

ERDC/ITL CR-04-xx



USACE CADD Center Project 03.036

**Functional Specification:
Quality Control & Metadata Building Tool for Managing
Scanning Projects**

Release Version 1.0
January 23, 2004

Prepared For:

Mr. Ralph Scheid
CEMVN-ED-SE
US Army Corps of Engineers, New Orleans District
P.O. Box 60267, New Orleans, LA 70160-0267
(504) 862-2995
Ralph.A.Scheid@MVN02.usace.army.mil

and

Mr. Blaise Grden
CEERD-ID
US Army Engineer Research and Development Center
CADD/GIS Technology Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6199
(601) 634-3581
Blaise.G.Grden@erdc.usace.army.mil

Prepared By:

InStep Software, LLC
55 E. Monroe
Chicago, IL 60603

Document Control

Change History

Issue No.	Status	Issue Date	Issuer	Comments
	Draft	Dec. 16, 2003	Karen Schellinger	Preliminary Document
0.1	Released	Dec 22, 2003	Ralph Scheid	Released for review with COE updates.
0.2	Released	Jan 5, 2004	Karen Schellinger	Added changes from COE ...
1.0	Final	Jan 23, 2004	Karen Schellinger	

Distribution

Karen Schellinger	Project Manager	InStep, Inc.
Michelle Sun	Program Developer	InStep, Inc.
Blaise Grden	ERDC Principal Investigator	USACE ERDC-ITL
Milton Richardson	ERDC Principal Investigator	USACE ERDC-ITL
Ralph Scheid	CW FWG Technical Manager for 03.036	USACE-MVN

Table of Contents

<i>Table of Contents</i>	3
<i>Introduction</i>	4
<i>System Requirements</i>	4
<i>Functional Description</i>	4
<i>Acceptance Testing</i>	16
<i>Tasks Performed</i>	16
<i>Project Schedule</i>	17

Introduction

This document is provided under USACE CADD Center Project 03.036 Historic & Digital Map Collection Methods & Online Retrieval Tool Development Phase I.

The purpose of this document is to provide an overview on the design and use of the Quality Control Metadata Building Tool for managing scanning jobs. The Quality Control Metadata Building Tool (hereafter referred to in this document as QCMBT) is intended for use during the physical scanning phase of a digital imaging project to aid the U.S. Army Corps of Engineers personnel and/or their contractors to record and review the progress of the scanning process. Its reporting capabilities can be used subsequent to the imaging process to identify a document or series of documents by querying the metadata assigned during the imaging process. Its collection of metadata may also be exported for input into an EDMS system as the digital documents created during the imaging process are migrated into a permanent repository.

System Requirements

Installation and use of QCMBT requires:

- Microsoft Office Suite of Products (including Microsoft Access)

No EDMS software or other applications are required for the use of QCMBT. The QCMBT Access database file is easily transported between physical locations. If contractors are involved in the imaging process the file may be made available to them for data entry, then become part of the electronic deliverable returned by the contractor along with the imaged documents. In the case where contractors have access to large scale imaging software that downloads directly into major EDMS systems this document should serve as the requirements to the imaging contractor for the specification of metadata fields.

Integration with MCABLS

QCMBT is fully compatible with MCABLS. MCABLS was designed as a means to import documents into Bentley's Digital Interplot Application.

MCABLS must be properly installed prior to exporting the contents of a QCMBT database to a MCABLS database. To install MCABLS see the "Metadata Collection and Bulk Loading System (MCABLS) Installation, Configuration, and User Guide".

Once MCABLS is installed correctly the export functionality within the tool can be used to export those portions of the metadata collected during the scanning phase that are used by the MCABLS software to the MCABLS database. The MCABLS program can then be invoked to enter additional EDMS fields and bulk load the documents into Bentley's Digital Interplot application. The metadata collected during the scanning/indexing process will not need to be manually re-entered in the MCABLS system.

Quality Assurance within the Workflow

QA should not be an external system designed to find failures, but rather an internal and integrated system that promotes continual monitoring of image quality throughout the workflow. Although the size of the project may dictate that the same person fulfills multiple roles, in general the approval process may include roles: "scanner/indexer", "quality assurance #1", "quality assurance #2", and "project acceptance". This ensures that an independent verification is done at each step of the process to ensure the integrity of that step.

The project manager would be responsible for designating the other members of the project team and would be responsible for final approval. The project manager selects the documents to be scanned and enters the job parameters for each "batch" of documents into the job parameter table using the QCMBT tool. A "batch" of documents consists of documents with similar physical characteristics and desired scanning parameters.

