move the created "Path" up the order
- "Move Down" - To move the created "Path" down the order

1.2. Roles and Users: -

Security of the data and giving appropriate accesses are one of the prime roles of software in a multi user environment. This helps in ensuring that we are in a position to track the work, changes and what each person in the environment is doing. In such a multi user setup, each user needs to be assigned a definitive set of options depending to the role he plays in the organizational setup. A role defines the extent of access the user will have to the options within the software. A multiple number of roles can be created within the digitization setup ex. Scanning Operator, QC Operator, Image Processing Operator, Supervisor, Administrator etc. Creation of such roles defines the hierarchy within the operations and gives access to the portions of the software that the respective people are authorized to and also ensures that the security issues are addressed.

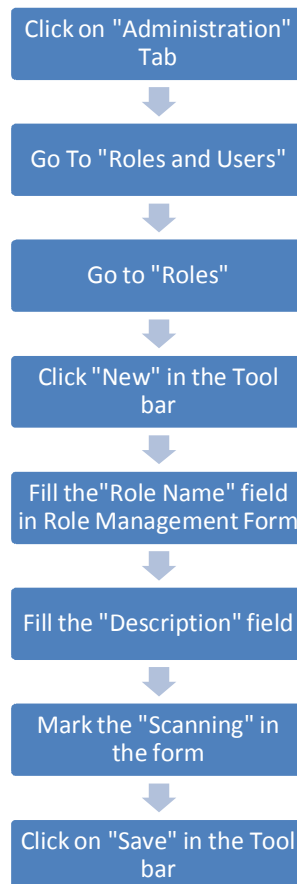
Only a person with the rights of an administrator can define and create roles. Once the roles have been created, users can be created and the respective roles applied.

Ex: if we have 3 Scanning operators John, James and Harry each one could be given a unique user id and password, since they would be doing the same task be assigned the same role. The illustration below will show you the fashion in which the above mentioned situation could be handled.

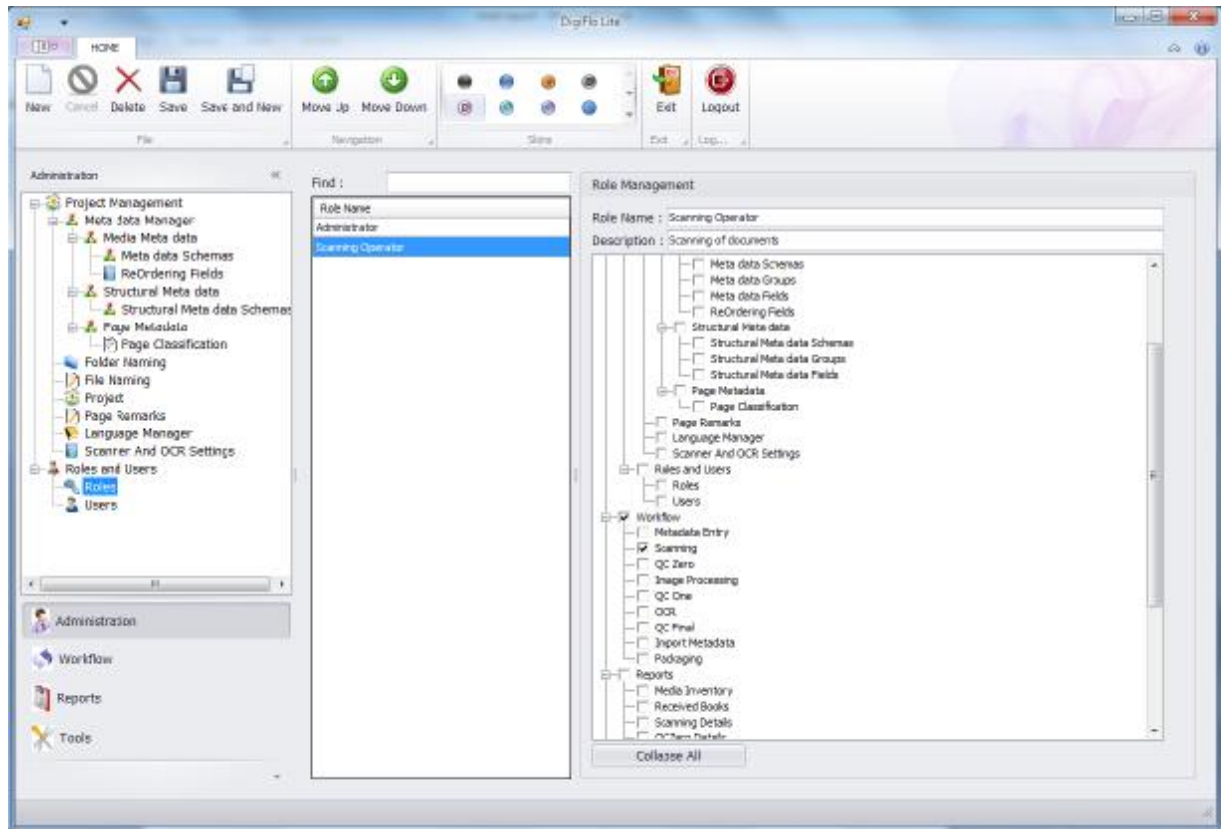
1.2.1 Role: -

Process of creating a role (Scanning operator) and assigning this role to the Users (John, James and Harry)

Process for creating a "Scanning Operator role" is as follows



The below figure shows the "Scanning" role has been created with scanning as his activity. By doing so, the scanning operator role can access only the Scanning activity. And the administrator has the right to give access to the other roles (scanning operator role).



Different operations that can be performed in this "Role" are: -

- "New" - This icon is used to get an empty form to create a new role.
- "Cancel" - To cancel the role before it has saved
- "Delete" - To delete the role even after saving the role
- "Save" - To save the role into the database
- "Save and New" - To get an empty form for creating a new User Role after successfully saving a record into the database.
- "Move Up" - To move the created roles up the order
- "Move Down" - To move the created role down the order
- "Expand all" - This operation is used to expand all the activities and the sub-activities in the Panel of the Role Management Form.
- "Collapse all" - This Operation will collapse all the activities and sub-activities in the panel of Role Management Form.

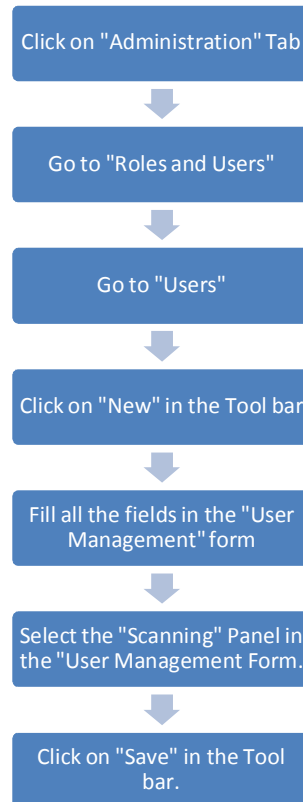
And After creating the Role, persons have to be assigned to these roles and these persons are called users.

The above example clearly shows that the Administrator has assigned three users (John, James and Harry) to the "scanning Operator" role.

1.2.2. Users: -

Process of Creating the "User" (John) is shown in the below flow-chart: -

Administration>Roles and Users>Users



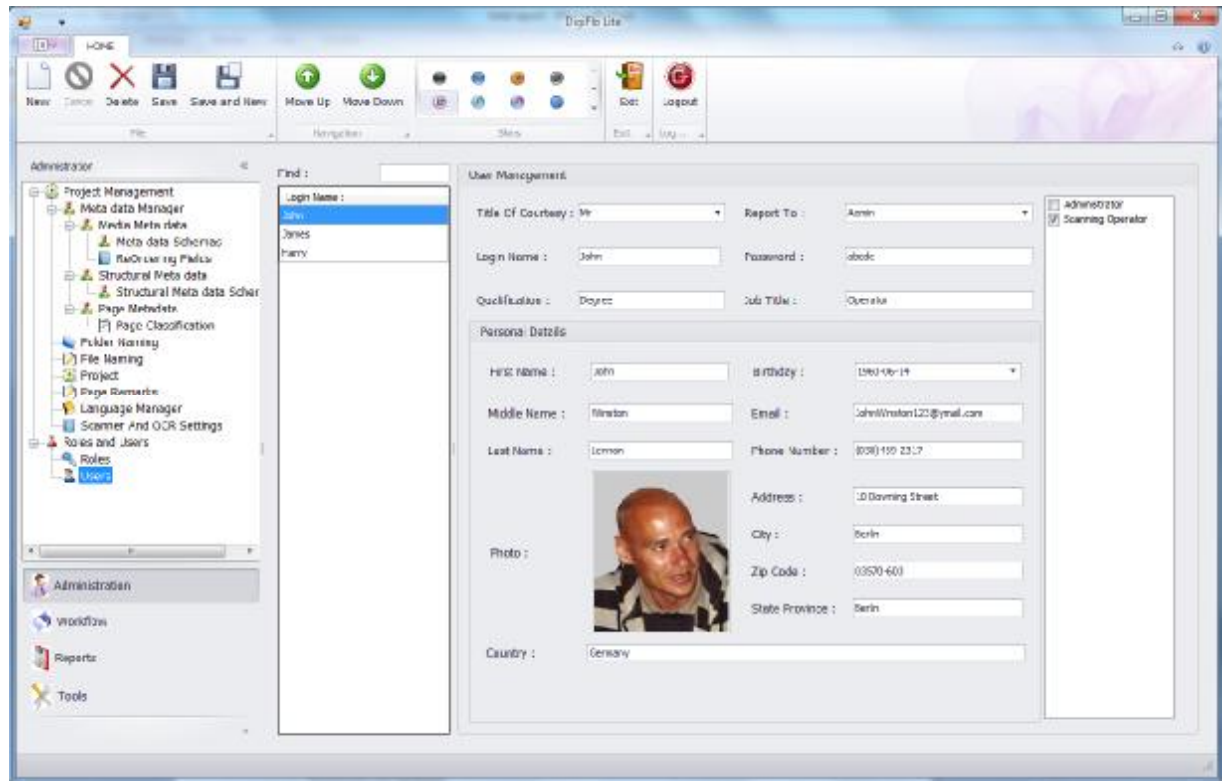
By entering all the fields in the "User Management" form, the User (John) will be saved in the database. And Login-name and Password will be created for the User (John), then the User (John) can access only his assigned roles (Scanning operator) role by using his login-name and password. And then John can perform his scanning activity.

And Different Users (John, James and Harry) are given different login-names and Passwords, even though they are doing same the role.

Fields that has to be filled in the "User Management" form: -

- "Title of the Courtesy" - Select from the drop-down (e.g.:- Mr.)
- "Report to" - Select the Authorizer from the drop-down (e.g.:- Admin)
- "Login name" - Enter a login name (e.g.:- John)
- "Password" - Enter a password. (e.g.:- abcde)
- The login name and password entered here for a user is used to access his roles.
- "Qualification" - Enter the educational qualification of the user.
- "Job Title" -Enter the designation of the user.
- Personal Details: -
- "First name" - Enter the first name of the user.
- "Middle name" - Enter the middle name of the user.
- "Last name" - Enter the last name of the user.
- "Photo" - Attach the photo of the user.
- "Birthday" - Select the birthday date of the user from the drop-down.
- "Email" - Enter the email of the user.
- "Phone Number" - Enter the phone number of the user.
- "Address" - Enter the address of the user.
- "City" - Enter the city of the user.
- "Zip Code" - Enter the code of the city.
- "State Province" - Enter the state province of the user.
- "Country" - Enter the country of the user.

The below figure shows User (John) has been created and Scanning Operator role has been assigned to him.



Different Operations that are performed in the "Users" are as follows: -

- "New" - This operation is used to get an empty form to create a new user.
- "Cancel" - To cancel the user before it has saved
- "Delete" - To delete the user even after saving the user
- "Save" - To save the user into the database after filling all the fields in the form.
- "Save and New" - To get an empty form for creating a new User after successfully saving a user record into the database.
- "Move Up" - To move the created users up the order
- "Move Down" - To move the created users down the order

2. Work flow: -

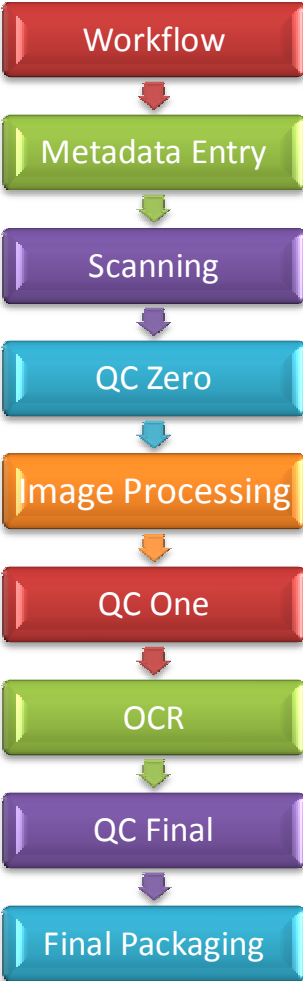
To execute any process, the activities in that process should be executed in a sequential order. Digitization process contains so many activities to be performed, ex-Metadata entry, scanning, Qc Zero, Image Processing, Qc One, OCR, QC Final. The workflow is the process of executing different activities in the sequential manner. The Administrator will define the workflow and sets the order of activities to be executed.

Different users are assigned to different roles in the process, and this workflow defines that with which activity the process has to be started, and with which activity the process has to be ended and in between the order of activities to be executed.

If the workflow is not defined then the users doesn't know which activity has to be started with and the order of activities to be continued with. And this will create a lot of confusion in the total process.

DigiFlo has defined certain workflow to execute the total process smoothly and to get a good output.

Workflow of Digitization process in DigiFlo: -



Here the process starts with Entering of Metadata for the given media and ends with the packaging of the output files. In between other activities are executed in the same order defined in the workflow in the above figure.

Ex: - If we have a book- "Glimpses Of World religions", which is to be digitized. The illustration below will show you how this book will be digitized by passing all the activities that are defined in the Workflow.

Digitization of the "Glimpses Of World religions" book will be executed step-by-step.

Metadata Entry: -

Metadata is the "data about data". Metadata is used for making the information searchable and retrievable. Metadata is used for managing the system for storage, processing and manipulation of information.

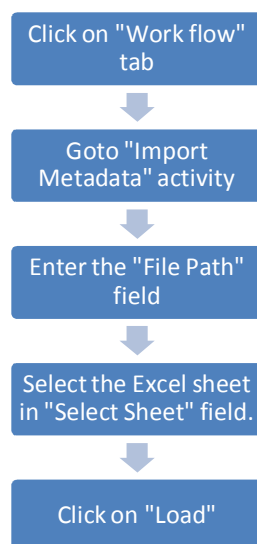
Now-a-days people need better ways to find and evaluate the information in the internet and the corporate intranet, and this is done by preparing and entering the metadata of a media in to the database.

Generally Metadata Entry is of two types, one is importing the saved data from system and another one is manual entry.

2.1. Import Metadata: -

This activity allows importing the metadata of the media from the system, if metadata is saved before. It is easy and saves lot of times.

Process of Importing the Metadata for the "Glimpses Of World religions" media from the system is as follows: -

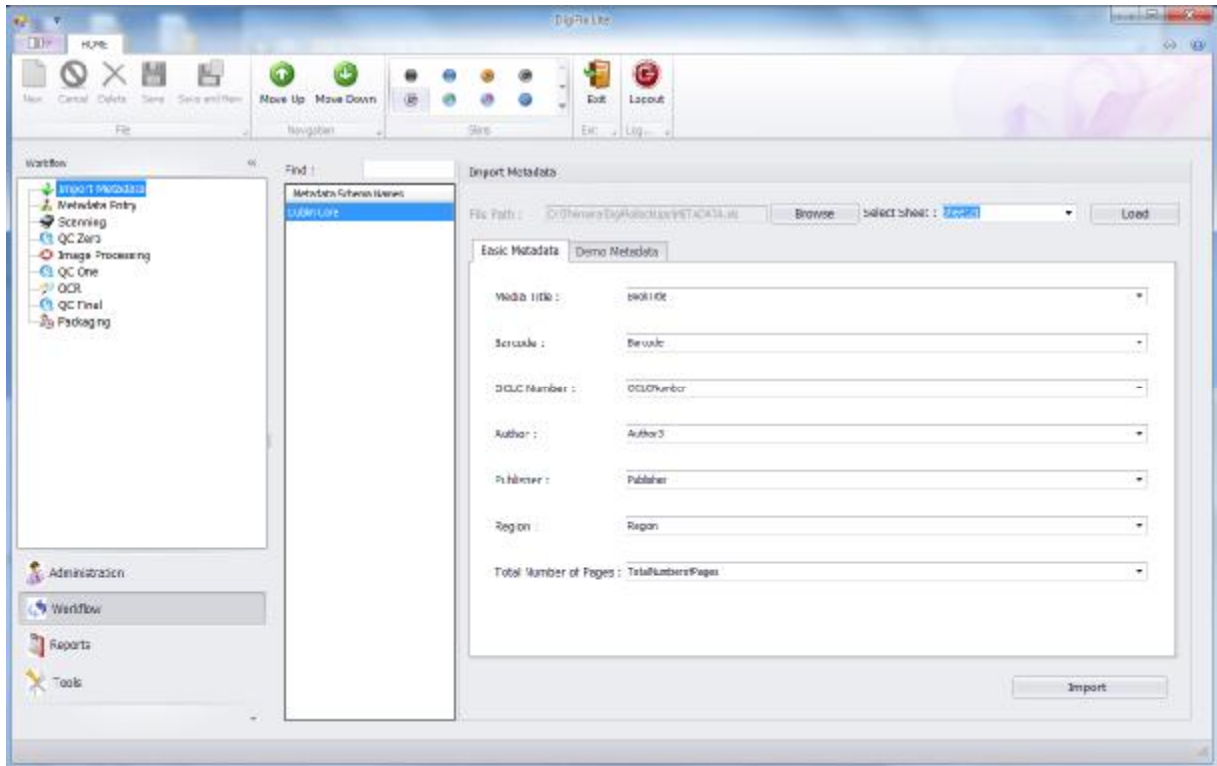


The Metadata will be imported either from "MS-Excel" or "XML" .

