# VQA Proficiency Testing Scoring Document for Quantitative HIV-1 RNA

The VQA Program utilizes a real-time testing program in which each participating laboratory tests a panel of five coded samples six times each year on a bi-monthly schedule. Panel performance is assessed using five technical scoring criteria: assay sensitivity, assay specificity, accuracy, precision, assay validity, will now include three non-technical criteria: data submission timeliness, query responsiveness, and data submissions that include PII/PHI.

## SCORING CRITERIA FOR 5-MEMBER REAL-TIME RNA PROFICIENCY TESTING PANELS

The VQA has determined that a minimum of 20 samples are needed to provide sufficient power for analyses. With only five samples on the new panels, no single panel could provide information on all of the scoring criteria. For example, a panel of five samples at the same nominal concentration would provide information on reproducibility and validity, provided the nominal concentration was above the range in which assay precision varies inversely with RNA concentration. However, this panel would not provide information on assay specificity or sensitivity. In most cases, statistical power to identify problems would also be quite limited given the small size of each panel. Therefore, proficiency scores will be based on the combined results from four consecutive panels, or 20 total samples. This approach boosts statistical power and allows performance to be monitored over an interval of six months rather than at one point in time.

The scoring criteria have been designed to reflect the way that RNA assays are used in clinical trials. Virus load measurements are currently conducted in real-time. Therefore, assessments of the total assay standard deviation, including both the intra-assay and inter-assay components, is used to monitor proficiency. In addition, testing for accuracy has been added to the performance criteria. Finally, sensitivity, specificity, assay validity, data timeliness and data query responsiveness are also used for measuring performance. Methods for assessing performance are discussed below.

## Technical Performance

### 1) Total assay SD:

As noted earlier, the measurement of change within a patient is not affected by inter-assay variation when all of the samples from that patient are assayed in a single batch. However, both intra-assay and inter-assay variation are important if samples are assayed in real time. The combination of the two components is defined here as total assay variation. When the original proficiency testing program for RNA assays was designed, the committee of virologists who worked with the VQA Program to develop performance criteria specified that the assays should be conducted with precision sufficient to detect a 5-fold difference between two samples in the same batch. This was interpreted to mean that the intra-assay standard deviation should provide at least 90% power to detect a five-fold difference, which led to the criterion of an intra-assay SD that was not statistically significantly greater than 0.15 $\log_{10}$. A new target standard deviation that takes account of both intra-assay and inter-assay variation is needed for the revised program. However, no mandate analogous to detecting five-fold differences has been provided for assay variation in real time testing. The statistical test for assessing performance also posed a difficulty. Under the original program, the intra-assay variance, which is the square of the intra-assay SD, was compared with the square of 0.15 $\log_{10}$ using a chi-square statistic. It is not possible to perform a similar test on the total assay standard deviation, because the square of the total assay standard deviation does not follow a chi-square

distribution. In fact, it does not have a closed form solution that would lead to simple calculation of a test statistic.

Another problem that arose was the need for replication within each five-member panel if the intra-assay component of variation was to be assessed. Under the original program, the intra-assay standard deviation was estimated from the variation in estimated RNA concentration among coded replicates at the same nominal concentration within a 20-member panel. Replication at even one nominal concentration within a five-member panel was considered too restrictive.

These problems were solved by replacing the original system for assessing assay precision with the following approach. (1) Performance is now assessed using $\log_{10}$ recovery, rather than $\log_{10}$ estimated RNA concentration, where recovery is the ratio of estimated to nominal concentration. This choice was based on the assumption that recovery is constant within the linear range of an assay. Under this assumption, recoveries can be combined across nominal concentrations within five-member panels to provide the replication needed for estimating the intra-assay component of variation. (2) In the absence of a mandate analogous to detecting a five-fold change, a new target standard deviation was developed by estimating expected values for the intra-assay and inter-assay standard deviations using data from recent proficiency panels. The expected values were combined to provide an expected value for the total assay SD. (3) The expected distribution of the total assay standard deviation for each set of four consecutive panels was determined by Monte Carlo simulation that will be based on the expected values. The simulation process is discussed below. Estimated total assay standard deviations that exceed specified cut points on this distribution will be flagged.

