



# **Systems Delivery and Project Portfolio Management (SDPPM)**

## **AFIS Renewal**

### **ATTACHMENT 1 TO APPENDIX G: BENCHMARK TEST PLAN**

**Last Updated Date:** 2016-06-21

**Status:** Final

**Version:** 2

**RDIMS Document No.:** 42561v10b



## TABLE OF CONTENTS

<b>1. INTRODUCTION.....</b>	<b>1</b>
1.1 General.....	1
1.2 Document Organization .....	1
1.3 Scope.....	1
1.4 Benchmark Phases.....	3
1.5 Overview of Benchmark Week.....	5
<b>2. PHASE 1 OPERATIONAL TESTS .....</b>	<b>8</b>
2.1 Objectives .....	8
2.2 Phase 1 Test Groups .....	8
2.3 Phase 1 Evaluation Criteria .....	8
2.4 Phase 1 Function and Pass / Fail Criteria.....	10
2.5 Benchmarking Operational Stability and User Interface.....	19
2.6 Phase 1 Scoring Matrix.....	20
<b>3. PHASE 2 BASIC PERFORMANCE TESTS .....</b>	<b>21</b>
3.1 Objectives .....	21
3.2 Phase 2 Test Groups .....	21
3.3 Phase 2 Evaluation Criteria .....	22
3.4 Phase 2 Scoring Matrix.....	22
<b>4. PHASE 3 BEST PRACTICES PERFORMANCE TESTS .....</b>	<b>24</b>
4.1 Objectives .....	24
4.2 Phase 3 Test Groups .....	24
4.3 Phase 3 Evaluation Criteria .....	25
4.4 Phase 3 Scoring Matrix.....	27
<b>5. OVERALL SCORE AND PROCEDURES .....</b>	<b>29</b>
5.1 Overall Benchmark Test Score .....	29
5.2 Benchmark Procedures .....	30
5.2.1 BENCHMARK LOCATION AND DURATION .....	30
5.2.2 BENCHMARK CONFIGURATION DATA .....	30
5.2.3 BENCHMARK EVALUATION TEAM .....	31
5.2.4 TEST DATABASES .....	31
5.3 Benchmarking Procedures.....	32
5.3.1 GENERAL .....	32
5.3.2 TEST STARTUP AND AUDIT.....	32
5.3.3 TEST OUTPUT DATA.....	33
5.4 Test Wrap Up.....	34
5.4.1 GENERAL .....	34
5.5 Exception Conditions .....	34
5.5.1 GENERAL .....	34

5.6	Equations and Calculations .....	35
5.6.1	GENERAL .....	35
5.6.2	KEY EQUATIONS.....	35
5.6.3	TENPRINT TO TENPRINT EQUATIONS.....	36
5.6.4	ULF EQUATIONS .....	37
5.7	Example Bidder's Overall Benchmark Score .....	38
5.7.1	PHASE WEIGHTINGS.....	38
5.7.2	PHASE WEIGHTINGS.....	39
6.	<b>GLOSSARY FOR BENCHMARK .....</b>	<b>44</b>

## TABLES

TABLE 1-1:	BENCHMARK PHASES.....	3
TABLE 1-2:	PHASE 1 TEST GROUPINGS .....	4
TABLE 1-3:	PHASE 2 AND 3 TESTS .....	4
TABLE 2-1:	PHASE 1 TEST OUTLINE.....	10
TABLE 2-2:	OPERATIONAL STABILITY AND USER INTERFACE EVALUATION CRITERIA.....	19
TABLE 2-3:	PHASE 1 SCORING MATRIX .....	20
TABLE 3-1:	PHASE 2 PERFORMANCE EVALUATION.....	22
TABLE 3-2:	PHASE 2 SCORING MATRIX .....	23
TABLE 4-1:	PHASE 3 MANUAL AND SYSTEM TRANSACTION TIME .....	25
TABLE 4-2:	PHASE 3 PERFORMANCE EVALUATION.....	27
TABLE 4-3:	PHASE 3 SCORING MATRIX .....	28
TABLE 5-1:	BENCHMARK TEST WEIGHTINGS.....	29
TABLE 5-2:	BACKGROUND FILE DESCRIPTIONS.....	31
TABLE 5-3:	SEARCH FILE DESCRIPTIONS .....	31
TABLE 5-4:	PHASE 1 WEIGHTS .....	38
TABLE 5-5:	PHASE 2 WEIGHTS .....	38
TABLE 5-6:	PHASE 3 WEIGHTS .....	39
TABLE 5-7:	SAMPLE PHASE 1 WEIGHTED SCORE .....	39
TABLE 5-8:	SAMPLE PHASE 2 WEIGHTED SCORE .....	40
TABLE 5-9:	SAMPLE PHASE 3 ADJUSTED SCORE .....	41
TABLE 5-10:	SAMPLE PHASE 3 OVERALL WEIGHTED SCORE.....	41
TABLE 5-11:	SAMPLE OVERALL WEIGHTED BENCHMARK SCORE.....	42
TABLE 6-1:	GLOSSARY OF TERMS .....	44

# **1. INTRODUCTION**

## **1.1 General**

1. This Attachment 1 to Appendix G 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 6 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 Appendix G 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 as many of the mandatory requirements as possible through its COTS solution,;
  - b. To have a VSS that supports the production performance speed for One-to-One (1:1) 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;

- d. The benchmark system must permit the measurement of the elapsed time required to search as well as manually edit ten print, palm and finger latent images and their feature sets, which will be measured; and
  - e. 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 calendar days. All background data sets must be converted in the Continental USA or Canada.
4. The background data sets will be provided in generically named Ten Print Packet (TPP) NIST packets and TIFF images. The NIST packets will include Type-1, Type-2, Type-4, Type 14 and Type-15 fields. The Type-2 fields identify the minimum RCMP defined fields that must be processed in order to effectively execute the benchmark tests such as DCN and VSS indicator. Attachment 3 to Appendix G identifies the planned format of the NIST packets and TIFF files. There may be multiple sets of prints for the same individual in the background data sets. The Bidder's solution is expected to file these prints under the same file number. The VSS indicator will identify prints that must also be recorded in the VSS renewal solution Benchmark system. TIFF images are expected to be saved with a unique identifier that is part of normal processing by the Bidder's solution as well as a unique identifier assigned by the RCMP.
5. The Bidder must include the list of file numbers, generated by the Bidder's solution, associated with all DCNs for TPPs that include the Type-2 field VSS indicator with the Benchmark test plan. These file numbers will be used in VSS test packets for 1:1 matching based on the file number.
6. Example test packets, which will also use a generic TPP, will be provided with the benchmark background set. The intended fields, including descriptions of the fields that are planned to be used in the test data NIST packets are included in Attachment 3 to Appendix G.
7. The Bidder's personnel will run the benchmark system during the execution of the Benchmark, to include reading electronic files, scanning forms, performing QA (Quality Assurance), marking and/or editing features (e.g. minutiae) on finger or palm images, and all other test steps appropriate to each phase. QA is a generic term, used herein, to reflect whatever functions/processes the Bidder's COTS provides to effectively process finger/palm prints.
8. 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. Only one Bidder's resource will operate a workstation at one time, with access to Bidder experts as required.
9. The Bidder's benchmark system is intended to meet three (3) 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 labour requirements (i.e. staff hours and manual processing).

