SYSTEMS DELIVERY AND PROJECT PORTFOLIO MANAGEMENT

ATTACHMENT 1 TO ANNEX H

RCMP AFIS RENEWAL BENCHMARK TEST PLAN

AFIS RENEWAL

Last Updated Date:	2015-09-21
Status:	Draft
WBS:	REB-11
Version:	0.1
RDIMS Document No.:	42561
Classification:	Protected A

Royal Canadian Gendarmerie royale
Mounted Police du Canada

Canada

RECORD OF AMENDMENTS

Version No.	RDIMS Ver.	Date	Comments	Author (s)

i

TABLE OF CONTENTS

1. INTRODUCTION

1.1 GENERAL

1. This Attachment 1 to Annex H describes the RCMP AFIS renewal benchmark test plan which must be considered as part of Stage 3 of the evaluation plan (i.e. Stage 3 - AFIS Renewal Benchmark Testing section). This attachment of the AFIS Renewal Evaluation Plan and Criteria describes the scope, objectives, evaluation criteria, and procedures for the AFIS renewal benchmark test plan. The glossary in section 5.8 defines the terms and abbreviations used in this document.

2. Each Bidder will submit a detailed Benchmark Plan with their proposal based on this document and additional information provided in Annex H regarding the benchmark testing (e.g. Stage 3 – AFIS Renewal Benchmark Testing, Section 3.4 paragraphs 4 and 5).

1.2 DOCUMENT ORGANIZATION

1. A high level description of the benchmark testing process is described followed by benchmark test phases. The remaining portion of the document describes the overall score and procedures including the equations used to evaluate the benchmark testing process.

1.3 SCOPE

1. The RCMP will perform benchmark tests on the COTS technology proposed by those Bidders who have satisfactorily passed Stages 1 and 2 of the AFIS Renewal RFP evaluation process. In the context of the AFIS Renewal Evaluation Plan and Criteria, a benchmark test is defined as a test performed in an optimized, controlled environment to show a system's COTS functional capabilities (operational, user interface, and administrative), performance (reliability, accuracy, relative reliability and speed) and value of best practices relative to resource costs. Results from evaluations conducted in this stage will form part of the overall technical score.

2. The benchmark system will consist of a suite of the Bidder's COTS AFIS technology that is representative of the components offered in the respective Bidder's proposal for the RTID AFIS project. The benchmark system is not expected to be a system with all of the functionality described in the Statement of Work (SOW); however, it is expected:

   a. To satisfy most of the mandatory requirements (e.g. AFIS production performance does not need to be satisfied in the benchmark test) to be included in the first release of the AFIS renewal (i.e. implementation Stage 1);

   b. To have a VSS that supports the production performance speed for one to one matches without a requirement for AFIS to interface with VSS;

   c. The benchmark system does not need to reflect the throughput or response times of the proposed AFIS renewal solution, but it must permit the measurement of the elapsed time required to manually edit tenprint, palm and finger latent images and their feature sets; and

d. The benchmark system must also include a demonstration of the LCMC and FRC.

3. The Bidder's personnel will load the background files on their AFIS system using the Bidder's documented Best Practices between the time the files are distributed and the beginning of the Benchmark. This period will be 30 working days. All background data sets must be converted in the Continental USA or Canada.

4. The background data sets will be provided in NIST packets and TIFF images.

5. The Bidder's personnel will run the benchmark system during the execution of the Benchmark, to include reading electronic files, scanning forms, performing QA, marking and/or editing features (e.g. minutia) on finger or palm images, and all other test steps appropriate to each phase.

6. Only the Bidder's personnel directly involved in the execution of the benchmark test and the Bidder's Program Manager will be allowed in the benchmark test execution area.

7. The Bidder's benchmark system is intended to meet three objectives:

a. Be a demonstration model that will permit the evaluation of the system's basic COTS functionality and human machine (user-experience) interactions;

b. Measure the performance (reliability, accuracy, relative reliability and speed) of the system in the Lights Out and Best Practice modes;

c. Measure the difference in performance between the Lights Out and Best Practice modes and rate it in terms of any increased manual labor requirements (i.e. staff hours and skill levels).

1.4 BENCHMARK PHASES

1. The benchmark tests will be run in three (3) phases at each location visited. The three (3) Phases are outlined below.

Phase 1 Operational Tests	A series of tests designed to demonstrate and evaluate capability to perform general AFIS operational and administrative functionality such as: ANSI/NIST compliance, repository management, user privilege management and user interface evaluation. As well as to verify the mandatory requirements. Two of the demonstrations will be rated for ease of use and other user-experience factors.
Phase 2 Basic Performance Tests	A series of tests designed to measure matcher performance (i.e., Reliability, Accuracy, and Relative Reliability) with no manual intervention. As well as, performance, including speed, for one to one matching for VSS.
Phase 3 Best Practices Performance Tests	The tests from Phase 2 will be rerun using all Bidder recommended best practices such as image and minutiae editing. In addition, Phase 3 will involve measuring and evaluating (against accuracy and reliability, etc.) the amount and skill level of any manual labor required to perform the recommended best practices on ten print tests.

Table 1 : Benchmark Phases

2. Table 2 : Phase 1 Test Groupings, below, shows the tests to be performed during Phase 1 of the Benchmark. The Bidder will follow the sequence defined in this document so that all Bidders can be evaluated on the same test flows. Table 3 : Phase 2 & 3 Tests, below, shows the tests to be performed during Phases 2 & 3 of the Benchmark.

3. The Phase 1 Tests will consist of numerous demonstrations of functionality and an evaluation of the human-machine (user-experience) interface. The Phase 1 Test Groupings are depicted in the following table. Each Phase 1 test will be rated Pass or Fail. The human-machine evaluation, on-going during Test Groups B & C of Phase 1 will be rated in accordance with the scoring described later in this document.

Test Group #	Test Group Activities
A	Create and modify user and administrator accounts then attempt to use old and new accounts. Set and adjust screen saver times and other similar features.
B	Scan and search tenprint records (best practices) to include verification of candidates, verification screen-printing, and fingerprint card / form printing; create ANSI NIST records (see glossary) for storage with appropriate compression. Add tenprint records to repository. Use Direct Scan capability to scan tenprint records with palms and add them to the repository for searching in Group C.
C	Scan and search finger and palm latent records (best practices) to include verification of candidates, verification screen-printing, and fingerprint card / form printing as well as results. Print match reports for files processed.
D	Dump ANSI/NIST records from Test Group B to USB external drive for independent verification of ANSI/NIST records by RCMP using Aware NIST Pack software.
E	Generate logs and reports for the period covered to that point since the start of Phase 1 (provide electronic output that can be printed as required)

Table 2 : Phase 1 Test Groupings

4. The Phase 2 and Phase 3 Tests are depicted in the following table with the number of search and background file records. Test performance will be measured in terms of Reliability, Accuracy, Relative Reliability and speed. The Phase 2 and Phase 3 performance will be compared and related to the scope of the skilled labor required for the Best Practices performance as part of the overall Phase 3 score.

Test	Test Name	Search File	Background File
A	Tenprint to tenprint	2,050 cards	285,572 cards
B	Tenprint to ULF	2,050 cards	2,000 impressions
C	Tenprint ID Flats to tenprint	500 cards	285,572 cards
D	Finger Latent to tenprint	94 impressions	285,572 cards

E	Palm Latent to known palms	50 impressions	21,404 impressions
F	Tenprint ID Flats to tenprint ID Flats (1 to 1)	30 impressions	200 impressions

Table 3 : Phase 2 & 3 Tests

5. Tests A & B within a Phase (for phases 2 and 3) will be run from a common input process as though they were operational prints being submitted for both a tenprint search and a tenprint to unsolved latent search. Tests A & B will be followed by Test C. The Latent tests for each phase (Tests D & E) will be run in parallel with the tenprint tests (A, B, & C). Test F will be run to test VSS capabilities as part of Phase 2 testing.

1.5   OVERVIEW OF BENCHMARK WEEK

1. The Benchmark Evaluation Team will arrive at the Benchmark site on Monday morning at approximately 8:30 AM local time. The agenda for the week will be as follows:

   a. Introductions

   b. Review of Bidder's detailed schedule and facility

      i. To include the identification of all locations and equipment planned for use in the Benchmark. (All RCMP team members).

   c. Parallel start-up activities: (no more than 2 hours)

      i. Audit of Benchmark Suite - Compare the actual hardware and software suite to that identified by the Bidder in their proposal (1 to 2 RCMP team members will audit).

      ii. Orientation Training of the RCMP Benchmark Evaluation Team (4 to 6 RCMP team members).

   d. Phase 1 tests – Benchmark clock starts at this point.

      i. Start-up system

      ii. Observation of Tests in Phase 1 – Pass / Fail

      iii. Observation of human-machine (user-experience) interface – rated

      iv. Verification at the end of Phase 1 that records added to the background files have been deleted.

   e. Phase 2 tests - Lights Out from the submittal to the search engine up to and including the unverified search results:

      i. Collection of search reports to support the later calculation of Reliability, Relative Reliability, Accuracy and Speed.

