

InterNational Committee for Information Technology Standards
INCITS Secretariat, Information Technology Industry Council (ITI)
1250 Eye St. NW, Washington, DC 20005
Telephone 202-737-8888; Fax 202-638-4922
Email: incits@itic.org

DOC: M1/05-0868
Date: December 14, 2005
Reply to: John W. M. Campbell
Phone: (613) 721 2275
Email: john@bionbiometrics.com

**Proposed Standing Document on Iterative “Round Robin” Testing of Biometric Products
for Interoperability Using Standardized Data Interchange Formats**

Iterative “Round Robin” Testing of Biometric Products for Interoperability Using Standardized Data Interchange Formats

Revision 1.0

December 14, 2005

Foreword

Note that this is a standing document and may be changed in the future by INCITS M1.

End users of systems incorporating biometrics are frequently concerned about interoperability. Many of the standards recently developed in the US by INCITS M1 have been intended to improve interoperability, either by allowing system components to communicate with one another (BioAPI) or by allowing different biometric systems to produce data in a commonly understood form (CBEFF and the data interchange formats). In order to understand to what extent the standards really do enable interoperability and to discover any ambiguities or issues in the standards that may hinder interoperability, it is necessary to perform iterative interoperability testing. This document discusses a methodology for iterative round robin testing of biometric products that may be used to analyze the interoperability of any set of products using a standardized data interchange format. It is noted that the formats developed by M1.3 (such as INCITS 378-2004, INCITS 385-2004, etc.) are the formats principally being considered here, but any other standardized data interchange format (such as those in the ISO/IEC 19794 series or the ILO SID-0002 fingerprint standard) would be amenable to the same methodology. It is also noted that this document does not attempt to define the specific details to be followed for technology or scenario testing, since these are adequately covered in the first three parts of INCITS 409-2005.

This document summarizes the current best practices for conducting an iterative interoperability test in a manner intended to provide the most useful feedback to a standards body.

Table of Contents

Table of Contents.....	3
Terms and Definitions	4
References.....	5
Test Framework	6
The Agents.....	6
Test Procedure	7
Test Types and Reporting methods.....	10
Verification for Access Control.....	10
Identification using Standardized Data Formats	12
Capture Device Equivalency.....	13

Terms and Definitions

Except as noted below, the terms and definitions from the standards listed below in the References section are sufficient for understanding this document. The specific terms listed below serve to better define certain concepts for the specific scope of this document.

interoperability

a measure of the ability of a biometric product to generate standardized data interchange formats that can be used for biometric matching by other systems or of the ability of a biometric system to use standardized data interchange formats generated by other systems for biometric matching. Interoperability is a specific subset of performance when different aspects of the biometric verification or identification process are performed by different biometric products

performance

a measure of the matching accuracy of a biometric product, typically defined by one or more metrics such as a detection error trade-off or a cumulative match characteristic

standardized data interchange format (SDIF)

a data format defined by a national or international standard that encodes biometric data in a specific form so that it can be properly understood by an application without referring to any external information beyond the standard and its normative references. Thus INCITS 381-2004 data records are a standardized data interchange format, but proprietary biometric data blocks encapsulated in a standardized CBEFF patron format are not

References

The following references are critical to the proper understanding and application of this document.

INCITS 377-2004, Finger Pattern Based Interchange Format

INCITS 378-2004, Finger Minutiae Format for Data Interchange

INCITS 379-2004, Iris Image Interchange Format

INCITS 381-2004, Finger Image-Based Data Interchange Format

INCITS 385-2004, Face Recognition Format for Data Interchange

INCITS 395-2005, Biometric Data Interchange Formats – Signature/Sign Data

INCITS 396-2005, Hand Geometry Interchange Format

INCITS 409.1-2005 Biometric Performance Testing and Reporting – Part 1: Principles and Framework

INCITS 409.2-2005 Biometric Performance Testing and Reporting – Part 2: Technology Testing and Reporting

INCITS 409.3-2005 Biometric Performance Testing and Reporting – Part 3: Scenario Testing and Reporting

ISO/IEC CD2 19795-4 Biometric Performance Testing and Reporting – Part 4: Performance and Interoperability Testing of Data Interchange Formats

Test Framework

Every round robin test for biometric interoperability using standardized data interchange formats would need to follow the same simple structure. Specifics of the test might differ and the various options are detailed in subsequent sections, but the agents and the overall procedure would not change from one test to the next.