## 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.

Table 1-1: Benchmark Phases	
<b>Phase 1 Operational Tests</b>	A series of tests designed to demonstrate and evaluate capability to perform general AFIS operational and administrative functionality such as: user privilege management / Operational Stability (OS), Ten Print/Latent (TP, LT) user interface evaluation and ANSI/NIST compliance through Direct File/Scan (DF), repository/file type management. As well as to verify the mandatory requirements identified as part of the Bidder's COTS solution within the context of the test.  Each demonstration will be rated for ease of use and other user-experience factors.
<b>Phase 2 Basic Performance Tests</b>	A series of tests designed to measure matcher performance (i.e. Reliability, Accuracy, and Relative Reliability) with no manual intervention; as well as, performance for 1:1 VSS matching.
<b>Phase 3 Best Practices Performance Tests</b>	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 (against accuracy and reliability, etc.) the processing time (i.e. manual labour and system processing time) required to perform the recommended best practices.

2. Table 1-2: 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 1-3: Phase 2 and 3 Tests, below, shows the tests to be performed during Phases 2 and 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". As well, the human-machine evaluation, on-going during Test Groups A through F, will be rated in accordance with the scoring described later in this document.

<b>Table 1-2: Phase 1 Test Groupings</b>	
<b>Test Group #</b>	<b>Test Group Activities</b>
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 ten print records (best practices) to include verification/certification of candidates, verification/certification screen-printing, and fingerprint card / form printing; create ANSI NIST records (see glossary) for storage with appropriate compression. Add ten print records to repository. Use Direct Scan capability to scan ten print 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/certification of candidates, verification/certification 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 Universal Serial Bus (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).
F	One-to-one (1:1) Ten print VSS like processing and post analysis like user interface.

4. The Phase 2 and Phase 3 Tests are depicted in the following table with the estimated 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 labour required for the Best Practices performance as part of the overall Phase 3 score.

<b>Table 1-3: Phase 2 and 3 Tests</b>			
<b>Test</b>	<b>Test Name</b>	<b>Search File</b>	<b>Background File</b>
A	Ten Print to Ten Print	2,500 cards	340,000 cards (25,000 TP cards will have Palms)
B	Ten Print to ULF	2,500 cards (80 with Reverse search indicator)	2,350 finger/palm impressions
C	Ten Print ID Flats to Ten Print	500 cards	340,000 cards
D	Finger Latent to Ten Print	100 impressions	340,000 cards
E	Palm Latent to Ten Print	50 impressions	340,000 cards (25,000 TP cards will have Palms)
F	Ten Print ID Flats to Ten Print	250 cards	30,000 cards

<b>Table 1-3: Phase 2 and 3 Tests</b>			
<b>Test</b>	<b>Test Name</b>	<b>Search File</b>	<b>Background File</b>
	ID Flats (1:1 VER/VSS)		

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

## 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), and
    - ii. Orientation overview for 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, and
    - 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; and
  - Reports for all searches on a USB external drive in an Excel spreadsheet to be provided to the RCMP. The RCMP will provide the format of this spreadsheet with the background data set;
- f. Verification at the end of Phase 2 that records added to the background files (none are anticipated) have been deleted;
- g. 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, and
    - Reports for all searches on a USB external drive in an Excel spreadsheet to be provided to the RCMP; The RCMP will provide the format of this spreadsheet with the background data set.
  - h. Collection of metrics on processing time (i.e. manual labour time and system processing time) to support later calculations of value. This will be a combination of log file data provided by the Bidder and stop watch timing by the benchmark evaluation team;
  - i. 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, and
    - iv. All sample, test and background data including original USB external drives are 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 the 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.

3. The RCMP Benchmark Team may use cameras, for photographs only, or screen captures to record on screen activities for later reference. The Bidder must provide a tool to perform screen captures as required.

## **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; and
  - c. Rate the Human-Machine Interface ease of use and the observed user experience.
  - d. Note: These Phase 1 tests are for demonstration purposes to allow the Bidder to show the functionality required for the RCMP to evaluate the solution. For example, a latent print could be processed as an ident to show the verification/certification process even though it is not an ident.

### **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 2-1: Phase 1 Test Outline. The Bidders will organize the Phase 1 tests into a series of appropriate scenarios as outlined in Table 1-2: 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 2-1: Phase 1 Test Outline as that table is what each Benchmark will be evaluated against.
3. All test data will be provided by the RCMP unless otherwise indicated. The test data will be in a paper form, TPP format or TIFF image format. The specific test data that will be provided is expected to allow the Phase 1 tests to be observed (e.g. test data with fingers out-of-sequence to enable the out-of-sequence processing to be observed).

### **2.3 Phase 1 Evaluation Criteria**

1. Pass / Fail Tests: Each of the functionalities (i.e. operational or administrative) listed in Table 2-1: Phase 1 Test Outline will be evaluated as an acceptable (i.e. Passed) or an unacceptable (i.e. Failed) demonstration of the functionality based on the criteria identified in Table 2-1: Phase 1 Test Outline.
2. Note: Other Performance metrics (e.g. Accuracy) are not to be considered in Phase 1.
3. The Operational Stability and Human-Machine (User-Experience) Interface Evaluation: For each of the Test Groups A through F, the operational stability and user experience will be evaluated for ease and simplicity of user interface and number of steps required. Each of these test groups will be rated in accordance with the criteria

listed in Table 2-2: Operational Stability and User Interface Evaluation Criteria based on the activity sets identified in subsection 2.4 Phase 1 Function and Pass / Fail Criteria.

## 2.4 Phase 1 Function and Pass / Fail Criteria

1. Phase 1 tests will be grouped into related activity sets previously identified; namely OS, TP, LT and DF. 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.

Table 2-1: Phase 1 Test Outline					
Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
A	1	OS	Demonstration of start-up of system from a written set of procedures.	<ul style="list-style-type: none"> <li>• Bidder will start all equipment from a set of written procedures provided with their test procedures.</li> <li>• When the Benchmark suite is restarted, Bidder will synchronize all devices with clocks to a common time which must be done to a Network Time Protocol (NTP) server automatically.</li> <li>• Bidder will check times for results of synchronization.</li> </ul>	<ul style="list-style-type: none"> <li>• The system starts according to the procedures.</li> <li>• Times are synchronized.</li> </ul>
A	2	OS	Demonstration of User Account Management Capabilities with the required granularity.	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Accounts will be used in Test Group A, Item 3. At least one (1) user created must have limited privileges that can be demonstrated (e.g. not authorized for Latent</li> </ul>	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Verify that accounts work correctly.</li> <li>• The granularity of the user management capabilities will be measured against User Management requirements in Annex</li> </ul>