         • Reports for first twenty-five (25) searches from each test - on paper and electronic.

         • Reports for all searches on a USB external drive in an Excel spreadsheet provided by the RCMP.

ii.   Verification at the end of Phase 2 that records added to the background files (none are anticipated) have been deleted.

b.   Phase 3 Tests – Best Practices from the submittal to the search engine up to and including the unverified search results:

i.   Collection of search reports to support the later calculation of Reliability, Relative Reliability, and Accuracy.

- Reports for first twenty-five (25) searches from each test - on paper and electronic.

- Reports for all searches on a USB external drive in an Excel spreadsheet provided by the RCMP.

ii.   Collection of metrics on manual labor time and skill levels to support later calculations of value.

c.   Wrap-up – Benchmark Clock stops when this set of tasks is completed or time runs out.

i.   Erasure of all background and search files and certification of the erasure by an authorized Officer.

ii.   Final USB external drive / report generation, if required.

iii.   Print out of logs to ensure all RCMP files and records as well as derived material (e.g., feature sets) have been deleted.

iv.   All sample, test and background data including original USB external drives to be returned. Certification by an authorized Officer that no copies remain with the Bidder, its employees, subcontractors, teammates, or any third party.

2.   Starting on the second day, each day will start with a review of the previous day's effort, the amount of elapsed Benchmark Time, and a status check on the schedule for the current day. The time in these meetings will not be counted against the Benchmark Time. The RCMP Benchmark Evaluation Team will record of start and stop times as well as determine the elapsed time. If the RCMP team chooses to stop the tests at any time for a caucus or any other reason, that time will also not be counted against the Benchmark Time. Testing will continue through lunch unless the Bidder elects, in its proposal, to have an interruption of testing during a lunch break. In that case the RCMP reserves the right to take appropriate steps to ensure no work continues unsupervised.

2. PHASE 1 OPERATIONAL TESTS

2.1 OBJECTIVES

1. The objective of the Phase 1 Tests will be to:

   a. Evaluate the performance of the COTS AFIS operational and administrative capabilities.

   b. Familiarize the RTID team with the AFIS system and its Human-Machine Interfaces and to facilitate understanding of the processes to be evaluated in Phases 2 and 3.

   c. Rate the Human-Machine Interface ease of use and the observed user experience.

2.2 PHASE 1 TEST GROUPS

1. The specific functionalities, how they are to be tested, and the individual Pass / Fail criteria for each are provided in Table 4 : Phase 1 Test Outline. The Bidders will organize the Phase 1 tests into a series of appropriate scenarios as outlined in Table 2 : Phase 1 Test Groupings.

2. In every case the Bidder will provide, in their Benchmark procedures, explicit references to the Phase 1 tests listed in Table 4 : Phase 1 Test Outline as that Table is what each Benchmark will be evaluated against.

2.3 PHASE 1 EVALUATION CRITERIA

1. **Pass / Fail Tests:** Each of the functionalities (operational or administrative) listed in Table 4 : Phase 1 Test Outline will be evaluated as an acceptable (Passed) or an unacceptable (Failed) demonstration of the functionality based on the criteria identified in Table 4 : Phase 1 Test Outline.

2. **Note:** other Performance metrics (e.g. Accuracy) are not to be considered in Phase 1.

3. **Human-Machine (User-Experience) Interface Evaluation:** For each of the Test Groups B and C, the *user experience* will be evaluated for ease and simplicity of user interface and number of steps required. Each of these two (2) test groups will be rated in accordance with the criteria listed in Table 5 : User Interface Evaluation Criteria.

2.4 PHASE 1 FUNCTION AND PASS / FAIL CRITERIA

1. Phase 1 tests will be grouped into related activity sets. Within each set there will be multiple tests run. The following table provides the guidance necessary for the Bidders, who are invited to perform a benchmark, to develop appropriate test procedures. All Bidders will submit, as part of their proposal, a Benchmark Plan that addresses these Groups and Tests.

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
A	1	Demonstration of startup of system from a written set of procedures.	Bidder will start all equipment from a set of written procedures provided with their test procedures. When the Benchmark suite is restarted, Bidder will synchronize all devices with clocks to a common time which must be done to an NTP server automatically. Bidder will check times for results of synchronization.	The system starts according to the procedures. Times are synchronized.
A	2	Demonstration of User Account Management Capabilities with the required granularity.	Bidder will add, change, delete persons, and their permissions. Bidder will attempt to use appropriate and inappropriate functions. Real names (e.g., Roger Tester) rather than names such as "Tester_1" should be used. Accounts will be used in Test Group A item 3.	Records are generated and a printout of all changes can be made. Reports are accurate representations of changes and times. Verify that no other users can perform these security functions. Verify that accounts work correctly.
A	3	Demonstration of operational stability when workstations are turned on and off, and other routine changes are made.	Bidder will perform shutdowns and restarts of tenprint and latent workstations with two (2) different user log-ins and functionality such as logging on as a sys admin with previously created accounts, logging out, shutting down and restarting the workstation, and having the new user log in and run any AFIS transaction (e.g., check verification queue for work) and then the Bidder logs back in and prints a transaction log before logging out.	All workstation shutdowns and restarts are accomplished with no anomalous system behavior. Logs accurately capture work performed and persons performing it as well as times and workstations used. Verify that accounts work correctly.
A	4	System Administrator ability to set time to screen saver initiation.	Bidder will change time outs [i.e., time to screen saver lock out] four (4) times to very short periods and the system left untouched to demonstrate that the screen saver is initiated at the appropriate time. Then the Bidder will use the	Screen saver locks up the workstations and requires a password or fingerprint to free it up. Time resets are controllable and accurate to within two (2) seconds of the value

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
			password and fingerprint (i.e. two factor authentication) to unlock the screen. Next pass, the Bidder will preclude a timeout by taking any action and seeing that it resets the clock.	entered. All input devices can be used to reset screen saver timeout (e.g., use of keyboard, a mouse, or touch screen is sufficient to avoid a timeout situation.)
B	5	Ability to scan, capture, format, and compress all 14 fingerprint images (WSQ 15:1) to include plain impressions and creating a TPRI and TPSEI NPS-NIST ICD compliant format.	Bidder will scan five tenprint cards (C-216)/ forms at 500 ppi and then each shall be segmented to capture 14 individual images correctly and compress them at 15:1 WSQ and direct file / direct scan to database. Data is to be saved in the database and on USB external drive (new USB external drive provided by the Bidder) for use in Group D (D17). The Bidder will enter the mandatory Type 2 demographic and biographic data and the text and images are to be properly labeled and stored as Type 1, 2, and 4 records on the USB external drive.	Quality of individual images must reflect such things as reasonable centering, minimum overlap, removal of "finger box" lines that do not intersect the fingerprint ridges, and the correct size. Verify in D17 that the NIST packet can be read by the Aware software, that the Type 1 and Type 2 records are correct and that the nominal compression ratio is WSQ 15:1.
B	6	Ability to receive, process and format palm print images to include creating ANSI/NIST Type 15 record compressed with WSQ 15:1.	Bidder will receive five palm print cards / forms (up to three images per submission, all from the same hand) at 500 ppi and compress them with WSQ at 15:1. The Bidder will enter the mandatory Type 2 demographic and biographic data and the text and images are to be properly labeled and stored as Type 1, 2, and 15 ANSI NIST records and written to a USB external drive (new USB external drive provided by the Bidder) for use in Group D (D18).	The palm records are correctly captured (upper, lower and writer's palms). Quality of individual images must reflect such things as reasonable centering and effective minutia. Verify in D18 that the NIST packet can be read by the Aware software, that the Type 1 and Type 2 records are correct and that the nominal compression ratio is WSQ 15:1.
B	7	Ability to perform lights out tenprint processing for five cards.	Bidder will selectively configure its AFIS to process 5 tenprint searches from electronic submissions (TPRI) through to match report (candidate list) without any manual intervention and will send responses to a verification station queue and then the Bidder will manually send the search results to a printer.	System can be selectively configured to run in a lights out mode. System can also be configured with threshold setting for each region identifying System Declared Hits and grey area potential hits. Searches are successfully

