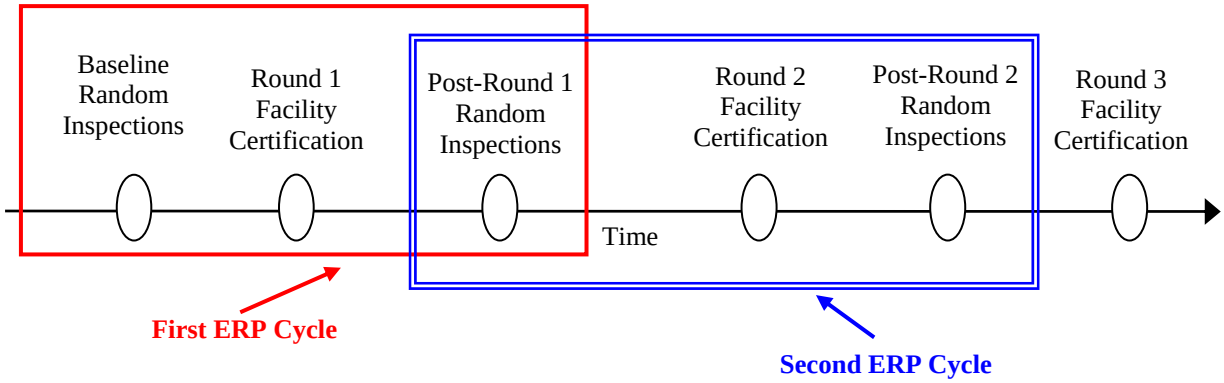
Appendix C: Template for Reporting Core ERP Measures

This appendix contains a complete list of measures for which the States ERP Consortium recommends states strive to report data. For each measure, this appendix presents examples and (when necessary) additional explanatory notes. Please note the following:

- **Prioritizing core measures:** The Consortium has identified 3 tiers of core measures:
  - **Tier I is the true core**—i.e., the set of measures that states that sign the Reporting Results Compact commit to measure and report on. These measures are listed in the table in **bold text** and are indicated by a ✸ symbol.
  - **Tier II includes aspirational measures**—i.e., outcome and cost-related measures that the Consortium strongly encourages states to report on wherever possible. Such measures may be more difficult to report on, but are considered very important in communicating ERP results. These measures are listed in the table in *italics* and are indicated by a ✧ symbol.
  - **Tier III contains additional measures** that the Consortium encourages states to report on when feasible, because they can be very helpful in communicating about ERP. These measures are listed in the table in gray text and are indicated by a ➴ symbol.
- **Data collection terminology:** The measures sometimes refer to "inspections" and "inspectors," as well as "site visits" and "data collectors," when describing the process of collecting data from randomly sampled facilities. These terms are used here interchangeably. The Consortium recognizes that states can effectively use different kinds of data collection approaches for their random samples, including regulatory inspections and site visits by non-regulatory personnel. The core descriptors template provides a place for states to identify their data collection approaches.
- **Reporting on metadata:** By signing the Reporting Results Compact, states commit to report statistical metadata (e.g., confidence level and confidence interval) associated with many of the measures derived from random inspection data. These metadata are typically generated by states conducting standard ERP statistical analyses, and have been chosen in order to accommodate a variety of recognized statistical approaches. Reporting recommendations for statistical metadata are provided in Appendix D. Unless otherwise noted, states are not asked to report underlying calculations, just the final measures.
- **Terminology for ERP milestones over time:** In a typical ERP, random inspections at the outset of the program (i.e., at baseline) are followed by an opportunity for facilities to submit self-certification forms. This self-certification period is followed in turn by another set of random inspections to measure group performance compared to the baseline. In this document, the term “cycle” refers to this process of random inspections, followed by facility certification and then follow-up random inspections. The term “round” is used to designate each new self-certification period. Figure 1, below, illustrates the first two cycles in an ERP over time. In the first ERP cycle, Baseline Random Inspections are followed by Round 1 Facility Certification, which is in turn followed by another set of random inspections. These "Post-Round 1 Random Inspections” also serve as the beginning of the second ERP cycle, when followed by Round 2 Facility Certification and a set of Post-Round 2 Random Inspections.

Figure 1: Terminology for ERP Milestones Over Time



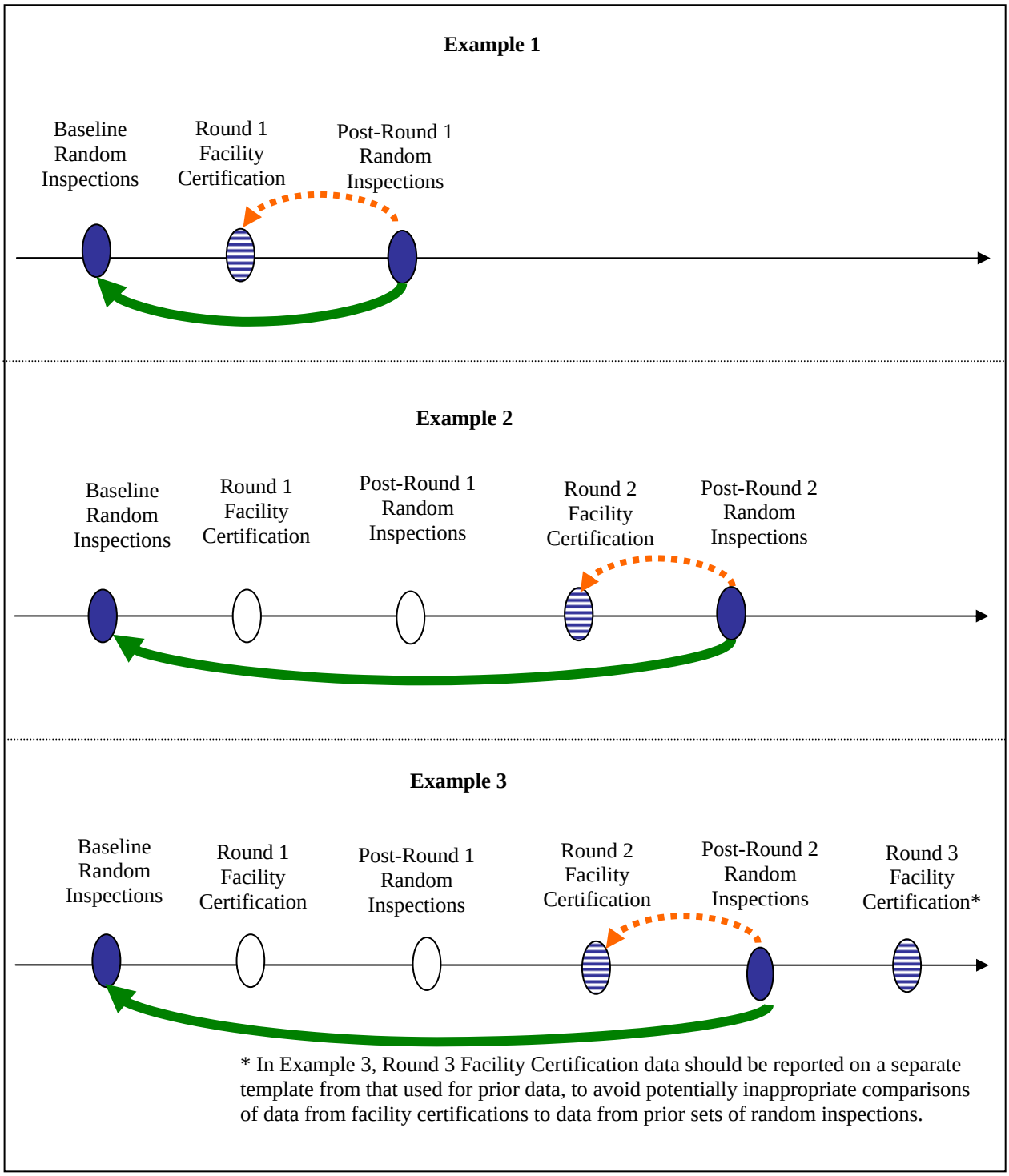
- **Timing of reporting:** States should report their baseline data and the most recent round of facility certification and post-certification inspection data, as shown in Figure 2 below. If a state has only analyzed baseline data, the state may choose to report only those data, before results are available from the first round of certification and post-Round 1 random inspections.