Once the image has been captured and processed it should be checked by the operator who undertook the work before it is allowed to move on within the workflow. It therefore becomes the responsibility of the operator (or team) to check each and every image (and associated metadata) before it is allowed to continue within the workflow. The process of signing off enables the production team to know that not only has every image been visually checked as being of acceptable quality, but also by whom and when it was signed off. This QA administrative metadata will form an image audit history. If a fault is recognized it should be rescanned immediately, or if the image cannot be corrected then the fault should be recorded in the appropriate error field of the QCMBT program.

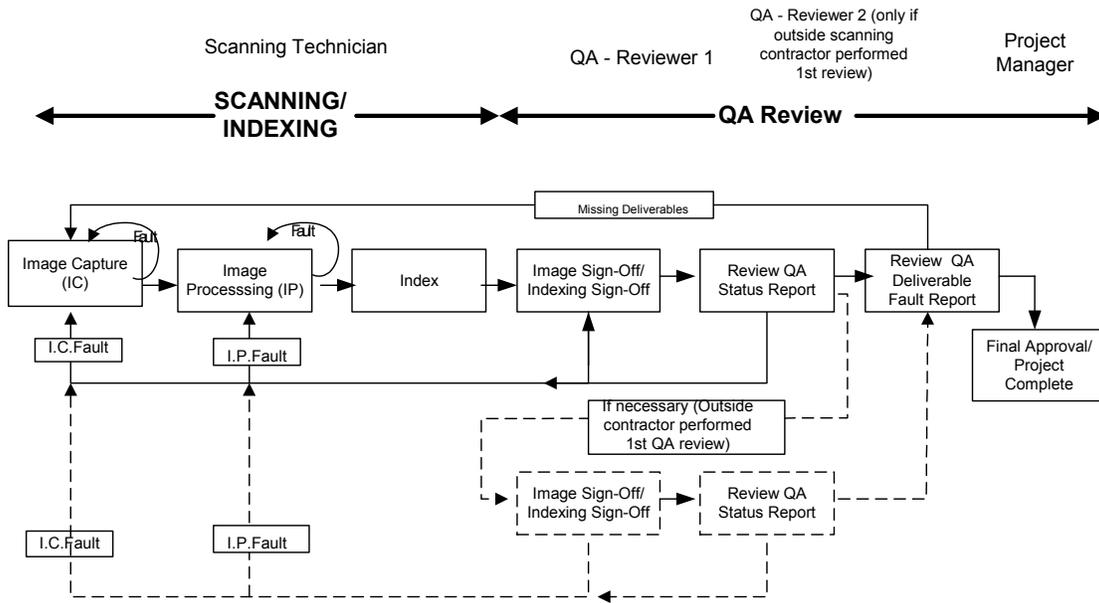
Faults that would be identified at the scanning stage would be basic capture faults: bad cropping, wrong orientation, incorrect exposure, out of focus, bad calibration of capture device, or dirty originals.

After the physical scanning, the scanner/indexer would enter the appropriate metadata for each document.

After the images have been created they should be inspected by a different person from the production operator to assess the fidelity of the images against the originals and to discover systemic defects or trends in poor operator technique. This should be done by the scanning suppliers for collections which are outsourced, and within collections for in-house projects. With outsourced scanning there should also be a selective review made when documents are returned from the supplier (this is purpose of the QA #2 role), so that there is always a measure of in-house review and approval. An additional check should be performed by the reviewer to verify that the metadata has been properly entered and matches correctly with both the original document and the imaged file. If the project is large this could be a selective review with the percentage of the collection reviewed set by the project manager at the outset as one of the project QA parameters.

The project manager has final approval over all documents, but the primary role of the project manager is to oversee the capture team, monitor the progress of the team and use the QA metadata that has been created by the sign off process and the fault reporting process to create progress reports. These QA reports will demonstrate both the rate of the work (completion dates) as well as the number of recorded faults.