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
				run [i.e., executed with no error messages and no anomalous system behavior] in lights out mode and responses are sent to a verification queue where they will be printed. The printed record accurately reflects the transaction results.
B	8	Ability to perform automated QA processes on submitted tenprints to include automated sequence checks, missing and amputations (AMP) detection, etc.	Bidder will submit RCMP provided tenprints then automatically (i.e., no manual intervention) evaluate each rolled image and automatically evaluate each rolled finger image for proper sequencing and missing fingers. Print the reports.	System automatically and correctly assesses and reports all sequence and missing fingers from the 20 tenprint submissions. Verify the results from the reports
B	9	Ability to perform on-screen QA review and edits on submitted tenprints based on automatically detected and operator-observed problems, from any workstation on the network.	Bidder will perform manual QA work on the same 20 tenprints from the previous test using three different QA workstations in parallel. Effort is to include centering, orienting, adjusting contrast and brightness, proper sequencing of tenprint records and substitution of plain images for rolled impressions as well as rolled to rolled to compensate for quality and sequence issues.	Operator can successfully perform basic QA functions [such as centering, sequence corrections, and orientation] on the tenprint images based on system recommendations or operator-initiated changes and save the results independent of which workstation was used. Individual changes can be undone, prior to any other changes being made, with a single command from a pull down menu or other simple means of user interface such as an undo button – all without resubmitting any images.
B	10	Ability to support the storage and searching of multiple fingerprint sets for an individual person (CAR Y & IMM).	Bidder will submit 12 sets of fingerprints [12 tenprint and 12 palm] with image sets for 4 different people. System will perform an AFIS search with each search print added to the repository, independent of it being a hit or not.	Correctly add the records to the appropriate person. Verify that the second and subsequent searches hit on the correct subject ID and that they are stored under that number in the

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
		Ability to delete tenprint records with an audit trail created and maintained in a fashion that permits review and selective printing.	Bidder will delete the initial fingerprint record for two such persons and verify that the person is still in the repository. The Bidder will print a repository maintenance report to reflect all additions and the deletions.	repository. Deletion of one of multiple records for an individual does not delete the individual completely. Record deletion audit trail is generated and the system administrator can selectively view using system reporting and print a report of records deleted by date, by user, and by Subject ID number; as well as view/print any other activity associated with the processing. The Subject IDs for the cases where a second record was deleted must still be in the active repository.
B	11	Ability to print fingerprint cards or forms (any fingerprint forms or cards) with an accurate representation of the original images in the correct positions along with provided Type 2 data field.	Bidder will be provided with four ID #s from the tenprint background set. The Bidder will print these fingerprint forms / cards directly from the AFIS workstation.	Tenprint forms can be printed from an application and an accurate representation of the original and all type 2 data provided. This requires that the form and printed material are aligned properly and the fingerprints are of EBTS Appendix F print quality representative of the original. Compare the printed copy to the original.
B	12	Ability to manage multiple tenprint priorities to include changing priorities on multiple transactions.	Bidder will submit 50 tenprint transactions into the AFIS where 100% are Priority 2 and then selectively change the last ten to Priority 1 without stopping the search engine itself in order to change the search priority. Search the transactions and print a log of search order to include priority.	Priority management processes permit priorities to be changed and subsequent processing is based on the revised priority list.

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
C	13	Ability to capture latent images and extract features.	Bidder will receive six finger-latent images at 1,000 ppi and the images are to be properly saved and minutiae extracted to create a feature set that will be displayed on screen. Bidder will submit each feature set for search through the AFIS; three with and three without manual intervention.	Results in 1,000 ppi images and the AFIS system supports minutiae extraction and feature set creation and display. Searches can be launched with or without editing by a latent examiner.
C	14	Ability to perform on-screen QA review and edits on scanned latents based on automatically detected and operator-observed problems, from any workstation on the network.	Bidder will receive and adjust orientation for three latent images using a single workstation. The bidder will perform manual QA work on all three latent images using three different workstations. At least one change at each workstation will be "undone" prior to the file being saved. The Bidder will reopen each saved file to see if the undone change was saved or not.	User can successfully perform basic QA functions, such as centering, and orientation, on latents images based on system recommendations or user-initiated changes and save the results independent of which workstation was used. Individual changes can be undone, prior to being saved, with a single command from a pull down menu or other simple means of user interface such as an undo button – all without rescanning any images.
C	15	Demonstration of latent image display and ability to manipulate to include zoom, B/W reversal, image flip, contrast adjustment, magnification, rotation, and brightness adjustment.	Bidder will manipulate five latent images on screen in each of the following processes: QA, manual encoding, and search results verification. The Bidder's Latent workstation will capture and track all of the actions.	Image manipulation audit trail is automatically generated, changes can be undone, and the Latent workstation demonstrates a set of image manipulation capabilities that satisfy the requirements of this AFIS Renewal RFP in terms of user interface and functionality such as zoom, B/W reversal, image flip, contrast adjustment, magnification, rotation, minutia editing and brightness adjustment.
C	16	Ability to fetch from the ULF, modify and then over-write original or create a new record in the unsolved	A total of 2 records will be selected for this test. Bidder will fetch an RCMP identified record, modify it and save it to the ULF under the original number. Bidder will fetch an RCMP identified record, modify it and save it to the	Verify by fetching both records from the ULF and that the modifications have been retained as entered.

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
		latent file (ULF).	ULF under a new number.	
D	17	Ability to direct file / direct scan searchable records. RCMP to read and evaluate data from Test Group B.	Bidder will provide five (5) tenprint search records on a USB external drive created in Group B (B5) tests. The NIST packet will be opened using the Aware software on the RCMP laptop.	All fields in samples can successfully be parsed and the original input data is in the correct fields. WSQ compression rate must be between 14 to 1 and 16 to 1, on average across all records in the three samples.
D	18	Ability to create Type 15 NIST records. RCMP to read and evaluate data from Test Group B.	Bidder will provide five (5) ANSI NIST palm search records on a USB external drive created in Group B (B6) tests. The NIST packet will be opened using the Aware software on the RCMP laptop.	All fields in samples can successfully be parsed and the original input data is in the correct fields. WSQ compression rate must be between 14 to 1 and 16 to 1, on average across all records in the three samples.
E	19	Ability to track, log, and report on all transactions performed, user account and database modifications, and any error conditions.	Bidder will print transaction and error logs after all Phase 1 Group A through Group C tests are completed. Output should be on paper and USB external drive. Transaction activities/events can be viewed and printed.	All observed transactions, changes to databases and user accounts, and any errors are in the log and can be easily viewed, deciphered and interpreted by the Benchmark Evaluation Team.
E	20	Demonstration of report generation capabilities for system operations and administrative functions.	Bidder will selectively print system utilization reports and repository size for at least three time periods within the elapsed time of Phase 1 Group A through C. This is to be done after Phase 1 Group C. Output is to be on paper and USB external drive in a PDF format.	Reports capture what transpired and are selectable and controllable at least to the extent that start and stop times and categories of items to be reported on are selectable by the user.
E	21	Demonstration of report on repository (tenprint, palm print, and ULF) capacity and percentage of capacity used to include a report on all records added in Phase 1.	Bidder will generate and print repository status report at the end of Phase 1 Group E. This is to cover images, features, and any other space-constrained data sets. Output is to be on paper and USB external drive in PDF format. Output to include a repository maintenance report to reflect all activity in Phase 1.	Reports selectable by repository (i.e., tenprint, ULF, and Palms) and cover repository size, available space, and percentage of total space used. Record audit trail is generated and the system administrator can selectively generate and print an accurate report of activity by date, by user, and by Record ID number.

Test Group	Item	Function	Test Approach	Pass / Fail Criteria
F	22	Demonstration of 1 to 1 matching UI for VSS capabilities on ID Flats.	Bidder will receive and process up to five (5) ID Flats (1 to 10 images) and match to background data as expected. Generate and print of activity and results.	Verify log file processing and verify UI provides capabilities that satisfy the requirements for VSS processing.