After selecting the file path, select the MS-Excel sheet number from where the metadata is imported. Then click on "Load" button to load all the information.

And now fields in the two tabs "Basic Metadata" and "Demo Metadata" has to be filled. And the words in the MS-Excel might be different from the fields in the form. So each field in the form has to be selected from the drop-down.

The below figure shows the "Import Metadata" form, and all the fields in the "Basic Metadata" and the "Demo Metadata" tabs are filled with the appropriate columns in the selected MS-Excel sheet.

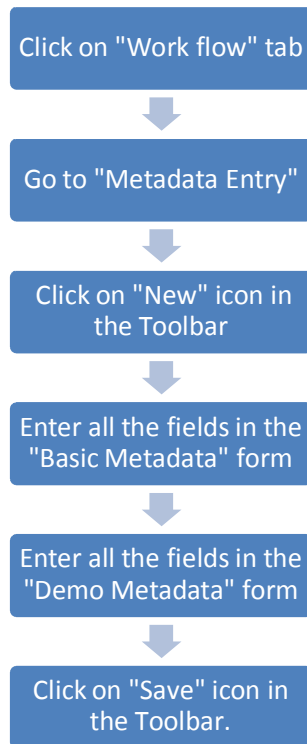


After that click on "Import" button to import all the information in the Excel sheet to the application.

2.2. Metadata Entry: -

If the Metadata is not there in the Excel sheets or in XML, then manually Metadata has to be entered for each media. So this "Metadata Entry" activity is used to enter the metadata of the each book manually.

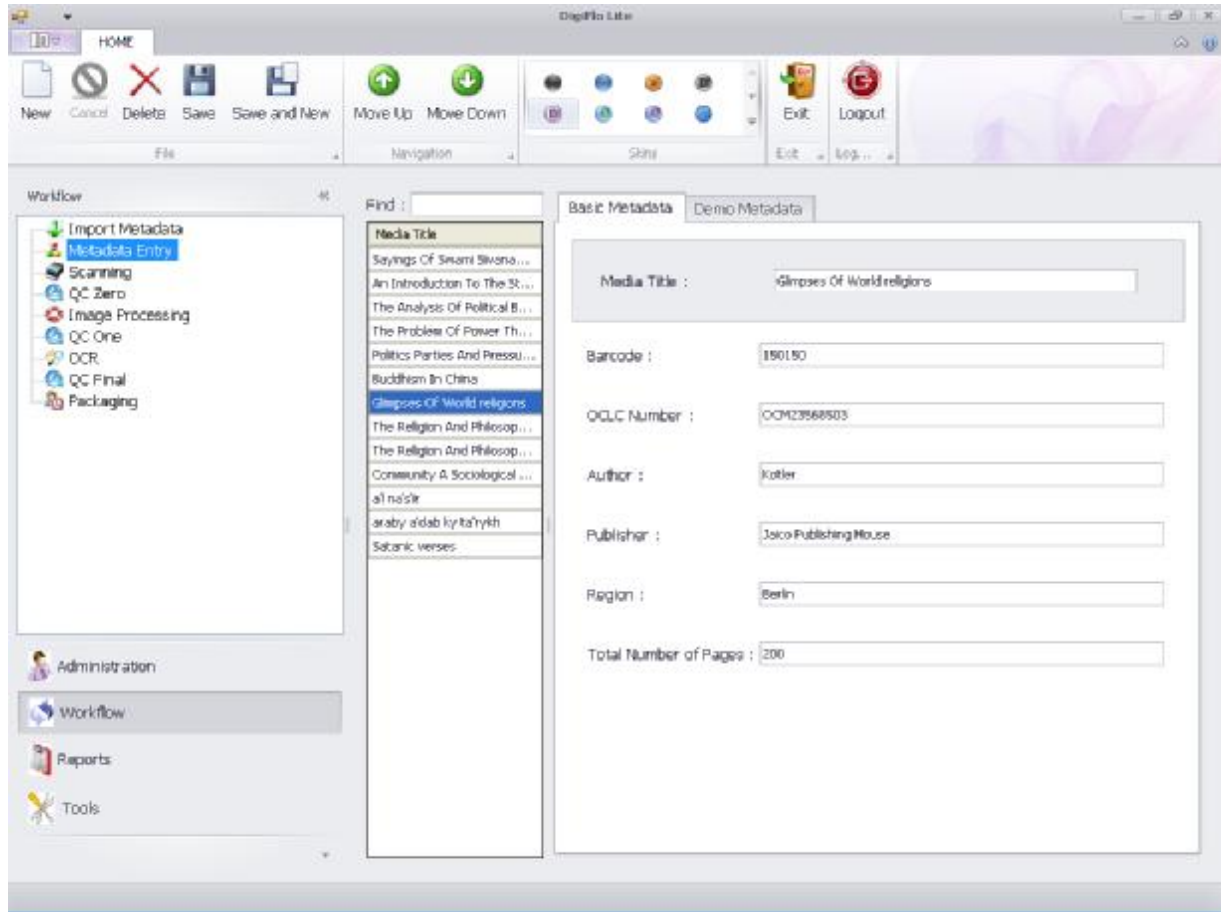
Process of entering the Metadata Entry for the "Glimpses Of World religions" media: -



The "Basic Metadata" form is the mandatory form and all the fields are to be filled. And these fields are used as parameters while creating the file_names and folder_names.

Different fields in the "Basic Demo" tab: -

- "Media Title" - Enter the name of the media. (Ex: - Glimpses Of World religions)
- "Barcode" - Enter the barcode of the media. (Ex: -150150)
- "OCLC Number" - Enter the OCLC number of the media. (Ex: - OCM23568503)
- "Author" - Enter the author of the media. (Ex: -Kotler)
- "Publisher" - Enter the publisher of the media. (Ex: - Jaico Publishing House)
- "Region" - Enter the region to which the media belongs to. (Ex: -Berlin)
- "Total Number of pages" - Enter the total number of pages that the media contains. (Ex: - 200)



The above figure shows that “Glimpses Of World religions” media has been created and all the fields in “Basic Metadata” form have been filled.

After filling all the fields in the “Basic Metadata” tab, then fields in the “Demo Metadata” tab has to be filled. This “Demo Metadata” tab has been created in the “Metadata Scheme” activity of the Administration” tab. Even the fields in the “Demo Metadata” tab are also created manually by the administrator.

Fields in the “Demo Metadata” tab: -

- “Scanning Centre” -Enter the place where scanning has to be done.
- “Received date and time” - Select the date and time from the drop-down.
- “Author 2” - Enter the second author of the media
- “Author 3” - Enter the third author of the media.
- “Subject” - Enter the subject on which the media has been written.
- “Book Size” - Enter the size of the media.
- “Scanner No” - Enter the scanner number on which scanning of this media is done.
- “Digital Republisher” - Enter the Publisher who is doing this digitization work.

- "Format" - Enter the format of the media
- "Identifier" - Enter the name of the Identifier.
- "Source" - Enter the name of the source from which the information have been taken.
- "Coverage" - Enter the jurisdiction under which the media has been written.
- "Copyrights" - Enter the copyrights
- "Digital Republishing Date" - Enter the date on which the digitization of this media is done.
- "Copyrights Owner" - Enter the name of the person or institution who owns the copyrights of this media.
- "Copyrights Expiry date" - Enter the date on which the copyrights will be expired.
- "ISBN No" - Enter the "International Standard Book Number" for the given media
- "Contributor" - Enter the name of the Contributor, who contributed information to the preparation of the media.
- "Total No of pages" - Enter the number of pages of the media.
- "Book Bank" - Enter the name of the Book Bank from where this media has been taken
- "Received From" - Enter the name of the person or the institution who have given the book for digitization.
- "Received by" - Enter the name of the person who has received the media.
- "Publisher Place" - Enter the place of the publisher.
- "Publishing Date" - Enter the date on which the media has been published.
- "Document type" - Enter the type of document (Ex: -book)
- "Language" - Enter the name of the language in which the media has been published.

Some fields in this "Demo Metadata" form are mandatory and some are not. This is done according the requirement of the customers.

If the Field in the "Demo Metadata" is mandatory, then a star (*) will be displayed on the field.

Different Operations that are performed in the "Metadata Entry" are as follows: -

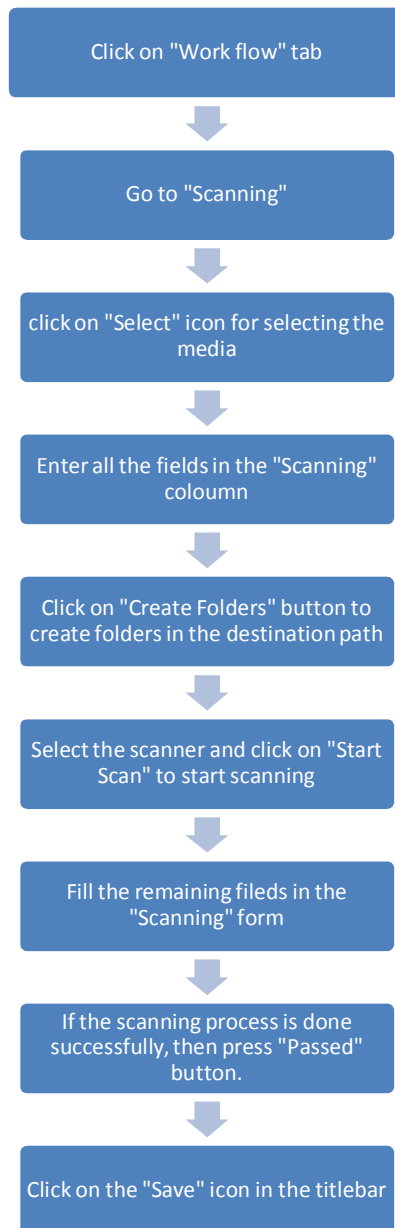
- "New" - This operation is used to get an empty form to create a new media.
- "Cancel" - To cancel the media before it has saved
- "Delete" - To delete the media even after saving the user
- "Save" - To save the media into the database after filling all the fields in the "Basic Metadata" form and "Demo Metadata" form
- "Save and New" - To get an empty form for creating a new media after successfully saving a previous media record into the database.
- "Move Up" - To move the created media up the order.
- "Move Down" - To move the created media down the order.

2.2. Scanning: -

After the metadata entry for the media, now the next step is scanning the given media. The process of converting the physical books and magazines into digital media such as images, electronic text, or electronic books (e-books) is called scanning. Book scanners are used for scanning of the media.

The scanned images are converted in to OTIFF format and stored in the database

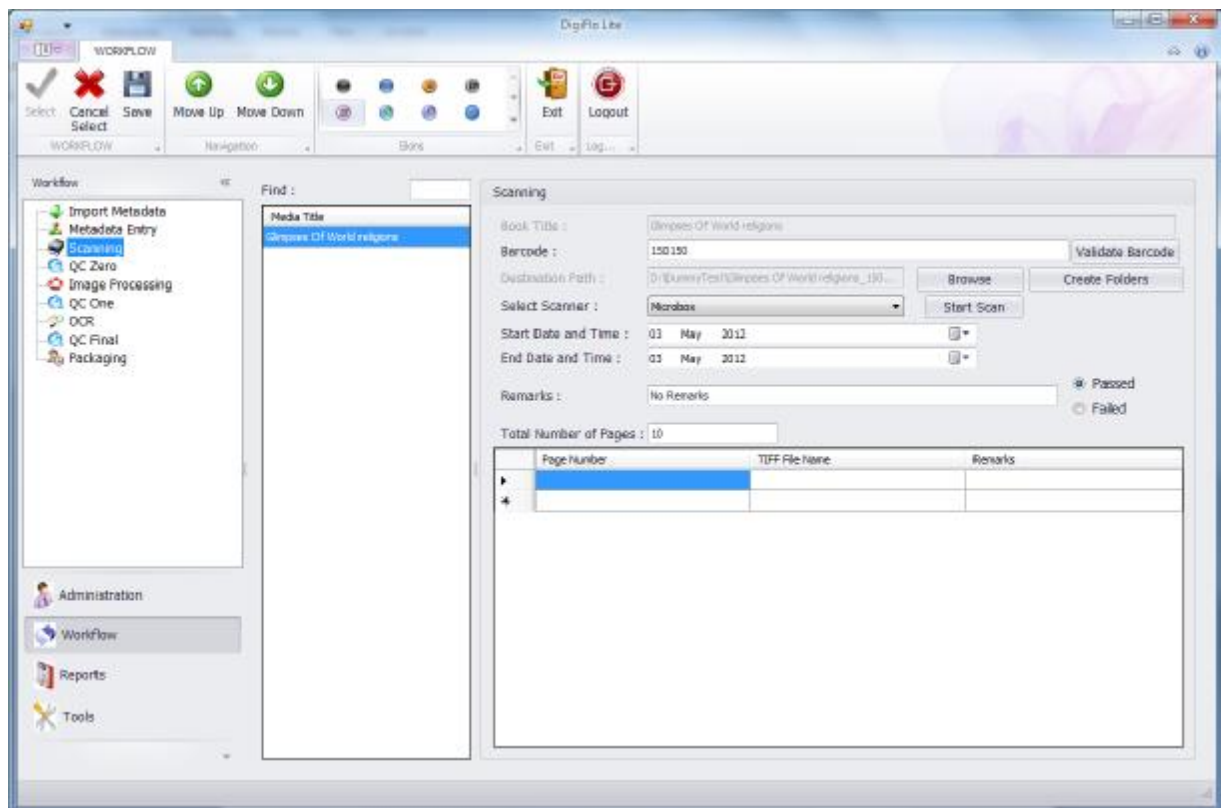
The process of Scanning the "Glimpses Of World religions" media: -



Different fields in the "Scanning" form: -

- “Book Title” - Enter the title of the media that has to be scanned.
- “Barcode” - Enter the barcode of the particular media. And the entered barcode can be verified by using the “Validate Barcode” button.
- “Destination path” - Enter the destination path manually or “Browse” button is used to browse the path for storing the scanned images.
- “Create Folders” - This button is used to create the folders in the database. The scanned images are stored in these folders only.
- “Select Scanner” - Select the scanner from the drop-down.
- “Start Scan” - This button is used to start the scanning process.
- “Start Date and Time” - Select the date and time from the drop-down.
- “End Date and Time” - Select the date and time from the drop-down.
- “Remarks” - This option is used to write any remarks that you find while doing the scanning operation.
- “Total Number of pages” - Enter the number of pages to be scanned.

The below figure shows that the scanning for the “Glimpses Of World religions” media has been done.



There is a Grid at the bottom of the Scanning form. This is used to mention any remarks if required. This Grid contains three columns; they are "Page Number", "TIFF File Name", and "Remarks".

It is mandatory to fire the scanner software with the DigiFlo application.

All the scanned images will be stored in the given "Destination Path", with the specified names. And these images are used in further steps.

If the scanning process is done successfully, then click on "Passed" button and click on "Save" icon in the toolbar, to pass this media into next step, i.e. QC Zero

After the scanning, if the errors are more, then select "Failed" button in the Scanning form, so that the media did not pass in to the next step, i.e. QC ZERO.

Different Operations that are performed in the "Scanning" are as follows: -

- "Select" - This operation is used to select the media that has to be scanned.
- "Cancel Select" - This operation is used to cancel the selected media.
- "Save" - This operation is used to save all the information in the fields of the scanning operation in to the database.
- "Move Up" - To move the media up the order.
- "Move Down" - To move the media down the order.

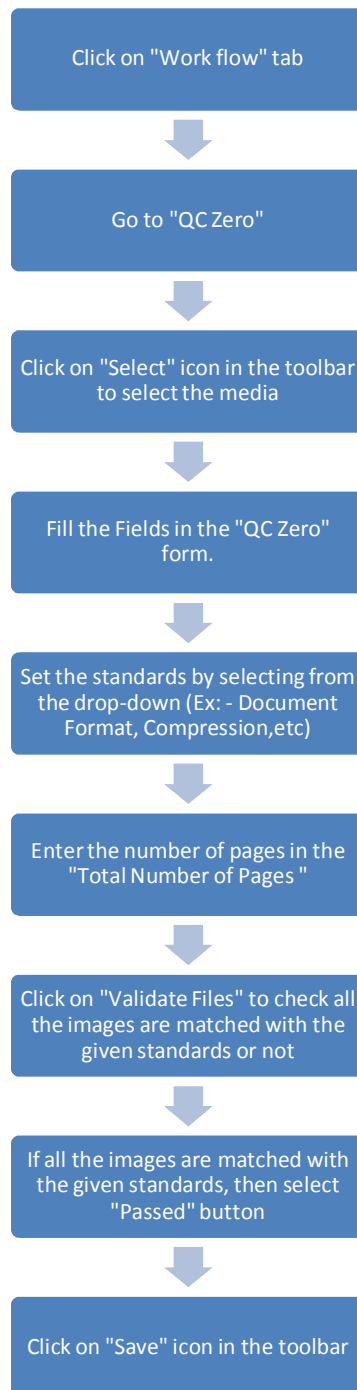
2.3. QC Zero: -

After the scanning of the media, the converted OTIFFs are processed to get good quality. In between this a verification has to be done to check no errors are found in the scanned images, this verification is called Quality Check Zero (QC Zero).

This Quality check has to be done according to the pre-defined standards. These standards are set by the customers who have given the media for digitization. And if all the standards are met by the scanned images, then this indicates that the QC Zero has been done successfully and the given media is passed to the next step, i.e. Image Processing.

If some images did not match with the given standards, then this indicates the QC Zero has been failed and again the media has been resent for scanning.

The process of doing the QC Zero for "Glimpses Of World religions" media: -



Different Fields that are to be filled in the "QC Zero" form: -

- "Media Title" - Name of the media that has to undergo Quality Check. Here when we select the media in "Media Title", the automatically the Media title will be filled.
- "Folder Path" - Select the path of the existing scanned documents. We can select this by using the "Browse" option.
- "Destination Path" - Select the Path for the Images that are to be stored after doing the QC Zero.