  Sometimes, states may wish to report data from a round of facility certification that has not yet been followed by post-certification inspections, such as Round 3 in Example 3 from Figure 2, below. In such cases, states should report these data separately from other, previous ERP data, to avoid potentially inappropriate comparisons of data from facility certifications to data from prior sets of random inspections.

- **How to report data on performance change:** States should report the result of subtracting the baseline inspection results from the most recent round of random inspections, as shown in Examples 1 and 2 in Figure 2, to ensure consistent interpretation across states. For instance, if a state observed 50.0% of facilities achieving an Environmental Business Practice Indicator (EBPI) during baseline inspections, and 70.0% of facilities achieving an EBPI during post-certification inspections, it would report a positive performance change of 20.0 percentage points (i.e., 70.0% minus 50.0%).

Figure 2: Examples of State ERP Reporting Over Time

(Note: States report data only for those sets of data with shaded ovals. Performance changes are determined by comparing the most recent random inspection data with baseline data, as indicated by solid, bold arrows. Comparisons between data from certifications and corresponding post-certification inspection data are designated by dashed arrows.)



- **Reporting historic data:** The Consortium does not expect that, in response to this Guide, states will revisit past analyses to conduct any new calculations for reporting purposes. When reporting past results (such as baseline results that occurred prior to the development of these core measures), states should provide baseline results for any core measures they have already calculated. Measures for which a state did not conduct calculations can be left blank, although states may wish to conduct and report new calculations for those missing measures, in order to better assess performance change vis-à-vis all these measures. If states reporting past results include measures that were calculated differently than the current core measures, states should explain how the measures were calculated. States that have been implementing ERP for a number of years may decide that at some point it makes sense to designate a certain year as a new baseline, and discontinue reporting data that predates the new baseline; in this case, states should explain that they have selected a new baseline year.

- **Guidelines for handling "not applicable" answers:** States should generally not include "not applicable" responses in the calculation of results for the core measures, with the exception of Core Measure #2, the rate of facility certification "Yes"/"NA"/"Blank" and inspector "No" responses on EBPIs. (See the entry for Measure #2 in the table below for more information on this exception.) States that deviate from this rule of thumb in certain instances are asked to note such instances and provide an explanation, to ensure appropriate interpretation.

  For instance, imagine a state is calculating the achievement rate for an EBPI, based on baseline inspections. It inspected 100 facilities overall, but the EBPI was not applicable to 50 of those facilities; 25 of the remaining 50 facilities were achieving the EBPI in question. The state calculates an achievement rate of 50.0% by dividing 25 achieving facilities by 50 applicable facilities.

  Handling non-applicable facilities by judging them as either achieving or not achieving EBPIs could create an unintended bias, and complicate comparability if different states use different approaches. In the example above, if the state decided to treat non-applicable facilities as achieving the EBPIs, the achievement rate would be calculated as 75/100, or 75.0%—a very different result that could be misinterpreted as indicating performance among applicable facilities is relatively high.

- **"Blanks" or non-responses:** States should not include "blank" responses in the calculation of results for the core measures, with the exception of Core Measure #2, the rate of facility certification "Yes"/"NA"/"Blank" and inspector "No" responses on EBPIs. (See the entry for Measure #2 in the table below for more information on this exception.) In this case, a "blank" (or "non-response") refers to an instance of having no answer or data recorded for a particular question on an inspector checklist or facility certification form. In that case, the state does not know whether the question is applicable to the facility and, if it is applicable, whether the facility is achieving or failing to achieve the item.

  For example, imagine a state is calculating the achievement rate for a particular EBPI. The state conducted 100 inspections, but only 75 responses were recorded for that EBPI. The state is able to reasonably assign a valid response for the 5 of the 25 blank items, through interviewing inspectors and/or deducing valid responses based upon other answers in the inspection checklist. Consequently, the state has 80 valid responses, 40 of which show achievement of the EBPI. The state calculates an achievement rate of 50.0% for that EBPI (40 achieving responses divided by 80 valid responses).

  Alternately, one could choose to treat those 20 blanks in a variety of ways—as achieving, non-achieving, 50.0% achieving (based on the valid responses), etc. Any of those approaches could inject bias into the results, and complicate comparability if different states use different approaches. States may wish to assess and characterize the potential for "non-response" bias in situations with relatively large numbers of blank responses. For instance, states could calculate "what-if" scenarios to understand the potential impact of substantial non-response bias. One way to do so would be to calculate the results both as if blanks were treated as 100.0% achieving and also as if blanks were treated as 0.0% achieving, which can provide upper and lower bounds for the potential bias. This guide recommends that, rather than making assumptions about what the responses would have been,

states leave out the blanks from the analysis, and qualitatively assess and characterize the potential for bias in situations with large amounts of blanks. States may choose to remove certain EBPIs or other measures from their analysis if there is reason to suspect strong non-response bias.

- **Abbreviations:** We use the following acronyms in this document:
  - EBPI = Environmental Business Practice Indicator
  - RTC plan = Return-to-compliance plan

- **Suggestions on reporting:** Please be concise, but if the template does not allow enough space to provide the information requested, provide attachments as needed. To allow more room for entering information, states should delete the examples and notes columns in the template before submitting it. States should submit their own completed templates as attachments to an email to the contacts listed below for the States ERP Consortium's Information Sharing Workgroup.[1]

  Thomas Armstrong
  thomas.armstrong75@cox.net

  Suzi Peck
  susan.peck@state.ma.us

[1] Contact information is current as of the publication date of this guide; states should check the Consortium website at: http://erpstates.org/consortiumContacts.aspx for up-to-date contact information.