The Agents

There are four principle agents in a round robin test. In certain cases, the same entity may take on the role of more than one agent. These are as follows:

Biometric Suppliers

These are the entities that will submit products to be tested. They may be companies, academic institutions, individuals or any other entity except the test organizer or the test lab. The suppliers must be willing to provide products that are conformant (in so far as they can determine) to the variant of the standard being tested and they must be willing to work with the test lab to resolve any conformance or interoperability issues that may arise with their product. They must be willing to support the iterative approach outlined in this document and modify their product in accordance with any final guidance that is produced by the test lab.

Standards Body

This is a group that develops biometric data interchange formats and wishes to understand what level of interoperability they support as well as receive feedback on possible changes to the standard that could help to improve interoperability. It has no commercial involvement in the testing and should not be provided with any results that identify individual suppliers. It is appropriate for the standards body to provide a group of experts to review the test methodology and results and provide feedback to the test organizer.

Test Lab

This is a reputable testing laboratory that has expertise in performance, conformance and interoperability testing using standardized data interchange formats. It must be able to safeguard the identity of the suppliers participating in the test and only reveal the data agreed upon by the suppliers and the test organizer to the test organizer itself, as well as the complete anonymous results to the standards body. The test lab must have the ability to evaluate conformance and interoperability problems and work with the suppliers to suggest solutions and eventually develop supplementary guidance or defect reports for the standard being tested as part of the feedback to the standards body. Thus, the test lab should ideally have expertise in the standardized data interchange format being tested as well as in the test methodology. The test lab is appointed by the test organizer.

Test Organizer

This is a reputable and independent organization that has the ability to organize the test. It must also be able to safeguard the identifying information that links individual biometric suppliers to their test results so that full but completely anonymous test results are provided to the standards

body. It is one of the duties of the test organizer to ensure that the standards body is kept completely independent of any commercial arrangements associated with the test.

Test Procedure

Phase 1 – Preparation

The test organizer will announce that the test is going to take place and biometric suppliers interested in being tested will be invited to submit their names and list the products that are to be tested by a specific deadline. The invitation to biometric suppliers should be as wide as possible and exclude no entity. Depending on the test, the product may or may not include a biometric capture device, but it will always include software, either stand-alone or embedded in a capture device, to enable the biometric data to be correctly formatted and stored in a standardized data interchange format. This software may also further process the biometric data to perform enrolment or matching using standardized biometric data interchange format records.

The test organizer will have proposed a particular variant of the standardized biometric data interchange format that is going to be tested (if that is appropriate) and also which type of test is to be performed (as described in subsequent sections). Based on the wishes of the biometric suppliers who commit to participate in the test, this may be modified, since the biometric suppliers may have a better understanding of the current market demand for a particular test type or data interchange format variant than the test organizer. After final agreement, the test organizer will communicate this to the suppliers and to the test lab and the biometric suppliers will send their products to the test lab.

Once the suppliers have submitted their products and the final form of the test is known (since scenario tests typically take longer than equivalently sized technology tests), the test lab will produce an estimate of the time scale for the test and the production of the final results. Due to the nature of the iterative test process, however, this will necessarily be an estimate.

Phase 2 – Conformance Testing

Only products that are conformant to the data interchange format standard that is the subject of the round robin test should be allowed to participate in the full test. Thus, the first stage is for the test lab to test all of the submitted products for conformance. If conformance testing methodology standards exist, even if only as drafts, then the test lab should follow those. If the drafts are in too early a stage or non-existent, then the test lab will have to develop its own conformance testing protocol for the base standard in question. This protocol should then be submitted as a technical contribution to the standards body so that it can be used in further development of a conformance testing methodology standard. As part of this process, supplier claims and second or third party testing should also be considered.

The minimum level of conformance testing required will be to verify the data structure and internal consistency of the standardized data interchange format records produced by any products that produce them. If any are determined not to be conformant, then a detailed report should be sent to the biometric supplier in question explaining exactly why their data interchange format records were not conformant. The supplier should then try to correct the errors and

provide a new version of their product. The test lab may assist the supplier in identifying the root causes of conformance problems. When the new product is received, it should be tested again and if it is still not conformant, another report generated and sent to the biometric supplier. This may occur multiple times but will be strictly terminated when the allotted time for completion of Phase 2 is reached. The goal is to ensure that by the end of Phase 2 as many products as possible produce data interchange format records in conformance with the standard.