Standards that are set for QC Zero: -

- "Document Format" - Select the document format from drop-down. And mark it in the box next to it.
- "Compression" - Select the Compression from the drop-down. And mark the box next to it.
- "Horizontal resolution" - Select the Horizontal Resolution from the drop-down. And mark the box next to it.
- "Vertical resolution" - Select the Vertical Resolution from the drop-down. And mark the box next to it.
- "Page Width" - Enter the page width for the selected media file
- "Page Height" - Enter the page height for the selected media file.
- "Min.File Size (in KB)" - Enter the Min.File Size (in KB) for selected Media file.
- "Max.File Size (in KB)" - Enter the Max.File Size (in KB) for the selected Media File

Select all the standards that are to be checked for the scanned images.

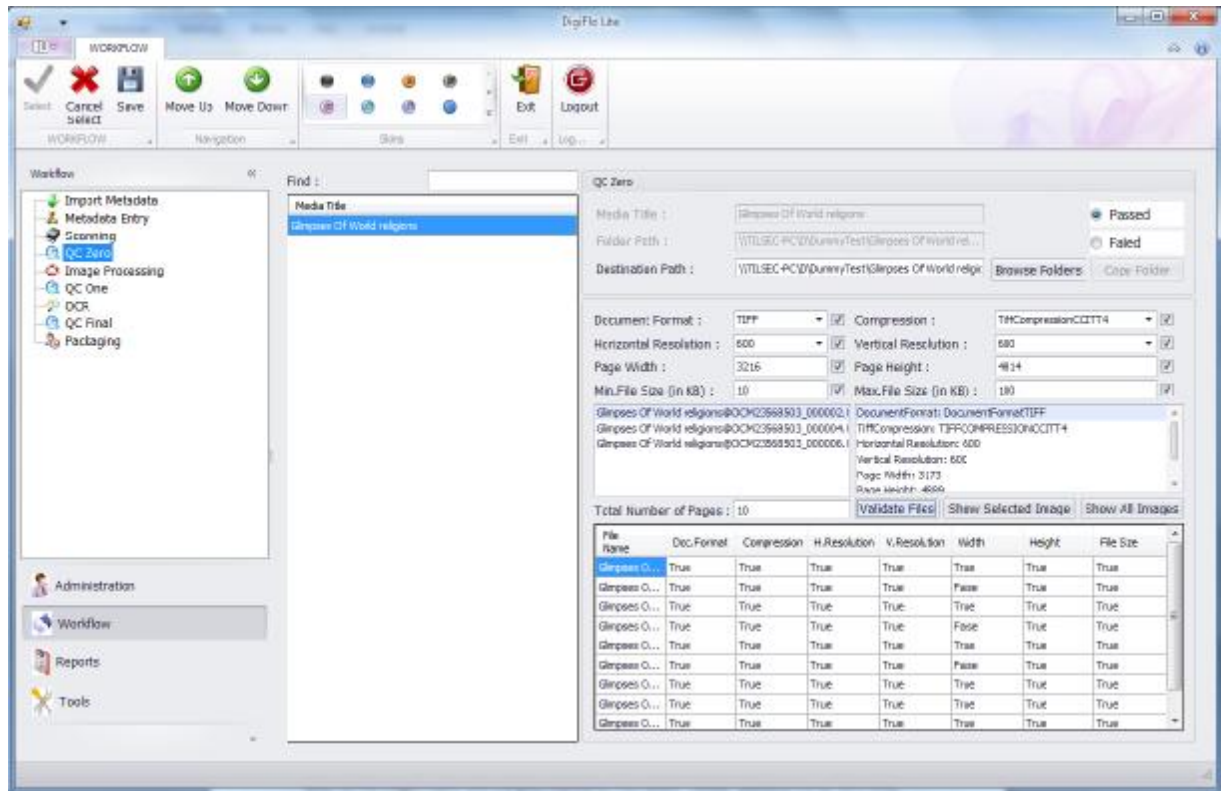
"Total number of Pages" - Enter the number of pages that have to undergo quality check.

Then click on "Validate Files" to see all the standards of the scanned images are matched with the given standards or not. And all the image filenames will be displayed in the grid at the bottom of the "QC Zero" form.

The Grid at the bottom of the "QC Zero" will contain all the Standards as the columns. (Ex: - Document Format, Compression, etc). If the image has matched with the certain standard then "TRUE" word will be displayed. And if the image did not match with the certain standard, then "FALSE" word will be displayed.

The Two grids just below the Standards, will give the information about the errors. The left grid will specify the filenames of the error images. And the Right grid will give the complete description about the error image.

The below figure shows the QC Zero for the "Glimpses Of World religions" media has been done



When all the images match with the given standards, then select “Passed” button in the QC Zero form, and click on “Save” icon in the toolbar to pass this media in to the next step, i.e. Image Processing.

If the images did not match with the given standards, then select “Failed” button and the given media did not pass to the next step.

Different Operations that are performed in the “QC Zero” are as follows: -

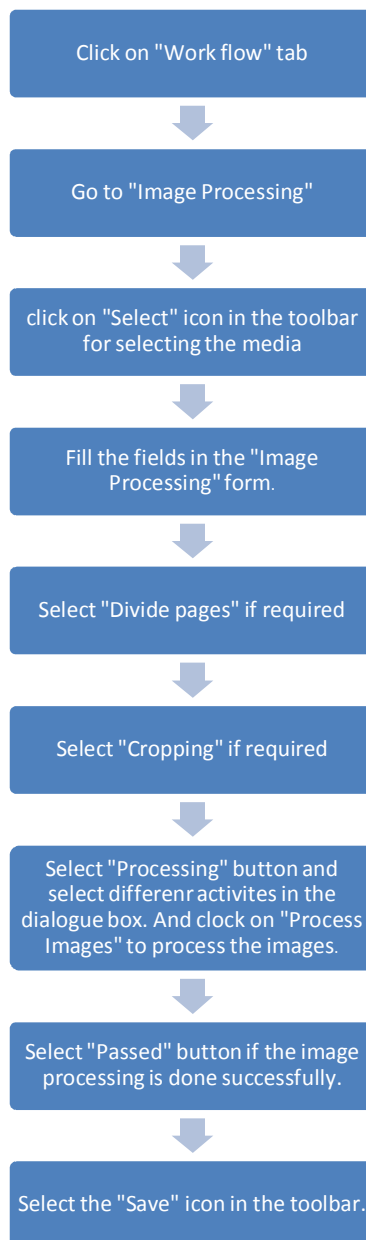
- “Select” - This operation is used to select the media that has to undergo QC Zero.
- “Cancel Select” - This operation is used to cancel the selected media.
- “Save” - This operation is used to save all the information in the fields of the QC Zero form in to the database.
- “Move Up” - To move the media up the order.
- “Move Down” - To move the media down the order.

2.4. Image Processing: -

Generally after the scanning of the media is done, some disturbances might be seen in some scanned images, and some scanned images will be blurred where the images are not seen properly. So to remove all the disturbances and to make the scanned images visible, the image processing activity is executed.

The image processing executes different activities like Auto crop, Auto Deskew, Auto Despeckle, Invert; divide the Images, Cropping the images, etc to make the image more visible.

The process of doing the Image Processing for "Glimpses Of World religions" media: -



Different fields that are to be filled in the "Image Processing" form: -

- "Book Title" - Name of the media that has to undergo "Image Processing".
- "Source Path" - Browse the source path from where we get the media files for image processing.
- "Destination Path" - Browse the Destination path to where these processed images are stored.
- "Start Date and Time" - Select the date and time for the Drop-down for starting the image processing.
- "End Date and Time" - Select the date and time from the drop-down for ending the image processing.
- "Divide Page" - This button is used to divide the pages horizontally or vertically.
- "Cropping" - Cropping button is used for the removal of the outer parts of an image, to make the image a good quality one.
- "Processing" - Processing button is used to perform different activities on the images.
- "Remarks" - Enter the remarks if required.
- "Total Number of Pages" - Enter the number of pages in the media that are to be processed.

The below shows the "divide" dialog-box has been opened to divide the required images horizontally or vertically: -



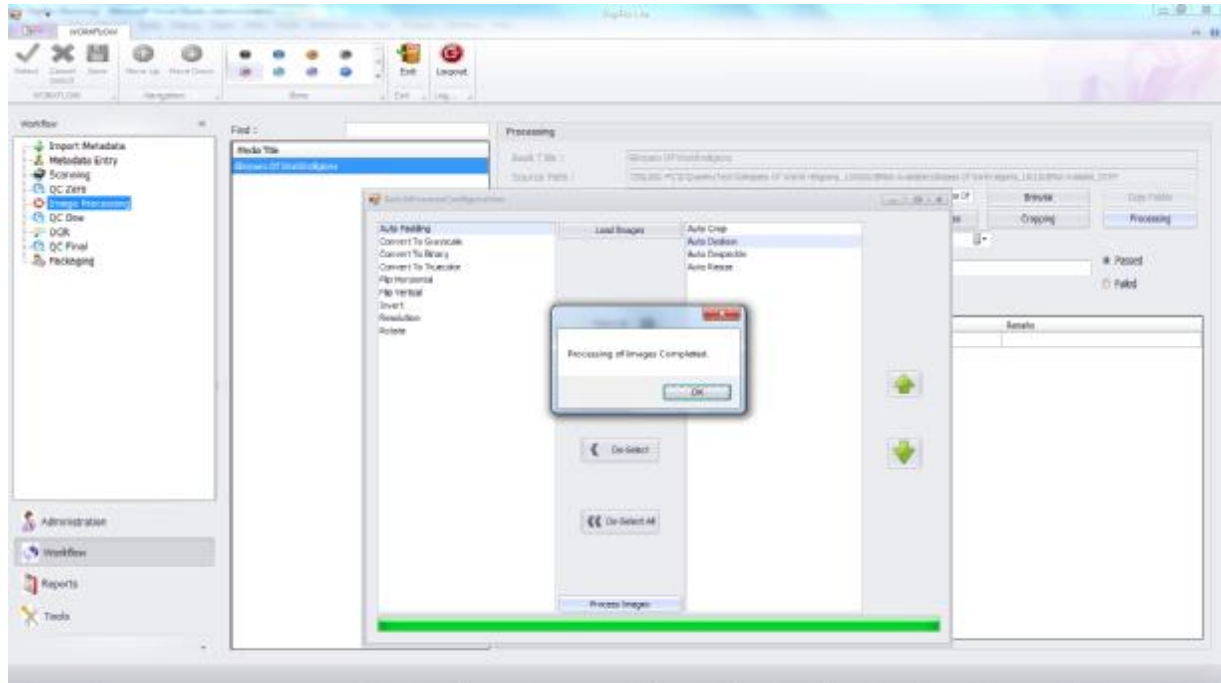
These divided images will be stored in the specified folders.

Image Processing:-

This image processing contains different activities which are performed on the images. They are

- Auto Deskew - The Automatic Deskewing command is ideal to equalize the image orientation of a document if it is placed slightly rotated on the layout section.
- Auto crop - This activity is used to remove the edges from the images in the scanned document.
- Auto Despeckle-Speckle indicates "Dots" . Despeckle means removing the dots. Here "Auto Despeckle" removes the dots (Noise) from the image and makes it more visible.
- Auto padding - If the size of the image file has reduced while Auto Cropping, then this activity is used to add the pieces of blank images to make the image file to its original size. So that all the image files will be of uniform size.
- Convert to Grayscale - This activity is used to convert from scanned colour images (24 bit or 32bit) to grayscale images (8bit).
- Convert to Binary - This activity is used to convert the given images in to Binary images (2bit). So these converted images will be in Black & White form.
- Convert to Truecolor - This activity is used to convert the Binary images or Grayscale images into original Color images.
- Flip Horizontal - This activity is used to convert the scanned image into Mirrored image and display it horizontally.
- Flip Vertical - This activity is used to convert the scanned image into Mirrored image and display it vertically.
- Invert - This activity is used to invert the colors. I.e. White to Black (or) Black to White. Generally by doing this activity, if the background is in white and the sentences are in black, then they will be converted as Background as Black and the sentences as White.
- Resolution - This activity is use to change the resolution of the image. This is used for making the image brighter. Generally the resolution is inversely proportional to the "Length and width of the image" . And in this DigiFlo software, even if the resolution of the image is increased, the "Length and Width" of the image is not changed, this indicates the DigiFlo application will keep all the scanned images in the uniform size.
- Rotate - This activity is used to rotate the scanned image

The below image will show different activities in the "Image Processing": -



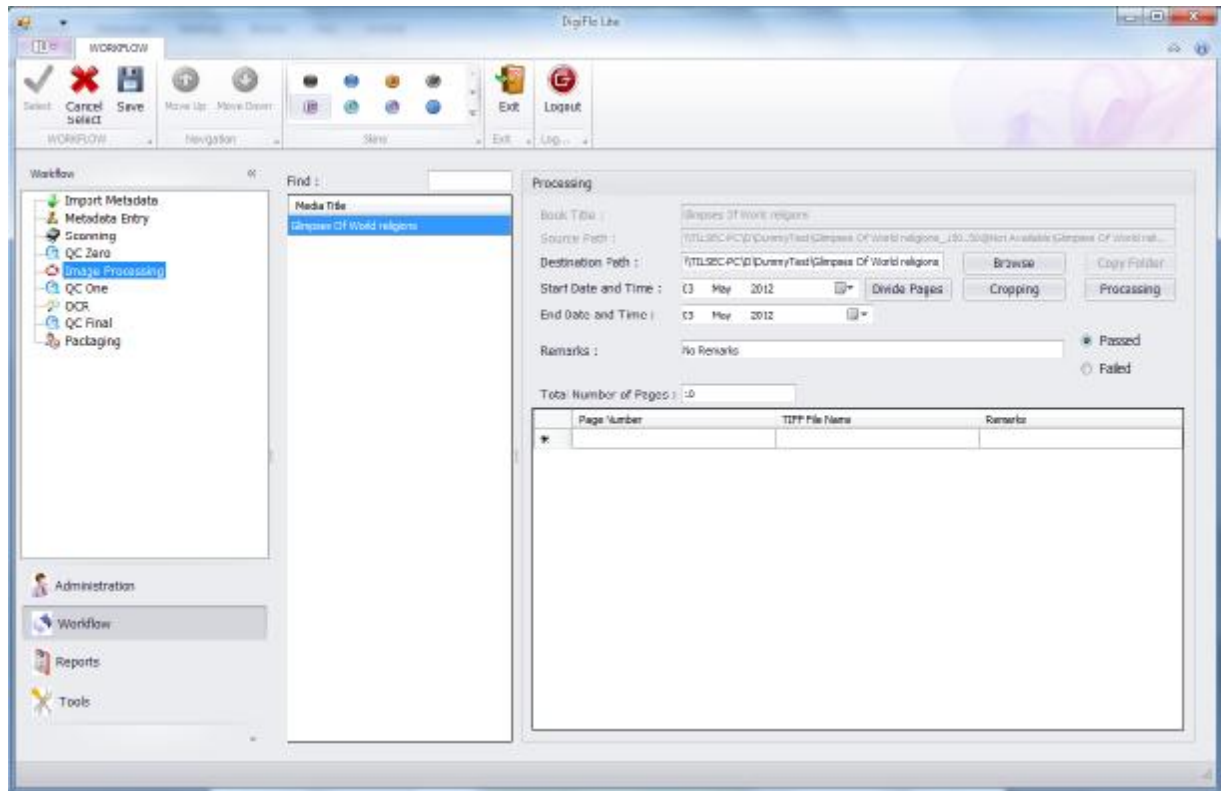
Different buttons in the "Processing" dialog-box are as follows: -

- "Load Images" - This is used to load the images from the source path
- "Select All" - This is used to select all the activities in the left grid and show them in the right grid.
- "Select" - This is used to select the activity individually, and the selected activities will be shown in the right grid.
- "De-Select" - This is used to De-Select the activities individually.
- "De-Select All" - This is used to De-Select all the activities from the right grid.
- "Process Images" - This is used to process all the selected activities and the out put of this is shown in the "Destination Path"

And the arrow marks in the dialog box are used to move the selected activities up and down according to the requirement.

All the images that are processed are stored in the destination path with the specified file names.

The below figure shows the Image Processing form for the "Glimpses Of World religions" media.



There is a Grid at the bottom of the “Image Processing” form. This is used to mention any remarks if required. This Grid contains three columns; they are “Page Number”, “TIFF File Name”, and “Remarks” .

If the Image processing is done successfully, then click on “Passed” button and click on “Save” icon in the toolbar, to pass this media into next step, i.e. QC One

After the Image Processing, if the errors are more, then select “Failed” button in the Image Processing form, so that the media did not pass in to the next step, i.e. QC One.

Different Operations that are performed in the “Image Processing” are as follows: -

- “Select” - This operation is used to select the media that has to be processed.
- “Cancel Select” - This operation is used to cancel the selected media.
- “Save” - This operation is used to save all the information in the fields of the “Image Processing” operation in to the database.
- “Move Up” - To move the media up the order.
- “Move Down” - To move the media down the order.

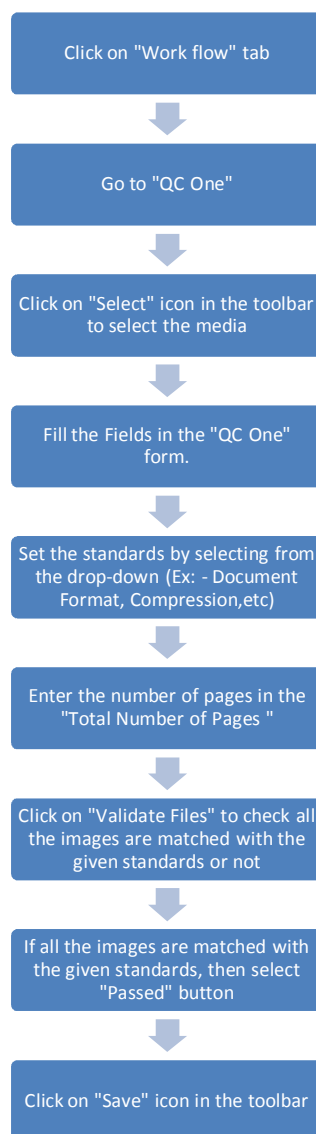
2.5. QC One: -

After the Image Processing of the media, the PTIFFs will be converted into editable text format by OCR activity. In between this verification has to be done to check no errors are found in the Processed images, this verification is called Quality Check One (QC One).