Appendix C: Template for Reporting Core ERP Measures

Table 1: General Information about this Report

State Reporting	
Lead Agency Implementing ERP	
ERP Sector/Group	
Types of Data Included in this Report and Year Data Collected [Mark an "X" for all options that apply]	❑ Baseline [Time period] ❑ Self-Certification [Time period] ❑ Post-Certification Inspection [Time period]
Date of This Report	
Status of Results (Draft or Final)	
Revision Number of this Document (first version of this document submitted should be indicated by #1; if there are subsequent revisions of the document submitted, they should be numbered sequentially)	
Individual Reporting Who Can be Contacted with Questions about Data Reported, including: Name, Organization, Phone Number, Email Address	

Table 2: Core Measures of the ERP

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
Key Statistical Metadata	**REFER TO APPENDIX D FOR INSTRUCTIONS**		
Confidence Level			Minimum 90%
Population Size (Baseline)			Population size at the time of the baseline random inspections.
Population Size (Post-Certification)			Population size as of the time of post-certification random inspections.
Sample Size (Baseline)			The number of valid inspections in the baseline random inspections.
Sample Size (Post-Certification)			The number of valid inspections in this round of inspections.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
Certification-Related Measures			
1. Final certification rate✸ ▪ Number of facilities submitting self-certifications (reported by itself, and as a percentage of all facilities). ▪ Whether the rate finalized before or after any post-certification follow-up activity by the agency intended to boost the certification rate.		*A state might report:* 800 shops (80.0% of the final universe of all auto body shops). Finalized 3 months after the certification deadline (2 months after sending out postcards reminding facilities of the deadline). *How the state calculated that result:* 600 of the state's 1,000 auto body shops submitted mandatory self-certification forms by the deadline. After the deadline, another 200 submitted forms, for a total of 800 a month after state follow-up had concluded. The state decided to count the late submissions toward the final certification rate. The final certification rate as a percentage of the final universe is calculated as 800 shops certifying divided by 1,000 shops in the universe.	• Report on the most recent round of certification. • Determine your own cut-off point or deadline for when the certification rate is deemed "final" for the purpose of this measure, but recognize that states typically choose to include all submitted certifications, even if late. Many states conduct some kind of enforcement or other follow-up after the certification deadline passes, in order to boost the certification rate. Please indicate whether the result is before or after any such follow-up activity. • This calculation is conducted the same way, regardless of whether the ERP has voluntary or mandatory certification.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
2. Rate of facility certification "Yes"/"NA"/"Blank" and inspector "No" responses on EBPIs✸ Percentage of EBPI responses on which an inspector, during random inspections, determined that the EBPI was not achieved, but the facility reported something else (i.e., facility reported achieving the EBPI or that the EBPI was not applicable to the facility, or the facility simply did not provide a valid answer).	<u>Post-Certification Inspections Compared to Previous Facility Certification:</u> Confidence interval:	*A state might report:* 10.0%. *The state would also report key statistical metadata associated with this figure, if available (see Appendix D).* *How the state calculated that result:* This state, with mandatory certification, had 10 EBPIs, and conducted 51 post-certification inspections. All of the inspected facilities had submitted certification forms so there were potentially 510 facility responses on EBPIs that could be checked by inspectors. Inspectors inadvertently left a total of 10 EBPI answers blank on checklists. Consequently, 500 valid inspector responses for EBPIs were recorded. These 500 responses indicated whether the inspector believed each EBPI was applicable and, if so, whether or not the facility was achieving that EBPI. Facility certifications for 50 of those 500 items showed non-achievement of the EBPI. The remaining 450 of those 500 items on certification forms claimed facilities were achieving EBPIs, or the facility had skipped the question or identified the question as non-applicable. For 50 of those 450 responses, inspectors found that the questions were applicable and that the facilities were not achieving the EBPI. The state calculates Measure #2 by dividing those 50 responses by the total valid inspector EBPI responses (500). Consequently, the state reports a 10.0% rate of facility certification "Yes"/"NA"/"Blank" and inspector "No" responses on EBPIs. This state did not calculate a margin of error for the measure, but doing so is possible and encouraged.	• This is just one measure of certification reliability, only reflecting instances that may merit significant attention from an agency. Further, while this measure provides information about the predictive value of certification forms, these discrepancies can have other causes than intentional facility falsification -- such as changes in facility performance over time, misunderstandings by facilities, poorly worded certification forms, etc. • Non-response and non-applicability are handled differently for this core measure than for other core measures. For inspection data, issues identified as not applicable by the inspector are included in the analysis as part of the overall number of potential EBPI responses; items left blank by inspectors (i.e., non-response items) are excluded from the analysis because comparison to facility responses cannot be conducted. For facility responses, the calculation includes "blank" and "non-applicable" facility responses, because those responses inform an assessment of the reliability of certification data. A key aspect of certification accuracy can be the extent to which any kind of facility response (or non-response) predicts actual performance, as observed through on-site inspections. • In a state with voluntary certification, the basis for these calculations is a comparison of self-certification responses to inspector responses at randomly sampled facilities that had chosen to certify. For instance, suppose the example state at left instead had a voluntary certification program, and 30 of the 51 random post-certification inspections were conducted at facilities that certified. Imagine the state identified 25 instances of inspector "no" responses on questions where facilities had certified "yes" or "N/A" or had left the question blank. In that case, Measure #2 would be equal to 8.3%. (8.3% is equal to 25 divided by 300 valid responses. There are 300 valid responses because inspectors collected valid data on all 10 EPBIs at each of 30 inspected facilities that had self-certified. • The provision of statistical metadata for this measure is currently optional. However, the Consortium encourages states to submit appropriate statistical metadata with this measure to add to the credibility of reported ERP results. The Consortium intends to provide states with guidance on how to appropriately conduct certain statistical calculations for this measure.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➢ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
3. Rate of self-disclosed noncompliance✸ Number of facilities self-disclosing one or more instance of noncompliance (reported by itself and as a percentage of all certifiers).		*A state might report:* 200 (25.0%) *How the state calculated that result:* The state has a universe of 1,000 facilities. 800 facilities submitted self-certification forms. 200 of the facilities that submitted self-certification forms indicated at least 1 instance of noncompliance. In reviewing the data and following up with facilities, the state discovers that 10 of these facilities should not have declared any noncompliance; i.e., they were in full compliance at the time of self-certification, but misunderstood or misreported the issue. These incorrect reports are *not* excluded from the calculation, which is intended to demonstrate the willingness of facilities to self-report. Consequently, the reported figure of 25.0% is calculated by dividing the total of 200 facilities that self-disclosed 1 or more instances of noncompliance by the total of 800 facilities that submitted self-certification forms.	• Report for the most recent round of certification. • This measure is intended to capture all facilities that disclosed at least one instance of noncompliance, whether or not those facilities submitted RTC plans. Include a facility even if the facility's self-disclosure has not been verified as accurate. • Conduct this calculation the same way regardless of whether the ERP has voluntary or mandatory certification.
4. Rate of RTC plan submission✸ Number of facilities submitting 1 or more RTC plans (reported by itself and as a percentage of all certifiers).		*The state in the example above (for Measure 3) might report:* 100 (12.5%) *How the state calculated that result:* In the example for Measure 3 (above), 100 of the 200 facilities that self-disclosed noncompliance submitted at least 1 RTC plan. 5 of those facilities submitted RTC plans that were incomplete (i.e., they did not address all violations disclosed on the self-certification form). The state reports that all 100 facilities (12.5% of all certifiers) submitted 1 or more RTC plans. The figure of 12.5% is calculated by dividing the 100 facilities that submitted RTC plans by the 800 facilities that submitted self-certification forms. Note that the 5 facilities that submitted incomplete RTC plans are counted toward the number of facilities submitting RTC plans; this measure does not evaluate completeness or adequacy of RTC plans.	• Report for the most recent round of certification. • The number of facilities listed under this measure will be a subset of the number of facilities self-disclosing noncompliance (reported under Measure 3) if some facilities self-disclosed noncompliance on their self-certification form, but did not submit a RTC plan.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
5. Rate of self-disclosing facilities submitting one or more RTC plans➴ Percentage of self-disclosing facilities submitting one or more RTC plans.		*The state in the example above (for Measure 4) would report:* 50.0% *How the state calculated that result:* In the example for Measure 4 (above), 100 of 200 facilities that self-disclosed noncompliance submitted at least 1 RTC plan. This means that the remaining 100 self-disclosers indicated on their self-certification forms that they were not in compliance at the time of certification, but did not demonstrate to the state that they have a plan for how they will return to compliance. The state reports 50.0% as the rate of self-disclosing facilities submitting 1 or more RTC plans. The figure of 50.0% is calculated by dividing 100 facilities that submitted RTC plans by 200 facilities that disclosed violations on their self-certification form.	Report for the most recent round of certification.