Table 4 : Phase 1 Test Outline

2.5   BENCHMARKING USER INTERFACE

1. For the Phase 1 Tests the following criteria will be used to evaluate the user interfaces. A single score between 0 and 100 will be assigned for Group B and another for Group C. The criteria and related scores will be based on the following:

Rating	User Interface Evaluation Criteria
100	High degree of automation and intuitive interfaces: simple (typically 1 click) selections with defaults. Few steps to complete a process – with no need to remember a value from one screen to enter it or select it on another (time tags, user name, and other strong clues are available when a selection has to be from a transaction list.) Few screens required to complete a serial process such as capture and prepare a latent image for search.
50	Easy-to-understand and intuitive interface but more user-intensive (multiple clicks or layers of options and pull downs). Need to recall or write down a value of transaction number from a previous screen. Multiple different screens to complete a serial task.
0	Hard-to-understand interface (location of mouse-able commands is counter intuitive) or an excessive number of steps are required for simple tasks. Data elements from previous steps are not always available for selection but rather require entry or prior knowledge to select them. Or failed to perform the function defined

Table 5 : User Interface Evaluation Criteria

2.6 PHASE 1 SCORING MATRIX

1. The following scoring matrix will be used to calculate the sub-scores for Phase 1.

Phase 1 Scoring Matrix			
Test	**Score**	**Calculated Value**	**Minimum Threshold**
All Phase 1 Functionalities	N/A	N/A	All Tests passed
User Interface	Score based on Table 5 : User Interface Evaluation Criteria	CV_1= Average score across Test Group B CV_2 = Average score across Test Group C	N/A

Table 6 : Phase 1 Scoring Matrix

3.   PHASE 2 BASIC PERFORMANCE TESTS

3.1   OBJECTIVES

1. Phase 2 tests are designed to measure the feature extraction and matcher performance. Test results will be evaluated in terms of reliability (S_1), accuracy (S_2) and relative reliability (RS_1) at a given selectivity, of the searches, absent manual intervention. Speed will also be measured for 1 to 1 matching required for VSS. For purposes of Phase 2 tests there will be no manual intervention associated with manipulation of either a forensic image (e.g., image centering) or a mathematical image representation (e.g., minutiae editing), nor manual substitution for low quality or out of sequence fingers. See the Glossary of terms in 5.8 for the full definition of "no manual intervention".

3.2   PHASE 2 TEST GROUPS

1. The Bidders will organize the Phase 2 tests into a series of appropriate scenarios, such as:

   a.   Tenprint receipt, storage and processing to include searches against background files. Search scenarios will include:

   i.   Tenprint searches (rolled and plain impressions) against the tenprint background (rolled and plain impressions). Hits are to be declared without benefit of manual verification and the top five candidates must be imported into the RCMP provided Excel spreadsheet.

   ii.   Tenprint searches (same records as above – from the same file ingest step) against the Unsolved Latent File (ULF). Hits are to be declared without benefit of manual verification and the top ten candidates must be imported into the RCMP Excel spreadsheet.

   iii.   ID Flat tenprint searches against the tenprint background file (rolled/plain and/or ID Flat impressions). Hits are to be declared without benefit of manual verification and the top five candidates must be imported into the RCMP Excel spreadsheet.

   b.   Latent receipt, storage and searching scenarios to include:

   i.   Latent searches against the tenprint background file. Hits are to be declared without benefit of manual verification and top ten candidates must be imported into the RCMP Excel spreadsheet.

   ii.   Palm latent searches against the palm print background file. Hits are to be declared without benefit of manual verification and top ten candidates must be imported into the RCMP Excel spreadsheet.

   c.   ID Flat 1 to 1 receipt, storage and matching in a VSS configured subsystem to include:

   i.   ID Flat 1 to 1 matches against the rolled/plain and/or ID Flat tenprint impressions in the background file. Hits are to be declared without benefit of manual verification and match/no match recorded as well as speed of matching; and results must be imported into the RCMP Excel spreadsheet.

3.3   PHASE 2 EVALUATION CRITERIA

1. Each of the functions listed in Table 7 : Phase 2 Performance Evaluation will be rated numerically based on the criteria identified in Table 6. In each case, the results of the rating will be scored in accordance with the matrices described in Table 7 to calculate the overall Phase 2 score. The following subsections provide directions on how to calculate the associated performance metrics.

	Phase 2 Performance Evaluation		
Test #	**Function Tested**	**Scored Areas**	**Rating Process**
A	Tenprint card searches	• Adjusted Reliability AS_1 • Accuracy S_2	Calculate Adjusted Reliability using a Selectivity setting of 1 and Acceptable Level of 98% and calculate Accuracy.
B	Tenprint card to ULF	• Adjusted Relative Reliability ARS_1	Calculate Adjusted Relative Reliability using a Selectivity setting of 10.
C	Tenprint ID Flat searches	• Adjusted Reliability AS_1 • Accuracy S_2	Calculate Adjusted Reliability using a Selectivity setting of 1 and Acceptable Level of 98% and calculate Accuracy.
D	Finger Latents to TP	• Adjusted Relative Reliability ARS_1	Calculate Adjusted Relative Reliability using a Selectivity setting of 10.
E	Palm latents to Palm	• Adjusted Relative Reliability ARS_1	Calculate Adjusted Relative Reliability using a Selectivity setting of 10.
F	Tenprint ID Flat searches (1 to 1)	• Adjusted Reliability AS_1 • Accuracy S_2	Calculate Adjusted Reliability using a Selectivity setting of 1 and Acceptable Level of 98% and calculate Accuracy.

Table 7 : Phase 2 Performance Evaluation

2. Note: An Acceptable Level is not to be confused with a Minimum Threshold in Phase 3. An Acceptable Level does not imply that the Bidder will be removed from the competition if this is not achieved unless the Acceptable Level is also the same as the Minimum Threshold for a specific test.

3.4   PHASE 2 SCORING MATRIX

1. The Bidder's Phase 2 performance ratings will be calculated in accordance with the scoring matrices in Table 8 : Phase 2 Scoring Matrix. For Phase 2 there is no minimum performance threshold; however, the minimum Acceptable Level is used to determine an adjusted reliability score which is used in the overall score.

2. **Tenprint to Tenprint tests:** Reliability, Adjusted Reliability, and Accuracy will be calculated using the equations in this Plan in section 5.6. One to one matching speed will also be verified as a mandatory performance requirement for VSS.

3. **Latent Tests:** Rather than just measure latent reliability scores we will use Relative Reliability values as they spread the scores out based on the position within the candidate list for any hits. Each change of position (e.g., from candidate position 4 to candidate position 5) in the candidate list for a hit causes a difference of 10 percent in the Relative Reliability score. By way of example – the average difference of one candidate position, with a selectivity of 10, would lead to a loss of 10 points from a Reliability of 100, 8 points from a Reliability of 80 and so on). Note that Accuracy is not a consideration in latent searches since Selectivity is greater than one. Adjusted Relative Reliability will be calculated as per section 5.6 equations and calculations.

	Phase 2 Scoring Matrix			
Test #	Function Tested	Score	Sub-Weights	Calculated Value (CV)
A	Tenprint cards against TP	• Adjusted Reliability AS_1 • Accuracy S_2	• 80% • 20%	$CV_A = AS_1*.8 + S_2*.2$
B	Tenprint cards against ULF	• Adjusted Relative Reliability ARS_1	• 100%	$CV_B = ARS_1$
C	Tenprint ID Flat search against TP (rolled/plain and/or ID Flat)	• Adjusted Reliability AS_1 • Accuracy S_2	• 80% • 20%	$CV_C = AS_1*.8 + S_2*.2$
D	Latent to TP background searches	• Adjusted Relative Reliability ARS_1	• 100%	$CV_D = ARS_1$
E	Palm latent to Palm background searches	• Adjusted Relative Reliability ARS_1	• 100%	$CV_E = ARS_1$
F	Tenprint ID Flat search against TP ID Flat (1 to 1)	• Adjusted Reliability AS_1 • Accuracy S_2	• 80% • 20%	$CV_C = AS_1*.8 + S_2*.2$

Table 8 : Phase 2 Scoring Matrix

4. PHASE 3 BEST PRACTICES PERFORMANCE TESTS

4.1 OBJECTIVES

1. Phase 3 tests are designed to measure the feature extraction and matcher performance using the Bidder's best practices with manual intervention permitted. Test results will be evaluated in terms of the accuracy, reliability, and relative reliability, for a given selectivity of the searches. The Benchmark Evaluation Team will evaluate the amount of manual intervention and expertise that will be required to operate the system at the best practice levels and assign both a time and, in the case of ten print, a skill level coefficient as defined in Section 4.3, Phase 3 Evaluation Criteria. The details of the tests and scoring are provided in Table 11 : Phase 3 Performance Evaluation and Table 12 : Phase 3 Scoring Matrix.

2. Note that wherever skill levels will be assessed, it will apply to ten print tests and latent tests.

4.2 PHASE 3 TEST GROUPS

1. Phase 3 Tests will be organized into two principal groups, exactly the same as those for Phase 2. The primary difference between the Phase 3 tests and those in Phase 2 is that the Bidder will employ its best practices as stated in its proposal. In addition, in each set of tests the amount of time applied as well as the skill level required (e.g., in editing data for searches), will be measured and factored into the overall Phase 3 score.

2. Note that all hits are to be declared without the benefit of manual verification even if the vendor's Best Practice includes this step.