Expected values for the intra-assay and inter-assay standard deviation were actually obtained for the Standard Monitor assay using data from the 'A' rounds of panels 016r through 019r. Using data from panels that received scores of C, the intra-assay SD was estimated to by 0.1196 $\log_{10}$ and the inter-assay SD was estimated to be 0.0831 $\log_{10}$. A recent update, using data from the 'A' rounds of panels 016r through 022r produced very similar values. Using the 'A' rounds of ultralow proficiency panels 004ru through 009ru, the intra-assay and inter-assay SD's for the Ultrasensitive Monitor assays were estimated to be 0.1196 $\log_{10}$ and 0.0894 $\log_{10}$ respectively. This analysis was also restricted to data sets with scores of C. The inter-assay SD was slightly higher on the Ultrasensitive Monitor assay than on the Standard Monitor assay because of some problems with low recovery in a few laboratories on a few rounds of testing. For example, problems with a centrifuge were identified at one site. Therefore, the values from the Standard Monitor assay were treated as empirically determined expected values for the intra-assay and inter-assay standard deviation. Taken together, they imply that the expected value for the total assay standard deviation is

$$(0.1196^2+0.0831^2)^{1/2} = 0.1456 \log_{10}.$$

As noted, these estimates were obtained using only the panels that received scores of C. Panels with scores of PC or P were eliminated for the following reasons. The most common problem encountered on proficiency panels is an elevated standard deviation. The elevated values appear as outliers in a plot of the frequency distribution of standard deviations. It did not seem appropriate to base the estimated standard deviations on a data set that included values that did not fit the overall distribution. Furthermore, the residuals from the linear model that were used to obtain the two standard deviations did not fit the assumption of normality when the complete data set was used to obtain the estimates. This failure was caused by extreme values

in the data sets with scores of PC or P; i.e. by outliers that did not fit the overall distribution of HIV RNA results.

The same expected values for the two components of variation will be used for all kits rather than separate expected values for each kit. This is a continuation of past practice.

The Monte Carlo simulations that are used to derive the distribution of the total assay SD from the expected intra-assay and inter-assay SDs proceed as follows. Suppose that the total assay standard deviation was evaluated using three samples from each of four consecutive panels (i.e. each panel also includes a negative sample or a sample near the limit of detection of the assay). There are five steps to the simulation process.

1. A random sample of size 4 is selected from a normal distribution with SD = 0.12 $\log_{10}$. This standard deviation represents expected intra-assay variation; i.e. the variation among results on the same panel.
2. Step 1 is repeated three more times to make four groups of 4 samples each.
3. A random sample of size 4 was selected from a normal distribution with SD = 0.084 $\log_{10}$.
4. The first value in step 3 is added to each of the four values in the first group, the second value in step 3 is added to each of the four values in the second group, and so on. The values generated in step 3 represent the added contribution of run-to-run, or inter-assay, variation to total assay variation; i.e. this component affects all four samples in a given group of 4 the same way but it varies among groups.
5. Estimate the inter-assay and intra-assay standard deviations from the 16 simulated values. Calculate the total assay SD from these estimates.

This process was repeated 10,000 times to generate a distribution for the total assay standard deviation (Table 1). The median was very close to the expected value that was defined above. The cut points at the 95$^{th}$ and 99$^{th}$ percentiles would be used to identify relatively minor and more serious problems with assay precision.

Table 1: Percentiles of the distribution of 10,000 simulated total assay standard deviations, assuming four proficiency panels with four samples each.

| MINIMUM | 25$^{th}$ | MEDIAN | 75$^{th}$ | 95$^{th}$ | 97.5$^{th}$ | 99$^{th}$ | 99.5$^{th}$ | MAXIMUM |
|---|---|---|---|---|---|---|---|---|
| 0.050 | 0.122 | 0.142 | 0.164 | 0.199 | 0.213 | 0.226 | 0.241 | 0.358 |

The results of a given set of simulations will depend upon the total number of samples included in an evaluation of reproducibility and the distribution of these values among panels. For example, different cut points would be obtained if the 16 samples were distributed among the four panels such that two panels had five each and the other two had three each.