For products that make use of the standardized data interchange format records for enrolment or biometric matching, the best solution is to perform an initial very limited interoperability test. In this case, a few samples of data interchange format records produced by each conformant product would be passed to the product being tested for correctly using conformant data interchange format records. If the product was able to read the records and perform its expected function (enrolment, verification, identification, etc.) then, regardless of the level of accuracy achieved, it is considered to have successfully completed Phase 2 of the test. The same iterative procedure of interaction between the test lab and the biometric supplier will be followed if the product can not correctly use conformant data interchange format records. In this case, however, the test lab may have to take a more active role in helping to understand why there is a problem, since only the test lab has access to all of the data interchange format records from all of the conformant products, and thus if some of them can be used and some can not, the test lab will be in the best position to determine why.

As part of this process, it may be necessary to share anonymized data interchange format records produced by products being tested with the biometric suppliers who have supplied products that seek to use them. If this happens, it will be the responsibility of the test lab to mask the CBEFF PID and any other identifying information within the standardized data interchange format records.

This phase ends when all of the products have successfully completed Phase 2 testing or when the time allotted for Phase 2 has elapsed, in which case only a limited subset of products will ordinarily proceed to the next phase. This phase may only be extended by mutual agreement of the test lab and the test organizer.

Phase 3 – “Round Robin” Interoperability Testing

As described below, the test lab will now conduct round robin interoperability testing as either a scenario or a technology test. The choice of scenario and/or technology testing, or some combination of the two, will be dependent on the particular application under consideration. If a scenario test is used, then all of the captured biometric samples, in as raw a form as possible, will be retained so that if subsequent iterations of the test are required, they may be done using offline testing with modified supplier products. This will not permit products that have the generation of the standardized data interchange format embedded into the capture sensor to be retested, however, but this seems like an acceptable compromise given the expense of multiple iterations of a scenario test.

As part of reviewing the results of the interoperability test, the test lab should also consider factors such as template size, transaction time or match throughput rate (depending on

applicability), failure to acquire (if applicable), failure to enrol, etc. As many of these factors as are relevant should be included in the test results.

After reviewing the interoperability results, the test lab may notice anomalies. At that point, it will need to have candid discussions with the biometric suppliers that gave rise to the anomalies. If potential areas of confusion or error in the data interchange format standard are discovered in these discussions, then the test lab will generate a guidance document and send it out to all of the participating suppliers. They will then have a fixed amount of time to modify their software appropriately. At the end of that time, the test lab will run the off-line (technology test) portion of the test again and generate new interoperability results. Depending on available time and resources, any further anomalies may lead to a repetition of this phase, but just as in Phase 2 this phase will definitively terminate after a specific period of elapsed time. This phase may only be extended by mutual agreement of the test lab and the test organizer.

Phase 4 – Publication of Results

The test lab will provide to the test organizer a complete set of anonymous interoperability results together with a list of any issues identified in the data interchange format standard (including areas where suppliers seem to have difficulty understanding or implementing it) and the final supplementary guidance document, if one was required, that was sent to the biometric suppliers. The test lab will also send a list of any new conformance test assertions or test methodologies it developed during the conformance testing phase. The test organizer will then send all of these to the standards body which will post them to its document register and use them as a basis to improve both the data interchange format standard and its companion conformance testing standard.

Phase 5 – Test Termination

Depending on the privacy agreement signed with any test crew used in a scenario test and the agreement between the test organizer and the biometric suppliers the biometric databases collected for the test will either be destroyed or remain with the test lab to simplify future testing. The latter course is recommended for simplified future testing.

Test Types and Reporting methods

Depending on the specific application being addressed and the particular data interchange format being considered, there are several different types of interoperability tests that would be relevant to a standards body and to the end user community. This section describes three of the most likely test types and explains how the interoperability results would be reported for each one. Although this list of test types is not exhaustive, it should cover the most frequently requested applications. It should also be noted that not all types of tests will apply to all data interchange formats.

