



Overview:

VersiDoc is a fully integrated in-house developed production platform. The development of VersiDoc stemmed from our need to meet a multitude of client requirements with a uniform production process and the recognition that there was no off the shelf package that meet that need. We set out to develop a tool to increase our productivity while maintaining high levels of quality control and accountability. By developing this platform in real time production with real life projects we believe it to be the best suited overall package for our needs.

This demo will walk you through the process we utilize from the time we receive a call from a new client with a project through the delivery of the final product back to the client.

Phase I – Specification

Job specification is always done by Project Coordinators assigned to the project prior to distribution to production.

Client Definition: The first step is to create an account for the client. Client information consists primarily of basic demographics about the client, including office locations, individual contact information such as phone number, office location, and email address. Information for multiple contacts is stored and used for subsequent client matters (*Projects*) and individual document productions (*Jobs*).

Project Definition: The project definition represents the highest unit of related work within a particular client, usually represented by a client-matter number. At this level we define parameters such as *Page Id* or *Bates Series*, *Document Group Series*, *File Name Series* and *Volume Series*; these definitions are used to continue sequential series from job to job or a new series can be defined for each job. Each of these series has associated identities, prefixes and formatting.

The *Document Group Series* has special functionality. Unlike the others series that are essentially counters, the *Document Groups Series* also defines a set of rules describing how the boundaries of documents relate to one another. In most cases these rules refer to the physical boundaries of paper documents or the structure of electronic documents such as emails. An example definition may state and enforce following structure:

1. Boxes must contain Redwelds which must contain Folders which must contain documents identified by a coding sheets
2. Boxes, Redwelds and Folders cannot be empty
3. A coding sheet must be present before every folder.
4. The image of a coding sheet is to be kept.
5. In the case of documents prepped with bar-coded slip sheets the slip sheets are not to be kept.
6. All but document boundaries must be explicitly terminated while document boundaries are implied by the beginning of the next document.

Job Definition: This is a unit of work that is defined by you (the client) Individual production instructions are recorded and production priorities are assigned based on your deadlines. A pickup order is generated initiating the process. Once the job is received a color code and/or bar code is applied to the media received associating it with the job instructions. At this point the initial instructions and the original media are reviewed. We contact you to answer any questions and then a final job specification is produced and a system generated e-mail is sent to the client for final approval. Once approved the job is submitted to the system.

Step one is to categorize the job; each category has its own set of specific instructions. These required instructions are designed to control the production and QC process to ensure a consistent deliverable.

Scanning – The scan definition screen identifies the parameters of a scan job as well as associated options. Some of the key parameters are

Parameter	Description
Prep Box	The box is prepped with slip sheets prior to scanning
Recognize Bates	Identify Bates numbers through the use of zonal OCR
Force Letter Size	Convert all image sizes to 8 ½ X 11
Reassemble	The documents are reassembled after scanning is complete
Loose Group Macro	A slip sheet is used to indicate that the following pages are all single page documents
Copies	Identifies the number of copies to make while scanning
Page ID	Identifies a page ID series to be used

In addition to basic parameters a series of production options are identified. These options consist of:

➤ Code	➤ Stamp
➤ OCR	➤ Export
➤ Print	➤ Scan-To-Print

Each one of these options drives option specific instructions.

Coding – Coded data can be captured at any defined document level (ex. redwelds, folders, document, etc.) The system also has the capability to recognize mark sensor sheets through the use of Optical Mark Recognition (OMR). Coded field definition consists of the following parameters:

Parameter	Description
Name	A value that identifies the coding field.
Static value	A value that will be used to pre-populate the field
Type	Type of data formatting / enforcement (ie: Date, number, text, list, etc)
Case	Identifies type of case to automatically apply to field (ie: Upper/Lower/Name/Title,etc)
Length	Maximum number if character to allow to be types
Required	Specifies a mandatory field
Read Only	System generate field – cannot be changed
Association	What series level should the coding be associated with

OCR – Optical character recognition (OCR) is used to capture the text of scanned documents and include it in the database load file. OCR coordinates are also available for inclusion for systems that support that functionality.

Stamp – Bates stamping is performed electronically by creating a new image with pre-assigned numbers endorsed at the specified location on the page. Legends and bulk redaction can also be performed during this step.