The total assay standard deviation (SD) is related to the intra-assay and inter-assay standard deviations as $S_T = (S_A^2+S_E^2)^{1/2}$ where $S_T$ is the total assay SD, $S_A$ is the intra-assay SD and $S_E$ is the inter-assay SD. The total assay SD is calculated for each data set generated within a laboratory using a given kit. The calculated targets for the 95$^{th}$ and 99$^{th}$ percentile for each combination of panels are used to determine the precision score for that laboratory/kit/panel combination.

2) Accuracy:

Accuracy is assessed by comparing average (mean) $\log_{10}$ recovery in a laboratory with expected $\log_{10}$ recovery. The average is calculated using the subset of samples that is used to assess the total assay standard deviation. Median $\log_{10}$ recovery across all laboratories in which the same kit was used on the panels in question is treated as the expected value. Recovery is used instead of estimated RNA concentration because recoveries can be combined across nominal concentrations within a panel or across successive panels to assess accuracy. The median across laboratories will be used as the expected value, instead of the mean or nominal concentration, for the following reasons. Use of a nominal concentration presupposes that expected recovery should be 100%. This is not reasonable if different kits on average produce somewhat different estimates. Means are more sensitive to the effects of outliers than medians and very large outliers have occasionally been identified on previous panels.

The test statistic for identifying accuracy problems is the distance, in standard errors, between average $\log_{10}$ recovery in a laboratory and expected $\log_{10}$ recovery: (R-M)/S where R is average $\log_{10}$ recovery in a laboratory, M is median $\log_{10}$ recovery across all laboratories and S is the standard error of the mean. A normal distribution is assumed for the test statistic. The standard error is derived from the inter-quartile range of the distribution of mean $\log_{10}$ recovery among laboratories; if the mean is normally distributed, then the inter-quartile range is 1.35 standard errors.

Accuracy statistics have been scored officially since September 2011. In some analyses, it has been noted that the standard error that is used for scoring accuracy does vary by round. If the standard error becomes too small, then there is an increased chance that data sets would be penalized for accuracy when, in fact, the deviations noted in the analysis were quite small. Per the VQA Advisory Board's (VQAAB) recommendation, an analysis investigating the impact of standard errors on the accuracy statistics on previous rounds of testing was conducted in April 2009. As a result of this investigation, a modified approach for the scoring accuracy was implemented. Under this modified approach, the observed standard error on a round of testing would be replaced by the historical median of 0.080517 when the observed SE was below the median; otherwise, the observed value would continue to be used

In the past, accuracy scoring for datasets generated with assays that do not have a minimum of seven data sets to generate an assay-specific median was provided for informational purposes only until further information regarding the performance characteristics of these assays could be obtained. This currently includes datasets generated with the In-House taqman (IT) assay and the BioMerieux NucliSens EasyQ (ME) assay. In an effort to officially implement accuracy scoring for all RNA proficiency testing data sets, VQAAB approved to permit the VQA to adopt new accuracy scoring criteria for datasets generated with these types of assays. Under this new criterion, median $\log_{10}$ recovery is compared to 1) the nominal concentration; 2) an overall cross-platform derived median using only data generated with "real-time pcr"; and 3) an overall cross-platform derived median using all data. The overall real-time cross-platform-derived median is determined using data only from all the commercially available real-time PCR kits (specifically, Abbott RealTime and Roche TaqMan). The overall cross-platform-derived median is determined using data from all the commercially available kits. All three medians are used for determining a final accuracy score. If the same score is assessed using 2 of the 3 medians, then that result is used for scoring (this could be a no-penalty scenario in which 2 out of the 3 p-values are > 0.05). If the results indicates three different scores (i.e. >0.05, between 0.05 and 0.01, or <0.01) then a penalty of between 0.05 and 0.01 is assessed, which is equivalent to a score of PC. These new scoring criteria will be used for any kit where fewer than 7 data sets

are available for the analysis (NOTE: if data are pending that will permit kit-specific median calculations, analyses will be put on hold until all the data are available for analysis).