Table 2-1: Phase 1 Test Outline					
Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
				processing).	B section 8.2.
A	3	OS	Demonstration of operational stability when workstations are turned on and off, and other routine changes are made.	<ul style="list-style-type: none"> <li>Bidder will perform shutdowns and restarts of ten print/latent workstations with two (2) different user log-ins and functionality such as logging on as a sys admin (sys admin account created prior to start of testing), logging out, shutting down and restarting the workstation, and having the new users log in and run any AFIS transaction (e.g. check verification/certification queue for work) and then the Bidder logs back in and prints a transaction log before logging out.</li> </ul>	<ul style="list-style-type: none"> <li>All workstation shutdowns and restarts are accomplished with no anomalous system behaviour.</li> <li>Logs accurately capture work performed and persons performing it as well as times and workstations used.</li> <li>Verify that accounts work correctly, including user with limited privileges unable to access specific functions.</li> </ul>
A	4	OS	System Administrator ability to set time to screen saver initiation.	<ul style="list-style-type: none"> <li>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.</li> <li>Then the Bidder will use the password and fingerprint (i.e. two factor authentication) to unlock the screen.</li> <li>Next pass, the Bidder will preclude a timeout by taking any action and seeing that it resets the clock.</li> </ul>	<ul style="list-style-type: none"> <li>Screen saver locks up the workstations and requires a password and fingerprint to free it up.</li> <li>Time resets are controllable and accurate to within two (2) seconds of the value entered.</li> <li>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).</li> </ul>
B	5	DF	Ability to scan, capture,	<ul style="list-style-type: none"> <li>Bidder will scan five (5) RCMP provided ten print</li> </ul>	<ul style="list-style-type: none"> <li>Quality of individual images must reflect</li> </ul>

**Table 2-1: Phase 1 Test Outline**

Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
			format, and compress all 14 fingerprint images (WSQ 15:1) to include plain impressions and creating a TPP (Bidder could use AFIS ICD TPRI packet as an example if desired).	<p>cards (C-216)/ forms at 500 pixels per inch (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 (D16).</p> <ul style="list-style-type: none"> <li>The Bidder will enter the TPP 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. These will be used in test B11.</li> </ul>	<p>such things as reasonable centering, effective minutia, minimum overlap, removal of “finger box” lines that do not intersect the fingerprint ridges, and the correct size.</p> <ul style="list-style-type: none"> <li>Verify in D16 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 average compression ratio is WSQ 15:1 (<math>\pm 10\%</math>).</li> </ul>
B	6	DF	Ability to receive, process and format palm print images to include creating ANSI/NIST Type 15 record compressed with WSQ 15:1.	<ul style="list-style-type: none"> <li>Bidder will receive five (5) RCMP provided TP cards (C216)/forms with palm print (up to six (6) images per submission, all from the same subject) at 500 ppi and compress them with WSQ at 15:1.</li> <li>The Bidder will enter the TPP 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</li> </ul>	<ul style="list-style-type: none"> <li>The palm records are correctly captured (upper, lower and writer’s palms).</li> <li>Quality of individual images must reflect such things as reasonable centering and effective minutiae.</li> <li>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 average compression ratio is WSQ 15:1 (<math>\pm 10\%</math>).</li> </ul>

**Table 2-1: Phase 1 Test Outline**

Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
				by the Bidder) for use in Group D (D17).	
B	7	TP	Ability to perform lights out ten print processing for five cards.	<ul style="list-style-type: none"> <li>Bidder will selectively configure its AFIS to process five (5) Bidder-provided ten print searches from electronic submissions (TPP) through to match report (candidate list) without any manual intervention and will send responses to a verification/certification station queue and then the Bidder will manually send the search results to a printer.</li> <li>Bidder's AFIS should show different thresholds for different regions, reverse search, different file types and search different target set.</li> </ul>	<ul style="list-style-type: none"> <li>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.</li> <li>Searches are successfully run (i.e. executed with no error messages and no anomalous system behaviour) in lights out mode and responses are sent to a verification / certification queue where they will be printed.</li> <li>The printed record accurately reflects the transaction results.</li> </ul>
B	8	TP	Ability to perform automated QA processes on submitted ten prints to include automated sequence checks, poor quality, missing and amputations (AMP) detection.	<ul style="list-style-type: none"> <li>Bidder will submit fifteen (15) RCMP provided ten prints (some with palms) and ID Flats then automatically (i.e. no manual intervention) evaluate each image for quality and automatically evaluate each rolled finger image for proper sequencing, poor quality and missing fingers with no indication the finger is missing.</li> </ul> <p>Print the reports.</p>	<ul style="list-style-type: none"> <li>System automatically and correctly assesses and reports all sequence, poor quality and missing fingers from the fifteen (15) ten print submissions. Verify the results from the reports.</li> </ul>
B	9	TP	Ability to perform on-	<ul style="list-style-type: none"> <li>Bidder will perform manual QA work on the</li> </ul>	<ul style="list-style-type: none"> <li>Operator can successfully perform</li> </ul>



**Table 2-1: Phase 1 Test Outline**

Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
			screen QA functions, review and edits on submitted ten prints based on automatically detected and operator-observed problems, from any workstation on the network.	<p>same fifteen (15) ten prints from the previous test using three (3) different QA workstations in parallel.</p> <ul style="list-style-type: none"> <li>Effort is to include centering, orienting, adjusting contrast and brightness, proper sequencing of ten print records and substitution of plain images for rolled impressions as well as rolled to rolled to compensate for quality and sequence issues.</li> </ul>	<p>basic QA functions (such as centering, sequence corrections, quality analysis and orientation) on the ten print images based on system recommendations or operator-initiated changes and save the results independent of which workstation was used.</p> <ul style="list-style-type: none"> <li>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.</li> </ul>
B	10	TP	<p>Ability to support the storage and searching of multiple fingerprint sets for an individual person (TPP NIST packets). Ability to delete ten print records with an audit trail created and maintained in a fashion that permits review and selective printing.</p>	<ul style="list-style-type: none"> <li>Bidder will submit RCMP-provided 12 sets of fingerprints (some with palms), with image sets for four (4) different people.</li> <li>System will perform an AFIS search with each search print added to the repository, independent of it being a hit or not.</li> <li>Bidder will delete the initial fingerprint record for two (2) such persons and verify that the person is still in the repository.</li> <li>The Bidder will print a repository maintenance report to reflect all</li> </ul>	<ul style="list-style-type: none"> <li>Correctly add the records to the appropriate person.</li> <li>Manually process to demonstrate the full verification/ certification process.</li> <li>Verify that the second and subsequent searches hit on the correct subject ID and that they are stored under that number in the repository.</li> <li>Deletion of one (1) of multiple records for an individual does not delete the individual completely. Record deletion audit trail is</li> </ul>

**Table 2-1: Phase 1 Test Outline**

Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
				additions and the deletions.	<p>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.</p> <ul style="list-style-type: none"> <li>The Subject IDs for the cases where a second record was deleted must still be in the active repository.</li> </ul>
B	11	TP	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.	<ul style="list-style-type: none"> <li>Bidder will print three (3) RCMP identified record ID #s from B5 test and two (2) ID # from B10 test.</li> </ul> <p>The Bidder will print these fingerprint forms / cards directly from the AFIS workstation.</p>	<ul style="list-style-type: none"> <li>Ten print 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.</li> <li>Compare the printed copy to the original.</li> </ul>
B	12	TP	Ability to manage multiple ten print priorities to include changing priorities on multiple transactions.	<ul style="list-style-type: none"> <li>Bidder will submit fifty (50) Bidder-provided ten print transactions with/without palms 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</li> </ul>	<ul style="list-style-type: none"> <li>Priority management processes permit priorities to be changed and subsequent processing is based on the revised priority list.</li> </ul>