This Quality check has to be done according to the pre-defined standards. These standards are set by the customers who have given the media for digitization. And if all the standards match with the processed images, then this indicates that the QC One has been done successfully and the given media is passed to the next step, i.e. Image Processing.

If some images did not match with the given standards, then this indicates the QC One has been failed and again the media has been resent for Image Processing.

The process of doing the QC One for "Glimpses Of World religions" media: -



Different Fields that are to be filled in the "QC One" form: -

- "Media Title" - Name of the media that has to undergo Quality Check. Here when we select the media in "Media Title", the automatically the Media title will be filled.
- "Folder Path" - Select the path of the existing scanned documents. We can select this by using the "Browse" option.
- "Destination Path" - Select the Path for the Images that are to be stored after doing the QC One.

Standards that are set for QC One: -

- "Document Format" - Select the document format from drop-down. And mark it in the box next to it.
- "Compression" - Select the Compression from the drop-down. And mark the box next to it.
- "Horizontal resolution" - Select the Horizontal Resolution from the drop-down. And mark the box next to it.
- "Vertical resolution" - Select the Vertical Resolution from the drop-down. And mark the box next to it.
- "Page Width" - Enter the page width for the selected media file
- "Page Height" - Enter the page height for the selected media file.
- "Min.File Size (in KB)" - Enter the Min.File Size (in KB) for selected Media file.
- "Max.File Size (in KB)" - Enter the Max.File Size (in KB) for the selected Media File

Select all the standards that are to be checked for the "Image Processed" images.

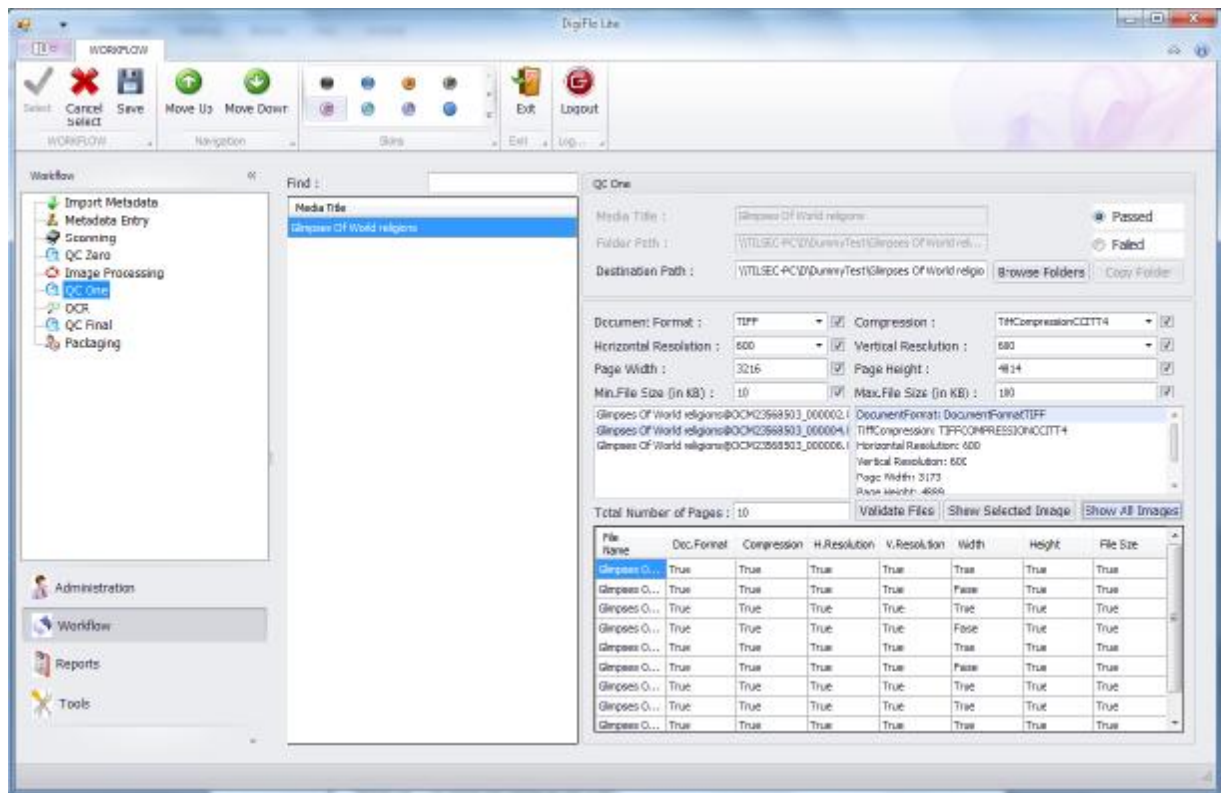
"Total number of Pages" - Enter the number of pages that have to undergo quality check.

Then click on "Validate Files" to see all the standards of the "Image Processed" images are matched with the given standards or not. And all the image filenames will be displayed in the grid at the bottom of the "QC One" form.

The Grid at the bottom of the "QC One" will contain all the Standards as the columns. (Ex: - Document Format, Compression, etc). If the image has matched with the certain standard then "TRUE" word will be displayed. And if the image did not match with the certain standard, then "FALSE" word will be displayed.

The Two grids just below the Standards, will give the information about the errors. The left grid will specify the filenames of the error images. And the Right grid will give the complete description about the error image.

The below figure shows the QC One for the "Glimpses Of World religions" media has been done



When all the images match with the given standards, then select “Passed” button in the QC One, and click on “Save” icon in the toolbar to pass this media in to the next step, i.e. OCR.

If the images did not match with the given standards, then select “Failed” button and the given media did not pass to the next step.

Different Operations that are performed in the “QC One” are as follows: -

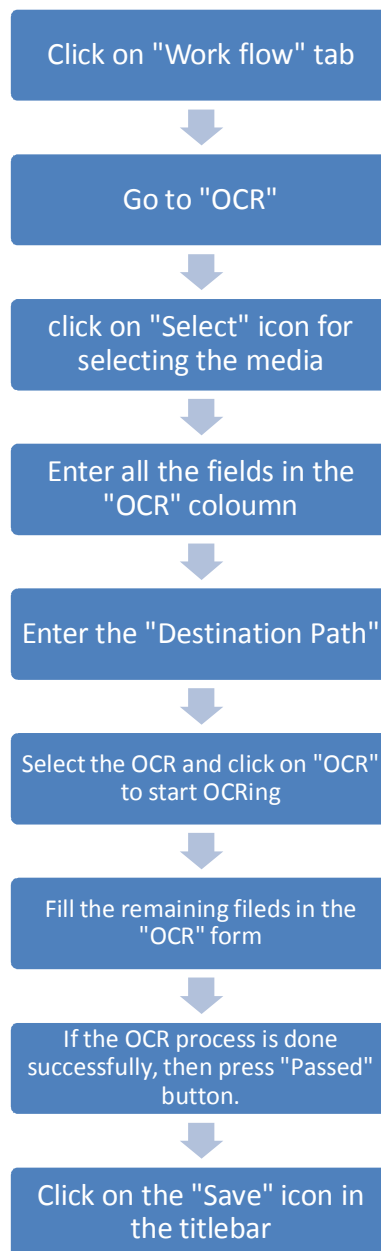
- “Select” - This operation is used to select the media that has to undergo QC One.
- “Cancel Select” - This operation is used to cancel the selected media.
- “Save” - This operation is used to save all the information in the fields of the QC One form in to the database.
- “Move Up” - To move the media up the order.
- “Move Down” - To move the media down the order.

2.6. OCR: -

OCR of a document is the process of converting the images of handwritten or printed characters in to Machine-Readable characters.

The images are made clear by using the "Image Processing ", and now by using the OCR these images are converted into editable text.

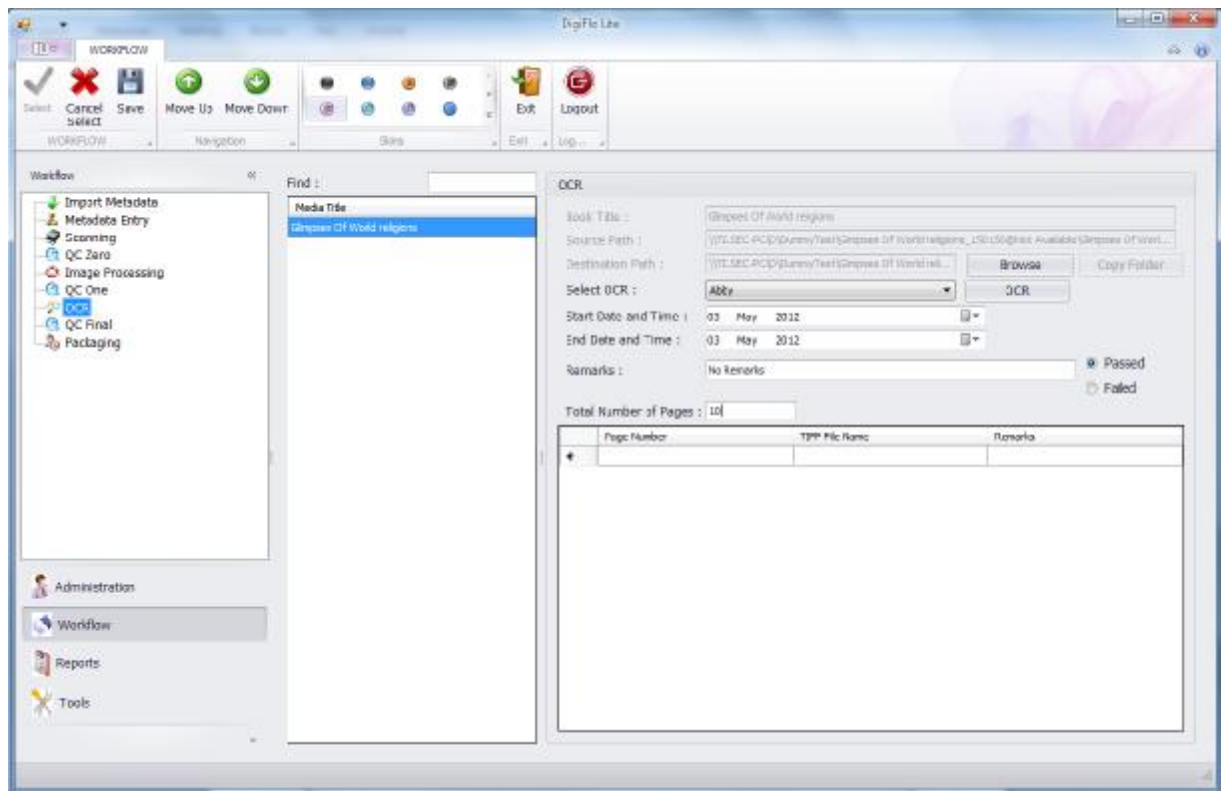
The process of OCR the "Glimpses Of World religions" media: -



Different fields in the "OCR" form: -

- “Book Title” - Enter the title of the media that has to be OCR
- “Source Path” - Browse the source path from where we get the media files for OCR
- “Destination Path” - Browse the Destination path to where these OCR images are stored.
- “Select OCR” - Select the OCR software from the drop-down.
- “Start Date and Time” - Select the date and time for the Drop-down for starting the image processing.
- “End Date and Time” - Select the date and time from the drop-down for ending the image processing.
- “Remarks” - Specify, if there are any errors in the OCR process
- “Total Number of Pages” - Enter the number of pages in the media

The below figure shows that the OCR process for the “Glimpses Of World religions” media has been done.



There is a Grid at the bottom of the OCR form. This is used to mention any remarks if required. This Grid contains three columns; they are “Page Number”, “TIFF File Name”, and “Remarks”.

It is mandatory to fire the OCR software with the DigiFlo application.

All the OCR images will be stored in the given "Destination Path", with the specified names. And these images are used in further steps.

If the OCR process is done successfully, then click on "Passed" button and click on "Save" icon in the toolbar, to pass this media into next step, i.e. QC Final

After the OCR, if the errors are more, then select "Failed" button in the Scanning form, so that the media did not pass in to the next step, i.e. QC Final.

After the OCRing the images are converted according to the requirements of the customers. They may be converted into "searchable PDF" or into "word document".

Different Operations that are performed in the "OCR" are as follows: -

- "Select" - This operation is used to select the media that has to be scanned.
- "Cancel Select" - This operation is used to cancel the selected media.
- "Save" - This operation is used to save all the information in the fields of the scanning operation in to the database.
- "Move Up" - To move the media up the order.
- "Move Down" - To move the media down the order.

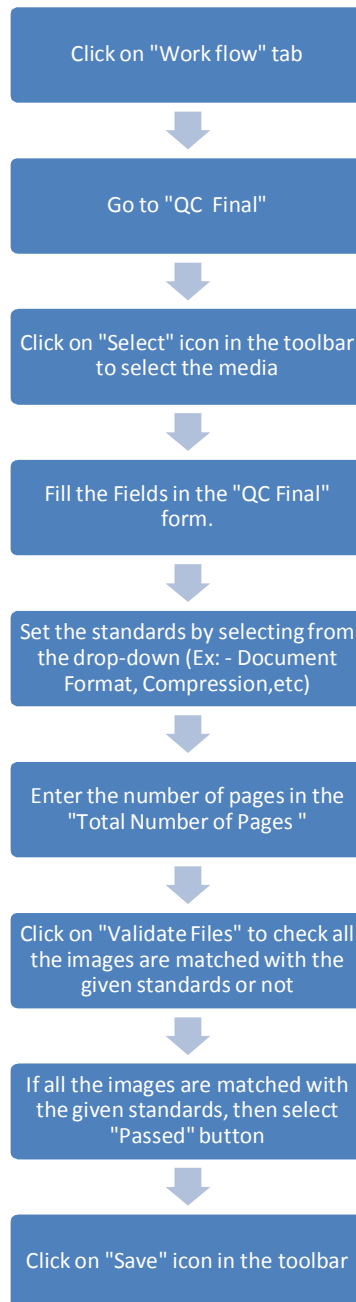
2.7. QC Final: -

After the OCR of the media, the total digitization of the media has been done. And now Packaging of these files is done to deliver this output to the customer. In between this verification has to be done to check no errors are found in the OCR images, this verification is called Quality Check Final (QC Final).

This Quality check has to be done according to the pre-defined standards. These standards are set by the customers who have given the media for digitization. And if all the standards match with the processed images, then this indicates that the QC Final has been done successfully and the given media is passed to the next step, i.e. Packaging.

If some images did not match with the given standards, then this indicates the QC Final has been failed and again the media has been resent for OCR.

The process of doing the QC Final for "Glimpses Of World religions" media: -



Different Fields that are to be filled in the "QC Final" form: -

- "Media Title" - Name of the media that has to undergo Quality Check. Here when we select the media in "Media Title", the automatically the Media title will be filled.
- "Folder Path" - Select the path of the existing scanned documents. We can select this by using the "Browse" option.
- "Destination Path" - Select the Path for the Images that are to be stored after doing the QC Zero.

Standards that are set for QC Final: -

- "Document Format" - Select the document format from drop-down. And mark it in the box next to it.
- "Compression" - Select the Compression from the drop-down. And mark the box next to it.
- "Horizontal resolution" - Select the Horizontal Resolution from the drop-down. And mark the box next to it.
- "Vertical resolution" - Select the Vertical Resolution from the drop-down. And mark the box next to it.
- "Page Width" - Enter the page width for the selected media file
- "Page Height" - Enter the page height for the selected media file.
- "Min.File Size (in KB)" - Enter the Min.File Size (in KB) for selected Media file.
- "Max.File Size (in KB)" - Enter the Max.File Size (in KB) for the selected Media File

Select all the standards that are to be checked for the "OCR" images.

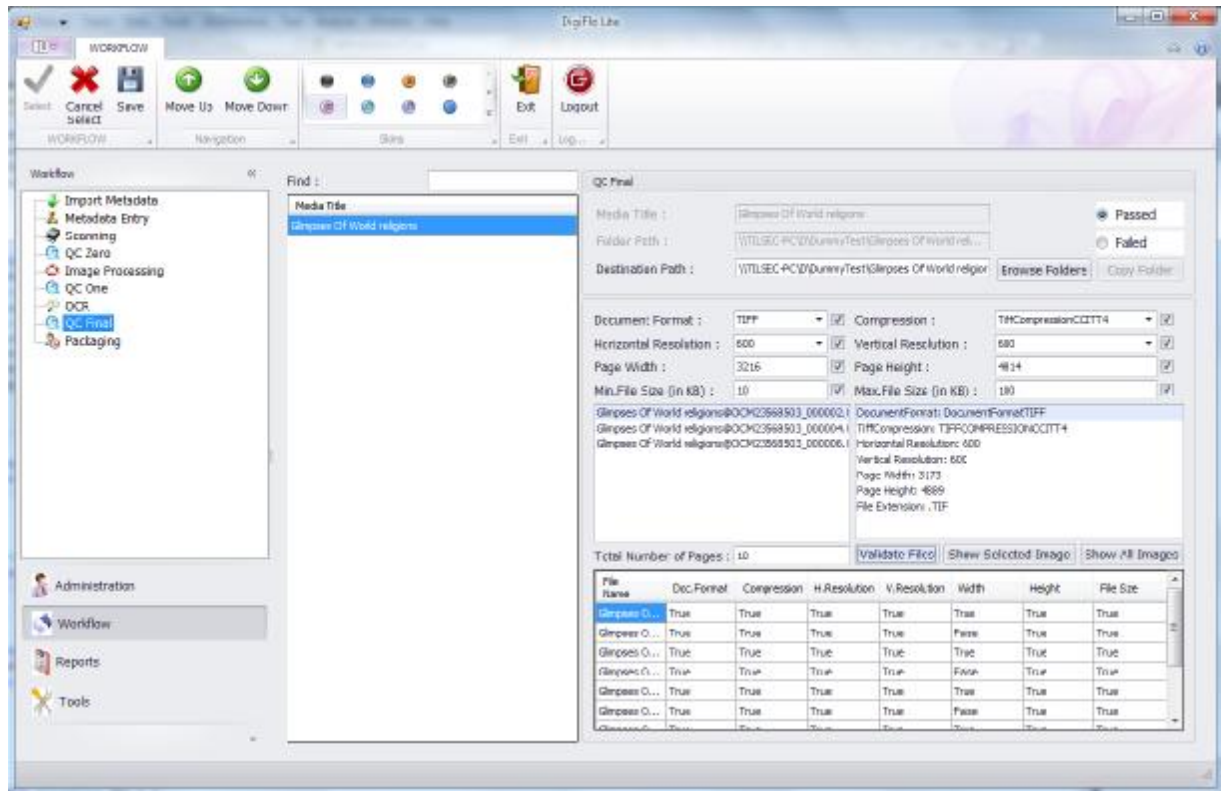
"Total number of Pages" - Enter the number of pages that have to undergo quality check.