Accuracy statistics that indicate average recovery is at least three but less than four standard errors from the expected value will be flagged as minor problems, while accuracy statistics that indicate a departure of at least four standard errors from the expected value will be flagged as major problems.

3) Sensitivity:

Sensitivity is assessed by determining the number of negative results obtained from replicates at the limit of detection for the assay in question. Typically, 3-5 replicates are used. The limit of detection is defined as the nominal concentration at which 95% of assay results are expected to be positive. A sample containing 50 copies/mL is currently used to assess the sensitivity of the Roche TaqMan and Abbott RealTime HIV-1 assays, even though their detection limits are 20 and 40 copies/mL, respectively. This nominal value is used because it represents the nominal value that was historically used for lower limits of detection in clinical trials. A sample containing 200 copies/mL is currently used to assess the sensitivity of the bioMerieux NucliSens EasyQ HIV RNA assay; a sample containing 50 copies/mL is currently used to assess the sensitivity of the internally developed taqman HIV RNA assay. These nominal values were chosen based on historical testing or pre-defined sensitivity as reported by the laboratory. The use of this same control provides comparable performance across the assay platforms, especially as the new assays become more sensitive. Qualitatively detected results are counted as a detectable result and are not flagged as a false negative result.

The maximum acceptable number of negative results is the number that would be exceeded no more than 5% of the time, given the number of samples involved and assuming independent binomial sampling with an underlying probability of a negative result of 0.05. Typically, one false negative result out of four replicates is not flagged as a problem; two false negative results out of four challenges is typically flagged as a minor problem and more than two false negative results out of four challenges is flagged as a major problem.

4) Specificity:

Specificity is assessed by monitoring the results obtained on samples that do not contain HIV. Each panel includes a small number of HIV-negative samples. Negative results (HIV RNA not detected) are expected from all of them; qualitatively detected HIV RNA is deemed a false positive result. One false positive result is flagged as a minor problem and two or more false positive results are flagged as a major problem.

5) Assay validity:

It is expected that all results should meet established criteria for assay validity. The criteria vary among kits. If an assay censors a sample within a run, then the sample must be repeated. Only the valid sample will be included in the analysis. Assay validity is also determined using the results for external controls included in the run. If an external run control is invalid (i.e. the VQA200 result falls outside the +/- 3SD range) then all the samples within that run are deemed invalid. A single invalid result is a minor problem and will result in a score of PC; two or more invalid samples are deemed a major problem and will result in a score of P.

## Non-Technical Performance

1) Data Timeliness:

A due date is assigned to each 5-member panel. Data must be received by the VQA data management group by the deadline or data will receive a penalty. A laboratory may request an exception using the VQA web utilities housed on FSTRF's website prior to the due date to request an extension. If no contact is made prior to the due date, and the data are received late once within four rounds of testing, then the proficiency testing score will be down-graded one level (i.e. a technical score of C would be assigned a score of $C1^T$, which equates to 2 points for a panel score). If data are received late for the second time in four rounds of testing, then the proficiency testing score will be down-graded two levels (i.e. a technical score of C would be assigned a score of $C2^T$, which equates to 4 points for a panel score).

2) Data Query Responsiveness (NEW):

Occasionally, a query is sent to a laboratory to clarify or fix a problem noted in their VQA submission. The VQA DMG will email the testing laboratory and ask them to follow up on the query. If the laboratory resolves the query within a week of receiving the original query (5 working days, excluding any holidays), then no penalty will be assessed. If no resolution is received, a second query will be sent to the testing laboratory and will include the VQA manager and affiliated network laboratory coordination center. If the query is not resolved within one week (5 working days excluding holidays) of the second query, then the data will be scored as a minor late query response. A third query will be sent to the testing laboratory, including the VQA manager, and the affiliated network laboratory coordination group. If the problem is still not resolved within one week (5 business days excluding holidays) then a major problem will be assessed. If a minor late query resolution is noted, the proficiency testing score will be down-graded one level (i.e. a technical score of C would be assigned a score of $C1^Q$, which equates to 2 points in a panel score; a technical score of PC would be assigned a score of $PC1^Q$ which equates to 4 points in panel score). If a major late query resolution is noted, then the proficiency testing score will be down-graded two levels (i.e. a technical score of C would be assigned a score of $C2^Q$, which equates to 4 points for a panel score). A major late query resolution will receive the maximum of 4 points for a panel score.