In each case, the end results will be summarized in the form of a two-dimensional matrix, where letters corresponding to each product will be labels for both the rows and the columns. The values in the cells are values of FRR at the specified level of FAR (as identified in the upper left corner of the matrix). Note that some products may only be represented in a row or a column and not both, depending on the biometric supplier's intent when submitting the product for the test. The entries in the matrix will represent a particular performance point when a standardized data interchange format record created by the product on the row is either used by the product in the column or used with a standardized data interchange format record created by the product in the column. In the case of verification performance, the row labels will represent products used to enrol and the columns will represent products used to verify. In the case of identification, the rows will represent the products that generated the galleries and the columns will represent the products that generated the probes. If the same product is used for both enrolment and verification then the performance is referred to as "native", otherwise it is "interoperable" performance. A typical matrix will look like the one shown below.

FMR = 0.1%	Verify on A	Verify on B	Verify on C	Verify on D
Enrol with A	x.x%	x.x%	x.x%	x.x%
Enrol with B	x.x%	x.x%	x.x%	x.x%
Enrol with D	x.x%	x.x%	x.x%	x.x%

Verification using Standardized Data Interchange Formats

In this case, the application is to enrol a person and store their biometric data in a standardized data interchange format record either on a token or in a database. Then, when the person presents themselves at a door, a computer terminal or some other access point, they are verified (1-to-1) against their own stored data interchange format record.

In this case, it is highly likely that the data capture device and the enrolment algorithm will be provided as part of an integrated product. Similarly, the verification capture device and the verification algorithm will usually be provided as a package. Note that there is no bias here towards enrolment of a biometric using a more or less processed form (template versus image, for instance) since each can be treated in the same way. Thus, in this test, each biometric product consists of a capture device paired with an algorithm for generating a standardized data interchange format record at enrolment or a capture device paired with an algorithm for

performing verification using the output of the capture device and a standardized data interchange format record. Many products may want to use the same capture device with two algorithms and thus be represented as both rows and columns in the results matrix, but some may not and thus the results matrix may be rectangular (rather than square).

This test is therefore best performed using a test crew for live capture of biometric samples that are stored in an unprocessed form for repeated analysis. Testing can then be performed as a technology test, offline and repeatably. Note that in addition to storing the interchange format record, metadata about the capture sequence is also retained to allow data analysis at the individual attempt level as well as at the transaction (multiple attempt) level. At this time all the information related to failures to acquire must also be retained. Failure to enrol is more complex, since it depends on the level of effort permitted and the process followed. It may also happen that samples which could not be enrolled with the products initially tested will be able to be enrolled with modified software provided in a subsequent technology test. Thus, it is useful to attempt to enrol the test crew during this initial phase, provided the enrolment procedure is carefully documented. In that case, failures to enrol are recorded, but wherever possible, the unprocessed capture device output should be retained so that failure to enrol can be re-evaluated during any subsequent technology tests. Since biometric products may have different capture settings during enrolment than during verification, it is also important to distinguish between those biometric samples captured for enrolment and those captured for verification, even if they were captured on the same device.

In the off-line analysis (technology test), the biometric samples captured for enrolment during the live capture phase are converted to standardized data interchange format records. Each one is then compared against biometric samples captured for verification from the same person in order to generate "genuine" match similarity scores. Similarly, the comparison will be performed using biometric samples captured for verification from different test subjects in order to generate "imposter" similarity scores. It is important to realize that depending on the application of the interoperability test, the verification samples may be converted to standardized data interchange formats before being matched, or they may be compared directly using the unprocessed output from the capture device or even a proprietary format generated from that output by the verification algorithm. Whichever method is selected, this must be reported in the results of the interoperability test.

In this test, there should be between two and four results matrices, depending upon the size of the test crew and the number of verification transactions. They will each provide the False Reject Rate at a different value of the False Accept Rate. The False Reject Rate will include the effects of failure to enrol and failure to acquire and will normally be computed on the basis of a transaction that involves only a single attempt. If the test organization has agreed with the biometric suppliers to use a different transaction basis because of its relevancy to the end user (such as the commonly used three attempts), then the test lab will need to group its data into transactions and compute the performance data on that basis. Normally, the FRR at FAR of 1% and FAR of 0.1% will be provided. Should the size of the crew and volume of captured data warrant it, the FRR can also be provided at a FAR of 0.01% and/or a FAR of 0.001%.

Recognizing that cost-effective testing will necessitate a moderate sized test crew, the quantity of data available to compute FTE (and to a lesser extent FTA) will be statistically limited. Therefore, in addition to reporting the composite FRR matrix values (identified above), the test lab will report the individual components used to build-up that FRR, those being the FTE, FTA and the rejection rate due to pure matching error. This will result in another set of two to four matrices, these having the False Non Match Rate (FNMR) for various values of the False Match Rate (FMR).