Then click on "Validate Files" to see all the standards of the "OCR" images are matched with the given standards or not. And all the image filenames will be displayed in the grid at the bottom of the "QC Final" form.

The Grid at the bottom of the "QC Final" form will contain all the Standards as the columns. (Ex: -Document Format, Compression, etc). If the image has matched with the certain standard then "TRUE" word will be displayed. And if the image did not match with the certain standard, then "FALSE" word will be displayed.

The Two grids just below the Standards, will give the information about the errors. The left grid will specify the filenames of the error images. And the Right grid will give the complete description about the error image.

The below figure shows the QC Final for the "Glimpses Of World religions" media has been done



When all the images match with the given standards, then select “Passed” button in the QC Final, and click on “Save” icon in the toolbar to pass this media in to the next step, i.e. Packaging.

If the images did not match with the given standards, then select “Failed” button and the given media did not pass to the next step.

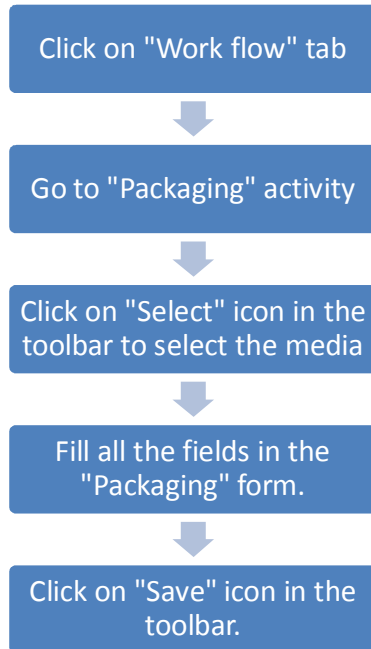
Different Operations that are performed in the “QC Final” are as follows: -

- “Select” - This operation is used to select the media that has to undergo QC Final.
- “Cancel Select” - This operation is used to cancel the selected media.
- “Save” - This operation is used to save all the information in the fields of the QC Final form in to the database.
- “Move Up” - To move the media up the order.
- “Move Down” - To move the media down the order.

2.8. Packaging: -

The process of combining all the files and also combining of files of different books, which are required by the customer in to a single folder, is called packaging. As the files of different media will be stored in different locations and it is very difficult to search them and combine them. So this “Packaging” activity will make the customer required files into one folder, so that the delivery will be easy.

Process of Packaging all the files: -

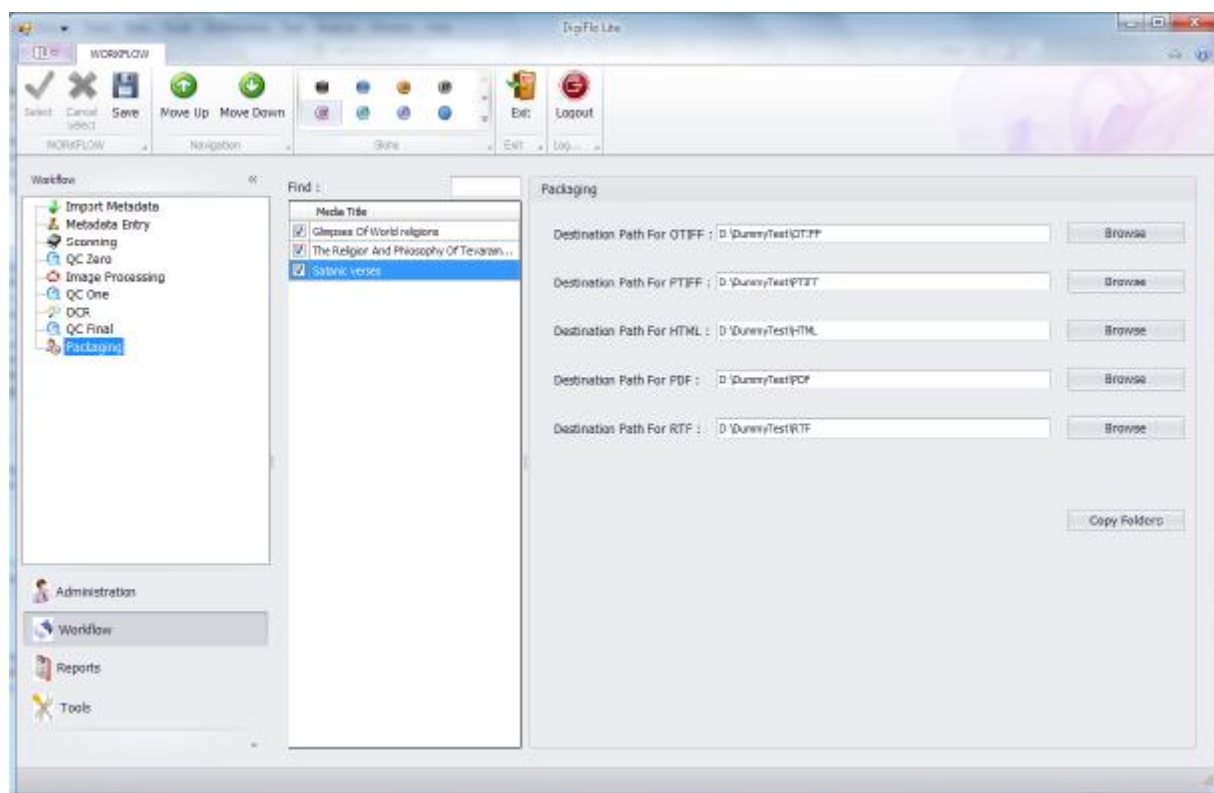


Different fields that are to be filled in the Packaging form: -

- “Destination Path for OTIFF” - Enter the path for the total OTIFF images that are to be delivered to the customer.
- “Destination Path for PTIFF” - Enter the path for the total PTIFF images that are to be delivered to the customer.
- “Destination Path for HTML” - Enter the path for the total HTML files that are to be delivered to the customer.
- “Destination Path for PDF” - Enter the path for the total PDF files that are to be delivered to the customer.
- “Destination Path for RTF” - Enter the path for the total RTF files that are to be delivered to the customer.

By filling all the fields and click on “Save” icon in the toolbar, the required files will be sent to their destination paths. And these files will be delivered to the customers.

The below figure shows the “Packaging” activity for the three books, and the files of the three books will be stored in the given destination path.



Different Operations that are performed in the “Packaging” are as follows: -

- “Select” - This operation is used to select the media that has to undergo “Packaging” activity.
- “Cancel Select” - This operation is used to cancel the selected media.
- “Save” - This operation is used to save all the information in the fields of the “Packaging” form in to the database.
- “Move Up” - To move the media up the order.
- “Move Down” - To move the media down the order.

3. Reports: -

Reports will track the process of the project. And the Report gives the complete information regarding each and every activity of the project. The Reports generated will be used for the future generation.

Generally the top management has no time to check the status of the each and every activity. So they expect some reports which will help them to know the status of the project every day. They can check the status by the reports. And the report contains the parameters required by the management.

Even the "Reports" tab in the "DigiFlo" also contains certain reports which are very useful for the management to know the status of the project.

Different Reports that can be generated are as follows,

- "Media Inventory"
- "Scanning Details"
- "QC Zero Details"
- "Processing Details"
- "QC One Details"
- "OCR Details"
- "QC Final Details"
- "Scanning Page Wise Remark Details"
- "Page Wise QC Zero details"
- "Processing Page Wise Remark details"
- "Page Wise QC One details"
- "OCR Page Wise Remark Details"
- "Page Wise QC Final details"

Different Operations that can be performed on the reports are: -

- "Search" - This icon is used to search for any word in the displayed report.
- "Open" - This icon is used to import the report from the database.
- "Save" - This icon is used to save the report in to the database.
- "Print" - Print icon is used to print the report
- "Quick Print" - This icon is used to print the report without any settings
- "Page Setup" - This icon is used to set the page while printing.
- "Scale" - This icon is used to adjust the displayed report and make it more visible.
- "Hand Tool" - This icon is used to drag the report.
- "Magnifier" - Zoom-in
- "Zoom-out" - This icon is used to zoom-out the report.
- "Zoom In" - This icon is used to zoom the report to make it more visible.
- "Color" - To change the color of the report.
- "Watermark" - To change the settings and make it more visible.
- "Export Document" - This icon is used to export the report in to certain formats. (Ex: - PDF, HTML, etc)
- "Send via E-Mail" - To send the report to the mail.

- “Exit” - come out of the report.

3.1. Media Inventory: -

This report contains all the media (books) which is under the process of Digitization. This report gives the information about each book and the status of each book in the Workflow.

- Different columns in this report are
- “Media Title” - Title of the book
- “Barcode” - Barcode of the book
- “Work Flow Stage” - It gives the status of the book in the “Workflow”, i.e. Scanning, QC One. If it shows “completed”, this indicates that the digitization of this particular book is completed.
- “OCLC Number” - OCLC Number of the book
- “Author” - Author of the book
- “Publisher” - Publisher of the book
- “Region” - Region where the book belongs to.

The below figure shows the “Media Inventory” Report

Media Title	Barcode	Work Flow Stage	OCLC Number	Author	Publisher	Region
Project Name : GEMO						
Songs Of David Givande	150145	Completed	024023588561	David Givande	The Givande	
An Introduction To The Study Of Clear Vain	150144	Completed	024023588487	Parsons, John Harb	Cambridge University	New York
The Analysis Of Social Behaviour In Hospital	150145	Completed	024023588488	Lawford, Harold C.	Routledge And Kegan	
Up a Beach						
The Problem Of Power The Reth Historical	150146	Scanning	024023588489	Lord Radcliffe Of	Scolar And Warburg	
Lectures						
For It's Politics And Progress Group	150147	Scanning	024023588505	Ray, J. V. O.	Thomas Y. Crowell	
Buddhism In China	150140	Scanning	024023588502	Seal, The Rev. C.	E. Arnold B. Young	
Clampet Of World Editors	150150	Completed	024023588503	Not Available	Julius Publishing	
The Religion And The Society Of Overseas	150151	Completed	024023588504	Bergamont, M. A.	University Of Malaya	Chennai
SSD Social History Of 10 Nations Around The World						
THE HISTORY OF THE EAST INDIA COMPANY	150152	Completed	024023588505	NOT AVAILABLE	UNIVERSITY MICRO	
Y. COMMISSION A SOCIETY OF BURY	150153	Completed	024023588506	NOT AVAILABLE	MAISONLINE	
WINE SV	150155	Scanning	024023588507	NOT AVAILABLE	COLLEGE	
WINE SV	150155	Scanning	024023588508	NOT AVAILABLE	COLLEGE	
WINE SV	150155	Scanning	024023588509	NOT AVAILABLE	COLLEGE	
WINE SV	150155	Scanning	024023588510	NOT AVAILABLE	COLLEGE	
WINE SV	150155	Scanning	024023588511	NOT AVAILABLE	COLLEGE	
WINE SV	150155	Scanning	024023588512	NOT AVAILABLE	COLLEGE	

The “Media Inventory” report can be exported to other formats like PDF and can be stored in the system.

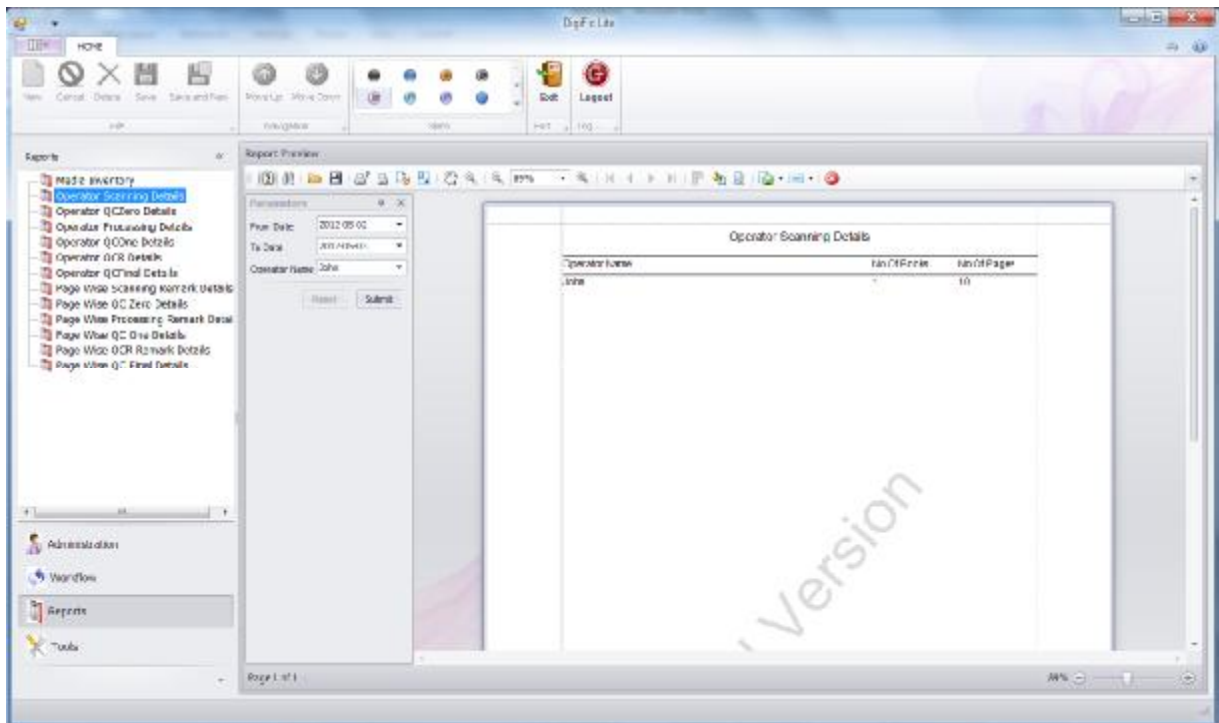
3.2. Scanning Details: -

This report gives the complete information regarding the scanning of the media. This report gives the information about the number of books scanned, number of pages scanned and the operator who has scanned the books.

The report can be generated date-wise and also "Operator-wise" .

To get the "Scanning Details" Report three fields has to be filled, they are as follows

- "From Date" - Select the date from drop-down.
- "To Date" - Select the date from drop-down.
- "Operator name" - enter the name of the scanning operator.



The "Scanning Details" report contains different columns, they are -

- "Operator Name"
- "Number of Books"
- "Number of Pages"

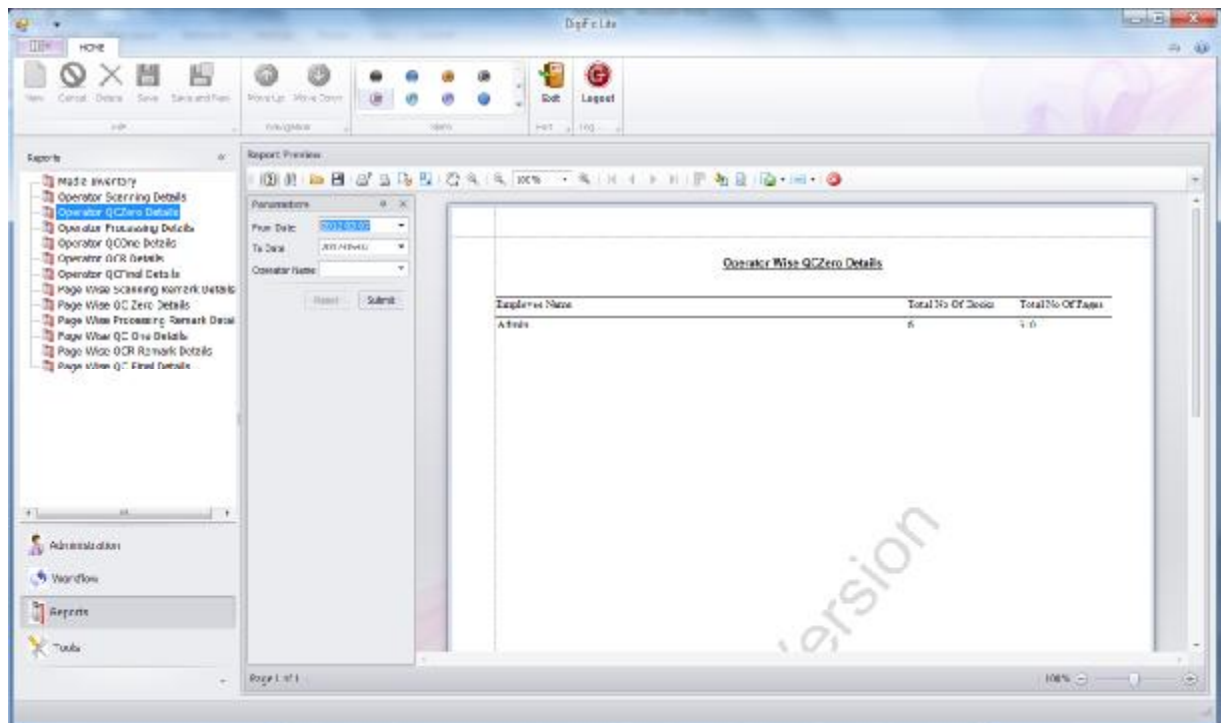
3.3. QC Zero Details: -

This report gives the complete information regarding the QC Zero of the media. This report gives the information about the number of books that have undergone QC Zero, number of pages that have undergone QC Zero and the operator who has done the QC Zero.