Table 2-1: Phase 1 Test Outline					
Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
				<p>search priority.</p> <ul style="list-style-type: none"> <li>Search the transactions and print a log of search order to include priority.</li> </ul>	
C	13	LT	<p>Ability to capture latent images and extract features.</p> <p>Ability to perform on-screen QA functions, review and edits on latents, based on automatically detected and operator-observed problems, from any workstation on the network.</p>	<ul style="list-style-type: none"> <li>Bidder will receive four (4) RCMP provided latent images at 1,000 ppi (two (2) finger and two (2) palm) and the images are to be properly saved and minutiae extracted to create a feature set that will be displayed on screen.</li> <li>Bidder will submit each feature set for search through the AFIS.</li> <li>Bidder will adjust orientation for latent images, using a single workstation. The bidder will then perform manual QA work on all four (4) latent images using three (3) different workstations.</li> <li>At least one (1) change at each workstation will be “undone” prior to the file being saved.</li> <li>The Bidder will reopen each saved file to see if the “undone” change was saved or not.</li> </ul>	<ul style="list-style-type: none"> <li>Results in 1,000 ppi images and the AFIS system support minutiae extraction and feature set creation and display.</li> <li>User can successfully perform basic QA functions, such as centering, orientation, black and white reversal and mirror on latents images based on system recommendations or user-initiated changes and save the results independent of which workstation was used.</li> <li>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.</li> </ul>
C	14	LT	<p>Demonstration of latent image display and ability to manipulate to include image flip, recalibrate size, contrast adjustment,</p>	<ul style="list-style-type: none"> <li>Bidder will manipulate four (4) latent images, from C13 tests, on screen in each of the following processes: QA, manual encoding, search results and the full verification/certification process.</li> </ul>	<ul style="list-style-type: none"> <li>Image manipulation audit trail is automatically generated, changes can be undone, and the Latent workstation demonstrates a set of image manipulation capabilities that satisfy</li> </ul>

<b>Table 2-1: Phase 1 Test Outline</b>					
<b>Test Group</b>	<b>Item</b>	<b>Test Area</b>	<b>Function</b>	<b>Test Approach</b>	<b>Pass / Fail Criteria</b>
			magnification, rotation, brightness adjustment and minutiae editing.	<ul style="list-style-type: none"> <li>Once the LT to TP search is complete, the Bidder will submit each latent impression for a LT to LT search as a continuation of the workflow</li> <li>The Bidder's Latent workstation will capture and track all of the actions.</li> </ul>	the requirements of this AFIS Renewal RFP in terms of user interface and functionality such as image flip, recalibrate size, contrast adjustment, magnification, rotation, brightness adjustment and minutiae editing. Manually process to demonstrate the full verification/certification process.
C	15	LT	Ability to fetch from the ULF, modify and then create a new record in the unsolved latent file (ULF).	<ul style="list-style-type: none"> <li>A total of two (2) previously submitted records will be selected for this test.</li> <li>Bidder will fetch an RCMP identified record, modify it and save it to the ULF under a new number.</li> </ul>	<ul style="list-style-type: none"> <li>Verify by fetching both records from the ULF and that the modifications have been saved as entered.</li> </ul>
D	16	DF	Ability to direct file / direct scan searchable records. RCMP to read and evaluate data from Test Group B.	<ul style="list-style-type: none"> <li>Bidder will provide five (5) ten print search records on a USB external drive created in Group B (B5) tests.</li> <li>The NIST packet will be opened using the Aware software on the RCMP laptop.</li> </ul>	<ul style="list-style-type: none"> <li>All fields in samples can successfully be parsed and the original input data is in the correct fields.</li> <li>WSQ compression rate must be 16.5:1 or less and on average 15:1 (± 10%) across all records in each sample (Refer to ICD for additional details).</li> </ul>
D	17	DF	Ability to create Type 15 NIST records. RCMP to read and evaluate data from Test Group B.	<ul style="list-style-type: none"> <li>Bidder will provide five (5) ANSI NIST palm search records on a USB external drive created in Group B (B6) tests.</li> <li>The NIST packet will be opened using the Aware software on the RCMP</li> </ul>	<ul style="list-style-type: none"> <li>All fields in samples can successfully be parsed and the original input data is in the correct fields.</li> <li>WSQ compression rate must be 16.5:1 or less and on average 15:1 (±</li> </ul>

<b>Table 2-1: Phase 1 Test Outline</b>					
<b>Test Group</b>	<b>Item</b>	<b>Test Area</b>	<b>Function</b>	<b>Test Approach</b>	<b>Pass / Fail Criteria</b>
				laptop.	10%) across all records in each sample (Refer to ICD for additional details).
E	18	OS	Ability to track, log, and report on all transactions performed, user account and database modifications, and any error conditions.	<p>Bidder will print transaction and error logs after all Phase 1 Group A through Group D tests are completed. Output should be on paper and USB external drive.</p> <ul style="list-style-type: none"> <li>Transaction activities/events can be viewed and printed.</li> </ul>	<ul style="list-style-type: none"> <li>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.</li> </ul>
E	19	OS	Demonstration of report generation capabilities for system operations and administrative functions.	<ul style="list-style-type: none"> <li>Bidder will selectively print system utilization reports and repository size for at least three (3) time periods within the elapsed time of Phase 1 Group A through D.</li> <li>This is to be done after Phase 1 Group D.</li> <li>Output is to be on paper and USB external drive in a PDF format.</li> </ul>	<ul style="list-style-type: none"> <li>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.</li> </ul>
E	20	OS	Demonstration of report on repository/file type (criminal, refugee, immigration) capacity and percentage of capacity used to include a report on all records added in Phase 1.	<ul style="list-style-type: none"> <li>Bidder will generate and print repository status report at the end of Phase 1 Group E.</li> <li>This is to cover images, features, and any other space-constrained data sets.</li> <li>Output is to be on paper and USB external drive in PDF format.</li> <li>Output to include a repository maintenance report to reflect all activity in Phase 1.</li> </ul>	<ul style="list-style-type: none"> <li>Reports selectable by repository/file type (i.e. criminal, refugee, immigration) and cover repository size, available space, and percentage of total space used.</li> <li>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.</li> </ul>

Table 2-1: Phase 1 Test Outline					
Test Group	Item	Test Area	Function	Test Approach	Pass / Fail Criteria
F	21	TP	Demonstration of 1:1 matching UI for VSS capabilities on ID Flats such as analysis, view of prints, quality indicators and matching scores.	<ul style="list-style-type: none"> <li>Bidder will process up to five (5) RCMP provided ID Flats (1 to 3 images) and match to VSS background data as expected.</li> <li>Generate and print of activity and results.</li> </ul>	<ul style="list-style-type: none"> <li>Verify log file processing and verify UI provides capabilities that satisfy the requirements for VSS processing such as analysis, view of prints, quality indicators and matching scores.</li> </ul>

## 2.5 Benchmarking Operational Stability and User Interface

- For the Phase 1 Tests the following criteria will be used to evaluate the operational stability and user interfaces. A single score between 0 and 100 will be assigned for Operational Stability (OS), Ten Print (TP), Latent (LT) and Direct File/Scan (DF) based on Group A through Group F tests. The criteria and related scores will be based on the following:

Table 2-2: Operational Stability and User Interface Evaluation Criteria	
Rating	Operational Stability and User Interface Evaluation Criteria
100	High degree of automation and intuitive interfaces: simple (typically one (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.

## 2.6 Phase 1 Scoring Matrix

1. The following scoring matrix will be used to calculate the sub-scores for Phase 1. The average score will be based on the average of all benchmark team evaluators.