KEY TO MEASURES: ✸ **= Tier 1, True Core;** ✧ *= Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
Inspection-Related Measures			
6. Achievement rate for each EBPI ✸ Percentage of randomly sampled facilities achieving each EBPI.	Baseline Random Inspections: (for each EBPI) Effective Sample Size: # Facilities Achieving Measure: % Facilities Achieving Measure: Confidence Interval: Type of Confidence Interval: Post Random Inspections: (for each EBPI) Effective Sample Size: # Facilities Achieving Measure: % Facilities Achieving Measure: Confidence Interval: Type of Confidence Interval: Performance Change: (for each EBPI) Effective Sample Size: Observed Difference: Confidence Interval (if calculated): Significance: **Note: It may be easier to submit the information for this measure in an attachment.**	*A state might report:* Baseline Random Inspections: EBPI #1: 50.0%; EBPI #2: 43.2% Post-Round 1 Random Inspections: EBPI #1: 66.7%; EBPI #2: 45.6% Performance Change: EBPI #1: +16.7 percentage points EBPI #2: +2.4 percentage points *The state would also report key statistical metadata associated with those figures (see Appendix D).* *How the state calculated those results:* The state has 10 EBPIs, and conducted 100 random inspections at baseline and 100 after certification. For EBPI #1, the state found in baseline inspections that the EBPI was applicable to all 100 facilities, and 50 of those 100 facilities were achieving that EBPI. In post-certification inspections, the state found that EBPI #1 was applicable to 90 facilities, and 60 of those 90 facilities were achieving that EBPI. For EBPI #1, the state reports that it observed a 50.0% achievement rate at baseline (50 facilities achieving the EBPI, divided by 100 relevant facilities). The state reports that it observed a 66.7% achievement rate post-certification (60 facilities achieving the EBPI, divided by 90 relevant facilities). The state reports an observed performance change of +16.7 percentage points, calculated by subtracting 50.0% from 66.7%. (Note that they report the "percentage point" change, not the percent change; the latter would be 33.0%, as 66.7% is 33.0% higher than 50.0%.). At both baseline and post-round 1 random inspections, the percentage of randomly sampled facilities achieving each EBPI is calculated as the number of facilities achieving the EBPI divided by the number of facilities for whom the EBPI is relevant. The change in achievement rate from baseline to post-round 1 random inspections is calculated as the difference between these 2 percentages.	• Report rate (and associated statistical metadata) for the baseline round of random inspections and for the most recent round of random inspections. Also report the change (with statistical metadata) in achievement rate from baseline to the most recent round of inspections. • Report whether changes are positive or negative.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
7. Summary of EBPI performance changes ✸ Number of EBPIs worsening/improving/not changing, and for those worsening/improving, number of each that are statistically significant.	Number of EBPIs improving (and number significant): Number of EBPIs worsening (and number significant): Number of EBPIs not changing:	*A state with 10 EBPIs might report:* Number of EBPIs improving (and number significant): 7 (3) Number of EBPIs worsening (and number significant): 2 (0) Number of EBPIs not changing: 1 (N/A) *How the state determined those results:* A state observed, during random inspections, that the achievement rate increased for 7 of its 10 EBPIs, and calculations showed that 3 of those increases can be considered statistically significant. The state observed decreased achievement rates on 2 of its 10 EBPIs, neither of which was statistically significant. The state also observed no performance change on 1 EBPI, which started at 100.0% performance; the concept of statistical significance is not relevant in this situation, because there is no observed difference.	• Report numbers for most recent round of random inspections versus baseline random inspections. • See page 14, Table 2, of EPA's ERP States Produce Results: 2007 Report to see how similar information has already been reported for multiple ERPs.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
8. Achievement rate across all compliance-related measures➴ (Note: this measure is commonly called a traditional compliance rate.) The percentage of randomly sampled facilities that are achieving all relevant compliance-related measures.	<u>Baseline Random Inspections:</u> Effective Sample Size: # of Facilities Achieving Measure: % of Facilities Achieving Measure: Confidence Interval: <u>Post Random Inspections:</u> Effective Sample Size: # of Facilities Achieving Measure: % of Facilities Achieving Measure: Confidence Interval: <u>Performance Change:</u> Effective Sample Size: Observed Difference Between Rates: Significant? (If significant) 1-Sided or 2-Sided? P-Value: Confidence Interval:	*A state might report:* <u>Baseline Random Inspections:</u> 35.0% <u>Post-Round 1 Random Inspections:</u> 50.0% <u>Performance Change:</u> +15.0 percentage points *The state also reports key statistical metadata associated with each of these figures (see Appendix D).* *How the state calculated those results:* The state conducted 100 random inspections at baseline and 100 random inspections after certification. At baseline, it found that 35 facilities had no instances of noncompliance (including all compliance-related EBPIs and all other compliance measures on the inspection checklist). During post-certification inspections, the state observed that 50 facilities had no instances of noncompliance. The state reports 35.0% (35 divided by 100) as the baseline result and 50.0% (50 divided by 100) as the post-certification result. In each case, the percentage of randomly sampled facilities that are achieving all relevant compliance-related measures is calculated as the number of facilities with no instances of noncompliance, divided by the number of facilities subject to compliance requirements. The state also reports a 15 percentage point performance change (i.e., 50.0% minus 35.0%, the difference between the baseline and post-certification achievement rates).	• Report rate (and associated statistical metadata) for the most recent round of random inspections and for the baseline round of random inspections, and report the change (with statistical metadata) from the baseline to the most recent round of inspections. Note that observed changes may reflect the cumulative impact of multiple ERP cycles. • Report whether changes are positive or negative. • Provide background information on the extent to which the state measured performance on all relevant compliance issues within each media category addressed by ERP. This information can be provided as part of answering the "Substantive Scope of ERP" descriptor. For instance, some ERPs may focus only on the highest-priority compliance issues within certain environmental media. In order for users to properly understand this metric, it is important to understand the comprehensiveness of the data collected. Because of potential state-to-state differences in this regard, users should note that cross-state comparisons of this measure may be particularly problematic. • The Consortium recommends that states calculating this measure also consider calculating Measure 14, the average facility score for all compliance-related measures. The average facility score, which can provide information on subtle changes in compliance, is complementary to the traditional compliance rate, which focuses on tracking "full compliance." See the entry for Measure 14 for more information.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➢ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