The following workflow diagram illustrates the scanning/indexing/QA process:

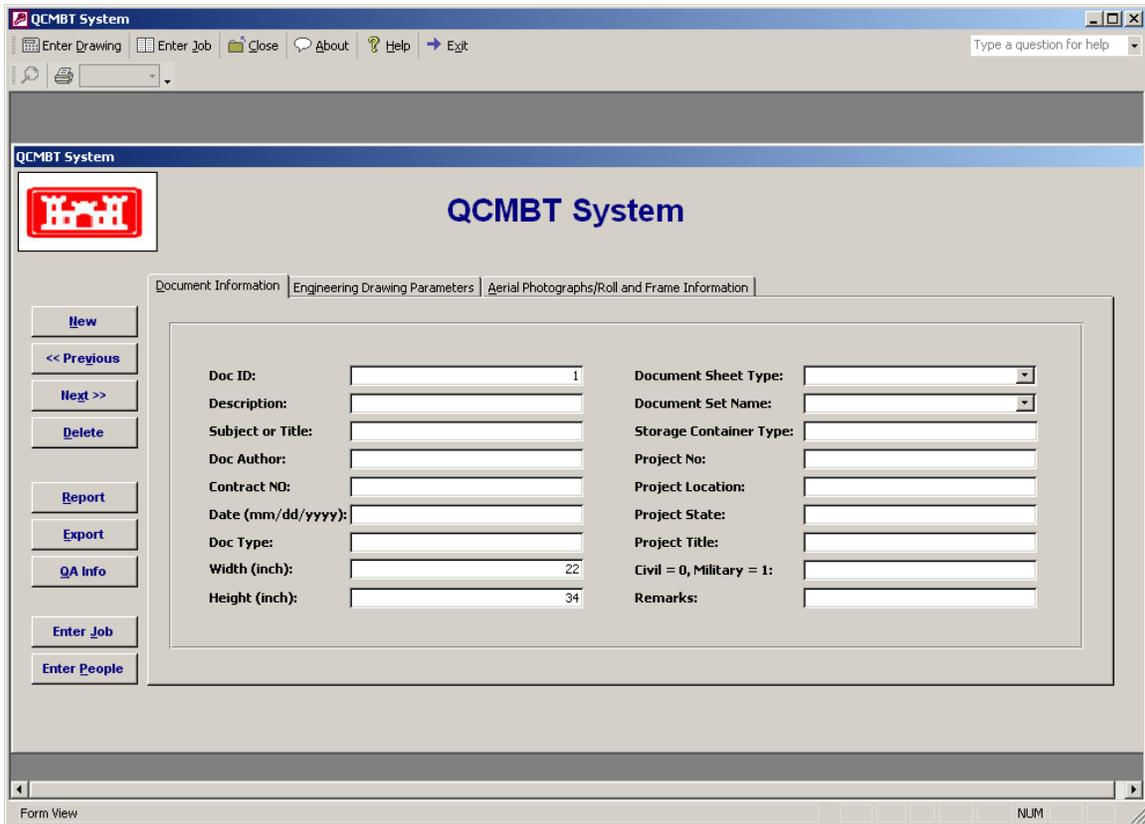


Functional Description

QCMBT Tables and User Interface

This section documents the steps users would take during a scanning project and the use of the tool during that process.

The tool has a main entry screen, which is shown below. This is the screen that will be presented upon initial entry.



An about box with the following information is available:

- Version 1.0
- Release Date
- Contact Info
- Contract Info
- CW FWG Reference

Step One – Enter Job Parameters

The QCMBT has one database that can support multiple document types. The user should double-click on the QCMBT.mdb file to start the tool. Microsoft Access will start with a data entry form appearing. PRIOR to start of scanning the scanner (or USACE project manager) uses the button entitled “Enter Job” to enter a number of job parameter “sets”, with a name identifier:

Each document set may define up to four scan types, as shown on the above screen shot. For each scan type desired the user should enter the following information:

DPI: the user may enter any number desired or select from the default values provided. Default values provided by the pull-down list include 150, 200, 300, 400, 600, 800, 1200.

Scan Bit Depth: the user may enter any number desired or select from default values provided. Default values provided by the pull-down list include 8, 16, 24.

Scan Format: the user may enter any format desired or select from default values provided. Default values provided by the pull-down list include TIF, JPG, MrSID, PDF.

Compression: the user may enter any compression desired or may select from default values provided. Default values provided by the pull-down list include No Compression, CCITT Group 3, and CCITT Group 4.

Color: the user must select either RGB or Grayscale

% Tolerance: The user may enter any value between 1-100. This value is used for non-compressed documents in the missing document QA report to check whether the size of the delivered document is within the calculated amount plus or minus this tolerance factor.

The Previous and Next buttons may be used to view and/or edit all the entries in the job parameter table.