Table 2-3: Phase 1 Scoring Matrix			
Test	Score	Calculated Value	Minimum Threshold
All Phase 1 Functionalities	N/A	N/A	All Tests passed
Operational Stability and User Interface	Score based on Table 2-2: Operational Stability and User Interface Evaluation Criteria.	$CV_1$ = Average score for OS $CV_2$ = Average score for TP $CV_3$ = Average score for LT $CV_4$ = Average score for DF	N/A

### 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, with no manual intervention. Speed will also be measured for 1:1 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), or manual substitution for low quality or out of sequence fingers. See the Glossary of terms in Section 6 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. Ten Print receipt, storage and processing to include searches against background files. Search scenarios will include:
    - i. Ten Print searches (rolled, plain and palm impressions) against the ten print background (rolled, plain, ID Flat and palm impressions). Hits are to be declared without benefit of manual verification and the top five (5) candidates must be imported into the RCMP-provided Excel spreadsheet.
    - ii. Ten Print searches (same records as above – from the same file ingest step) against the ULF including finger and palm. Hits are to be declared without benefit of manual verification and the top ten (10) candidates must be imported into the RCMP Excel spreadsheet.
    - iii. ID Flat Ten Print searches against the ten print background file (rolled, plain, and ID Flat impressions). Hits are to be declared without benefit of manual verification and the top five (5) candidates must be imported into the RCMP Excel spreadsheet.
  - b. Latent receipt, storage and searching scenarios to include:
    - i. Latent searches against the ten print background file. Hits are to be declared without benefit of manual verification and top ten (10) 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 (10) candidates must be imported into the RCMP Excel spreadsheet.
  - c. ID Flat 1:1 receipt, storage and matching in a VSS configured subsystem to include:
    - i. ID Flat 1:1 matches against the rolled/plain and/or ID Flat ten print impressions in the VSS background data. 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

- Each of the functions listed in Table 3-1: Phase 2 Performance Evaluation will be rated numerically based on the criteria identified in Table 3-1. In each case, the results of the rating will be scored in accordance with the matrices described in Table 3-2 to calculate the overall Phase 2 score. The following subsections provide directions on how to calculate the associated performance metrics.

Table 3-1: Phase 2 Performance Evaluation			
Test #	Function Tested	Scored Areas	Rating Process
A	Ten Print card searches	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	Calculate Adjusted Reliability using a Selectivity setting of 1 and Acceptable Level of 98% and calculate Accuracy.
B	Ten Print card to ULF	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	Calculate Adjusted Relative Reliability using a Selectivity setting of 10.
C	Ten Print ID Flat searches	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	Calculate Adjusted Reliability using a Selectivity setting of 1 and Acceptable Level of 98% and calculate Accuracy.
D	Finger Latents to TP	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	Calculate Adjusted Relative Reliability using a Selectivity setting of 10.
E	Palm Latents to Palm	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	Calculate Adjusted Relative Reliability using a Selectivity setting of 10.
F	Ten Print ID Flat searches (1:1)	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	Calculate Adjusted Reliability using a Selectivity setting of 1 and Acceptable Level of 98% and calculate Accuracy.

- 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

- The Bidder's Phase 2 performance ratings will be calculated in accordance with the scoring matrices in Table 3-2: 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.
- Ten Print to Ten Print tests: Reliability, Adjusted Reliability and Accuracy will be calculated using the equations in this Plan in Section 5.6. 1:1 matching speed will also

be verified as a mandatory performance requirement for VSS. Prints from the same individual, provided in the background data set, are expected to be filed under the same file number. Any successful hits during Ten print testing to multiple file numbers will be penalized by losing half (0.5) point. That is, if the background data set prints for the same individual were misfiled under more than one file number and they were successfully identified during the test, a half point reduction will be applied for each successful misfiled hit. For example, if background data for the same individual is filed under three (3) different file numbers and during the test the Bidder's solution correctly identifies all three DCNs under three different file numbers; instead of three (3) points for three hits, the Bidder would receive two (2) points ( $1 + 0.5 + 0.5$ ) for one (1) hit and two (2) hits to misfiled prints.

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%, where an average position of 5 would reduce a Reliability of 100% to Relative Reliability of 60%.
4. Note: Accuracy is not a consideration in latent searches since Selectivity is greater than one (i.e. the Selectivity applied against the candidate position reduces the Latent Reliability in manner similar to Accuracy). Adjusted Relative Reliability will be calculated as per Section 5.6 equations and calculations.

<b>Table 3-2: Phase 2 Scoring Matrix</b>				
<b>Test #</b>	<b>Function Tested</b>	<b>Score</b>	<b>Sub-Weights</b>	<b>Calculated Value (CV)</b>
A	Ten Print cards against TP	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	<ul style="list-style-type: none"> <li>80%</li> <li>20%</li> </ul>	$CV_A = AS_1 \cdot .8 + S_2 \cdot .2$
B	Ten Print cards against ULF	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	<ul style="list-style-type: none"> <li>100%</li> </ul>	$CV_B = ARS_1$
C	Ten Print ID Flat search against TP (rolled/plain and/or ID Flat)	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	<ul style="list-style-type: none"> <li>80%</li> <li>20%</li> </ul>	$CV_C = AS_1 \cdot .8 + S_2 \cdot .2$
D	Latent to TP background searches	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	<ul style="list-style-type: none"> <li>100%</li> </ul>	$CV_D = ARS_1$
E	Palm Latent to Palm background searches	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	<ul style="list-style-type: none"> <li>100%</li> </ul>	$CV_E = ARS_1$
F	Ten Print ID Flat search against TP ID Flat (1:1)	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	<ul style="list-style-type: none"> <li>80%</li> <li>20%</li> </ul>	$CV_F = AS_1 \cdot .8 + S_2 \cdot .2$

## **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 processing time (i.e. manual labour and system processing time) that will be required to operate the system at the best practice levels and assign a time level coefficient as defined in Section 4.3, Phase 3 Evaluation Criteria. The details of the tests and scoring are provided in Table 4-2: Phase 3 Performance Evaluation and Table 4-3: Phase 3 Scoring Matrix.

### **4.2 Phase 3 Test Groups**

1. Phase 3 Tests will be organized into two (2) 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 will be measured and factored into the overall Phase 3 score.
2. Note: All hits are to be declared without the benefit of manual verification or decision by the operator, even if the vendor's Best Practice includes this step. For example, the Bidder cannot automatically encode, see the results of the search, and then include additional minutia before submitting. That is, the Bidder only has one opportunity to perform the search:
  - a. For Ten Print, the search prints can be adjusted as part of Manual QA (e.g. segmentation, orientation, sequence) and then submitted for search; and
  - b. For Finger/Palm Latents, if the Bidder's best practice only includes manually encoding the search print, then the Bidder must plot the minutiae and then submit for search; or
  - c. For Finger/Palm Latents, if the Bidder's best practice includes automatically encoding the search print followed by adding/removing minutia, then the Bidder must encode the minutiae directly after automatic encoding and then submit for search prior to seeing any results.