9. Media-specific achievement rates across all compliance-related measures➢ The percentage of randomly sampled facilities that are achieving all relevant compliance-related measures within each environmental medium being tracked.	<u>Baseline Random Inspections:</u> Effective Sample Size: # of Facilities Achieving Measure: % of Facilities Achieving Measure: Confidence Interval: <u>Post Random Inspections:</u> Effective Sample Size: # of Facilities Achieving Measure: % of Facilities Achieving Measure: Confidence Interval: <u>Performance Change:</u> Effective Sample Size: Observed Difference Between Rates: Significant? (If significant) 1-Sided or 2-Sided? P-Value: Confidence Interval:	*The state in the example above (for Measure 8) might report:* <u>Baseline Random Inspections:</u> • Air: 35.0% • Hazardous waste: 84.0% • USTs: 49.0% *The state also reports results for post-certification inspections and performance changes over time, and key statistical metadata associated with each of these figures (see Appendix D).* *How the state calculated those results:* In the example for Measure 8 (above), the ERP addresses 3 environmental media programs: air, hazardous waste, and underground storage tanks. In its 100 random baseline inspections, the state discovered that 35 facilities had no air violations, 42 facilities had no hazardous waste violations, and 49 facilities had no UST violations. Compliance with hazardous waste regulations is relevant for only 50 facilities, but compliance with air and with UST regulations is relevant to all facilities. At baseline, the state reports a 35.0% (35/100) achievement rate across all compliance-related measures for air, 84.0% (42/50) for hazardous waste, and 49.0% (49/100) for USTs. In each case, the percentage of randomly sampled facilities that are achieving all relevant compliance-related measures is calculated as the number of facilities with no instances of noncompliance for that particular medium, divided by the number of facilities subject to compliance requirements within that medium.	• Consider calculating facility scores for compliance measures on a media-specific basis, if it would be helpful to you. Facility scores can complement these achievement rates by showing finer gradations in performance and changes in performance over time. • Report whether changes are positive or negative.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ⩛ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
10. Average facility score for all EBPIs ✸ The percentage of all relevant EBPIs being achieved, on average, by randomly sampled facilities.	<u>Baseline Random Inspections:</u> Effective Sample Size: Average Facility Score: Confidence Interval: Standard Deviation: <u>Post Random Inspections:</u> Effective Sample Size: Average Facility Score: Confidence Interval: Standard Deviation: <u>Performance Change</u> Effective Sample Size: Observed Difference Between Averages: Significant? 1-Sided or 2-Sided? P-Value: Confidence Interval:	*A state might report:* <u>Baseline Random Inspections:</u> 72.3% <u>Post-Round 1 Random Inspections:</u> 81.3% <u>Performance Change:</u> +9.0 percentage points *The state also reports key statistical metadata associated with each of these figures (see Appendix D).* *How the state calculated those results:* The state conducted 100 random inspections at baseline and 100 random inspections after certification. The state recorded each inspected facility's performance with regard to 10 EBPIs. For each facility, the state calculated a facility score across all EBPIs. For instance, if a facility was achieving 9 out of 10 EBPIs, it would get a score of 90.0%. A facility achieving 6 out of 8 EBPIs (with 2 EBPIs not relevant to that facility) would receive a score of 75.0%. For each round of inspections, the state reports the average of all the individual facility scores (calculated by summing the percentages from each facility and dividing by the number of facilities, in this case, 100). The state also reports the percentage point change between baseline and post-certification average facility scores.	• Report score (with statistical metadata) for the baseline round of random inspections and the most recent round of random inspections, as well as the change (with statistical metadata) from the baseline to the most recent round of inspections. Note that observed changes may reflect the cumulative impact of multiple ERP cycles. • Report whether changes are positive or negative.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
11. Distribution of facility scores for all EBPIs➴ Distribution of facility scores for all EBPIs among randomly sampled facilities.	<u>Baseline Random Inspections:</u> 0-9.9%: X% of facilities 10-19.9%: Y% 20-29.9%: Z% 30-39.9%: Etc. 40-49.9%: 50-59.9%: 60-69.9%: 70-79.9%: 80-89.9%: 90-100%: Total number of randomly sampled facilities: Minimum score: Maximum score: Median score: <u>Post-Round 1 Random Inspections:</u> 0-9.9%: A% of facilities 10-19.9%: B% 20-29.9%: C% 30-39.9%: Etc. 40-49.9%: 50-59.9%: 60-69.9%: 70-79.9%: 80-89.9%: 90-100%: Total number of randomly sampled facilities: Minimum score: Maximum score: Median score:	*The state from the example above (for Measure 10) might report the following results:* <u>Baseline Random Inspections:</u> • 0-9.9%: 7% of facilities • 10-19.9%: 10% • 20-29.9%: 3% • Etc. *The state would also provide a distribution for post-certification results.* *How the state calculated those results:* Recall that an individual facility score is the number of EBPIs achieved by that facility divided by the number of EBPIs relevant to that facility. For each round of inspections, the state calculates a facility score for each facility. Then the state calculates the percentage of randomly inspected facilities with a score of 0%-9.9%, the percentage of facilities with a score of 10%-19.9%, etc., up to the percentage of facilities with a score of 90%-100%. The state then reports the percentage of facilities in each decile (i.e., in each ten-percentage-point range of scores). The state would also report the minimum, maximum and median facility scores. The median facility score is the middle score (i.e., there are an equal number of facility scores above and below the median).	Report information for the baseline and the most recent round of random inspections.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
12. Average facility score for compliance-related EBPIs➴ The percentage of all relevant compliance-related EBPIs being achieved, on average, by randomly sampled facilities.	<u>Baseline Random Inspections:</u> Effective Sample Size: Average Facility Score: Confidence Interval: Standard Deviation: <u>Post Random Inspections:</u> Effective Sample Size: Average Facility Score: Confidence Interval: Standard Deviation <u>Performance Change:</u> Effective Sample Size: Observed Difference Between Averages: Significant? 1-Sided or 2-Sided? P-Value: Confidence Interval:	*The state from the example above (for Measure 10) might report:* <u>Baseline Random Inspections:</u> 60.1% <u>Post-Round 1 Random Inspections:</u> 76.0% <u>Performance Change:</u> +15.9 percentage points *The state also reports key statistical metadata associated with each of these figures (see Appendix D).* *How the state calculated those results:* Using the example for the average facility score for all EBPIs, the state undertook the same approach to calculating and reporting, but only used the results from compliance-related EBPIs. For instance, if 5 of the 10 EBPIs were compliance-related, and a facility was achieving only 3 of the 5 compliance-related EBPIs but all of the other 5 EBPIs, that facility would get a facility score for compliance-related EBPIs of 60.0% (3/5), whereas its facility score for all EBPIs would be 80.0% (8/10).	• See notes under Measure 10 (Average facility score for all EBPIs). • In the example at left, the average facility score for compliance-related EBPIs is lower than the average facility score associated with all EBPIs (shown for Measure 10). This indicates that the average facility is doing better on its voluntary EBPIs than its compliance-related EBPIs at baseline and at post-certification. • Report whether changes are positive or negative.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ⩛ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
13. Distribution of facility scores for compliance-related EBPIs⩛ Distribution of facility scores for all compliance-related EBPIs among randomly sampled facilities.	<u>Baseline Random Inspections:</u> 0-9.9%: X% of facilities 10-19.9%: Y% 20-29.9%: Z% 30-39.9%: Etc. 40-49.9%: 50-59.9%: 60-69.9%: 70-79.9%: 80-89.9%: 90-100%: Total number of randomly sampled facilities: Minimum score: Maximum score: Median score: <u>Post-Round 1 Random Inspections:</u> 0-9.9%: A% of facilities 10-19.9%: B% 20-29.9%: C% 30-39.9%: Etc. 40-49.9%: 50-59.9%: 60-69.9%: 70-79.9%: 80-89.9%: 90-100%: Total number of randomly sampled facilities: Minimum score: Maximum score: Median score:	*The state from the example above (for Measure 10) might report the following results:* <u>Baseline Random Inspections:</u> • 0-9.9%: 6% of facilities • 10-19.9%: 9% • 20-29.9%: 4% • Etc. *The state would also provide a distribution for post-certification results.* *How the state calculated those results:* Using the example for the distribution of facility scores for all EBPIs, the state undertook the same approach to calculating and reporting, but based individual facility score calculations only upon the results from compliance-related EBPIs.	See notes under Measure 11 (Distribution of facility scores for all EBPIs).