3.) Data Submissions Containing PHI and/or PII (NEW):

PHI (protected health information), PII (personally identifiable information), or any information that could link submitted data to an individual participant must not be included in any data files submitted to the VQA for analysis. Examples of identifiers that must not be included are provided below in Appendix 1.

*Note: it is at the discretion of the VQA data management group to reject data sets that are presumed to contain PHI/PII*.

When submitting VQA data, the laboratory will be required to attest that the submission does not contain PHI or PII. If, during QA of the data, the ***VQA DMG*** finds PHI or PII, a penalty will be assessed. All data that contains PII or PHI will be expunged from the data base and laboratories will be required to remove the information and resubmit their data. Data resubmitted after the due date, without an exception request, will also receive a late penalty score.

If a data set contains PHI/PII, the proficiency testing score will be downgraded one level for a first offense (i.e. a technical score of C would be assigned a score of $C1^P$ for the first incident, which equates to 2 points for a panel score). For subsequent offenses within 4 rounds of testing, the proficiency testing score will be downgraded an additional level (i.e. a technical score of C would be assigned a score of $C2^P$, which equates to 4 points for a panel score). Laboratory directors, network coordination centers and the Division of AIDS will be notified if a laboratory submits data containing PHI/PII.

*Note: Any combination of penalties (technical or non-technical) will result in a maximum panel score of 4 points (e.g. $C1^{QP}$, $PC1^T$, etc).*

## PROFICIENCY SCORING FORMAT FOR THE REAL-TIME TESTING SCHEME

The scoring process is illustrated by the example in Table 2. The five columns in the table represent the nominal concentrations on five consecutive 5-member panels that are designed to assess proficiency on the Roche COBAS AmpliPrep / COBAS TaqMan HIV RNA assay. The first four panels would be used to generate the first proficiency score. Specificity would be assessed using the negative samples, sensitivity would be assessed using the samples with a nominal value of 50 copies/mL, while and accuracy and reproducibility would be assessed using the samples with nominal concentrations of 100 copies/ml or higher from all four rounds. The next proficiency score would be based on the combined data from rounds 2 through 5 once the results of round 5 were available.

Table 2: Five hypothetical 5-member panels. Entries are nominal concentrations.

| Round 1 | Round 2 | Round 3 | Round 4 | Round 5 |
|---|---|---|---|---|
| 0 | 300,000 | 3,000 | 15,000 | 0 |
| 15,000 | 0 | 50 | 3,000 | 10,000 |
| 50 | 15,000 | 300,000 | 0 | 50 |
| 3,000 | 30,000 | 0 | 50 | 400,000 |
| 300,000 | 50 | 30,000 | 30,000 | 60,000 |

While a single five-member panel can provide only limited information, it could provide important information if performance problems are severe. Performance on individual panels will therefore be monitored. Laboratory personnel and the VQAAB will be kept apprised of results.

## DETERMINATION OF PROFICIENCY SCORES

A score of PC will be assigned if a minor problem is identified with one of the seven scoring criteria. A P is assigned if minor problems are identified on at least two criteria or if a major problem is identified on at least one criterion. Minor problems consist of:

- a total assay SD that exceeds the 95$^{th}$ percentile but not the 99$^{th}$ percentile on the Monte Carlo simulations per round;
- an accuracy score that indicates that average $\log_{10}$ recovery differs from expected $\log_{10}$ recovery by at least 3 but less than 4 standard errors per round;
- a rate of negative results at the limit of detection that would be met or exceeded with probability $\leq 0.05$ but $>0.01$ per round;
- a single false positive per round;
- a single invalid result per round;
- one late submission in four rounds of testing (this penalty does not carryover, though multiple late submissions within 4 rounds can result in a major penalty);

- one minor late query response within 4 rounds of testing (this penalty does not carryover, though multiple late submissions within 4 rounds can result in a major penalty); or
- one submission containing PHI, PII, or other identifying information within 4 rounds of testing (this penalty does not carryover, though multiple late submissions within 4 rounds can result in a major penalty).