### 4.3 Phase 3 Evaluation Criteria

1. Each of the performance factors listed in Table 4-2: 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 (1). In each case, the results of the rating will be entered into the scoring matrix described in Table 4-3: Phase 3 Scoring Matrix to calculate the overall Phase 3 score.
2. The RCMP prefers a solution with no digit determination; therefore, the use of digit determination will not be allowed in the Phase 3 Test D (Finger Latent to Ten Print) 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 (S1), Accuracy (S2), Relative Reliability (RS1), Adjusted Reliability (AR1) and Adjusted Relative Reliability (ARS1). However, there will be one (1) additional metric applied to assess the human resources impact. The following are factors used in the equations to determine this impact:
  - a. **Transaction Time (TT)** is the average transaction time per transaction for best practices in seconds, including all manual intervention and system processing time, for the tests 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 completion of the last transaction to the Search result 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 Ten Print, Finger Latent and Palm Latent, as appropriate that will be applied in the calculation of the Bidder’s score; and
  - b. For Benchmark planning purposes, the Bidder should expect approximately five hundred (500) Ten print test transactions to be timed and all Latents test transactions to be timed (approximately one hundred (100) finger Latents and fifty (50) palm Latent).

Table 4-1: Phase 3 Manual and System Transaction Time					
Ten Print		Finger Latent		Palm Latent	
Average Transaction Time (manual & system processing) (seconds)	Transaction Time Coefficient (TTC)	Average Transaction Time (manual & system processing) (seconds)	Transaction Time Coefficient (TTC)	Average Transaction Time (manual & system processing) (seconds)	Transaction Time Coefficient (TTC)
0–5	100%	0-120	100%	0-180	100%
5.5	99%	123	99%	184.5	99%
6.0	98%	126	98%	189	98%

<b>Table 4-1: Phase 3 Manual and System Transaction Time</b>					
<b>Ten Print</b>		<b>Finger Latent</b>		<b>Palm Latent</b>	
<b>Average Transaction Time (manual &amp; system processing) (seconds)</b>	<b>Transaction Time Coefficient (TTC)</b>	<b>Average Transaction Time (manual &amp; system processing) (seconds)</b>	<b>Transaction Time Coefficient (TTC)</b>	<b>Average Transaction Time (manual &amp; system processing) (seconds)</b>	<b>Transaction Time Coefficient (TTC)</b>
6.5	97%	129	97%	193.5	97%
7.0	96%	132	96%	198	96%
7.5	95%	135	95%	202.5	95%
8.0	94%	138	94%	207	94%
8.5	93%	141	93%	211.5	93%
9.0	92%	144	92%	216	92%
...	...	...	...	...	...
25	60%	240	60%	360	60%
...	...	...	...	...	...

- c. For every 0.5 seconds of Ten Print Average Transaction Processing Time above 5 seconds, the Ten Print Transaction Time Coefficient (TTC) will decrease by 1%;
- d. For every three (3) seconds of Finger Latent Average Transaction Processing Time above 120 seconds, the Latent Transaction Time Coefficient (TTC) will decrease by 1%; and
- e. For every four and half (4.5) seconds of Palm Latent Average Transaction Processing Time above 180 seconds, the Latent Transaction Time Coefficient (TTC) will decrease by 1%.

<b>Table 4-2: Phase 3 Performance Evaluation</b>			
<b>Test #</b>	<b>Function Tested</b>	<b>Scored Areas</b>	<b>Rating Process</b>
A	Ten Print card searches	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	Calculate Adjusted Reliability and Accuracy using a Selectivity setting of 1 and an Acceptable Level of 98%. Then determine the Transaction Time Coefficient (TTC).
B	Ten Print card to ULF	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	Calculate Adjusted Relative Reliability using a Selectivity setting of 10. Then determine the Transaction Time Coefficient (TTC).
C	Ten Print ID Flat searches	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	Calculate Adjusted Reliability and Accuracy using a Selectivity setting of 1 and an Acceptable Level of 98%. Then determine the Transaction Time Coefficient (TTC).
D	Finger Latents to TP background	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	Calculate Adjusted Relative Reliability using a Selectivity setting of 10. Then determine the Transaction Time Coefficient (TTC).
E	Palm latent to Palm background	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	Calculate Adjusted Reliability using a Selectivity setting of 10. Then determine the Transaction Time Coefficient (TTC).

## 4.4 Phase 3 Scoring Matrix

1. The Bidder's performance rating will be calculated in accordance with the following Table for Phase 3.
2. Prints from the same individual, provided in the background data set, are expected to be filed under the same file number. Any successful hits during Ten print testing to multiple file numbers will be penalized by losing half (0.5) point. That is, if the background data set prints for the same individual were misfiled under more than one file number and they were successfully identified during the test, a half point reduction will be applied for each successful misfiled hit. For example, if background data for the same individual is filed under three (3) different file numbers and during the test the Bidder's solution correctly identifies all three DCNs under three different file numbers; instead of three (3) points for three hits, the Bidder would receive two (2) points ( $1 + 0.5 + 0.5$ ) for one (1) hit and two (2) hits to misfiled prints.

<b>Table 4-3: Phase 3 Scoring Matrix</b>				
<b>Test #</b>	<b>Function Tested</b>	<b>Score</b>	<b>Sub-Weights</b>	<b>Calculated Value (CV)</b>
A	Ten Print card to TP	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	<ul style="list-style-type: none"> <li>80%</li> <li>20%</li> </ul>	$CV_A = (AS_1 * TTC) * .8 + (S_2 * .2)$
B	Ten print card to ULF	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	<ul style="list-style-type: none"> <li>100%</li> </ul>	$CV_B = ARS_1 * TTC$
C	Ten print ID Flat to TP	<ul style="list-style-type: none"> <li>Adjusted Reliability <math>AS_1</math></li> <li>Accuracy <math>S_2</math></li> </ul>	<ul style="list-style-type: none"> <li>80%</li> <li>20%</li> </ul>	$CV_C = (AS_1 * TTC) * .8 + (S_2 * .2)$
D	Finger Latent to TP searches	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	<ul style="list-style-type: none"> <li>100%</li> </ul>	$CV_D = ARS_1 * TTC$
E	Palm latent to TP Palm searches	<ul style="list-style-type: none"> <li>Adjusted Relative Reliability <math>ARS_1</math></li> </ul>	<ul style="list-style-type: none"> <li>100%</li> </ul>	$CV_E = ARS_1 * TTC$

## 5. OVERALL SCORE AND PROCEDURES

### 5.1 Overall Benchmark Test Score

1. The following approach will be used to weigh 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 5-1: Benchmark Test Weightings.
3. The individual scores for each of the three (3) Benchmark Phases will be calculated per this Plan and then weighted in accordance with the table below. Then the three (3) 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).