KEY TO MEASURES: ✸ **= Tier 1, True Core;** ✧ *= Tier 2, Aspirational Measures;* ⩛ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
14. Average facility score for all compliance-related measures ⩛ The percentage of all relevant compliance-related measures being achieved, on average, by randomly sampled facilities.	<u>Baseline Random Inspections:</u> Effective Sample Size: Average Facility Score: Confidence Interval: Standard Deviation: <u>Post Random Inspections:</u> Effective Sample Size: Average Facility Score: Confidence Interval: Standard Deviation <u>Performance Change:</u> Effective Sample Size: Observed Difference Between Averages: Significant? 1-Sided or 2-Sided? P-Value: Confidence Interval:	*The state from the example above (for Measure 10) might report:* <u>Baseline Random Inspections:</u> 54.6% <u>Post-Round 1Random Inspections:</u> 71.6% <u>Performance Change:</u> +17.0 percentage points *The state also reports key statistical metadata associated with each of these figures (see Appendix D).* *How the state calculated those results:* Using the example for the average facility score for all EBPIs, the state undertook the same approach to calculating and reporting, but used results for all compliance-related EBPIs and other compliance-related measures tracked by inspectors. For instance, consider a situation where 5 of the 10 EBPIs were compliance-related, and the facility was achieving all 5 of these EBPIs. The facility was also achieving 50 out of 95 other relevant compliance-related practices. That facility was therefore achieving 55 out of 100 possible compliance-related practices, for a facility score of 55.0%.	• See notes under Measure 10 (Average facility score for all EBPIs). • Note that this measure will be the same as Measure 12 (Average facility score for compliance-related EBPIs) if all compliance-related measures tracked by a state are also EBPIs. However, compliance EBPIs for most states are typically a subset of all compliance-related measures. Consequently, the average facility score for all compliance-related measures will oftentimes be different than the average facility score for just compliance-related EBPIs. • This measure is recommended as a good complement to Measure 8, (Achievement rate across all compliance-related measures), which is also known as a traditional compliance rate. The traditional compliance rate reflects the percentage of facilities that are in full compliance with all measured issues. The average facility score for all compliance-related measures shows the percentage of measured issues with which the average facility is in compliance. As such, the average facility score may show gradations in compliance that are not detected when using the traditional compliance rate approach. For instance, take the example facility at left that was achieving 55 of 100 relevant compliance-related practices. This facility would be considered non-compliant for the purposes of a traditional compliance rate, but is almost certainly doing better than a facility achieving 5 of 100, and doing worse than a facility achieving 95 of 100. The average facility score takes these variations into account in helping to understand the extent to which both partial and full compliance are being achieved. • Report whether changes are positive or negative.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures;* ➢ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
15. Distribution of facility scores for all compliance-related measures➢ Distribution of facility scores for all compliance-related measures among randomly sampled facilities.	<u>Baseline Random Inspections:</u> 0-9.9%: X% of facilities 10-19.9%: Y% 20-29.9%: Z% 30-39.9%: Etc. 40-49.9%: 50-59.9%: 60-69.9%: 70-79.9%: 80-89.9%: 90-100%: Total number of randomly sampled facilities: Minimum score: Maximum score: Median score: <u>Post-Round 1 Random Inspections:</u> 0-9.9%: A% of facilities 10-19.9%: B% 20-29.9%: C% 30-39.9%: Etc. 40-49.9%: 50-59.9%: 60-69.9%: 70-79.9%: 80-89.9%: 90-100%: Total number of randomly sampled facilities: Minimum score: Maximum score: Median score:	*The state from the example above (for Measure 10) might report the following results:* <u>Baseline Random Inspections:</u> • 0-9.9%: 3% of facilities • 10-19.9%: 13% • 20-29.9%: 6% • Etc. *The state would also provide a distribution for post-certification results.* *How the state calculated those results:* Using the example for the distribution of facility scores for all EBPIs, the state undertook the same approach to calculating and reporting, but based the individual facility score calculations upon the results from all compliance-related measures (including both EBPIs and non-EBPI measures).	• See notes under Measure 11 (Distribution of facility scores for all EBPIs). • Note that this measure will be the same as Measure 13 (Distribution of facility scores for compliance-related EBPIs) if the only compliance-related measures tracked by a state are also EBPIs. However, most states draw a relatively small number of high-priority compliance-related EBPIs from a more complete list of all compliance-related measures, meaning the results for measures 13 and 15 will likely be different.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ⩛ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
16. Aggregate achievement rate for all EBPIs⩛ Percentage of all relevant EBPIs being achieved, across all randomly sampled facilities.	Baseline Random Inspections: Post Random Inspections: Performance Change:	*A state might report:* Baseline Random Inspections: 52.6% Post-Round 1 Random Inspections: 73.7% Performance Change: +21.1 percentage points *The state also reports key statistical metadata associated with each of these figures, if available (see Appendix D).* A state conducted 100 random inspections at baseline and 100 random inspections after certification. The state recorded facility performance on 10 EBPIs. If each EBPI had been relevant to all facilities, the state would have recorded 1,000 possible responses tracking EBPI achievement; however, 1 EBPI was only applicable at 50 of the facilities in each round, so there were 950 possible responses in each round. At baseline, the state found that facilities were achieving EBPIs 500 of those 950 possible times (52.6%). Post-certification, the state found EBPI achievement 700 times out of a possible 950 (73.7%). The state would report 52.6% as the observed baseline result, 73.7% as the observed post-certification result, and 21.1 percentage points (the difference between these 2 percentages) as the observed performance improvement. If available, the state also reports key statistical metadata associated with the percentage point change (see Appendix D), necessary for drawing inferences for the performance of the universe as a whole.	• Report rate for the most recent round of random inspections and for the baseline round of random inspections. • Report the change from baseline to the most recent round of inspections. Note that observed changes may reflect the cumulative impact of multiple ERP cycles. • The provision of statistical metadata for this measure is currently optional, because the Consortium recognizes that statistical calculations associated with this measure have not, to date, been common across ERPs. However, the Consortium encourages states to submit appropriate statistical metadata with this measure, if available. Such metadata can add to the credibility of reported ERP results. The Consortium intends to provide states with guidance on how to appropriately conduct certain statistical calculations for this measure. • This measure is complementary to the average facility score for all EBPIs (Measure 10). If a state's EBPIs apply to every facility, Measure 16 here will be mathematically equal to Measure 10; however, this is not typical because, generally, some EBPIs will not apply to some facilities. For example, suppose that roughly half of the the ERP universe included some relatively complex facilities for which all of 20 EBPIs were relevant, and the remaining half of facilities were relatively simple ones for which only a few EPBIs were relevant. In this case, the simple facilities might be achieving the vast majority of the EBPIs that apply to them, but the complex facilities might be achieving only a minority of their EPBIs. In this case, the aggregate achievement rate for all EBPIs could be substantially less than the average facility score. The differences between these 2 measures could suggest a more nuanced interpretation of the ERP data, in this case alerting policy makers that facilities greater numbers of EBPIs are performing more poorly than those facilities with smaller numbers of EBPIs. • Report whether changes are positive or negative.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ⩛ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
Measures of Environmental and Public Health Outcomes[2]			
17. Rate of managing/ controlling certain environmental aspects ✧ Amounts of emissions, waste, discharges, or chemicals/materials used that are properly controlled or managed (reported as a raw number and as percentage of total).	(for each medium) <u>Baseline Random Inspections:</u> <u>Post Random Inspections:</u> <u>Performance Change:</u>	*A state might report the following for baseline:* 50.0% of the 30,000 pounds of annual hazardous waste generated by randomly inspected facilities is being properly managed. *The state would also provide the following methodological information*: During baseline random inspections for our auto body ERP, our inspectors determined whether each facility was properly identifying, managing and disposing of all hazardous wastes. The inspectors also estimated each facility's annual amount of hazardous waste generated. We found that 60.0% of randomly inspected facilities were properly controlling or managing all of their hazardous waste. Taken together, these facilities generated 15,000 pounds of hazardous waste annually. The other 40.0% of facilities, which were found to be failing in some regard in terms of proper hazardous waste management, also generated a total of 15,000 pounds of hazardous waste annually. The estimated annual hazardous waste generation among all randomly inspected facilities is therefore 30,000 pounds. Consequently, we report that 50.0% of annual hazardous waste generated is being properly managed (even though 60.0% of inspected facilities are properly managing their waste). That 50.0% figure is calculated by dividing the amount of waste being properly managed (15,000 pounds) by the amount of waste being generated overall (30,000 pounds). We are not currently able to report statistical metadata (such as confidence level and confidence interval) associated with these random inspection results, so that population performance cannot be inferred. However, we do intend to conduct statistical analysis on this outcome measure as part of analyzing our post-certification results.	• Measure could reflect the total of a contaminant of concern at this type of facility, or a contaminant of concern related to a particular practice or process. Be very careful to explain that the measure is not based upon a facility-wide analysis, so as to not to misrepresent the data; also, provide an estimate of how much of the facility's contaminant of concern is *not* accounted for in the analysis, and the extent to which observed outcome changes may be counteracted by other processes (e.g., such as discussing whether performance is improving or declining on other practices related to this contaminant). • Report sufficient detail on the methodology, context, data sources and calculations so that reviewers can draw their own conclusions about the validity of the measure. For instance, be clear about the extent to which calculations are based upon inspector determinations, facility-provided information, and/or other sources. • Given the high profile of this kind of outcome measure, exercise special caution when reporting measures based upon data from random samples. Providing statistical metadata can lend credibility to reported results. If choosing to report results that are not statistically significant, seriously consider providing additional supporting information. For instance, implementation of RTC plans may provide evidence of a positive performance change, even when sample sizes are too small to show that the change is statistically significant.