It is worthwhile to note that there are also three possible ways of generating all of the results matrices. The method selected should be reported as part of the test results. They are briefly described below.

Method 1 – The threshold providing a FAR (or FMR) at the desired operating point for a particular product is calculated from purely native similarity scores. This threshold is then applied to all verifications using the same product to generate a single column of the output matrix. If that product does not participate as an enrolment product, then all verifications from all enrolment products are used to generate the threshold. This represents a situation in which the verify thresholds are set by the manufacturer and they don't know from what source standardized data interchange formats will be coming.

Method 2 – This is very similar to Method 1 except that all thresholds are set based on the similarity scores from all verifications using all enrolment products. In this case, the manufacturer's also have to predetermine the threshold but they are using the information from the interoperability test to do so.

Method 3 – A separate detection error trade-off curve is calculated for each combination of enrolment and verification product and the FRR (or FNMR) at the FAR (or FMR) corresponding to each desired operating point is calculated directly. This represents the situation where each supplier would know the source of each standardized data interchange format and would be able to adjust the threshold to maximize interoperability.

Identification using Standardized Data Formats

In this case, the application is to take data captured and stored in standardized data interchange format records and use it to perform biometric identification against databases of standardized data interchange format records. The key here is to find out whether the standardized data interchange format records are appropriate for large scale identification.

In this scenario, it is likely that the databases would be relatively large and thus it is best if they already exist, allowing the test to be run as a pure technology test. The test organization will need to inform the biometric suppliers of which databases will be used and will need to provide them with samples of those databases. Each biometric supplier must then provide an algorithm that will encode an image (or series of images or other sensor output) from each database into a standardized data interchange format record. Each supplier must also provide an algorithm that will provide a ranked candidate list of the top 10 matching candidates from a gallery of standardized data interchange format records when presented with a standardized data

interchange format record as a probe. Since each supplier is providing an algorithm that can be used to encode both probes and galleries, the results matrix will always be square.

There will be two results matrices for each database used in the test. In each matrix, the columns represent the product whose algorithm encoded the probe samples into standardized data interchange formats and the rows represent the product whose algorithm encoded the gallery databases and matched the probes against the galleries. More complex options are possible where the encoding of the galleries and the identification function are separated, but these are more difficult to interpret and are less likely to occur in reality. The content of the first matrix would be the probability of correct identification of the probe at Rank 1 and the content of the second matrix would be the probability of correct identification of the probe at Rank 10. It would also be useful to report the failure to enrol rate of each product on each database when attempting to enrol both probes and galleries.

Capture Device Equivalency

In this case, the main goal is to determine the interoperability of different biometric capture devices that produce standardized data interchange formats, normally of a less processed form. In order to do this, some biometric operation must be performed upon the standardized data interchange formats. Depending on the agreement between the test organizer and the biometric suppliers, this may be either verification or small gallery sized identification. Either one should be of benefit to the standards body.

This test is more complex, because the biometric suppliers being tested will only provide a capture device and an algorithm to generate a standardized data interchange format from the capture device. Thus, this test must involve a live capture phase and the suppliers may not be able to adapt to any changes recommended in subsequent guidance documents without providing new hardware, thus necessitating another live capture phase. It also requires a group of other biometric suppliers to provide biometric verification or identification algorithms that will use the standardized data interchange format generated by the capture devices being tested. None of these suppliers should have capture products being tested, and there should be a group of them so that the results of the algorithm can be averaged to prevent there from being bias towards or unfair recognition of any single supplier. Alternatively, one or more reference matching algorithms provided by government labs or academic institutions could also be used. In this case, it is not the absolute accuracies that are important, but only the differences in accuracy when the enrolment and verification (or probe and gallery) biometric samples come from different capture devices.

In this case, the results matrices will always be square because every supplier submits both a capture device and an encoding algorithm and there is no differentiation between encoding for enrolment, verification, galleries or probes. The form of the results matrices is exactly the same as either the verification or identification cases above (depending on which test is being done) except that the biometric matching algorithm is actually unrelated to either the row or column product and may often be multiple algorithms from different suppliers, with the contents of the results matrices simply being the mean of the results from the individual algorithms.