The report can be generated "Date-wise" and also "Operator-wise" .

To get the "QC Zero Details" Report three fields has to be filled, they are as follows

- "From Date" - Select the date from drop-down.
- "To Date" - Select the date from drop-down.
- "Operator name" - enter the name of the QC Zero operator.



The above figure shows the report has been taken "Operator-Wise" .

The "QC Zero Details" report contains different columns, they are -

- "Employee Name"
- "Total Number of Books"
- "Total Number of Pages"

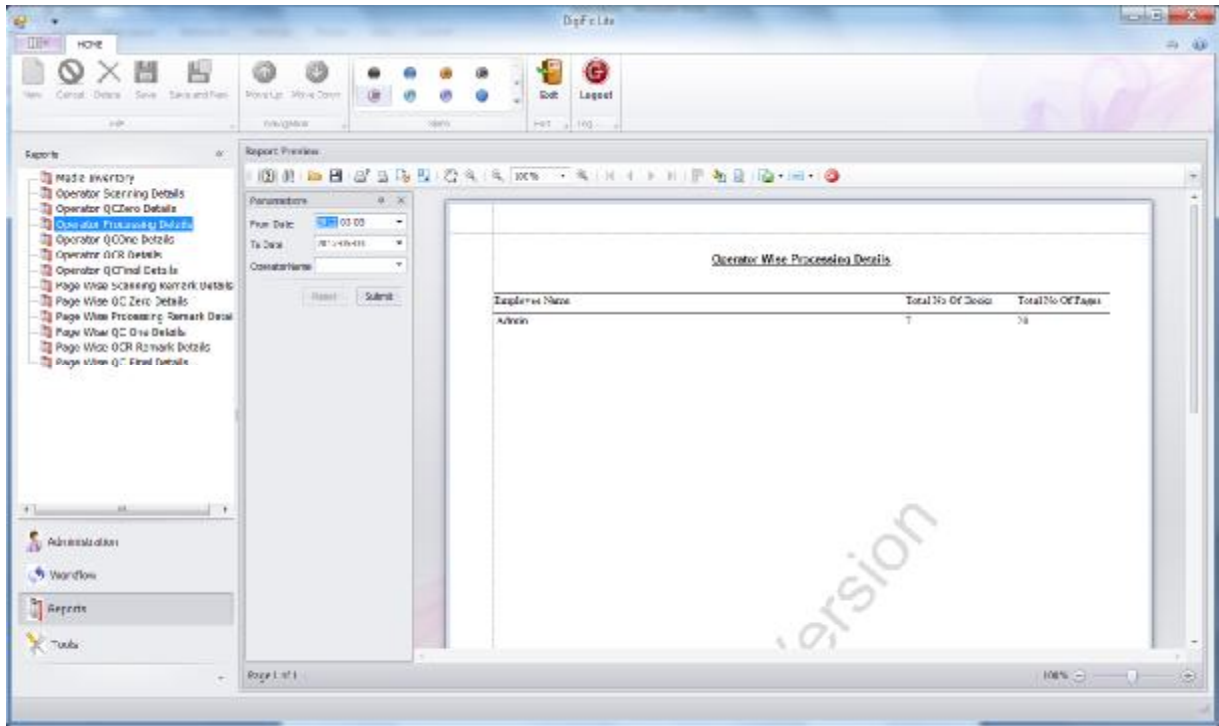
3.4. Processing Details: -

This report gives the complete information regarding the processing of the media. This report gives the information about the number of books that are processed, number of pages that are processed and the operator who has processed the books.

The report can be generated date-wise and also "Operator-wise" .

To get the "Processing Details" Report, three fields has to be filled, they are as follows

- "From Date" - Select the date from drop-down.
- "To Date" - Select the date from drop-down.
- "Operator name" - Enter the name of the Processing Operator.



The above figure shows the report has been taken "Operator-Wise".

The "Processing Details" report contains different columns, they are -

- "Employee Name"
- "Total Number of Books"
- "Total Number of Pages"

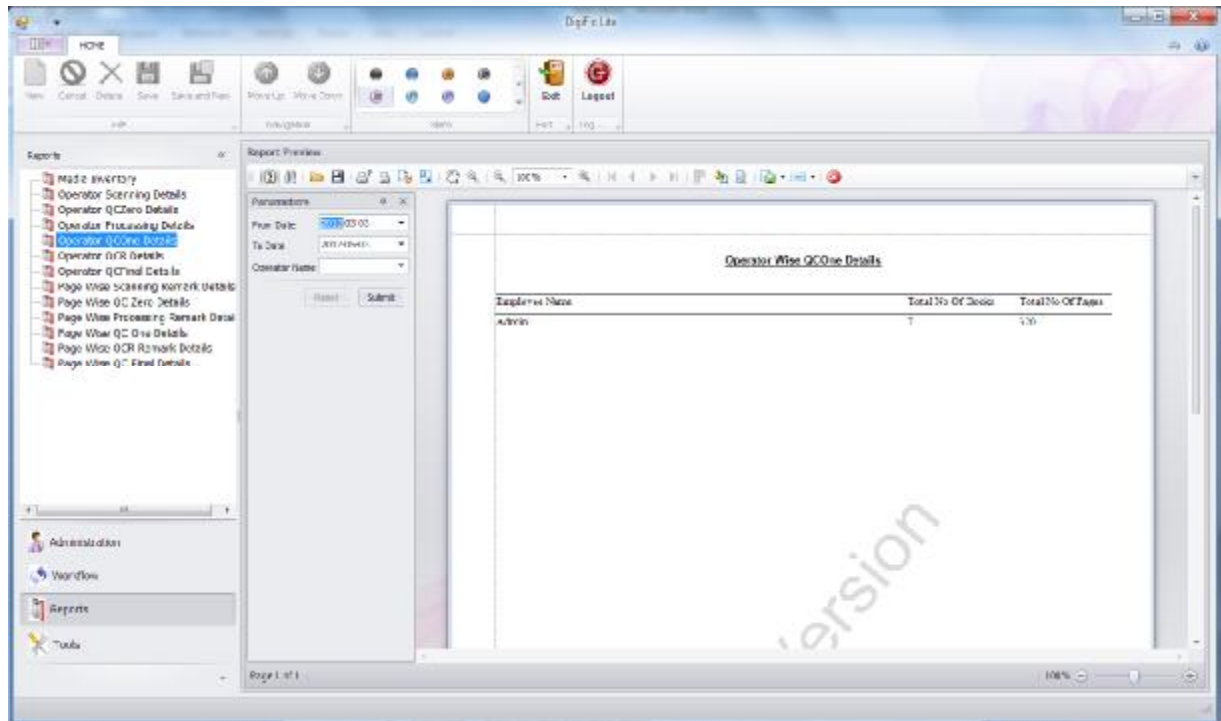
3.5. QC One Details: -

This report gives the complete information regarding the "QC One" of the media. This report gives the information about the number of books that have undergone QC One, number of pages that have undergone QC One and the operator who has done the QC One.

The report can be generated "Date-wise" and also "Operator-wise".

To get the "QC One Details" Report three fields has to be filled, they are as follows

- "From Date" - Select the date from drop-down.
- "To Date" - Select the date from drop-down.
- "Operator name" - enter the name of the QC One operator



The above figure shows the report has been taken “Operator-Wise” .

The “QC Zero Details” report contains different columns, they are -

- “Employee Name”
- “Total Number of Books”
- “Total Number of Pages”

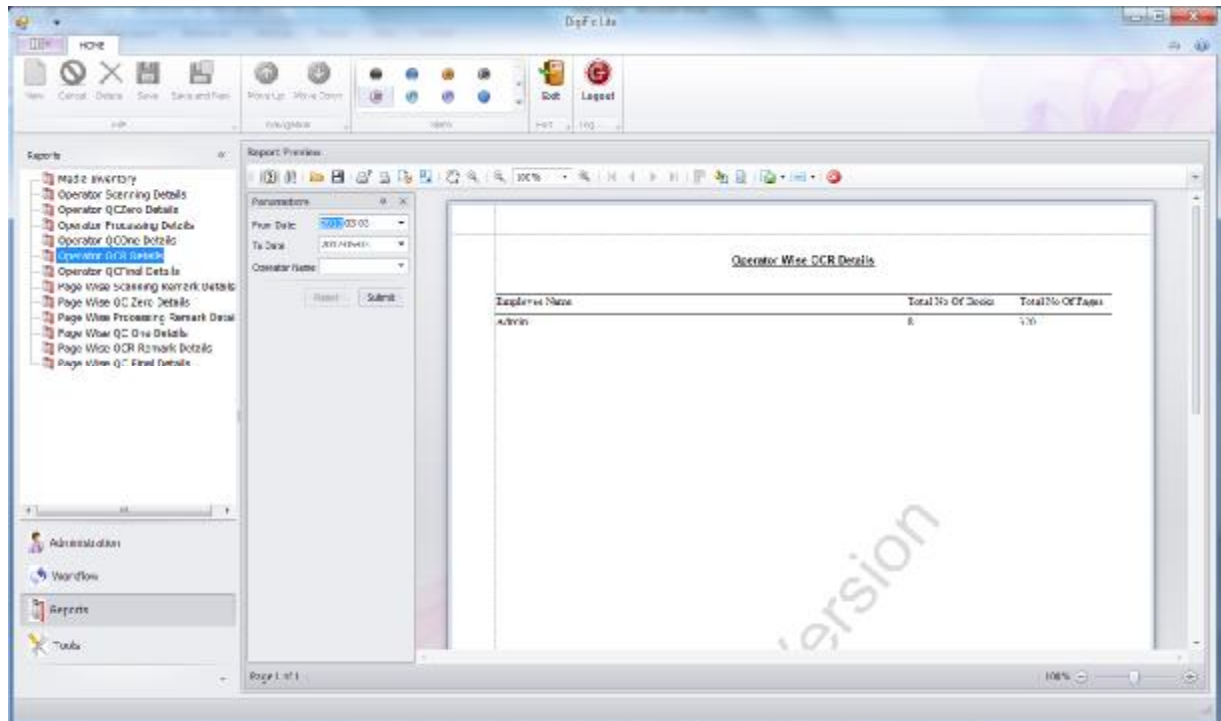
3.6. OCR Details: -

This report gives the complete information regarding the OCR of the media. This report gives the information about the number of books that have undergone OCR, number of pages that have undergone OCR and the operator who has done OCR of books.

The report can be generated date-wise and also “Operator-wise” .

To get the “OCR Details” Report, three fields has to be filled, they are as follows

- “From Date” - Select the date from drop-down.
- “To Date” - Select the date from drop-down.
- “Operator name” - Enter the name of the OCR Operator.



The above figure shows the report has been generated “Operator-Wise” .

The “OCR Details” report contains different columns, they are -

- “Employee Name”
- “Total Number of Books”
- “Total Number of Pages”

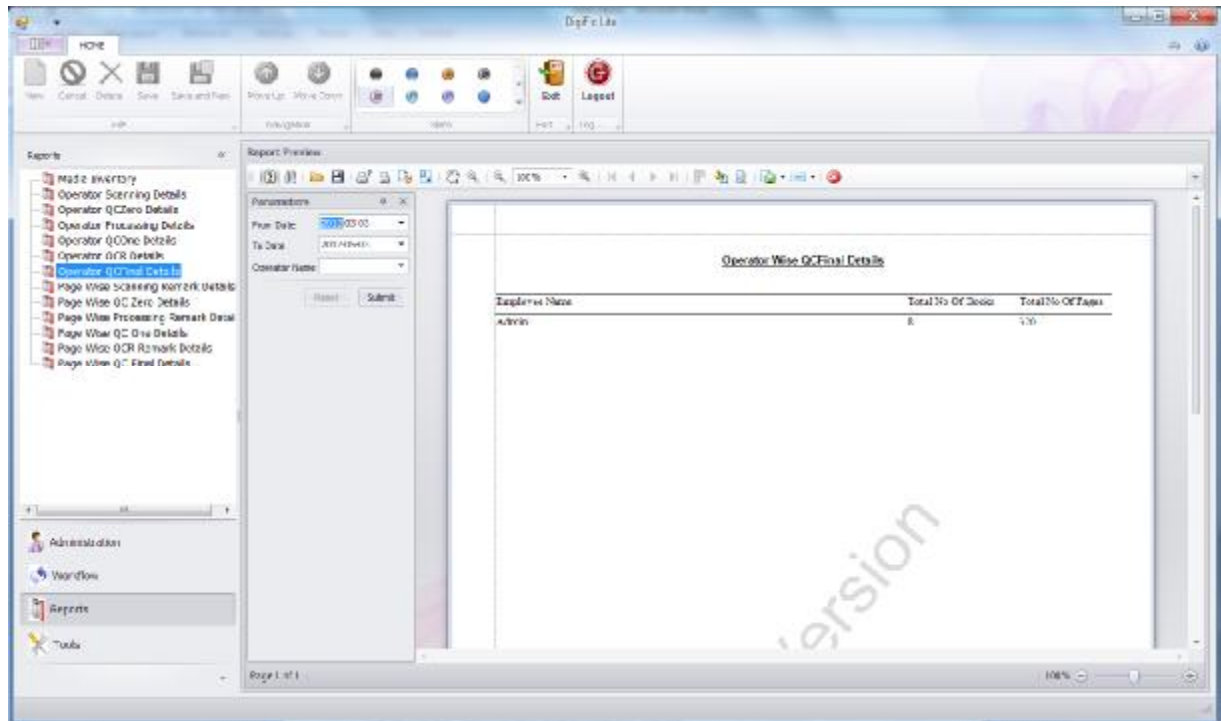
3.7. QC Final Details: -

This report gives the complete information regarding the “QC Final” of the media. This report gives the information about the number of books that have undergone QC Final, number of pages that have undergone QC Final and the operator who has done the QC Final.

The report can be generated “Date-wise” and also “Operator-wise” .

To get the “QC Final Details” Report three fields has to be filled, they are as follows

- “From Date” - Select the date from drop-down.
- “To Date” - Select the date from drop-down.
- “Operator name” - enter the name of the QC Final operator



The above figure shows the report has been taken "Operator-Wise".

The "QC Final Details" report contains different columns, they are -

- "Employee Name"
- "Total Number of Books"
- "Total Number of Pages"

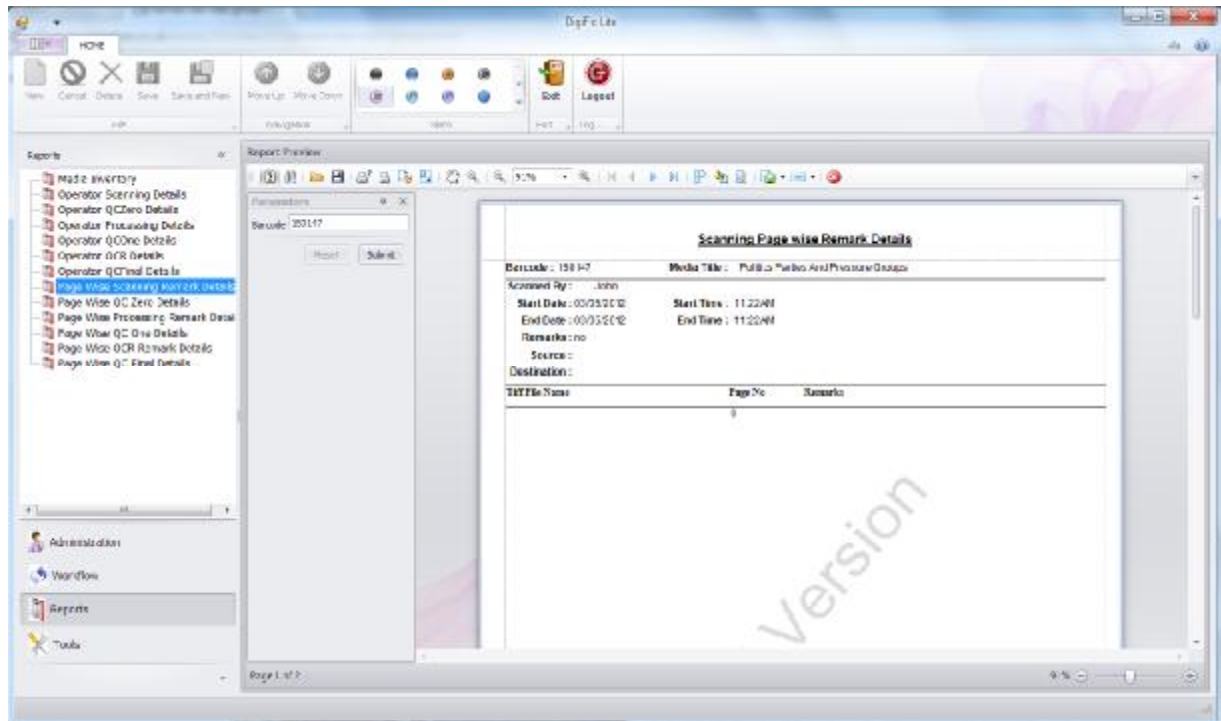
3.8. Scanning Page Wise Remark Details: -

This report gives the complete information regarding the Page Wise remarks while executing scanning operation.

This report gives the status about the remarks that has occurred during the scanning of the pages. And the administrator can re-check the remarked pages to see the errors have been rectified or not.

To get the report of the "Scanning Page Wise Remark Details", "Barcode" of the media has to be entered. So all the pages in that media which have remarks will be shown in the report.

The below figure shows the "Scanning Page Wise Remark Details" report has been generated by entering "Barcode" of the particular media.



This report contains all the details of the media, i.e. Barcode, Media Title, Scanned By, Start Date, Start Time, End Date, End Time, Remarks, Source, and Destination.