[2] Measures related to environmental outcomes may not be possible for all sectors or for all ERPs, but they are top priorities for communicating the story of ERP. The Consortium sees these measures as aspirational and heartily encourages states to gather and report information on environmental outcome measures whenever possible. The Consortium provides limited guidance on these measures here but intends to provide supplemental guidance in the future.

KEY TO MEASURES: ✸ **= Tier 1, True Core;** ✧ *= Tier 2, Aspirational Measures;* ⍱ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
18. Level of group emissions, waste, discharges, or chemical usage related to certain environmental aspects ✧ Amounts of emissions, waste, discharges or materials/chemical usage (reported as a raw number and with a normalizing factor).		*Massachusetts could report an example like the following:* Statistical analysis suggests Massachusetts can be 95% confident that increases in the proportion of printers using UV ink have led to estimated annual VOC emissions reductions of between 157 and 8011 tons per year. *The state would also provide the following methodological information:* During baseline random inspections for our printers ERP, we observed that 1 of 32 sampled facilities (3.1%) were using so-called ultraviolet (UV) ink, which is cured with ultraviolet light and emits few or no volatile organic compounds (VOCs), a class of pollutants that can cause respiratory problems and often contributes to the formation of ground-level ozone. MA observed in our third round of random inspections that 5 of 22 sampled facilities (22.7%) were using UV ink. EPA used these results to estimate that the true proportion of printers using UV ink increased by between 1.0 and 38.2 percentage points (at a 95% confidence level). EPA used these lower and upper bounds as inputs into a VOC emissions estimator developed for Massachusetts by a contractor, the methodology of which is described in more detail in an attachment. The attachment also presents complete statistical metadata for the observed proportions of facilities using UV ink, as well as EPA's approach to converting VOC estimates from gallons to tons. Readers should recognize that there are other important potential sources of VOC emissions at these facilities, so this figure should not be taken to represent all changes in VOC emissions. *More information on this example can be found in Section 3.3 and the Appendix of EPA's ERP States Produce Results: 2007 Report. In the same place, readers can find another example of this kind of measure focusing on the total amount of a contaminant of concern (perchloroethylene) dry cleaning facilities use, emit, and dispose of -- rather than just the amount from 1 facility process.*	• Report a normalizing factor to indicate changes in production -- e.g., the number of facilities in the universe or a measure of total production across all facilities at the baseline and during the most recent round of inspections. The Consortium intends to provide future guidance on the use of normalization factors. • See additional notes under Measure 17 (Rate of managing/controlling certain environmental aspects).

KEY TO MEASURES: ✸ **= Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
19. Ecological, occupational and public health impacts related to certain environmental aspects of the group ✧ Impacts of the group on measures of ecological, occupational and public health (reported as a raw number and with a normalization factor).		*A state with an ERP targeting polluted runoff in a particular watershed might report:* We estimate that annual discharges of nitrogen by the targeted facilities have been reduced by between 5 tons and 35 tons. Using these figures and a recognized water quality impact model, we have estimated that resulting reductions in total nitrogen levels in Polluted Lake range from negligible to as much as 1,200 mcg per liter. Analysis of recent, limited water quality samples from Polluted Lake are consistent with these estimates: the samples show nitrogen levels 300 to 2,000 micrograms per liter lower than past known levels. These results provide some confidence that the lake is getting cleaner over time; more work is required to better correlate these results to our ERP activity and understand any potential role of factors unrelated to ERP. A detailed description of our observations and methodology (including our approach to normalizing the data) is provided as an attachment.	• Report a normalizing factor to indicate changes in production -- e.g., the number of facilities in the universe or a measure of total production across all facilities at the baseline and during the most recent round of inspections. The Consortium intends to provide future guidance on the use of normalization factors. • See additional notes under Measure 17 (Rate of managing/controlling certain environmental aspects).