4.3 PHASE 3 EVALUATION CRITERIA

1. Each of the performance factors listed in Table 11 : Phase 3 Performance Evaluation will be rated numerically based on the criteria identified in the Table. Note that Accuracy is not a consideration in latent searches since Selectivity is greater than one. In each case, the results of the rating will be entered into the scoring matrix described in Table 12 : Phase 3 Scoring Matrix to calculate the overall Phase 3 score.

2. Note that in order to simulate a latent search in a database size approximating 3 Million records, the use of digit determination will not be allowed in the Phase 3 Test D (Finger Latent to Tenprint) even if it is normally used in the Best Practice of the Bidder.

3. The equations for calculating the performance metrics for Phase 3 are the same as those that are used in Phase 2 for Reliability (S_1), Accuracy (S_2), Relative Reliability (RS_1), Adjusted Reliability (AR_1) and Adjusted Relative Reliability (ARS_1). However, there will be two additional metrics applied to assess the human resources impact. The following are factors used in the equations to determine this impact.

4. **Transaction Time (TT)** is the average transaction preparation time per transaction for best practices in seconds, for the test under consideration. This time will be measured from the reading of the first electronic transaction from the USB external drive provided by the RCMP to the submittal of the last transaction to the Search queue for a block of search prints. Best practices must be performed in a "first in, first out" sequential order at each workstation. Below is an example of the Transaction Time Coefficients (**TTC**) for Tenprint and Latent, as appropriate, that will be applied in the calculation of the Bidder's score.

Tenprint		Latent	
Average Transaction Prep Time (seconds)	Transaction Time Coefficient (TTC)	Average Transaction Prep Time (seconds)	Transaction Time Coefficient (TTC)
0-18	100%	0-120	100%
19.8	99%	123	99%
21.6	98%	126	98%
23.4	97%	129	97%
25.2	96%	132	96%
27	95%	135	95%
28.8	94%	138	94%
30.6	93%	141	93%
32.4	92%	144	92%
…	…	…	…
90	60%	240	60%
…	…	…	…

Table 9 : Phase 3 Prep and Transaction Time

1. For every 1.8 seconds of Tenprint Average Transaction Prep Time above 18 seconds, the Tenprint Transaction Time Coefficient (TTC) will decrease by 1%.

2. For every three seconds of Latent Average Transaction Prep Time above 120 seconds, the Latent Transaction Time Coefficient (TTC) will decrease by 1%.

3. **Skill Level (SL)** is a rating that reflects the skill level and experience that a fingerprint examiner would require in performing Tenprint Card, Tenprint Card to ULF and Tenprint Plain searches as part of the Phase 3 Best Practices. To determine this, the Benchmark Evaluation Team will identify the tasks performed and categorize the overall rating for each test as **Low**, **Medium** or **High** impact based on the list of activities on the next page. Below is the Skill Level Coefficient (**SLC**) that will be applied in the calculation of the Bidder's score for each rating for Tenprint:

4. **A 20% penalty will be applied to the proportion of transactions requiring a high skill level during the timed portion of the test. If a Bidder uses medium skill level on some transactions and high skill level practices on other transactions then the frequency of both medium and high skill transactions will be combined and treated as high level. If high skill levels are never applied for the test in question, then the penalty will not be applied.**

5. **A 20% bonus will be applied to the proportion of transactions with lights out or low skill level practices applied during the timed portion of the test.**

Skill Level Coefficient (SLC)	
Rating	**100% + Bonus - Penalty**
Low/None	**Bonus: +20%** x Frequency
Medium	**Penalty: 0%** (if high skill level is never used) **Penalty: -20%** x Frequency (if there are one or more high skill level transactions in the timed tests)
High	**Penalty: -20%** x Frequency

Table 10 : Phase 3 Skill Level Coefficient

Phase 3 Performance Evaluation			
Test #	**Function Tested**	**Scored Areas**	**Rating Process**
A	Tenprint card searches	• Adjusted Reliability AS_1 • Accuracy S_2	Calculate Adjusted Reliability and Accuracy using a Selectivity setting of 1 and an Acceptable Level of 98. Then determine the Transaction Time Coefficient (TTC) and Skill Level Coefficient (SLC).
B	Tenprint card to ULF	• Adjusted Relative Reliability ARS_1	Calculate Adjusted Relative Reliability using a Selectivity setting of 10. Then determine the Transaction Time Coefficient (TTC) and Skill Level Coefficient (SLC).
C	Tenprint Plain searches	• Adjusted Reliability AS_1 • Accuracy S_2	Calculate Adjusted Reliability and Accuracy using a Selectivity setting of 1 and an Acceptable Level of 98%. Then determine the Transaction Time Coefficient (TTC) and Skill Level Coefficient (SLC).
D	Finger Latents to TP	• Adjusted Relative Reliability ARS_1	Calculate Adjusted Relative Reliability using a Selectivity setting of 10. Then determine the Transaction Time Coefficient (TTC).
E	Palm latent to Palm background	• Adjusted Relative Reliability ARS_1	Calculate Adjusted Reliability using a Selectivity setting of 10. Then determine the Transaction Time Coefficient (TTC).

Table 11 : Phase 3 Performance Evaluation

4.4 PHASE 3 SCORING MATRIX

1. The Bidder's performance rating will be calculated in accordance with the following Table for Phase 3.

Phase 3 Scoring Matrix				
Test #	Function Tested	Score	Sub-Weights	Calculated Value (CV)
A	Tenprint card to TP	• Adjusted Reliability AS_1 • Accuracy S_2	• 80% • 20%	$CV_A = (AS_1*TTC*SLC)*.8 + (S_2 *.2)$
B	Tenprint card to ULF	• Adjusted Relative Reliability ARS_1	• 100%	$CV_B = ARS_1*TTC*SLC$
C	Tenprint Plain to TP	• Adjusted Reliability AS_1 • Accuracy S_2	• 80% • 20%	$CV_C = (AS_1*TTC*SLC)*.8 + (S_2 *.2)$
D	Latent to TP searches	• Adjusted Relative Reliability ARS_1	• 100%	$CV_D = ARS_1*TTC$
E	Palm latent to Palm searches	• Adjusted Relative Reliability ARS_1	• 100%	$CV_E = ARS_1*TTC$

Table 12 : Phase 3 Scoring Matrix

5. OVERALL SCORE AND PROCEDURES

5.1 OVERALL BENCHMARK TEST SCORE

1. The following approach will be used to weight the individual Benchmark Phases and determine the overall Benchmark Test Score.

2. For those Bidders who successfully pass Phases 1, 2 and 3 of the benchmark, the Weighted Benchmark Score for Stage 3 will be calculated using the weights identified in Table 13 : Benchmark Weights.

3. The individual scores for each of the three Benchmark Phases will be calculated per this Plan and then weighted in accordance with the table below. Then the three Phases will be integrated using the Relative Phase Weights at the bottom of the following Table. This set of calculations will produce an overall Benchmark Score for each Bidder. Note that under the Test # column, "X" refers to the applicable phase (i.e.: either Phase 2 or Phase 3).

Benchmark Test Weightings					
Test #	Function	Phase 1	Phase 2	Phase 3
	Phase 1			
	Phase 1 Demos (Pass / Fail) Threshold = must pass all.			
	Tenprint User Ease of Use (Rated) Group B	50%		
	Latent User Ease of Use (Rated) Group C	50%		
		100%		
	Phases 2 & 3			
X–A	Tenprint cards to TP background		35%	25%
X–B	Tenprint cards to finger ULF		15%	20%
X–C	Tenprint plain to TP background		15%	15%
X–D	Latent search records to TP background		20%	30%
X–E	Palm latent search records to Palm background		10%	10%
X–F	Tenprint ID Flat searches (1 to 1) background		5%	
	Phase Totals:	**100%**	**100%**	**100%**
	Relative Phase Weights	**20%**	**40%**	**40%**

Table 13 : Benchmark Weights

5.2 **BENCHMARK PROCEDURES**

5.2.1 **BENCHMARK LOCATION AND DURATION**

1. Each Bidder invited to perform a Benchmark will update the detailed Benchmark Test Procedures submitted with its Proposal and submit to the PWGSC Contracting Authority a minimum of fifteen (15) working days prior to its scheduled Benchmark test for review and approval.

2. Bidders will propose a Continental USA or Canadian vendor facility, where the RCMP evaluation team will have access to its benchmark system for the purpose of monitoring / conducting the benchmark tests. The location will be identified in the Bidder's written proposal. The Bidder will provide the RCMP Benchmark Evaluation team with a private conference room with high-speed Internet access for at least 3 simultaneous connections as well as a printer for the evaluation team's use.

3. The benchmark tests will be conducted in less than one workweek (consisting of no more than four sequential nine-hour days) for each Bidder.