VersiDoc Cluster Service - Both OCR and stamping are performed by the VersiDoc Cluster Service. The VersiDoc Cluster Service is a bank of 45 CPUs controlled by a centralized cluster manager whose purpose is to distribute individual images to the cluster engines for processing (OCR or Stamping). By dynamically distributing this work to all available engines, OCR and image endorsement are done very rapidly and efficiently.

Print – There are 2 print options available in the VersiDoc system, *Real-time Printing* and *Batch* printing. Real-time printing is specified by selecting 1 or more copies to make during scan definition. Batch printing is performed after scanning, validation, and Bates stamping if it is selected. Batch printing can be configured to print colored slip-sheets corresponding to any or all document groups (ex. folders, redwelds, etc.).

Export – Imaging jobs consisting of scanned pages and data, in the form of database load files, are exported from the system directly to CD or to the file system for electronic delivery. The export parameters consist of the following:

Parameter	Description
Image File Format	Single-page TIFF, Multi-page TIFF, PDF
Load File Type(s)	Summation, Concordance, DocuMatrix, etc. There is the option identify what fields are to be included in each load file. Multiple load files can be put on 1 CD
Root Directory	When an image path is exported, this identifies the root path used.
Image Orientation	Should the image always be portrait or if OCR processing was done should the OCR dictate the orientation of the image
Remove Spaces from file Names	Forces all file names to be contiguous characters and if there were spaces that are to be removed, what character if any should replace the spaces.
File Distribution Method	Specifies the directory structure used to distribute the images. For example, in the <i>Box</i> distribution scheme a directory is created to represent each physical box.
File Name Series	File names can match bates numbers or a separate file naming series can be used.
Volume Series	Specifies the volume label and series of CD/DVD media to be created.
Max Media Size	Specifies the size of a volume (ex. 650MB = CD, 4608MB = DVD)
Client CD Copies	Specifies the number of CDs to automatically burn for delivery. An additional copy is always created for archival purposes. All CDs created by the VersiDoc system are produced using CD burning machines (Rimage) that automatically burn and label the CDs with pertinent job information (Date, Client/matter, document & page counts, etc.)

Quality Control Definition: A QC methodology is designed for each and every job based upon the specific client deliverables and the processes defined for the job. Each step in the production process has its own tests, both programmatic and manual, which must be passed before the next step can be initiated. In addition, subsequent to scanning there is a Validation task in which several tests are performed. Finally, a Project Coordinator reviews and approves every assignment to verify that it was done completely and correctly.

The following operations are performed and tested during the Validation task:

Operation	Description
Manual Bates Indexing	If the scanned documents contain Bates numbers that could not be captured using zonal OCR (ex. hand labeled numbering) operators manually key the bates numbers by viewing each page.
Image vs. Original Test	If the scanned documents do not have preexisting numbers, a page for page, image for image check can be performed to verify that all pages were scanned. Any missing pages can be inserted during this step.
Image Orientation Check	Upside down pages can be rotated if any are detected and this is required by the job specification.
Image Quality Check	Image quality is reviewed at this time and poor images are replaced or corrected if possible,
Bates Sequence Test	If there are preexisting Bates numbers, once they have been captured all gaps, duplicates and out of order documents are reported and corrected or noted, depending on the job instructions.
Document Group Test	The rules defined in the document group series (see above) are all programmatically verified.

In addition to the validation step, the following are examples of quality control tests for the other steps defined above:

Scanning: Document scanning is typically done on multifunction copy/scan/print devices (ex. Canon ImageRunner 105) During the scan operation the documents are disassembled, scanned and reassembled one at a time and the scanned images are displayed for the specialist to review on screen before moving on to the next document.

Coding: In addition to the date, case and field type validation that is programmatically enforced while documents are coded, the VersiDoc system supports blind double key verification for all coded fields. If this option is selected 2 different document specialists code the documents and the values are individually compared to each other. Only matching values are allowed into the deliverable. Values that do not match are reviewed and routed to another specialist or the manager.

OCR: Every page submitted to the OCR cluster is attempted a minimum of 2 attempts. Any pages that could not be OCRed are logged and reviewed.