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ⩛ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
Measures of Costs of Implementing ERP[3]			
20. Agency level of effort, first cycle ✧ Total number of staff hours expended in developing and implementing the first ERP cycle.	Level of Effort: Methodology for Calculating Level of Effort:	*A state ("State X") might report:* Level of Effort: 10,000 hours of staff time between January 2004 and August 2007. Methodology for Calculating Level of Effort: Based on timesheets of all state employees involved in the ERP. Prorated employee leave is not incorporated in this estimate; only actual hours worked are reflected. *How the state calculated that result:* The state has a timesheet system under which all employees must attribute time spent on a variety of projects. Early on in the ERP process, the state begins tracking ERP activity of all types -- policy, inspections, planning, analytical, data management, grant management, etc. The state finds that 10,000 hours of staff time were charged to timesheets from the beginning of ERP planning, through the post-certification random inspections and finishing the primary analysis of ERP results. To some degree, the choice of the end date is based on the judgment of the state as to when first-cycle ERP activities ended, especially when the ERP will be continuing into subsequent cycles. *Another state ("State Z") might report:* At least 10,000 hours of staff time between February 2006 and January 2008. Result is derived as follows: 3 staff people were assigned to work on ERP since inception. Our state does not formally track time spent on individual projects, but we are able to estimate that these individuals averaged about 75% of their time working on the ERP, and each staff person worked full time (2,080 hours per year, including leave time). In addition, we assigned 2 inspectors full time for 2 months to conduct baseline and post-certification inspections. Consequently, for the 2 years involved in planning and implementing the first round of ERP , we estimate that we invested	Provide background information on the following: • Kinds of activities tracked, • Starting and ending dates for the estimate (to allow an understanding of the amount of time over which level of effort was applied), • Approach to tracking level of effort (e.g., a formal timesheet system used by all employees, or a more informal approach), and • Any other key assumptions, such as whether leave time was included in the calculations.

[3] Measures related to ERP costs may not be possible for all sectors or for all ERPs, but they are top priorities for communicating the story of ERP. The Consortium sees these measures as "aspirational" and suggests states gather and report information on program costs, whenever possible. The Consortium provides limited guidance on these measures here but intends to provide supplemental guidance in the future.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ✓ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
		a minimum of 10,000 hours of staff time (3 workers x 75% time x 2 years x 2,080 hours plus 2 inspectors x 8 weeks x 40 hours per week). This estimate accounts for information technology support, which was provided by 1 of the 3 staff assigned to the project, but does not account for occasional involvement by senior management and our legal staff. Including those staffing categories could add up to 5% to the overall estimate.	
21. Agency level of effort, subsequent cycles ✧ Total number of staff hours expended in developing and implementing later ERP cycles being reported.	Level of Effort: Methodology for Calculating Level of Effort:	*State X from the example above (for Measure 20) might report:* Level of Effort: We utilized 4,000 hours of staff time in our second ERP cycle, which started in September 2007 with preparations for our round 2 facility certification and ended in August 2008 with analysis of the results of post-round 2 random inspections. Methodology for Calculating Level of Effort: This figure includes some staff time actually spent on additional analysis of the results from our first ERP cycle; for simplicity, we are attributing those hours to the second ERP cycle. The figure also does not represent incidental time spent on this ERP by senior management and our legal staff (but their involvement was significantly reduced relative to the first ERP cycle). Overall, actual hours spent may be approximately 5% lower the 4,000 hours reported. *How the state calculated that result*: The state has a timesheet system under which all employees must attribute time spent on a variety of projects. Early on in the ERP process, the state began tracking ERP activity of most types -- policy, inspections, planning, analytical, data management, grant management, etc.	• Avoid double counting. For instance, the first post-certification round of random inspections should be counted only toward the first ERP cycle, not the second ERP cycle as well. • See additional notes under Measure 20 (Agency level of effort, first cycle).

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
Miscellaneous Measures			
22. Other benefits of ERP✸ Are there any other benefits of ERP -- to your agency, to the public, to business etc. -- that you would like to share?		*Maryland could have reported the following for its auto repair ERP:* The Maryland Department of the Environment (MDE) focused its ERP on auto repair and body shops in the Park Heights neighborhood of Baltimore, a low-income area with a high minority population. Regulators were concerned about "environmental justice" issues in this neighborhood because shops were considered by many residents to be a blight on the community and were in close proximity to schools and facilities for health care, child care and elder care. Project partners randomly surveyed residents after the project to gauge how they felt about ERP's accomplishments. They found that 47% of a large sample of residents knew about the project, and 87% of those believed the project was working to improve the neighborhood's environmental conditions. Further, the survey indicated that residents perceived a number of neighborhood improvements over the timeframe of the project. These improvements include perceived reductions in discarded parts and tires, used oil spills, nighttime noise, smoke, and odors. (More details on the survey are found in MDE's attachments.) The success of this project shows that, while ERP may be most cost-effective on a larger scale, it can help create meaningful and noticeable improvements at the local level as well. *This example was presented on page 23 of EPA's ERP States Produce Results: 2007 Report; most of the text from that report is repeated verbatim here.*	Consider providing sufficient information that outsiders will understand the results you report here and be able to use them in telling stories about the effectiveness of ERP. For instance, quotes should ideally be provided with clear attribution of the person making the statement, the organization that person is affiliated with, that person's role in the organization, organization location, and the date. E.g., Jane Martinez, Owner, Auto Paint Works of Dover, Delaware (August 2005). Some examples of benefits or outcomes that states may want to report here include: • Support from businesses, using quotes from prominent trade associations or from individual business operators as evidence. • Examples of how the use of ERP tools may have been transformational for the way the agency approaches problems; e.g., by using random sampling to characterize problems before taking action. • Feedback from the public (in the form of quotes or survey results) on perceived benefits of ERP to the public.

KEY TO MEASURES: ✸ = **Tier 1, True Core;** ✧ = *Tier 2, Aspirational Measures*; ➴ = Tier 3, Additional Measures.

Measure Name and Definition	**Info Reported by State**	**Example** *[Delete this column before submitting template]*	**Notes** *[Delete this column before submitting template]*
23. Other key measures you have identified ✸ Open-ended placeholder question.		*Massachusetts could have reported:* Evidence suggests that data from ERP self-certification forms helps us effectively target follow-up inspections. A preliminary examination of 4 rounds of random and targeted inspections, across 3 different ERPs, revealed that enforcement actions resulted from 63% of targeted inspections, as opposed to 42% of random inspections -- a difference of 21 percentage points. The analysis did not specifically examine the seriousness of the violations, but anecdotal information from inspectors suggests that more serious violations are found during targeted inspections. We did not conduct statistical tests on or develop confidence intervals for these data. *This example was presented on page 20 of EPA's ERP States Produce Results: 2007 Report; some of the text from that report is repeated verbatim here.*	• This measure is intended to allow states to provide unanticipated, but useful information in each category. Please consider including results that are otherwise unreported but that you believe are important in communicating key outcomes of your ERP. • If applicable, report the measure (with statistical metadata, if applicable) for the baseline round of random inspections, for the most recent round of random inspections, and report the change (with statistical metadata) from the baseline to the most recent round of inspections.