The "Processing Page Wise Remark Details" report contains three columns. And the page remarks are displayed in these three columns in detail. The three columns are as follows

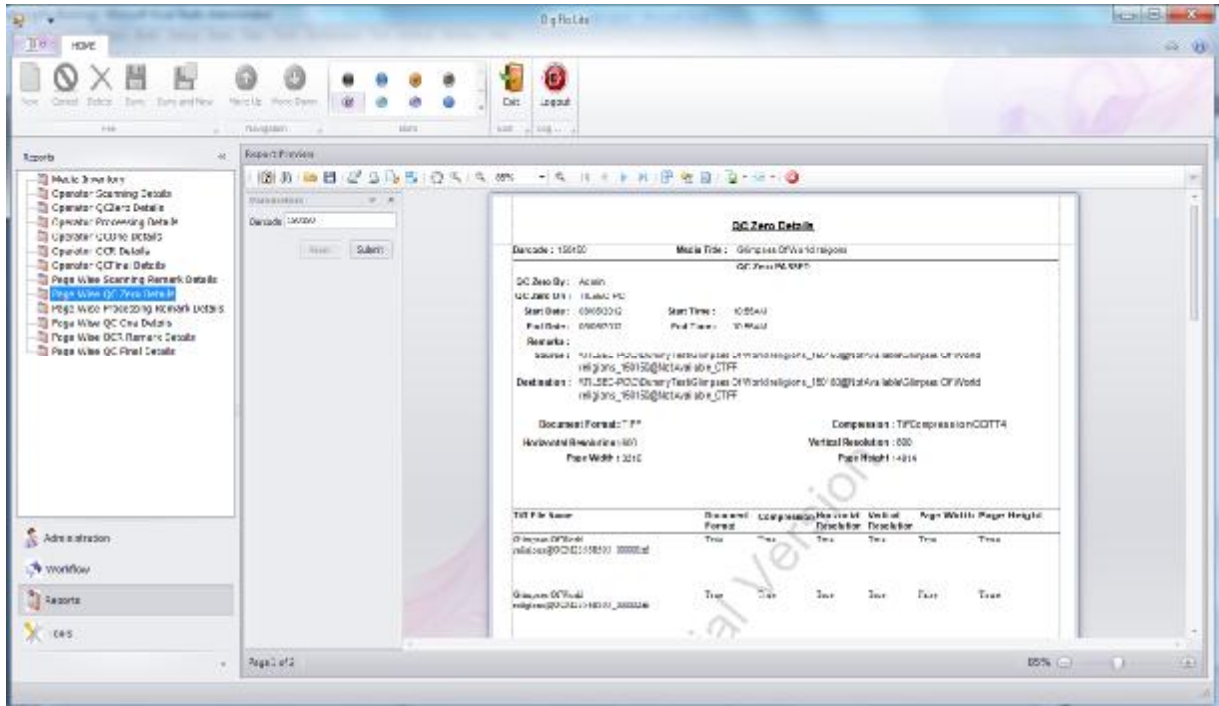
- "Tiff File Name"
- "Page No"
- "Remarks"

3.9. Page Wise QC Zero Details: -

This report gives the information regarding the QC Zero activity page wise. And this report gives the page wise status of the media during the QC Zero activity. By knowing the status the administrator can recheck the files and can send them to re-scanning if required.

To get the report of the "Page wise QC Zero Details", "Barcode" of the media has to be entered.

This report contains all the details of the media, i.e. Barcode, Media Title, QC Zero By, QC Zero On, Start Date, Start Time, End Date, End Time, Remarks, Source, and Destination. And also the required standards that are set. (Ex:-Document Format, Compression)



This “Page Wise QC Zero Details” contains different standards as columns. Columns in this report are

- “Tiff File Name”
- “Document Format”
- “Compression”
- “Horizontal Resolution”
- “Vertical Resolution”
- “Page Width”
- “Page Height”

The result will be displayed as “True” and “False” .

- “True” indicates the specific image file matches with the standard.
- “False” indicates the specific image file doesn’t match with the standard.

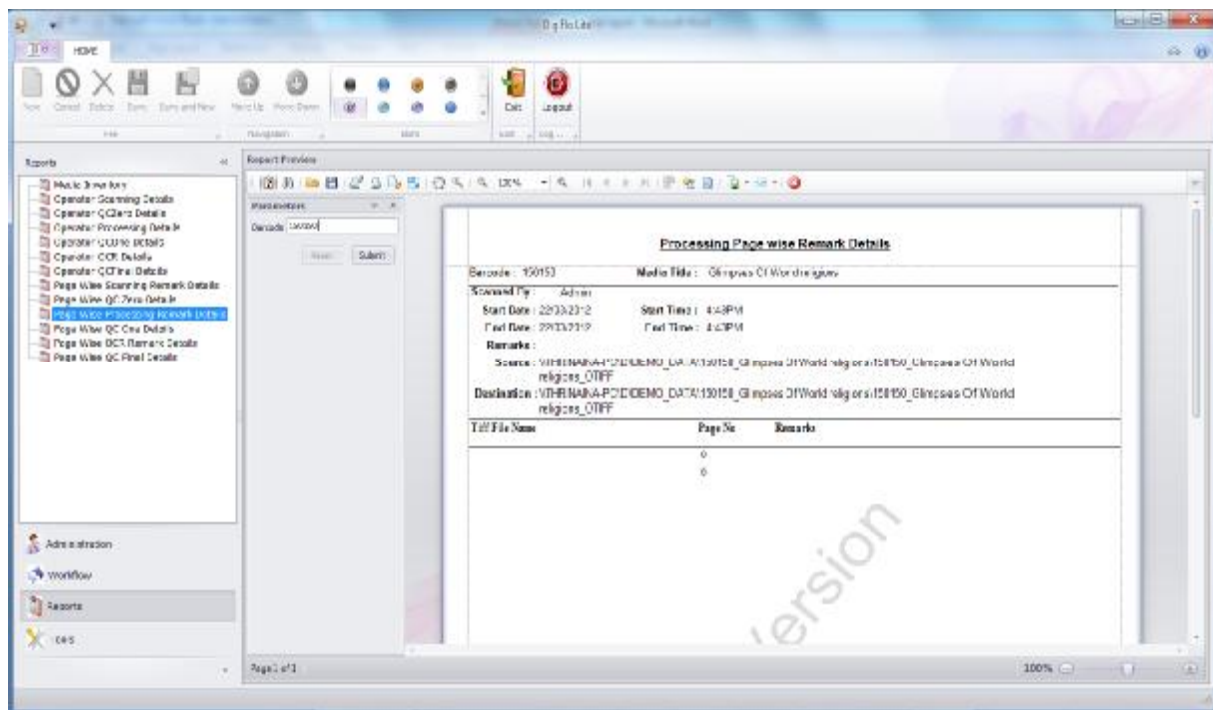
3.9. Processing Page Wise Remark Details: -

This report gives the complete information regarding the Page Wise remarks while executing Image Processing operation.

This report gives the status about the remarks that have occurred during the Processing of the pages. And the administrator can re-check the remarked pages to see the errors have been rectified or not.

To get the report of the “Processing Page Wise Remark Details”, “Barcode” of the media has to be entered. So that all the pages in that media which have remarks will be shown in the report.

The below figure shows the “Processing\ Page Wise Remark Details” report has been generated by entering “Barcode” of the particular media.



This report contains all the details of the media, i.e. Barcode, Media Title, Scanned By, Start Date, Start Time, End Date, End Time, Remarks, Source, and Destination.

The “Processing Page Wise Remark Details” report contains three columns. And the page remarks are displayed in these three columns in detail. The three columns are as follows

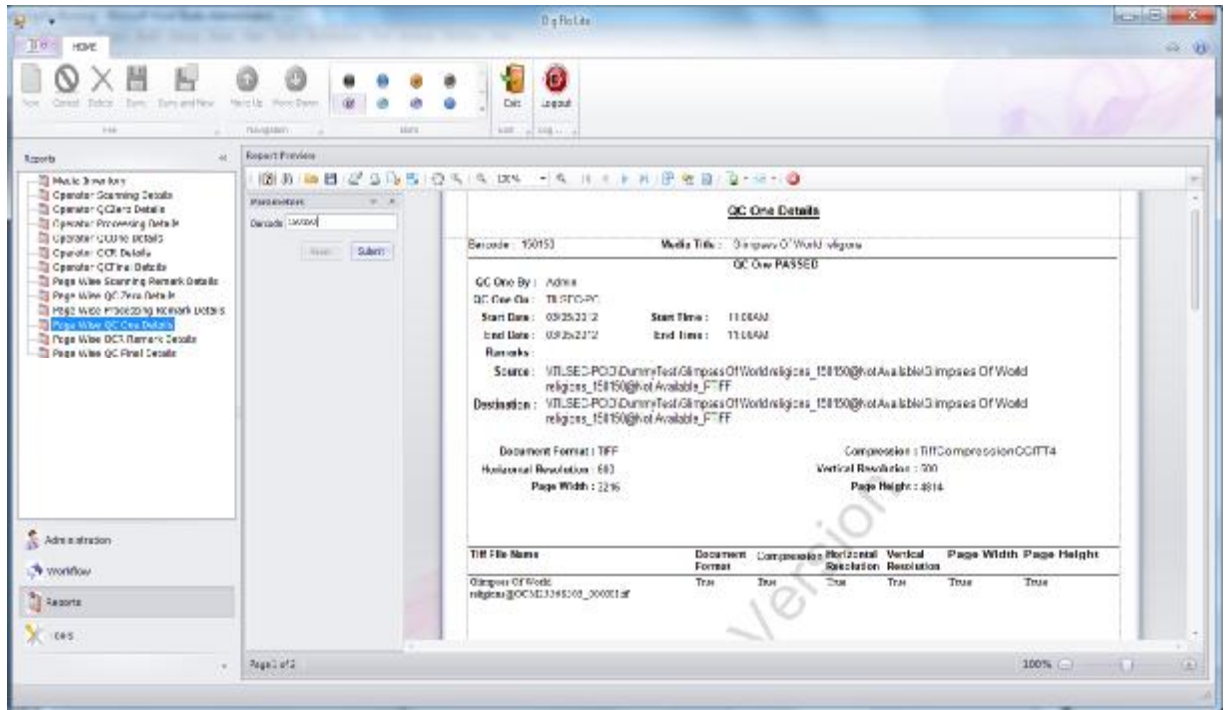
- “Tiff File Name”
- “Page No”
- “Remarks”

3.9. Page Wise QC One Details: -

This report gives the information regarding the QC One activity page wise. And this report gives the page wise status of the media during the QC One activity. By knowing the status the administrator can recheck the files and can send them to for processing again if required.

To get the report of the “Page wise QC One Details”, “Barcode” of the media has to be entered.

This report contains all the details of the media, i.e. Barcode, Media Title, QC One By, QC One On, Start Date, Start Time, End Date, End Time, Remarks, Source, and Destination. And also the required standards that are set. (Ex:-Document Format, Compression, etc)



This "Page Wise QC One Details" contains different standards as columns. Columns in this report are

- "Tiff File Name"
- "Document Format"
- "Compression"
- "Horizontal Resolution"
- "Vertical Resolution"
- "Page Width"
- "Page Height"

The result will be displayed as "True" and "False".

- "True" indicates the specific image file matches with the standard.
- "False" indicates the specific image file doesn't match with the standard.

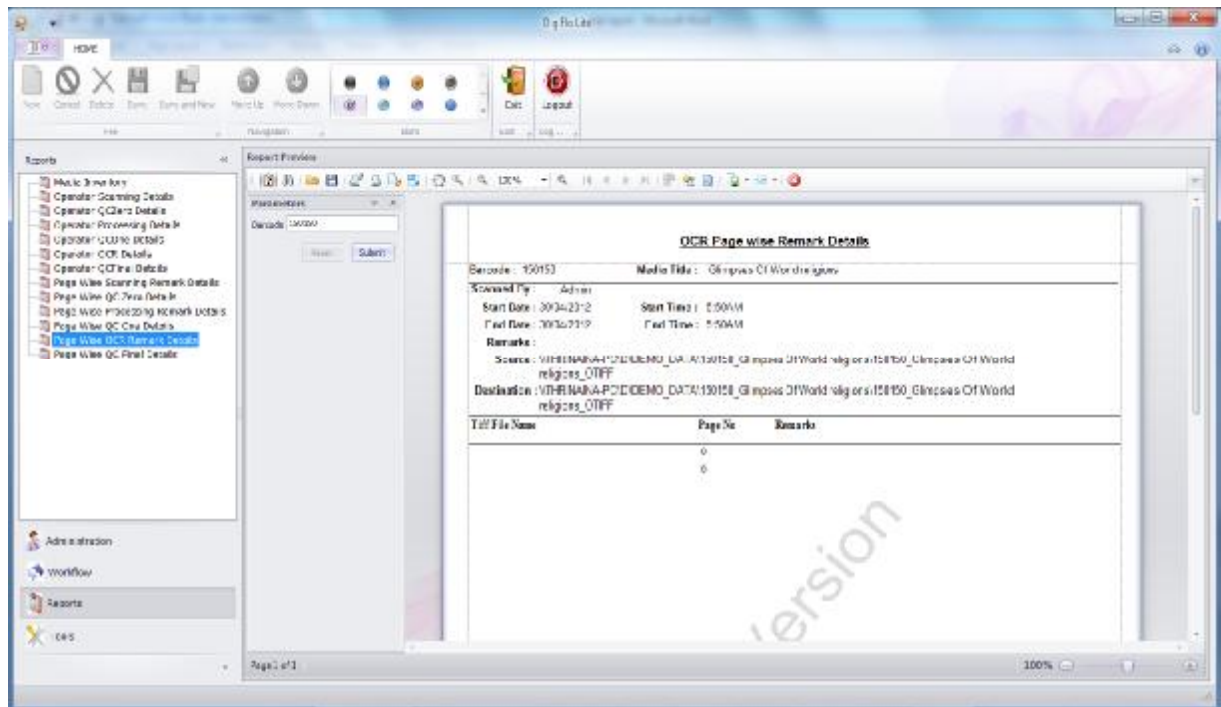
3.10. OCR Page Wise Remark Details: -

This report gives the complete information regarding the Page Wise remarks while executing OCR.

This report gives the status about the remarks that have occurred during the OCR of the pages. And the administrator can re-check the remarked pages to see the errors have been rectified or not.

To get the report of the “OCR Page Wise Remark Details”, “Barcode” of the media has to be entered. So that all the pages in that media which have remarks will be shown in the report.

The below figure shows the “OCR Page Wise Remark Details” report has been generated by entering “Barcode” of the particular media.



This report contains all the details of the media, i.e. Barcode, Media Title, Scanned By, Start Date, Start Time, End Date, End Time, Remarks, Source, and Destination.

The “OCR Page Wise Remark Details” report contains three columns. And the page remarks are displayed in these three columns in detail. The three columns are as follows

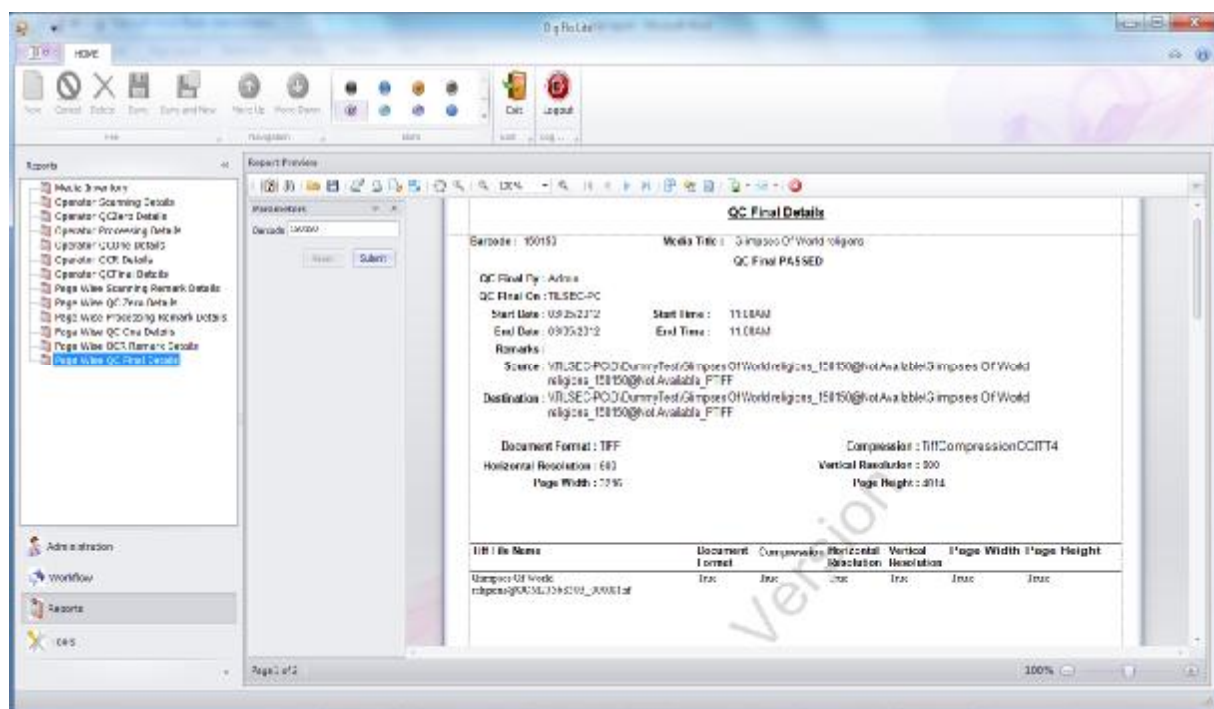
- “Tiff File Name”
- “Page No”
- “Remarks”

3.9. Page Wise QC Final Details: -

This report gives the information regarding the QC Final activity page wise. And this report gives the page wise status of the media during the QC Final activity. By knowing the status the administrator can recheck the files and can send them to for OCR again if required.

To get the report of the "Page wise QC Final Details", "Barcode" of the media has to be entered.