Stamping: The assignment of numbers to individual pages is done prior to the actual application of the numbers to the images. This assignment is done programmatically based on the Page ID series selected at the outset of the job. Since this number is stored at the client matter level, gaps or duplicated numbers are not possible. In addition, each time an image is modified (ex. orientation changed in validation or image endorsed in stamp) a new revision is created for the entire box. The creation and storage of these revisions allows the box to be *rolled back* to any previous state if necessary.

Printing: Each document slip-sheet printed during batch printing includes a sequential number to allow the user to confirm that all documents have been printed in the correct order. In addition, since the image count is known, the meter reading on the print device is compared to the expected count and verified.

Export: Prior to exporting the deliverable to the file system or burning to CD the following tests are performed:

Test	Description
TIFF Formatting	TIFFs are verified to be Group IV, 300 DPI
Compression Decompression Test	Images are all programmatically opened and closed to verify successful compression and decompression
Page Size Test	Only pre defined page sizes are allowed (ex. Letter, Legal, 11X17, etc.)
Image Cross Reference Test	Load file to image references are all verified to ensure that all referenced images are present and all images present are referenced.
Load File Validation	Load file formatting is always validated and reviewed by a Project Coordinator. (ex. XML formatted load files re opened in an XML parser)
Document Group Test	The rules defined in the document group series (see above) are all programmatically verified.

Once all of the above specification is complete an e-mail containing all of the instructions is sent by the system to the client for verification. Upon verification, the job is sent to production.

Phase II – Production

Once a job specification is complete, the original items (ex. boxes) are added to the job and a grid is created which displays each item and indicates which steps need to be performed:

Label & Type	Copy	Stamp (Paper)	Runback	Prepare	Scan	Import	Convert	Validate Box	Generate Bates	Code	Stamp (Image)	OCR	Finalize Data	Generate Load Files	Print	Assemb.	Export
Original Box BOX 001					Yellow												
Original Box BOX 002					Yellow												
Original Box BOX 003					Yellow												
Original Box BOX 004					Yellow												
Original Box BOX 005					Yellow												
Image Box BOX 001								Blue	Blue	Blue	Blue	Blue	Blue	Blue			
Image Box BOX 002								Blue	Blue	Blue	Blue	Blue	Blue	Blue			
Image Box BOX 003								Blue	Blue	Blue	Blue	Blue	Blue	Blue			
Image Box BOX 004								Blue	Blue	Blue	Blue	Blue	Blue	Blue			
Image Box BOX 005								Blue	Blue	Blue	Blue	Blue	Blue	Blue			
Box Copy BOX 001-1															Blue		
Box Copy BOX 002-1															Blue		
Box Copy BOX 003-1															Blue		
Box Copy BOX 004-1															Blue		
Box Copy BOX 005-1															Blue		

In the example below;

- Original Box 001 has been scanned and approved by the manager.
- Original Box 002 has been scanned but not approved by the manager.
- Original Box 003 is in progress.
- Original Box 004 has been assigned but has not been started.
- Original Box 005 has not been assigned.
- Image Box 001 is ready for assignment.

Document Production Console 0.3.72 ("administrator") - [Job: [4] (World Copy0001_001)]

File Window Help

Refresh Open Wizard Reopen Job Close Job Print Work Order Print Counts Open Field Map Release Series Locks

Detail Task Grid

Label & Type	Copy	Stamp (Paper)	Runback	Prepare	Scan	Import	Convert	Validate Box	Generate Bates	Code	Stamp (Image)	OCR	Finalize Data	Generate Load Files	Print	Assemb.	Export
Original Box: BOX 001																	
Original Box: BOX 002					done												
Original Box: BOX 003					⚙️												
Original Box: BOX 004																	
Original Box: BOX 005																	
Image Box: BOX 001																	
Image Box: BOX 002																	
Image Box: BOX 003																	
Image Box: BOX 004																	
Image Box: BOX 005																	
Box Copy BOX 001-1																	
Box Copy BOX 002-1																	
Box Copy BOX 003-1																	
Box Copy BOX 004-1																	
Box Copy BOX 005-1																	

This display allows the manager to always know where a job is in the overall production process as well as what he or she needs to do while simultaneously being the interface to perform the actual tasks.