5.2.2 **BENCHMARK CONFIGURATION DATA**

1. The Bidder's Benchmark configuration will be as follows:

   a. Hardware configuration for the benchmark system is to include anticipated average match rates in comparisons-per-second both with and without factoring in any filters (for example, sex, pattern, age, etc.) for tenprints, latent fingerprints, and latent palms. This must be the same hardware baseline as proposed for the AFIS renewal solution on a smaller scale with the expectation that the Bidder's proposed solution includes more than just software (i.e. hardware, special purpose boards). If the proposed solution is software-only, then the benchmark configuration is expected to be the same as the proposed solution. For example, if the proposed VSS renewal solution, which requires slower matching speed, is a software only solution then the benchmark configuration is expected to the same as the proposed solution. If not, the Bidder must explain why it is not part of their proposal. As well, if the proposal is based on a software-only solution the benchmark system cannot run on special purpose hardware/boards or alternatively, if the proposal is based on special hardware/boards, the benchmark solution cannot be software only.

   b. Software configuration information (version #, release #) for the benchmark system, it must be the same or nearly identical software product(s) version number (i.e. it does not have to be the anticipated delivery release) as proposed for the AFIS renewal.

2. The Benchmark suite will support Laser and color laser printing of all FP Application screens, directly from a button or pull-down menu in the fingerprint application, without the use of any screen capture utilities or operating system screen print functions. These printers must be capable of printing fingerprint cards and forms according to the FBI Appendix F quality and be representative of the original form / card.

3. ***The RCMP reserves the right to review all of the proposed benchmark configurations and the other relevant parts of each written proposal to determine if the benchmark system conforms to the technology base proposed for the AFIS functionality of the RTID project.***

5.2.3   BENCHMARK EVALUATION TEAM

1. The Benchmark Evaluation Team will consist of approximately six (6) representatives from the RCMP and its contractors and PWGSC. The PWGSC procurement officer will be responsible for ensuring that the benchmark is conducted in accordance with the procedures specified in the RFP, interpreting the RFP where clarification is required, and acting as the Benchmark spokesperson in consultation with the RCMP Benchmark Evaluation Team Leader. The PWGSC procurement officer will also be responsible for collecting all of the benchmark test results and providing a copy for the record. The RCMP Benchmark Evaluation Team Leader will be responsible for coordinating the evaluation team's activities and representing the RCMP interests.

5.2.4   TEST DATABASES

1. The same background and search data will be used at each benchmark location. Thirty working days in advance of each Benchmark Test, the RCMP will provide the respective Bidder with USB external drives containing the background file data.

BACKGROUND FILE DESCRIPTIONS		
File #	**Forensic Data Type**	**Format / Media**
1	Tenprint cards, Tenprint ID Flat cards	ANSI-NIST packet on USB external drives
2	Latent Finger prints – ULF file	TIFF on USB external drives
3	Palm Prints	TIFF on USB external drives

Table 14 : Background File Descriptions

2. If the Bidder enters its own identifiers in each record, it will also retain the RCMP identifiers for use in the search reports.

SEARCH FILE DESCRIPTIONS		
File #	**Forensic Data Type**	**Format / Media**
4	Tenprint cards	ANSI-NIST USB external drive
5	Tenprint ID Flats	ANSI-NIST USB external drive
6	Latent Fingerprints	TIFF USB external drive
7	Latent Palm Prints	TIFF USB external drive

Table 15 : Search File Descriptions

3. At the time of the tests, the Benchmark Evaluation Team will provide the four (4) search files (Files 4, 5, 6, & 7) identified in Table 15 : Search File Descriptions. The Bidder will demonstrate that all search data and related mathematical representations are erased from each benchmark as soon as the full Benchmark has been run.

5.3   BENCHMARKING PROCEDURES

5.3.1   GENERAL

1. The minimal RCMP benchmark procedures are described in the following sections. The Bidder will be responsible for updating these procedures to reflect their configurations and capabilities and to reflect the scenarios they choose to execute for each Phase. The Bidder procedures will be of sufficient detail that the RCMP Benchmark Evaluation Team can follow them step by step and observe that the proper input was provided and that the expected results were or were not achieved. The Bidder will provide a mapping between the steps and tests in this Benchmark Plan and the Bidder's detailed Benchmark Plan provided with its proposal.

2. The Bidder will operate the benchmark equipment for all Phases. Prior to the start of Phase 1 the Bidder will provide an orientation overview of the benchmark facility and equipment. They will also provide the RCMP personnel with a minimal set of training on the use of the individual workstations as part of the orientation.

3. The three Phases will be run in sequential order, with the prior Phase completed before any subsequent Phase starts. At the end of each phase, the test results shall be printed, annotated by the Bidder as to the date, site, Bidder, phase, etc. and signed by the Bidder and the RCMP Evaluation Team Leader. The RCMP will consider a print out to have not been printed if it is not in the print queue at the end of the related test.

4. If, during any phase of the test, the benchmark system, other than workstations, needs to be restarted then any incomplete test(s) associated with that phase will be rerun. All of the data from any unfinished tests will be set aside and will not be considered in the Benchmark Evaluation. If any system or subsystem has to be reconfigured, other than through a changing of settings an application screen, then that entire Phase will be rerun.

5. If a software configuration change has to be made that includes new software being loaded then the Phase will have to be restarted. If this happens during Phase 3 then Phase 2 and Phase 3 will both have to be restarted. All of the data, from the initial running of any Phases that have to be restarted, will be set aside and will not be considered in the Benchmark Evaluation.

6. No additional time will be allocated to the Bidder should a restart be required.

5.3.2   TEST STARTUP AND AUDIT

1. The benchmark system will be brought down to a fully "turned off" condition, both logically and physically, prior to the start of Phase 1. This means that all hardware to be used in the Benchmark will be physically turned off when the Benchmark Evaluation Team arrives to include printers, workstations, servers, disk arrays, matchers, networks, communications gear, rack-based cabinets, scanners, monitors, cameras, and RAID.

2. Prior to the start of Phase 1 tests, the Benchmark Evaluation Team will audit the hardware/software configuration using a checklist provided by the Bidder in its written Technical Proposal. This audit is intended to confirm that the Benchmark location uses the same technology as proposed by the Bidder.

3. To facilitate the audit the Bidder will clearly identify and label all hardware, workstations, and cables with the same titles used in the checklist. The excessive use of proprietary abbreviations is to be avoided; wherever possible names of components should be spelled out.

4. The Bidder's operator will start the benchmark system up as soon as the audit is finished. If the audit shows that the configuration is not based on the proposed technology, then the RCMP evaluation team leader will document the differences and a determination as to Pass or Fail for the benchmark will be determined. The RCMP and PWGSC will make the determination while the evaluation team continues with the benchmark in question.

5. The Benchmark Evaluation Team will audit the digital repository prior to the first test in each Phase, to confirm that the RCMP test data has been loaded and is the only repository file for the benchmark test. The Benchmark Evaluation Team will complete a copy of the checklist. It will indicate repository confirmation or discrepancies. The checklist will be filed with the Benchmark Evaluation report.

6. Failure of any of these audits may be cause for failure of the Benchmark Test Stage. The RCMP along with the PWGSC will be the determining authority.

5.3.3   TEST OUTPUT DATA

1. The benchmark system will produce the following information for each search both in a printed record and in an ASCII data file (i.e. importable into Excel) to be written to a removable storage device for subsequent evaluation by the RCMP. The USB external drive record will be a compilation of all the searches in a particular test rather than separate data sets on a case-by-case basis.

2. A separate file will be created on the removable storage media for each of the tests and the file name will be encoded to reflect the Phase (2 or 3) and Test numbers (e.g. 2B). The files will contain the data elements as defined in the sample matrix (i.e. spreadsheet) to be provided by the RCMP.

3. If one or more of these parameters are not automated then the Bidder will keep detailed records on paper and provide them with the report generated at the end of each Phase.

5.4 **TEST WRAP UP**

5.4.1   **GENERAL**

1. After each Phase, the operator will delete all of the search data, the associated results and the log files from the benchmark system.   A certification, signed by the Bidder's authorized officer, will be submitted to the Benchmark Evaluation Team Leader to attest to the fact that all RTID benchmark data has been removed from the all systems, on-line and backup storage, RAM, work queues, etc. All Sample, Test and Background data and USB external drives are to be returned and the Bidder must certify that no other copies exist. Furthermore, after the end of each Phase the Team will note any messages or issues associated with record deletion.

2. Original forms and notes of all team members will be initialed and dated by the appropriate team member and archived by the Benchmark Evaluation Team Leader until the final Benchmark Evaluation report has been completed and signed off by the Benchmark Evaluation Team Leader.