The job parameter database table would enable multiple “batches” per database and would contain the following fields:

JobParameters : Table		
	Field Name	Data Type
🔑	Document_Set_Name	Text
	Archival_DPI	Number
	Archival_Bit_Depth	Number
	Archival_Scan_Format	Text
	Archival_Color	Text
	Archival_Compression	Text
	Archival_Tolerance	Text
	Browse_Pixels	Number
	Browse_Bit_Depth	Number
	Browse_Scan_Format	Text
	Browse_Color	Text
	Browse_Compression	Text
	Browse_Tolerance	Text
	Dist_Pixels	Number
	Dist_Bit_Depth	Number
	Dist_Scan_Format	Text
	Dist_Compression	Text
	Dist_Color	Text
	Dist_Tolerance	Text
	Other_Compression	Text
	Other_Scan_Format	Text
	Other_Bit_Depth	Number
	Other_Color	Text
	Other_Pixels	Number
▶	Other_Tolerance	Text

During scanning the scanner selects which job parameter set this document belongs. This allows the report logic (discussed later) to check for associated documents (whatever is defined for that job class) and to use the parameters to perform the size checking after deliverables have been received. The job parameters must all be assigned prior to the start of scanning.

Step Two – Enter People and Roles

Prior to scanning the project manager should establish roles and team members assigned to those roles. Use the Enter People button on the main form to bring up the entry form, below:

People and Roles				
USER_NAME	SCANNER	QC1	QC2	PROJMGR
Ralph	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
▶	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Record: of 2

Enter each member of the team and the roles assigned to that team member. This information is stored in the People table.

PEOPLE : Table	
Field Name	Data Type
ID	AutoNumber
USER_NAME	Text
SCANNER	Yes/No
QC1	Yes/No
QC2	Yes/No
PROJMGR	Yes/No

Step Three – Scan and Index

The scanner/indexer would scan each document in turn and fill in information for each document scanned. The metadata collected by QCMBT falls into three primary categories:

- Document metadata fields
- Type-specific document fields
- Quality assurance metadata fields

The first two are found in the document attributes table and are discussed in this step, the third is discussed in Step Three – Quality Assurance Review.

The following data fields are to be completed by the person whose role has been designated as “scanner/indexer”. This is the information that is gathered for each document without regard to type. The scanner/indexer does not need to enter all fields, the fields that are required are marked. Fields that are the same as the Corps’ EDMS model are named, typed and sized to match the model.

Data Element Name	Data Type	Req?	Size	Comment
DOC_ID (K)	Text	Y	25	System-generated
DOC_HEIGHT	Integer	Y		
DOC_WIDTH	Integer	Y		
STORAGE_CONTAINER_TYPE	Text	N	50	
DOC_SUBJECT_OR_TITLE	Text	Y	255	
DOC_TYPE	Text	N	25	
DOC_DESC	Text	N	50	
DOC_REMARKS	Text	N	50	
DOC_AUTHOR	Text	N	50	
DOC_CONTRACT_NO	Text	N	20	
DOC_DATE	Date(4 digit yr)	N	11	
CIVIL_MIL_FLAG	Yes/no	N		
PROJECT_NO	Text	N	20	
PROJECT_LOCATION	Text	N	50	
PROJECT_STATE	Text	N	2	
PROJECT_TITLE	Text	N	255	
JOB_PARAMETER_KEYID (FK)	Numeric	Y		
DOC_SHEET_TYPE	Text	N		

Doc_ID.	This is a system-generated unique identifier
Doc_Height	The height in inches of the document.
Doc_Width	The width in inches of the document.
Doc_Subject_or_Title.	If the document has a title, that would be entered or captured here.
Doc_Type	The type of document, e.g., map, drawing, specification, correspondence, etc.
Doc_Desc	A broader description of the document, if desired
Doc_Remarks	Any remarks applicable to this document
Doc_Author.	Document's author if known
Doc_Contract_No.	Contract number if the document is related to a contract.
Doc_Date.	Date of the original document.
Civil_Mil_Flag	Flag, Civil or Military
Doc_Project_No.	The number assigned to a project
Project Location.	The physical location of the project
Project Title.	The official title of a project.
Project State.	The state within which the project is located.
Job Parameter Key	The name of the job parameter set that applies to this document
Doc_Sheet_type	The sheet type: None, Letter, Legal, A (ansi), B(ansi), C(ansi), D(ansi), E(ansi), F, A0(iso), A1(iso), A2(iso), A3(iso), A4(iso), A5(iso), Aerial(9x9), Aerial(9x18) If the user enters anything other than none the appropriate width and length fields for the document are filled in.