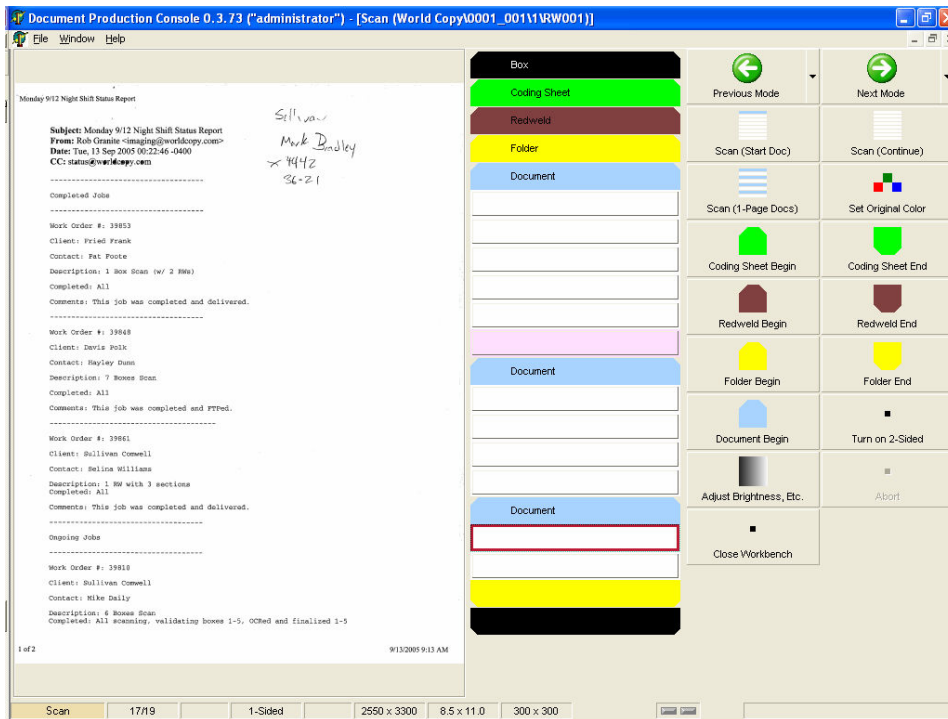
In the following section, each task and its interface will be described and its features and benefits explained. The tasks supported by the system are:

Task	Description
Scan	Paper capture through touch screen interface
Validate Box	Perform QC and capture Bates if required
Generate Bates	Programmatically assigns bates numbers to pages
Code	Code data from scanned documents, coding sheets, or mark sensor sheets
Stamp (Image)	Distributed electronic endorsement of images with pre-assigned Bates values
OCR	Distributed Optical Character Recognition
Finalize Data	Compiles document level data for future export
Generate Load Files	Transforms compiled data into requested load file formats
Export	Creates electronic deliverable on CD or file system
Print	Batch printing
Import	Imports native files and extracted data for conversion to TIFF and export
Convert	Distributed conversion of native files to TIFF

There are some tasks that occur outside the scope of the system but are, nevertheless tracked by the system. These include:

Task	Description
Copy	Document duplication on copy machines
Stamp (Paper)	Off-line paper Bates stamping
Runback	Duplication of off-line Bates stamped documents
Prepare	Off-line preparation of documents with slip sheets for scanning
Assemble	Off-line reassembly of scanned documents

Scan: When a document specialist logs in and begins a scan assignment the Workbench interface automatically opens as seen below:

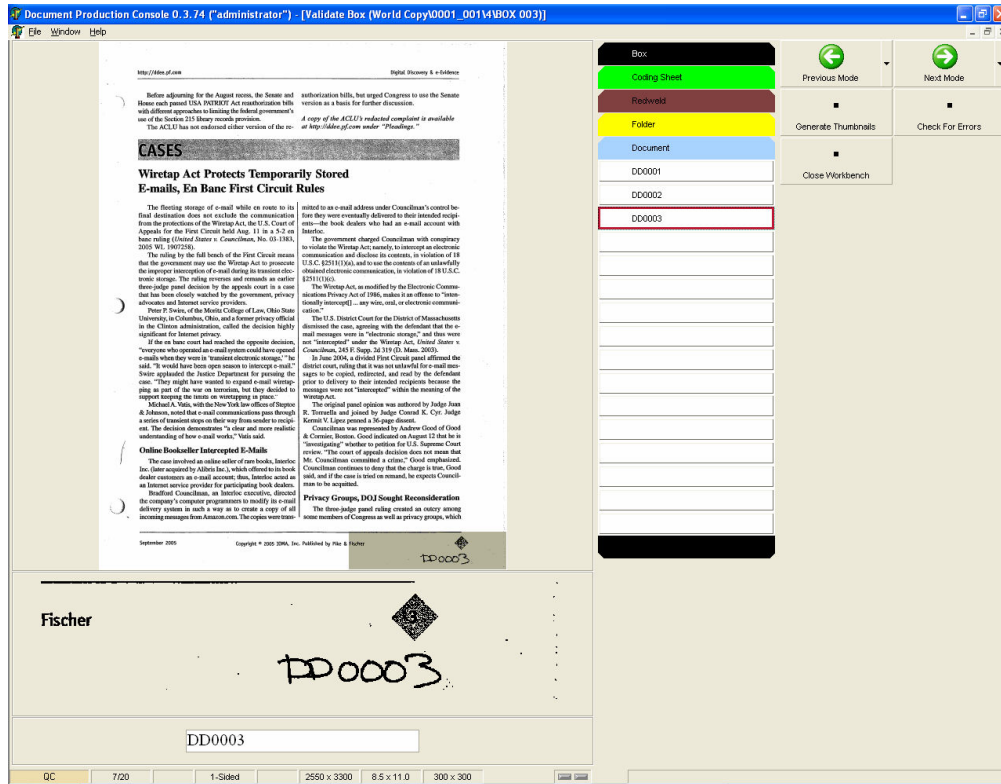


The Workbench interface is used for scanning and validation. The window is divided into 3 main sections:

- Image Viewer – The left side of the window displays all scanned images as they are captured.
- Page List – The middle section of the window displays a row for each page and document group captured. Document groups are displayed in different colors and their name is shown in the row. If Bates numbers exist and are being captured at scan time the value would be displayed in this section.
- Button List – The right side of the window is the button interface. There are 8 button sections available in Workbench. Each of these sections is designed for use with a touch screen monitor. They are:
 - Administration
 - Enable Prepared/Raw scanning – toggles scan mode for prepared or non-prepared originals
 - Scan Blank Test Page – Tests scanner recognition of blank pages
 - Change Blank Color – Changes the display color in the Page List of detected blanks.
 - Image on Right – Moves image viewer to right side of screen.
 - Show Subimage – Displays blown up region of image. For use when recognizing existing Bates numbers during scanning.
 - Show Bates Editor – Shows recognized Bates values
 - Resynchronize images – Synchronizes images on local machine with server.
 - Information – Displays image information.
 - Color – Sets the paper color of the originals scanned. These pages can be blown back on their original color paper or, in the case of color originals, replaced with a color copy.
 - Import – Imports images directly from file system with image list.
 - Manipulate
 - Delete – Deletes selected page(s)
 - Toggle Blank Status – Marks or unmarks pages recognized as blanks.
 - Rotate Permanently – Rotates and saves scanned image.
 - Select/Deselect to End
 - Navigate
 - First/Last Page of box
 - Screen Up/Down
 - Next/Previous Page
 - Next/Previous Patch Page
 - Next/Previous Blank Page
 - QC
 - Generate Thumbnail Images (Used in Validation task)
 - Check for Errors (Used in Validation task)
 - Scan (shown above)
 - Scan (Start Doc) – Begins a new document
 - Scan (Continue) – Continues to add pages to the current document
 - Scan (1-Page Docs) – Scans the current batch as single page documents or 2 page documents in the case of double-sided scanning.
 - Begin/End Document Group(s) – Color-coded begin and end buttons for each document group defined for the job (ex. Folder, Redweld)
 - Document Begin – Inserts document brake at current position in the page list.
 - Turn on/off 2-sided – Toggles scanner between simplex & duplex scanning.
 - Adjust Brightness, etc. – Adjusts brightness, contrast & threshold for the current batch
 - Scan (Prepared) – Scans prepared documents and recognizes bar-coded document group patch pages.

Validate: When scanning is complete and approved by the manager the next step in the production line is Validation. Again, Workbench is used for validation and during this task all of the Quality Control checks listed above are performed. Since the same Workbench interface is used for both scanning and validation, insertion of missed pages is done consistently.

If manual Bates capture is required, the Workbench interface is changed slightly to display the Subimage Viewer and Bates Editor.