5.5 **EXCEPTION CONDITIONS**

5.5.1   **GENERAL**

1. If all Phases cannot be completed successfully (i.e., all steps are executed – independent of the matcher results accuracy) within four consecutive workdays (for purposes of this evaluation – a workday will consist of up to nine hours), then that benchmark and the entire Benchmark Evaluation stage for that Bidder will be rated as "Failed". The only exceptions will be for causes outside of the control of the Bidder such as acts-of-God, war, terrorism, or widespread power outages in which case PWGSC, in consultation with the RCMP, will establish a revised schedule based on the situation.

5.6 **EQUATIONS AND CALCULATION**

5.6.1   **GENERAL**

1. This section contains the equations used to calculate Reliability, Relative reliability, and Accuracy as well as to measure the value and human resource cost of the Best Practices.

5.6.2   **KEY EQUATIONS**

1. The following shows the key equation calculations:

$$\textbf{Reliability} = \frac{\text{number of mates found}}{\text{number of known mates in the database}} \times 100\%$$

$$\textbf{Adjusted Reliability} = 100\% - (100\% - \text{Reliability}) \times \frac{40}{(100 - \text{Acceptable Level})}$$

The TP Reliability Acceptable Levels are as defined in Table 7 : Phase 2 Performance Evaluation for Phase 2 and Table 11 : Phase 3 Performance Evaluation for Phase 3. The Adjusted Reliability will be set to 0 where the formula results in a negative value.

$$\textbf{Latent Reliability} = \frac{\text{number of mates in the candidate lists}}{\text{number of known mates for search prints in the database}} \times 100\%$$

$$\textbf{Relative\_reliability} = \text{Reliability} \times (1 - (\frac{PN - 1}{\text{Selectivity}}))$$

Where PN is the average position number of the first correct candidate.

Adjusted Relative Reliability = Relative Reliability + (100% - Highest Relative Reliability), where *Highest Relative Reliability* means the Relative Reliability of the Bidder that achieved the Highest Relative Reliability for the respective test.

Skill Level Coefficient = 100%+(20% Bonus*Low Skill Frequency)-(20% Penalty*High Skill Frequency)

Frequency = Number of transactions to which the skill level was applied divided by the total number of timed transactions for the particular test (represented as a percentage).

High Skill Frequency includes Medium Skill Frequency where High Skill Frequency is greater than 0 for a particular test.

$$\textbf{Accuracy} = (1 - (\frac{\text{\# of False Matches}}{\text{number of search prints}})) \times 100\%$$

5.6.3 TENPRINT TO TENPRINT EQUATIONS

1. The following shows the equation calculations for tenprint to tenprint searches:

    Reliability and Accuracy shall be calculated for Test A and C in **Phases 2 & 3** using the following equations:

    NCH = Number of correct hits (i.e. mates found) with no false drops. Correct declared hits on multiple instances of the same person will be measured. If correct matches are made that were previously unknown they shall each count as a correct hit for purposes of this calculation to reflect the added value the search algorithm provides.

    NAH = Number of background records that are anticipated to be hit.

    $$S_1 = \frac{NCH}{NAH} \times 100\% = \textbf{Reliability score.}$$

    For example, if there are 2050 searches to evaluated and 817 background records have a single mate in the background file then the NAH = 817. If 802 search hit each of their mates NCH = 802. Thus the reliability S_1 = 802 divided by 817 which is 98.16%.

    $$AS_1 = 100\% - (100\% - S_1) \times \frac{(40)}{(100 - \text{AcceptableLevel})} = \textbf{Adj Reliability Score}$$

    If the Acceptable Level is 98% for the test in question and the Reliability score achieved is 98% then AS_1 = 60%. This allows small differences in matching reliability to have mathematical relevance in the evaluation calculations. In the example above the AS_1 = 63.28%.

NFM = Number of False Matches (i.e. one or more false candidates declared a Hit).

NSP = Number of search prints.

$$\mathbf{S_2} = (1 - \frac{\mathrm{NFM}}{\mathrm{NSP}}) \times 100\% = \textbf{Accuracy score.}$$

For example, if there are 2050 searches to evaluated and seven (7) search prints with single mates in the background file hit against non-mated records then the NFM = 7. The number of search prints is NSP = 2050; therefore, the Accuracy = (1 - (7 divided by 2050) or (1 - .0034) or 99.66%.

5.6.4   TENPRINT TO ULF EQUATIONS

1. The following shows the equation calculations for tenprint to ULF searches:

Reliability shall be calculated as follows:

NCH = Number of correct hits with a selectivity of ≤ 10. Correct multiple Unsolved Latent Fingerprint instances of the same person in the candidate list will be measured. If correct matches are made that were previously unknown they shall each count as a correct hit for purposes of this calculation to reflect the added value the search algorithm provides.

NAH = Number of ULF records that are anticipated to be hit.

$$\mathbf{S_1} = \frac{\mathrm{NCH}}{\mathrm{NAH}} \times 100\% = \textbf{Reliability score.}$$

$$\mathbf{RS_1} = S_1 \times (1 - \frac{\mathrm{PN} - 1}{\mathrm{Selectivity}}) = \textbf{Relative Reliability score.}$$

For example, if there are 30 search records with single mates in the ULF and 5 with two mates in the ULF then the NAH = 40. If 30 mates are hit then the NCH = 30. Then the Reliability = 30 divided by 40 = 75%.

If there is a mix of candidates in the first position and in later positions in the candidate list then the PN will be the average position which will be greater than the NCH. The position in the list will be totaled. For example, if 25 were in the first position, 4 were in the third position and 1 was in the fifth position then the PN = (25 + 12 +5) / 30 = 1.40. With a Selectivity = 10, then Relative Reliability = the 75% Reliability calculated above multiplied by (1 – (PN-1) divided by the Selectivity)) = 75% * (1 – (1.4-1)/10) = 1 – (1/10) = 72%.

$\mathbf{ARS_1} = \mathbf{RS_1}$ + (100%- Highest $\mathbf{RS_1}$ for that test by any bidder) = Adj Rel Reliability Score

If the Highest Relative Reliability = 80% in Phase 2 for the test in question and the Relative Reliability score achieved is 72% then $\mathbf{ARS_1}$ = 92%. This allows the bidder with the Highest Relative Reliability to achieve an $\mathbf{ARS_1}$ = 100% and allow a relatively comparable percentage for all other bidders. As well, this allows latent test to within a comparable range to tenprint test; otherwise the tenprint results, which are expected to be near 100%, would skew the overall score.

5.6.5 FINGERPRINT LATENT TO TENPRINT EQUATIONS

1. The equation calculations for fingerprint latents to tenprint searches use the same logic as the fingerprint latent to tenprint searches.

5.6.6 PALM LATENT TO TENPRINT PALM EQUATIONS

1. The equation calculations for palm latents to tenprint palm searches use the same logic as the fingerprint latent to tenprint searches.

5.6.7 ONE TO ONE VSS ID FLAT TO TENPRINT AND/OR ID FLAT EQUATIONS

1. The equation calculations for one to one ID Flat to tenprint and/or ID Flat searches use the same logic as the tenprint to tenprint searches.

5.7 EXAMPLE BIDDER'S OVERALL BENCHMARK SCORE

5.7.1 PHASE WEIGHTINGS

1. This section contains example of the phased weighting used in the sore calculations.

   EXAMPLE:

   Calculation of Bidder's Overall Benchmark Score: The Bidder's overall score for the Benchmark is calculated in the following manner. By way of example, the formulas will be illustrated using sample scoring for a Bidder.

   Weightings: The three tables below summarize the weightings that are applied to each Phase:

   Phase 1:

Test Group	Description	Weight
B	Tenprint Workflow	50%
C	Latent Workflow	50%

Table 16 : Phase 1 Weights

Phase 2:

Test	Description	Reliability	Accuracy	Relative Reliability	Calculated Value
A	TP Cards to TP	80%	20%		100.00%
B	TP to ULF			100%	100.00%
C	TP Plain to TP	80%	20%		100.00%
D	Finger Latent to TP			100%	100.00%
E	Palm Latent to Palms			100%	100.00%

Table 17 : Phase 2 Weights

Phase 3:

Test	Description	Reliability	Accuracy	Relative Reliability	Calculated Value
A	TP Cards to TP	80%	20%		100%
B	TP to ULF			100%	100%
C	TP Plain to TP	80%	20%		100%
D	Finger Latent to TP			100%	100%
E	Palm Latent to Palms			100%	100%

Table 18 : Phase 3 Weights

5.7.2   PHASE WEIGHTINGS

1. Calculate Phase 1 (User Interface) Weighted Score

   Steps:

   a. Each test in Group B and C is assigned a rating based on the criteria in Table 4 (User Interface Evaluation Criteria);

   b. The score for Tenprint Workflow and Latent Workflow is calculated based on the formula in Table 6 : Phase 1 Scoring Matrix;

   c. The assigned score for each of the scored tests in Phase 1 is weighted in accordance with Table 13 : Benchmark Weights at 50% each; and

   d. Add the two weighted scores together to determine the Overall Score for Phase 1.