The Corps has identified two document types for which they would like additional metadata collected, engineering drawings and aerial photographs. Each document type has fields specific to that document type.

The following type-specific metadata should be entered by the scanner/indexer if the document is an aerial photograph:

Table 1 - Metadata for Aerial Photographs/Roll and Frame attributes

Data Element Name	Data Type / Size	Comment
roll_no	Text(20)	SDSFIE compatible
date_taken	Number	SDSFIE Format YYYYMMDD
photo_scale	integer	
film_type_id	Text (1)	
film_format_id	Text (1)	
sensor_class_id	Text (1)	
filter_type_id	Text (8)	
forward_lap	integer	
frame_no	integer	
Flight_line_name	Text(10)	
Orthorec_d	Y/N	

roll_no: Roll no. of the film can from which the print was made. This field is compatible with matching SDSFIE attribute.

date_taken: Date on which the photo was taken. This field has been renamed and changed to a 10-digit integer field (from a date) field to be compatible with SDSFIE attribute.

photo_scale: 1: xxxxxx feet.

film_type_id: Type of film – choose from following list:
 1 b & w infared
 2 color infared
 3 color
 4 b & w
 5 thermal
 6 panchromatic
 9 other (enter in remarks)

film_format_id: Film format type – choose from following list:
 1 2.76" x 2.76" (70mm x 70mm)
 2 4.5" x 4.5" (11cm x 11cm)
 3 9.0" x 9.0" (23cm x 23cm)
 4 9.0" x 18.0" (23cm x 46cm)
 5 7.0" x 7.0" (18cm x 18cm)
 6 7.0" x 9.0" (18cm x 23cm)
 7 6.0" x 8.0" (15cm x 20cm)
 8 1.0" x 1.38" (25mm x 35mm)
 9 other (enter in remarks)

filter_type_id: Color of filter used – choose from following list:
 yellow yellow
 orange orange
 red red
 cyan cyan (blue-green)
 clear clear

forward_lap: The percent of forward lap from photo frame to photo frame.

sensor_class_d: Sensor class – choose from following list:
 1 vertical cartographic (implies stereoscopic overlap)
 2 vertical reconnaissance (implies monoscopic coverage)
 3 side looking airborne radar (SLAR)
 4 oblique
 9 other (enter in remarks)
 10

frame_no: The frame number of the image
flight_line_name Character 10, The flight line that the photo is on.
Orthorec_d A yes/no value whether or not the aerial photograph has been ortho rectified.

The following type-specific metadata should be entered by the scanner/indexer if the document is an engineering drawing:

Table 2 - Metadata for Engineering Drawings/Maps

Data Element Name	Data Type / Size	Comment
drawing_no	Text (20)	
sheet_no	integer	
total_no_sheets	integer	
spec_no	integer	
contract_no	Text(40)	
Dwg_designer	Text (50)	
Dwg_drafter	Text (50)	
Dwg_checker	Text (50)	
Dwg_approvedby	Text (50)	
Dwg_approval_date	Text (50)	
Scale	Text (50)	
Revision	Text (50)	
Revision_Date	Date	
Dwg_Revisor	Text(50)	
Dwg_loc1	Text (50)	
Dwg_loc2	Text (50)	
Dwg_loc3	Text (50)	

drawing_no Drawing number
sheet_no Sheet number or plate no.
total_no_sheets Total number of sheets in the drawing set
spec_no Specification number that relates to the drawing plate (if known or written on the plate)
contract_no Contract number (if known)
Dwg_designer **Name of the person who** designed the drawing
Dwg_drafter **Name of the person who** drew the drawing
Dwg_checker **Name of the person who** checked drawing
Dwg_approvedby **Name of the person who** approved drawing
Dwg_approval_date Approval Date
Scale Scale
Revision Revision No., if a revision
Revision Date Date drawing revised.
Dwg_Revisor Person who revised the drawing.

(The following three fields are for any location information applicable to the map/drawing, examples could be drainage/basin/river, city/county/state, reservoir, building, etc. These fields can be left blank if not applicable.)