Major problems consist of

- an total assay SD that exceeds the 99$^{th}$ percentile on the Monte Carlo simulations;
- an accuracy score that indicates that average $\log_{10}$ recovery differs from expected $\log_{10}$ recovery by at least 4 standard errors per round;
- a rate of negative results at the limit of detection that would be met or exceeded with probability $\leq$0.01 per round;
- two or more false positives per round;
- two or more invalid results per round;
- two late submissions within four rounds of testing;
- one major late query response during a round of testing or two minor late query responses within 4 rounds; or
- two or more submissions containing PHI, PII, or other identifying information.

If a score of P is assigned and the problem can be traced to the most recent 5-member panel, then a recoded version of that panel will be sent to the laboratory. Results of the repeat assays will be used in all subsequent scoring. Suppose, for example, that a P is assigned because of problems on the fourth of the five hypothetical panels in Table 2. The score for the first four consecutive panels would be based on the first run of round 4. The score for rounds 2 through 5 would be based on the second run of round 4 and the first runs of rounds 2, 3 and 5. If a PC is assigned and the problem was caused by the most recent five-member panel, then the panel will be repeated at the discretion of the Laboratory Director.

If the problem exists over multiple panels and repeating the last five member panel will not resolve the problem, then the laboratory will be given the option of completing a re-qualification panel (20-member panel). This will permit the laboratory to replace all previous data for future analyses. As with a new laboratory running a pre-qualification panel, the 20 member panel data will be included in the analysis as 5-member panel data are added to the analysis. The 20-member panel data will be dropped from the analysis once four 5-member panels have been completed.

*Note: only intra-assay precision statistics will be used when analyzing a 20-member panel. Total assay standard deviations will be implemented once the first 5-member panel has been added to the analysis.*

After the analysis is complete, and the scores have been approved by the VQAAB, a report will be sent to each laboratory via email. This summary will include the decoded data from that laboratory as received by the VQA, performance on each scoring criterion and an approved score. A list of the scores for all laboratories will be circulated via email after the VQAAB review. The laboratory must contact Dan Zaccaro via e-mail (dzaccaro@rti.org), telephone (919-541-6310) or FAX (919-541-5966) if there is a discrepancy in the report received from RTI.

Appendix 1: Excerpt from the "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII): Recommendations of the National Institute of Standards and Technology" by Erika McCallister, Tim Grance, and Karen Scarfone (special publication 800-122).

http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf

## 2. Introduction to PII

One of the most widely used terms to describe personal information is PII. Examples of PII range from an individual's name or email address to an individual's financial and medical records or criminal history. Unauthorized access, use, or disclosure of PII can seriously harm both individuals, by contributing to identity theft, blackmail, or embarrassment, and the organization, by reducing public trust in the organization or creating legal liability. This section explains how to identify and locate PII[14] maintained within an organization's environment and/or under its control, and it provides an introduction to the Fair Information Practices. Sections 3 and 4 discuss factors for assigning PII impact levels and selecting safeguards, respectively. Section 5 discusses incident response for breaches involving PII.

### 2.1 Identifying PII

PII is —any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.ll[15]

To *distinguish* an individual[16] is to identify an individual. Some examples of information that could identify an individual include, but are not limited to, name, passport number, social security number, or biometric data.[17] In contrast, a list containing only credit scores without any additional information concerning the individuals to whom they relate does not provide sufficient information to distinguish a specific individual.[18]

To *trace* an individual is to process sufficient information to make a determination about a specific aspect of an individual's activities or status. For example, an audit log containing records of user actions could be used to trace an individual's activities.