Test Group	Description	Score	Weight	Weighted Score
B	Tenprint Workflow	75.00	50%	37.50
C	Latent Workflow	75.00	50%	37.50
Overall Phase 1 Weighted Score				75.00

Table 19 : Sample Phase 1 Weighted Score

2. Calculate Phase 2 (Light Out) Weighted Score

Steps:

a. Each test in phase 2 is completed with the scoring results totaled based on Table 7 : Phase 2 Performance Evaluation, Table 8 : Phase 2 Scoring Matrix and equations describe in section 5.6.

b. Note: Highest Relative Reliability = 80%

Phase 2 Tests	Description	Reliability	Adjusted Reliability	Accuracy	Relative Reliability	Adjusted Relative Reliability	Calculated Value	Weights	Weighted Score
2A	TP Cards to TP	98.16%	63.28%	99.66%			70.56	35.00%	24.70
2B	TP to ULF	75.00%			72.00%	92.00%	92.00	15.00%	13.80
2C	ID Flat to TP	99.47%	89.36%	100.00%			91.49	15.00%	13.72
2D	Finger Latent to TP	45.05%			42.31%	62.31%	62.31	20.00%	12.46
2E	Palm Latent to Palms	48.00%			46.00%	56.00%	56.00	10.00%	5.60
2F	ID Flat to TP 1 to 1	99.47%	89.36%	100.00%			91.49	5.00%	4.57
								Total	74.85

Table 20 : Sample Phase 2 Weighted Score

3. Calculate Phase 3 (Best Practices) Overall Weighted Score

Steps:

a. Each test in phase 3 is completed with the scoring results totaled based on Table 11 : Phase 3 Performance Evaluation, Table 12 : Phase 3 Scoring Matrix and equations describe in section 5.6;

b. Determine the **Transaction Time Coefficient** (TTC) based on the values in Table 9 : Phase 3 Prep and Transaction Time; and

c. Determine the **Skill Level Coefficient** (SLC) based on the values in Table 10 : Phase 3 Skill Level Coefficient.

d. Note: Highest Relative Reliability = 90%

Phase 3 Tests	Description	Reliability	Adjusted Reliability	Accuracy	Relative Reliability	Adjusted Relative Reliability
3A	TP Cards to TP	98.16%	63.28%	99.66%		
3B	TP to ULF	75.00%			72.00%	82.00%
3C	ID Flat to TP	99.47%	89.36%	100.00%		
3D	Finger Latent to TP	73.63%			72.75%	82.75%
3E	Palm Latent to Palms	56.00%			55.20%	65.20%

Table 21 : Sample Phase 3 Adjusted Score

Phase 3 Tests		Time Adjusted Reliability	Accuracy	Time Adjusted Relative Reliability	Skill Level Coefficient (SLC)	Calculated Value	Weights	Weighted Score
3A	TP Cards to TP	63.28%	99.66%		120.00%	80.68	25.00%	20.17
3B	TP to ULF			82.00%	120.00%	98.40	20.00%	19.68
3C	TP Plain to TP	89.36%	100.00%		120.00%	105.79	15.00%	15.87
3D	Finger Latent to TP			82.75%		82.75	30.00%	24.82
3E	Palm Latent to Palms			65.20%		65.20	10.00%	6.52
Overall Phase 3 Weighted Score								87.06

Table 22 : Sample Phase 3 Overall Weighted Score

4. Calculate the Overall Benchmark Weighted Score

Steps:

a. Calculate the Benchmark Weighted Scores for each Phase in accordance with Table 13 : Benchmark Weights; and

b. Calculate the Overall Weighted Benchmark Score by adding the three weighted scores as shown below.

Phase	Function	Weighted Phase Score	Benchmark Weight	Benchmark Weighted Score
1	Phase 1 A-E Demos (Pass / Fail)	N/A	N/A	N/A
1	User Interface	76.56	20.00%	15.31
2	Lights Out	74.85	40.00%	29.94
3	Best Practice	87.06	40.00%	34.83
	Overall Weighted Benchmark Score			**80.08**

Table 23 : Sample Overall Weighted Benchmark Score

5.8   GLOSSARY FOR BENCHMARK

5.8.1   TERMS AND DEFINITIONS

1. The following Table provides the basic definitions used in this Benchmark Plan.

Term	Definition
Accuracy	The probability of no false matches being made (i.e., that the AFIS system will **not** match a fingerprint to the wrong record in the repository.) Typical values of accuracy for tenprint to tenprint searches are in the high 98% range. Accuracy is 1 minus the probability of a false match. **Note:** Accuracy is not a valid measure of latent performance where the Selectivity is greater than one (1). Shown in this document as S_2.
ANSI/NIST	American National Standards Institute – National Institute for Standards and Technology. For the purposes of this benchmark both the current ANSI/NIST standard and the FBI EBTS V7 implementation will be used.
Best Practice	All of the techniques the vendor recommends to maximize performance such as tenprint image box checking and adjusting, marking portions of images as not searchable, correcting auto pattern classification, pruning or adding minutiae, and selective image enhancement. For purposes of this Benchmark no manual or automated resubmittal of a previous image will be permitted with or without different minutiae and other search parameters.
Drop (or candidate)	A record returned from a fingerprint search for consideration of being a hit.
Edge Enhancement	Image processing techniques to make edges within an image more pronounced to the eye. The techniques include sharpness and modulation transfer function compensation.
EBTS	Electronic Biometric Transmission Specification – the FBI's implementation of the ANSI/NIST Standard for the Interchange of Fingerprint Images
False hit (or false drop or False Match)	A returned candidate that does not match the search print. False Drop Rate (FDR) is the average number of non-match candidates returned by the AFIS when there is a hit.
Hit	A hit occurs when an AFIS system returns a previously enrolled candidate that is a correct mate of the search fingerprint(s).
ICD	Interface Control Document (e.g., AFIS ICD, NPS-NIST- External ICD)
Lights Out	See "No Manual Intervention"
Magnification	Normally a set rescaling of an image at a set ratio to the original image (e.g., 2X, 3X, 4X.)
Mate (or Match)	A mate to a fingerprint is another, different impression (or image) of the same finger. A mating fingerprint form (or card) is another form containing

Term	Definition
	fingerprints from the same person.
No Manual Intervention (Lights Out)	No manual centering; no manual finger substitution or resequencing; no manual tenprint image rotation; no manual latent image rotation other than initial placement under a camera or on a scanner (can be centered and captured under the camera via manual adjustments made in response to an image viewed on the latent workstation, before any processing takes place); no manual classification, no manual placement of core/delta markers, no manual ridge counting, and no manual minutiae editing or manual image processing (e.g., manually selected enhancements) other than placing a box around the latent print via the latent workstation. This does not preclude any automated quality control; any automated verification (e.g., via second or third stage matching algorithms) processes; or any other techniques as long as they do not require an attended workstation and / or any feedback from a person. No re-submittal of a previously searched image/record will be permitted.
PN	The average position number in the set of searches performed as part of a test. Only the first hit is factored into this measurement even if there are multiple hits against the ULF.
QA	Quality assurance
Reasonable centering	Accurate as measured with a ruler.
Relative Reliability	The higher the true match is on the candidate list the more value that provides. Therefore, the position on the candidate list will be factored into the **Relative Reliability** score for latent searches [finger and palm] and searches against the ULF by multiplying the calculated Reliability score by (1 minus [the position number (**PN**) minus 1 / the Selectivity value of 10]). If the average position number is 5, for example, then the maximum relative reliability score that can be achieved is 60%. Selectivity will be a constant value of ten (10) for each Bidder. Shown in this document as RS_1.
Reliability	The probability that an AFIS system will find the correct matching fingerprint if it is in the repository. Typical values of reliability for tenprint systems are in the high 98% range when the False Match Rate (Type 2 errors) is set to zero. If the search print finds multiple hits then the number of hits will be factored into the numerator in the appropriate equation. Shown in this document as S_1.
Selectivity	The number of candidates that are returned from a search to be evaluated by a tenprint technician or a latent examiner to determine matches. The number reflects how far through the candidate list an examiner should plan to look when verifying candidates.
Zoom	A continuous change of magnification based on user control. It permits the viewing of a small area of an image at a higher magnification or a larger area at a lower magnification with a continuous range of scale.

Table 24 : Glossary of Terms