Dwg_loc1 Location of information in drawing/map
Dwg_loc2 Location of information in drawing/map
Dwg_loc3 Location of information in drawing/map

Step Four – Quality Assurance Review

After scanning and indexing is complete the scanner/indexer would review the images produced and the metadata entered. They would then fill in the first four fields in the table (this information is on tab labeled “Quality Assurance Information” on the data entry form). The reviewer(s) would then review the scans and the metadata collected, and in accordance with the workflow above either approve or return the document for rescanning/re-indexing. The fields are stored for each individual document in the

The following table indicates what data fields would be completed during the imaging process and whose responsibility it is to complete those fields:

Data Element Name	Data Type	Size	Completed By
Doc ID			Auto
SCAN_filename			Scanner/indexer
SCAN_TYPE			(Master, Distrib, Browse, Other) Scanner/indexer
DOC_STORAGE_CONTAINER	Alpha	50	(Volume ID where file is stored) Scanner/indexer
SCAN_COMPLETE	Boolean	1	Scanner/Indexer
SCAN_ERROR	Alpha	35	Could be entered by any role
SCAN_DATE	Date(4 digit yr)	11	Scanner/Indexer
SCANNER_NAME	Alpha	25	Scanner/Indexer (Domain table lookup)
SCAN_1_APPROVED	Boolean	1	QA – Reviewer 1(Domain table lookup)
SCAN_APPROVAL_1_DATE	Date(4 digit yr)	11	QA – Reviewer 1
SCAN_APPROVAL_1_NAME	Alpha	25	QA – Reviewer 1(Domain table lookup)
SCAN_2_APPROVED	Boolean	1	QA – Reviewer 2
SCAN_2_DATE	Date(4 digit yr)	11	QA – Reviewer 2
SCAN_APPROVAL_2_NAME	Alpha	25	QA – Reviewer 2(Domain table lookup)
PROJ_MGR_APPROVAL	Boolean	1	Project Manager(Domain table lookup)
PROJ_APPROVAL_DATE	Date(4 digit yr)	11	Project Manager
PROJ_MGR_NAME	Alpha	25	Project Manager
Remarks			

Reporting

Two reports will be available to the QCMBT user(s) through the control button labeled “Reports” on the primary data entry window.



The first will list each document id and show the 13 metadata fields associated with quality assurance. This can be used by any user for status updates since it details where the document is in the quality control process (who has checked it and when). The second will be a missing document report.

It will be used either by the project manager or his designated QA reviewer to assess the deliverable returned from the scanner/indexer. It will list the total number of documents in the database, the total number of documents delivered on CDs or removable drives, a list of the files that are in the database but not found on any of the CDs or removable drives, and a list of the files that were found on the CDs or removable drives that are not found in the database, and a list of documents that have failed the size check.

The size check will be done only on documents that are both in the database and found on the CDS or removable drives delivered and also are delivered in a non-compressed TIF format according to the entries made in the database. The size check will use the following formula:

$$\text{Size (MB)} = L \times R \times W \times R \times \text{BD} / 1024 \times 1024 \quad \text{where}$$

L = length of document(in)
R = resolution(dpi)
W = width of document(in)
BD = bit depth / 8

A percent tolerance, entered during the job parameter setup prior to the start of scanning, will be used to determine if the archival image meets the size requirement. This equation is applicable only to non-compressed tiff images.

It will be possible to print each report if desired. Both reports will be generated using standard Microsoft Access reporting commands and presentation.

The second report (missing deliverables) uses several temporary tables within the database that are not relevant to the user but are used internally by the code while generating the report.

Data Export

It will be possible to export the data to MCABLS using the "Export data" button on the primary database form. The user will be asked the location of the MCABLS database, which must be accessible and not write-protected.



It will also be possible to export the Document Attributes table to either an Excel file (*.xls) or a text file (*.txt) using the "Export data" button. These formats can then be used to load the information about the scanned documents into EDMS applications. All data exporting will be generated using standard Microsoft Access export commands. The user will be queried for a filename and will be given an indication of the success or failure of the export.

Acceptance Testing

USACE ERDC and/or USACE New Orleans District will perform acceptance testing on the QCMBT Tool. The testing will be based on the requirements and assumptions listed in this document.

InStep will provide a draft Acceptance Test Form/Checklist with the QCMBT beta. The COE is responsible for the final Acceptance Test Form/Checklist used since a detailed checklist requires knowledge of the specifics of the documents, equipment, and personnel available for testing.

Tasks Performed

The following items are to be performed by InStep, Inc. personnel or by the COE's personnel as noted:

- USACE will demonstrate the use of the tool on actual data to ensure that the requirements are met.
- The COE will take no more that two weeks to verify the documents and data in the development data source.

Revised Project Schedule

	Duration	Start	End
Project Scoping			
InStep functional specification			1/5/2004
COE sign-off in functional specification document	5 days	1/5/2004	1/9/2004
Delivery of Beta Version of QCMBT			1/23/2004