*Linked* information is information about or related to an individual that is logically associated with other information about the individual. In contrast, *linkable* information is information about or related to an individual for which there is a possibility of logical association with other information about the individual. For example, if two databases contain different PII elements, then someone with access to both databases may be able to link the information from the two databases and identify individuals, as well as access additional information about or relating to the individuals. If the secondary information source is present on the same system or a closely-related system and does not have security controls that effectively segregate

the information sources, then the data is considered linked. If the secondary information source is maintained more remotely, such as in an unrelated system within the organization, available in public records, or otherwise readily obtainable (e.g., internet search engine), then the data is considered linkable.

[14] Even if an organization determines that information is not PII, the organization should still consider whether the information is sensitive or has organizational or individual risks associated with it and determine the appropriate protections.

[15] GAO Report 08-536, *Privacy: Alternatives Exist for Enhancing Protection of Personally Identifiable Information*, May 2008, http://www.gao.gov/new.items/d08536.pdf.

[16] The terms ―individual‖ and ―individual's identity‖ are used interchangeably throughout this document. For additional information about the term *individual*, see Appendix B.

[17] These data elements are included in a list of identifying information from the Identity Theft and Assumption Deterrence Act of 1998, Public Law 105-318, 112 Stat. 3007 (Oct. 30, 1998).

[18] Information elements that are not sufficient to identify an individual when considered separately might nevertheless render the individual identifiable when combined with additional information. For instance, if the list of credit scores were to be supplemented with information, such as age, address, and gender, it is probable that this additional information would render the individuals identifiable.

Organizations are required to identify all PII residing within their organization or under the control of
their organization through a third party (e.g., a system being developed and tested by a contractor). Organizations should use a variety of methods to identify PII. Privacy threshold analyses (PTAs), also referred to as initial privacy assessments (IPAs), are often used to identify PII.[19] Some organizations require a PTA to be completed before the development or acquisition of a new information system and when a substantial change is made to an existing system. PTAs are used to determine if a system contains PII, whether a Privacy Impact Assessment (PIA) is required, whether a System of Records Notice (SORN) is required, and if any other privacy requirements apply to the information system. PTAs are usually submitted to an organization's privacy office for review and approval. PTAs are comprised of simple questionnaires that are completed by the system owner in collaboration with the data owner.

PTAs are useful in initiating the communication and collaboration for each system between the privacy officer, the information security officer, and the information officer. Other examples of methods to identify PII include reviewing system documentation, conducting interviews, conducting data calls, using data loss prevention technologies (e.g., automated PII network monitoring tools), or checking with system and data owners. Organizations should also ensure that retired hardware no longer contains PII and that proper sanitization techniques are applied.[20]

## 2.2 Examples of PII Data

The following list contains examples of information that may be considered PII.

- Name, such as full name, maiden name, mother's maiden name, or alias
- Personal identification number, such as social security number (SSN), passport number, driver's license number, taxpayer identification number, patient identification number, and financial account or credit card number[21]
- Address information, such as street address or email address

- Asset information, such as Internet Protocol (IP) or Media Access Control (MAC) address or other host-specific persistent static identifier that consistently links to a particular person or small, well- defined group of people
- Telephone numbers, including mobile, business, and personal numbers
- Personal characteristics, including photographic image (especially of face or other distinguishing characteristic), x-rays, fingerprints, or other biometric image or template data (e.g., retina scan, voice signature, facial geometry)
- Information identifying personally owned property, such as vehicle registration number or title number and related information
- Information about an individual that is linked or linkable to one of the above (e.g., date of birth, place of birth, race, religion, weight, activities, geographical indicators, employment information, medical information, education information, financial information).

[19] Some organizations have similar processes in place and do not call them PTA or IPA. For example PTA/IPA templates, see http://www.usdoj.gov/opcl/initial-privacy-assessment.pdf or http://www.dhs.gov/xlibrary/assets/privacy/privacy_pta_template.pdf.

[20] For more information on media sanitization, see NIST SP 800-88, *Guidelines for Media Sanitization*, http://csrc.nist.gov/publications/nistpubs/800-88/NISTSP800-88_rev1.pdf.

[21] Partial identifiers, such as the first few digits or the last few digits of SSNs, are also often considered PII because they are still nearly unique identifiers and are linked or linkable to a specific individual.