Table 5-1: Benchmark Test Weightings				
Test #	Function	Phase 1	Phase 2	Phase 3
	<b>Phase 1</b>			
	Phase 1 Demos (Pass / Fail) Threshold = must pass all			
	Operational Stability	10%		
	Ten print User Ease of Use (Rated) Group B	40%		
	Latent User Ease of Use (Rated) Group C	40%		
	Direct File/Scan & ANSI NIST Compliance	10%		
		100%		
	<b>Phases 2 and 3</b>			
<b>X-A</b>	Ten print cards to TP background		35%	25%
<b>X-B</b>	Ten print cards to finger/palm ULF		15%	20%
<b>X-C</b>	Ten print ID Flat to TP background		15%	15%
<b>X-D</b>	Finger Latent search records to TP background		20%	30%
<b>X-E</b>	Palm Latent search records to TP background		10%	10%
<b>X-F</b>	Ten print ID Flat searches (1:1) to VSS background		5%	
	Phase Totals:	<b>100%</b>	<b>100%</b>	<b>100%</b>
	Relative Phase Weights	<b>20%</b>	<b>40%</b>	<b>40%</b>



## **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 the Bidder's 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 three (3) simultaneous connections as well as a printer for the evaluation team's use.
3. The benchmark tests will be conducted in one (1) work week (consisting of no more than five (5) 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 ten prints, latent fingerprints, and latent palms, if the Bidder's system requires the use of filters. 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).
  - b. If the proposed solution is software-only for VSS, then the benchmark configuration must be the same as the proposed solution. For example, if the proposed VSS renewal solution is a software only solution then the benchmark configuration must be the same as the proposed solution, without the HA capabilities (i.e. if only one (1) server with matching software supports the Benchmark requirements, then only one (1) server would be required). 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; and.
  - c. Software configuration information (Version #, Release #) for the benchmark system 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 printing requirements as stated herein, 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. The background and search data will reflect current RCMP operational data. Thirty (30) calendar days in advance of each Benchmark Test, the RCMP will provide the respective Bidder with USB external drives containing the background file data.

Table 5-2: Background File Descriptions		
File #	Forensic Data Type	Format / Media
1	Ten Print cards with and without palms, Ten Print ID Flat cards	ANSI-NIST packet on USB external drives
2	Latent Finger/Palm prints – ULF file	TIFF on USB external drives

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

Table 5-3: Search File Descriptions		
File #	Forensic Data Type	Format / Media
3	Ten Print cards with and without palms	ANSI-NIST USB external drive
4	Ten Print ID Flats	ANSI-NIST USB external drive
5	Latent Fingerprints	TIFF USB external drive
6	Latent Palm Prints	TIFF USB external drive

3. At the time of the tests, the Benchmark Evaluation Team will provide the four (4) search files (Files 3, 4, 5 and 6) identified in Table 5-3: 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 an overview on the use of the individual workstations as part of the orientation.
3. The three (3) 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 on 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 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 digital repository 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 USB external drive 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 external drive 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 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 five (5) consecutive workdays (for purposes of this evaluation – a workday will consist of up to nine (9) 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 Calculations

### 5.6.1 GENERAL

1. This section contains the equations used to calculate Reliability, Adjusted Reliability, Latent Reliability, Relative Reliability, Adjusted 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:

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

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

The TP Reliability Acceptable Levels are as defined in Table 3-1: Phase 2 Performance Evaluation for Phase 2 and Table 4-2: Phase 3 Performance Evaluation. The Adjusted Reliability will be set to 0 where the formula results in a negative value.

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

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

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

Adjusted Relative Reliability = Relative Reliability + (90% – Highest Relative Reliability), where *Highest Relative Reliability* means the Relative Reliability of the Bidder that achieved the Highest Relative Reliability for the respective test. This adjustment will only be applied if all Bidder's are below 90%.

### 5.6.3 TENPRINT TO TENPRINT EQUATIONS

1. The following shows the equation calculations for Ten Print to Ten Print searches which includes the One-to-Many (1:N) rolled/plains or ID flats Ten Print searches and 1:1 ID Flat Ten Print searches:

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

**NCH** = Number of correct hits (i.e. mates found). 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.

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

For example, if there are 2050 searches to evaluate 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%.

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

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).

**NAFM** = Number of Allowed False Matches = 0.

$$S_2 = \left(1 - \frac{NFM}{NAFM + 1}\right) \times 100\% = \text{Accuracy score.}$$

For example, if one (1) search print hits against non-mated records then the NFM = 1. The number of allowed false matches is NAFM = 0; therefore, the Accuracy =  $(1 - (1/0+1))$  or (0) or 0%. Any negative values will be evaluated as zero (0).

## 5.6.4 ULF EQUATIONS

1. The following shows the equation calculations for Ten Print to ULF searches, Latent finger to TP finger and Latent Palm to Ten Print Palm:

**Reliability** shall be calculated as follows:

**NCH** = Number of correct hits with a selectivity of  $\leq 10$ . Correct multiple ULF instances for reverse search or multiple instances of the same person for Latent to Ten Print 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.

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

$$RS_1 = S_1 \times \left(1 - \frac{PN - 1}{\text{Selectivity}}\right) = \text{Relative Reliability score}$$

For example, if there are 30 search records with single mates in the ULF and 5 with two (2) 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 totalled. 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  $(1 - (PN-1) \text{ divided by the Selectivity}) = 75\% * (1 - (1.4-1)/10) = 1 - (1/10) = 72\%$ .

**ARS<sub>1</sub> = RS<sub>1</sub> + (90% - Highest RS<sub>1</sub> 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 **ARS<sub>1</sub> = 82%**. This allows the bidder with the Highest Relative Reliability to achieve an **ARS<sub>1</sub> = 90%** and allow a relatively comparable percentage for all other bidders. As well, this allows latent test to be within a comparable range to ten print test; otherwise the ten print results, which are expected to be near 100%, would skew the overall score.



## 5.7 Example Bidder's Overall Benchmark Score

### 5.7.1 PHASE WEIGHTINGS

1. This section contains an example of the phased weighting used in the score 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 (3) tables below summarize the weightings that are applied to each Phase:

Table 5-4: Phase 1 Weights		
Test Group	Description	Weight
OS	Operational Stability	10%
TP	Ten Print Workflow	40%
LT	Latent Workflow	40%
DF	Direct File/Scan & ANSI NIST	10%

Table 5-5: Phase 2 Weights					
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 ID Flat to TP	80%	20%		100.00%
D	Finger Latent to TP			100%	100.00%
E	Palm Latent to Palms			100%	100.00%
F	ID Flat to TP 1:1	80%	20%		100.00%

<b>Table 5-6: Phase 3 Weights</b>					
<b>Test</b>	<b>Description</b>	<b>Reliability</b>	<b>Accuracy</b>	<b>Relative Reliability</b>	<b>Calculated Value</b>
<b>A</b>	<b>TP Cards to TP</b>	<b>80%</b>	<b>20%</b>		100%
<b>B</b>	<b>TP to ULF</b>			<b>100%</b>	100%
<b>C</b>	<b>TP ID Flat to TP</b>	<b>80%</b>	<b>20%</b>		100%
<b>D</b>	<b>Finger Latent to TP</b>			<b>100%</b>	100%
<b>E</b>	<b>Palm Latent to Palms</b>			<b>100%</b>	100%

## 5.7.2 PHASE WEIGHTINGS

### 1. Calculate Phase 1 (Operational Stability and User Interface) Weighted Score

Steps:

- Each test in Groups A through F is assigned a rating based on the criteria in Table 2-2: Operational Stability and User Interface Evaluation Criteria;
- The score for each of Operational Stability, Ten Print Workflow, Latent Workflow and Direct File / Direct Scan is calculated based on the formula in Table 2-3: Phase 1 Scoring Matrix;
- The assigned score for each of the scored tests in Phase 1 is weighted in accordance with Table 5-1: Benchmark Test Weightings; and
- Add the four (4) weighted scores together to determine the Overall Score for Phase 1.