This report contains all the details of the media, i.e. Barcode, Media Title, QC Final By, QC Final On, Start Date, Start Time, End Date, End Time, Remarks, Source, and Destination. And also the required standards that are set. (Ex:-Document Format, Compression, etc)



This "Page Wise QC Final Details" contains different standards as columns. Columns in this report are

- "Tiff File Name"
- "Document Format"
- "Compression"
- "Horizontal Resolution"
- "Vertical Resolution"
- "Page Width"

- "Page Height"

The result will be displayed as "True" and "False" .

- "True" indicates the specific image file matches with the standard.
- "False" indicates the specific image file doesn't match with the standard.

4. Tools: -

Tools tab is used to do some extra operation in the project, which are not included in the workflow of the project.

Tools tab in the DigiFlo contains different operations which are executed on the image files to get the output required by the customer. Some operations in this Tools tab are not there in the workflow of the Digitization, but are required while executing the project.

This "Tools" tab contains four operations. They are as follows,

- "Batch Processing"
- "Split Files"
- "Merge File"
- "Rename Files"

4.1. Batch Processing: -

The Batch Processing contains the different activities which are processed on the images to make the images more quality ones. All the activities in the batch processing are present in the "Image Processing" of Workflow tab.

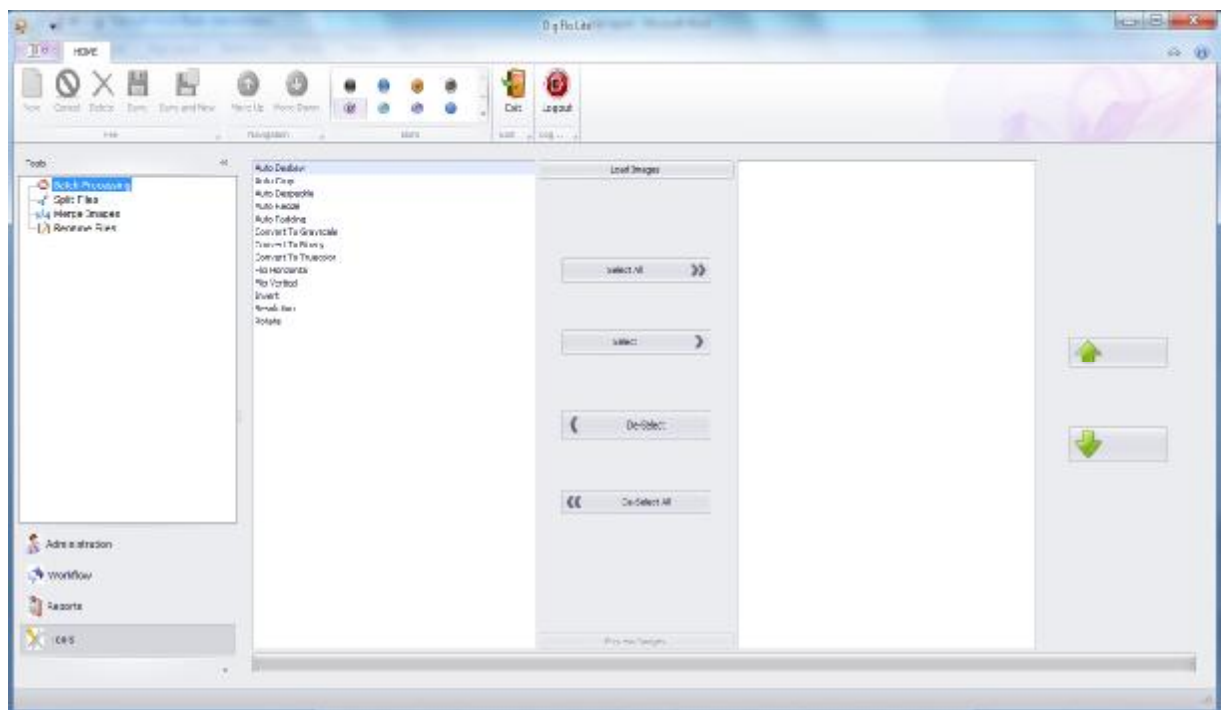
Some files are missed while executing the "Image Processing" and now if those images have to undergo the activities in the "Image Processing", then it is very difficult to go to Workflow. So this "Batch Processing" helps in processing these activities on those images according to the requirement.

Different activities in the "Batch Processing" are as follows -

- Auto Deskew - The Automatic Deskewing command is ideal to equalize the image orientation of a document if it is placed slightly rotated on the layout section.
- Auto crop - This activity is used to remove the edges from the images in the scanned document.
- Auto Despeckle -
- Auto padding - If the size of the image file has reduced while Auto Cropping, then this activity is used to add the pieces of blank images to make the image file to its original size. So that all the image files will be of uniform size.
- Convert to Grayscale - This activity is used to convert from scanned colour images (24 bit or 32bit) to grayscale images (8bit).

- Convert to Binary - This activity is used to convert the given images in to Binary images (2bit). So these converted images will be in Black & White form.
- Convert to Truecolor - This activity is used to convert the Binary images or Grayscale images into original Color images.
- Flip Horizontal - This activity is used to convert the scanned image into Mirrored image and display it horizontally.
- Flip Vertical - This activity is used to convert the scanned image into Mirrored image and display it vertically.
- Invert - This activity is used to invert the colors. I.e. White to Black (or) Black to White. Generally by doing this activity, if the background is in white and the sentences are in black, then they will be converted as Background as Black and the sentences as White.
- Resolution - This activity is use to change the resolution of the image. This is used for making the image brighter. Generally the resolution is inversely proportional to the "Length and width of the image". And in this DigiFlo software, even if the resolution of the image is increased, the "Length and Width" of the image is not changed, this indicates the DigiFlo application will keep all the scanned images in the uniform size.
- Rotate - This activity is used to rotate the scanned image

The below figure shows the "Batch Processing" form. And different activities are shown.



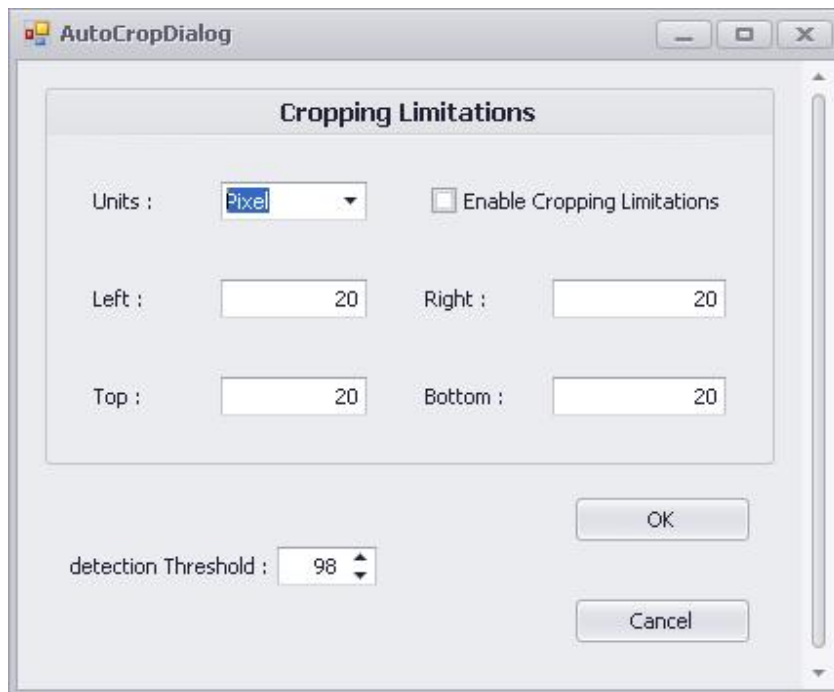
This "Batch Processing" form contains different operation, they are as follows,

- "Load Images" - This button is used to load the required images that have to undergo "Batch processing"
- "Select All" - This is used to select all the activities in the left grid and all the activities will be reflected in the right grid.
- "Select" - This is used to select the activity individually, and the selected activities will be reflected in the right grid.
- "De-Select" - This is used to De-Select the activities individually.
- "De-Select All" - This is used to De-Select all the activities from the right grid.
- "Process Images" - This is used to process all the selected activities and the out put of this is shown in the "Destination Path"

And the two arrows on the right of the form are used to move the activities up and down according to the requirement.

The activity dialog box will be opened, when we double click on the activity in the right grid. This activity dialog box is used to adjust the settings of the activity. (Ex: - AutoCropDialog Box)

The below figure shows the "AuroCropDialog box".



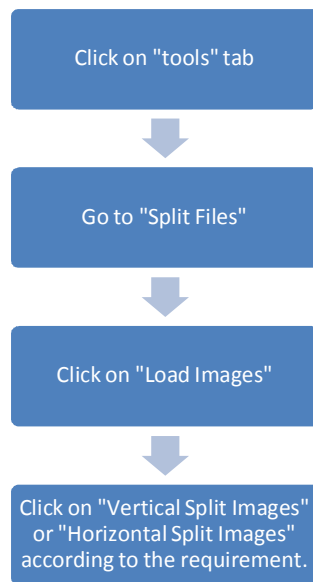
This Dialog Box contains different setting of the "Auto Crop" activity. And these settings can be adjusted according to the requirement.

4.2. Split Files: -

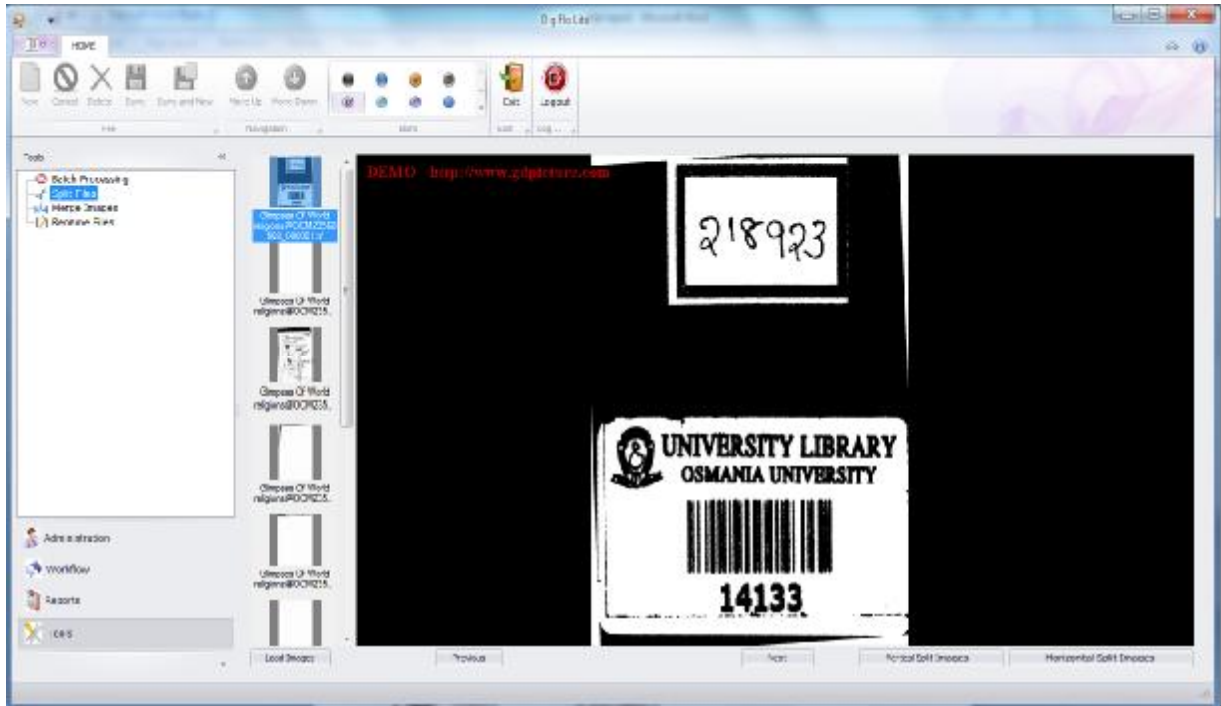
This Operation is used to split the files according to the requirement of the customers. And at some times the scanning operator might combine two pages at the time of scanning and now by using the "Split files" operation the scanned image will split. The image files can be Split horizontally or vertically according to the requirement.

Ex: -If the customer asked to split an image horizontally. The illustration below will show the process of "split files".

The Process of "Split files": -



The below figure shows the screen of "Split Files" operation with the image file, which has to be split.



After splitting the image horizontally, the given image file will be split into two files. The below two figures (Figure-1 and Figure - 2) are the images that have split from the given image.

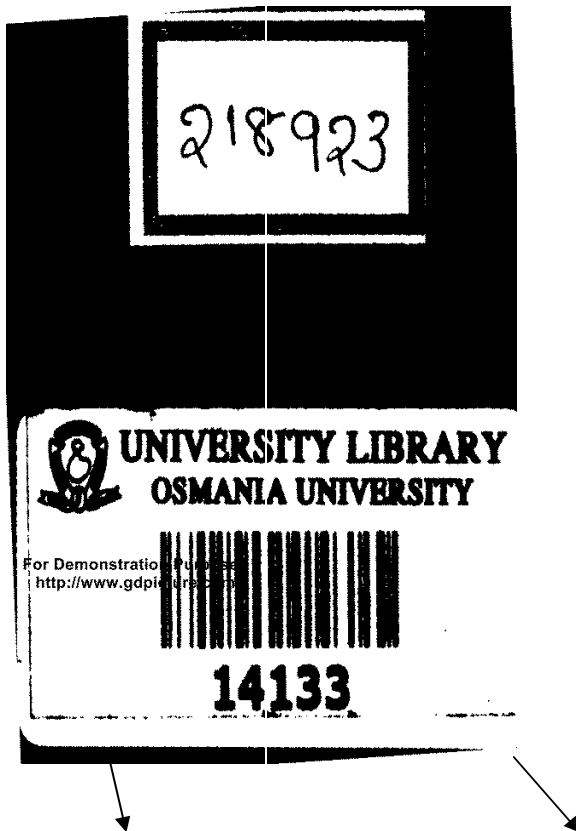


Figure - 1

Figure - 2

4.3. Merge Files: -

This operation is used to merge the files into one file. Some times customers ask for the files in different media to merge and submit as a single file. So, at that time, this "Merge files" Operation will merge the files.

Some times the scanning operator might have divided the page will doing the scanning for his convenience, and now while submitting the images to the customer; the images have to be merged. So, this "Merge Files" operation is used to merge those files.

Ex: - Two pages which are scanned and converted into images. Now the customer asks to merge those files in to one file. The illustration below will show the process of merging the file

The Process of merging the files: -

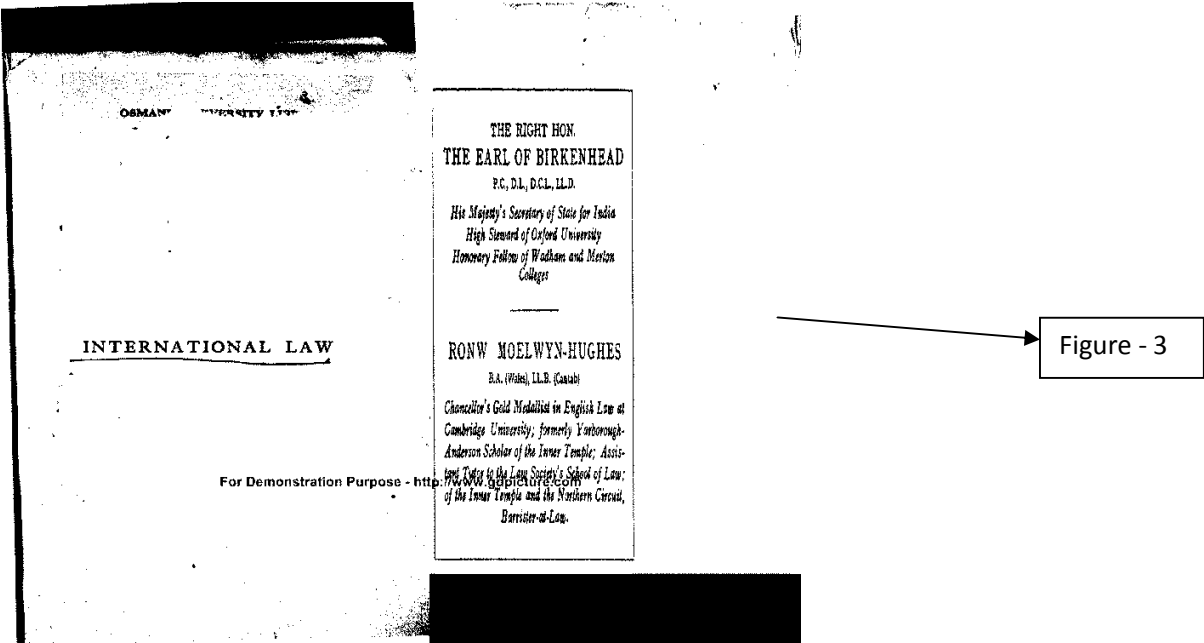


The below figure shows the screen of the "Merge Files" operation: -

Figure - 1

Figure - 2

Then Click on "Merge Images" to merge the two images. The below figure(Figure - 3) shows the merged image

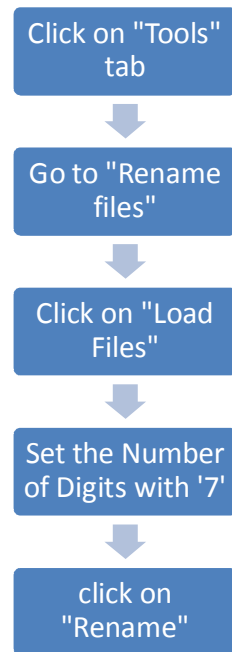


4.4. Rename Files: -

This Operation is used to rename the files if required. Here the "Rename files" operation is providing the option of changing the number of zeros in the file name.

Ex: - The File name of one image is "Glimpses Of World religions@OCM23568503_0000001", and due to some reasons the customer wants to change the file name by changing the number of zeros. And the customer wants the file name as "12345_23456_0001". The illustration below shows the process of changing the Renaming the file.

The Process of "Rename Files" operation: -



The Left box in the “Rename Files” form shows all the loaded image files which are not renamed. And the Right Box in the “Rename files” form shows all the image files which are renamed.

The below figure shows the “rename Files” Operation: -