The Subimage Viewer is the section directly below the image viewer in the above example. The zoom region is highlighted in gray on the image and is blown up and displayed in a gray background for easy viewing.

The Bates Editor is directly below the Subimage Viewer and is used to enter the Bates value displayed in the Subimage Viewer. Document Specialists need only to type in the first Bates value and upon hitting the Enter key, the next page is displayed with the next Bates value automatically populated.

Generate Bates: This task is performed by the system as opposed to a Document Specialist. Bates generation can happen in 3 modes:

- Manual – Each box's Bates generation is executed by the manager in any order.
- First Completed – When a box has passed validation it is immediately assigned bates numbers.
- Job Item Label – Boxes are processed in the order of their labels (ex. Box001, Box002, etc.)

Code: The coding form is dynamically generated and displays the image on the right side of the screen and the defined fields on the left. Since coding can occur at any document group level, the current level is displayed in the appropriate color at the top of the screen and only the appropriate fields are enabled.

The coding system also has the capability to recognize, using OMR, a mark sensor sheet to automatically populate the fields. Below is an example of the coding form recognizing a mark sensor sheet:

Stamp (Image): Electronic endorsement is a function of the VerisDoc Cluster Service. Since each page is pre-assigned its number, when a box is submitted for stamping the Cluster Server dynamically distributes the box to however many Cluster Nodes are available. For example, if a box contains 2500 pages and there are 50 Cluster Nodes online each Cluster Node will be given 50 pages to endorse with the assigned number. The cluster node received the image and the instructions from the server, created the new image and submits it back to the server. Once a box is complete it is recompiled by the server and imported into the database as a new revision of that box. This versioning allows the opportunity to produce numbered and unnumbered versions of the same documents if necessary.

OCR: Optical Character Recognition (OCR) is also a Cluster Service function. Images submitted for OCR are dynamically distributed in the same way as stamping. A cluster of approximately 50 nodes has the capability to process over 1 million pages in 24 hours.

Finalize Data: Finalization of data is the compilation of all information from the page level to the document level. This task is performed by the system and consists of the following:

- Merging the OCR data for all of the pages of a document into one field at the document level.
- Storing the beginning and ending Bates or page IDs of a document.
- Merging coded data from all appropriate levels (ex. folders, redwelds, etc.) to the document level.
- Creating document page counts
- Generating Document IDs if necessary.
- Creating CD or DVD volumes by calculating the disk space required of the images and data of each box.

Generate Load Files: Load file generation is a system task that simply converts the finalized data created above into the selected and defined load file formats for the job (ex. Concordance, Ipro, Documatrix, etc.)

Export: Export is a system task that actually creates the CD or DVD, by use of an automated CD/DVD recording/printing unit. In addition to exporting the deliverable data to the file system the export task also burns the specified number of media defined for the job plus an additional copy for archival purposes. Each CD is automatically labeled with the following information:

- Creation Data
- Client Name
- Client Matter
- Volume Label
- Image Type
 - Single Page TIFF
 - Multi Page TIFF
 - PDF
- Load File Type(s)
- Document Count
- Page Count
- Work Order Number

Print: Batch printing allows users to select which, if any, document group slip-sheets to print from which trays as well as printing pages on their original colored paper if it was set at scan time. For example, if blue paper were loaded in tray 1 of the printer, yellow paper in tray 2, legal paper in tray 3 and white letter paper in the side paper deck the document specialist could print folder slip-sheets from the yellow tray, documents originally on blue paper from the blue tray, legal documents from the legal tray and letter pages from the side paper deck. Document slip-sheets can be configured to contain any information available in the database, for example:

- Bates number(s)
- Folder information
- Sequence information
- Box number
- Client matter
- Etc.

Import: This feature imports, from the Document Extraction System, native files and their extracted metadata into the system for conversion to TIFF and/or printing. The system can import:

- Microsoft PSTs
- Lotus Notes NSF's
- Native files
 - Word
 - Excel
 - PowerPoint
 - Text
 - Images (GIF, JPEG, BMP, etc.)
 - Etc.

Convert: Convert is another VersiDoc Cluster Service. In the same way images are submitted for OCR and stamping, the cluster service distributes native files to however many nodes are available for conversion to TIFF and import back into the database to be processed in the same ways any scanned documents can be processed.