<b>Table 5-7: Sample Phase 1 Weighted Score</b>				
<b>Test Group</b>	<b>Description</b>	<b>Score</b>	<b>Weight</b>	<b>Weighted Score</b>
<b>OS</b>	<b>Operational Stability</b>	<b>90.60</b>	<b>10%</b>	9.06
<b>TP</b>	<b>Ten Print Workflow</b>	<b>75.00</b>	<b>40%</b>	30.00
<b>LT</b>	<b>Latent Workflow</b>	<b>75.00</b>	<b>40%</b>	30.00
<b>DF</b>	<b>Direct File/Scan &amp; ANSI NIST</b>	<b>75.00</b>	<b>10%</b>	7.50
<b>Overall Phase 1 Weighted Score</b>				<b>76.56</b>

## 2. Calculate Phase 2 (Lights Out) Weighted Score

### Steps:

- Each test in Phase 2 is completed with the scoring results totaled based on Table 3-1: Phase 2 Performance Evaluation, Table 3-2: Phase 2 Scoring Matrix and equations described in Section 5.6.
- Note: Highest Relative Reliability = 80% for TP to ULF; therefore, the example Bidder score is increased by 10%;
- Note: Highest Relative Reliability = 70% for Finger Latent to TP; therefore, the example Bidder score is increased by 20%;
- Note: Highest Relative Reliability = 80% for Palm Latent to Palms; therefore, example Bidder score is increased by 10%; and
- Note: Accuracy = 1 false match; therefore, accuracy = 0%.

Table 5-8: Sample Phase 2 Weighted Score									
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%	0%			50.62	35.00%	17.72
2B	TP to ULF	75.00%			72.00%	82.00%	82.00	15.00%	12.30
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 1:1	99.47%	89.36%	100.00%			91.49	5.00%	4.57
<b>Total</b>									<b>66.37</b>

## 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 4-2: Phase 3 Performance Evaluation, Table 4-3: Phase 3 Scoring Matrix and equations describe in Section 5.6;
- b. Determine the Transaction Time Coefficient (TTC) based on the values in Table 4-1: Phase 3 Manual and System Transaction Time; and
- c. Note: Highest Relative Reliability = 80% for all Latent searches; therefore, the example Bidder score is increased by 10%; and
- d. Note: Accuracy = 0 false matches; therefore, accuracy = 100%.

Table 5-9: Sample Phase 3 Adjusted Score						
Phase 3 Tests	Description	Reliability	Adjusted Reliability	Accuracy	Relative Reliability	Adjusted Relative Reliability
3A	TP Cards to TP	98.16%	63.28%	100.00%		
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 5-10: Sample Phase 3 Overall Weighted Score							
Phase 3 Tests	Description	Time Adjusted Reliability	Accuracy	Time Adjusted Relative Reliability	Calculated Value	Weights	Weighted Score
3A	TP Cards to TP	63.28%	100.00%		70.56	25.00%	17.66
3B	TP to ULF			82.00%	82.00	20.00%	16.40

Table 5-10: Sample Phase 3 Overall Weighted Score							
Phase 3 Tests	Description	Time Adjusted Reliability	Accuracy	Time Adjusted Relative Reliability	Calculated Value	Weights	Weighted Score
3C	TP ID Flat to TP	89.36%	100.00%		91.49	15.00%	13.72
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							79.12

4. Calculate the Overall Benchmark Weighted Score

Steps:

- a. Calculate the Benchmark Weighted Scores for each Phase in accordance with Table 5-1: Benchmark Test Weightings; and
- b. Calculate the Overall Weighted Benchmark Score by adding the three (3) weighted scores as shown below.

Table 5-11: Sample Overall Weighted Benchmark Score				
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	Operational Stability and User Interface	76.56	20.00%	15.31
2	Lights Out	66.37	40.00%	26.55

<b>Table 5-11: Sample Overall Weighted Benchmark Score</b>				
<b>Phase</b>	<b>Function</b>	<b>Weighted Phase Score</b>	<b>Benchmark Weight</b>	<b>Benchmark Weighted Score</b>
3	Best Practice	79.12	40.00%	31.65
	<b>Overall Weighted Benchmark Score</b>			<b>73.51</b>

## 6. GLOSSARY FOR BENCHMARK

- The following table provides the basic definitions used in this Benchmark Plan.

Table 6-1: Glossary of Terms	
Term	Definition
<b>Accuracy</b>	The probability of no false matches being made (i.e. that the AFIS system will <b>not</b> match a fingerprint to the wrong record in the repository.) Typical values of accuracy for ten print to ten print searches are expected to be 100%. A false match is an extremely serious error. It is expected that no false matches should occur in Benchmark testing or in a production AFIS system. <b>Note:</b> Accuracy is not a valid measure of latent performance where the Selectivity is greater than one (1). Shown in this document as $S_2$ .
<b>ANSI/NIST</b>	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 V10 implementation will be used.
<b>Best Practice</b>	All of the techniques the vendor recommends to maximize performance such as ten print image box checking and adjusting, correcting auto pattern classification, deleting or adding minutiae, manual centering, manual finger substitution or resequencing, manual minutiae editing, manual image processing 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.
<b>Candidate</b>	A record returned from a fingerprint search for consideration of being a hit.
<b>Correct Hit</b>	A correct hit occurs when an AFIS system returns a previously enrolled candidate that is a correct mate of the search fingerprint(s).
<b>Edge Enhancement</b>	Image processing techniques to make edges within an image more pronounced to the eye. The techniques include sharpness and modulation transfer function compensation.
<b>EBTS</b>	Electronic Biometric Transmission Specification – the FBI's implementation of the ANSI/NIST Standard for the Interchange of Fingerprint Images.
<b>False hit (or False Match)</b>	A returned candidate that does not match the search print.
<b>ICD</b>	Interface Control Document (e.g. AFIS ICD, NPS-NIST – External ICD).
<b>Lights Out</b>	See “No Manual Intervention”.

<b>Table 6-1: Glossary of Terms</b>	
<b>Term</b>	<b>Definition</b>
<b>Magnification</b>	Normally a set rescaling of an image at a set ratio to the original image (e.g. 2X, 3X, 4X).
<b>Mate (or Match)</b>	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 fingerprints from the same person.
<b>No Manual Intervention (Lights Out)</b>	<p>No manual centering; no manual finger substitution or resequencing; no manual ten print image rotation; 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.</p> <p>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.</p> <p>No re-submittal of a previously searched image/record will be permitted.</p>
<b>PN</b>	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.
<b>QA</b>	Quality assurance
<b>Reasonable Centering</b>	Accurate as measured with a ruler.
<b>Relative Reliability</b>	<p>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 <b>Relative Reliability</b> score for latent searches [finger and palm] and searches against the ULF by multiplying the calculated Reliability score by (1 minus [the position number (<b>PN</b>) minus 1 / the Selectivity value of 10]).</p> <p>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.</p> <p>Shown in this document as RS<sub>1</sub>.</p>
<b>Reliability</b>	The probability that an AFIS system will find the correct matching fingerprint if it is in the repository. Typical values of reliability for ten print systems are in the high 98% range when the False Match Rate 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 <sub>1</sub> .
<b>Selectivity</b>	The number of candidates that are returned from a search to be evaluated by a ten print technician or a latent examiner to determine



Table 6-1: Glossary of Terms	
Term	Definition
	matches. The number reflects how far through the candidate list an examiner should plan to look when verifying candidates.
<b>Zoom</b>	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.