EPA Capacity Assesment Report

March 28, 2022

EPA

Table of Contents

Overview

Introduction

The Foundations for Evidence-Based Policymaking Act (Evidence Act) requires Chief Financial Officer Act agencies to conduct a Capacity Assessment to appraise their ability and infrastructure to carry out evidence-building activities. The Environmental Protection Agency's (EPA's) approach to the Capacity Assessment can be broadly described in two phases:

- ✓ **The initial phase focuses on assessing EPA's ability to answer the priority questions in the Agency Learning Agenda.**
- ✓ **The second phase focuses on assessing EPA's skills, organizational structure, resources, expertise, and infrastructure to meet Agency Learning Agenda goals, as well as to implement the Evidence Act across the Agency.**

EPA's Current Context

EPA's ability to pursue its mission to protect human health and the environment depends upon the availability and quality of data and evidence that support and inform environmental policies, decisions, guidance, and regulations. As a science-based organization, EPA is committed to developing and using evidence to achieve its mission. Evidence-building activities are governed by a myriad of EPA and governmentwide policies, standards, and guidance to promote the quality, reliability, and accuracy of the information EPA develops and/or uses to inform policy and decision-making. These include (but are not limited to) EPA's Peer Review Policy and Handbook for internal and external review of scientific products, EPA's Information Quality Guidelines, EPA's Policy and Procedures on Protection of Human Subjects in EPA Conducted or Supported Research, EPA's Plan to Increase Access to Results of EPA-Funded Scientific Research, EPA's Guidelines for Preparing Economic Analysis, and EPA's Scientific Integrity Policy. EPA has also drafted a "Policy for Evaluations and Other Evidence-Building Activities" for release by April 2022.

The Evidence Act builds on longstanding principles of good governance and asks that agencies ensure the use of data and evaluation to support program performance and the improvement of operations. Relatedly, EPA has longstanding performance measurement efforts incorporated throughout the Agency's work. Performance measurement is a part of the Agency's strategic plan development, annual planning and budgeting, operations and implementation, and accountability and results processes to inform decision-making. The Agency also has a history of using Lean Kaizen tools integrated with performance measurement to advance a culture of using data and visual management to support business process improvement and day-to-day operations.

However, the Evidence Act provides an opportunity for EPA to reconsider its capacity to use evaluation, data, statistics, research, analysis, and other evidence-building activities to support policymaking.

In response to the Evidence Act, EPA seeks to reestablish a centralized evaluation function to support and coordinate Agency evaluations as well as to build capacity for evaluation and other Evidence Act activities across the organization. This Capacity Assessment will aid EPA's efforts to identify staffing and resource capabilities to implement the Evidence Act over the long-term. Identifying Agency strengths and opportunities for improvement will help set priorities, catalyze action, enable decisions that advance the robust use of data and evaluation, and support the routine development and use of evidence in decision-making.

INITIAL ASSESSMENT:

EPA's Ability to Answer the Questions in the Agency Learning Agenda

Status

EPA has initiated an assessment of the extent to which the agency has the necessary resources—expertise, capability, funding, data, technology, partners, organizations, and extramural vehicles—to answer the questions in three of the Agency's Learning Agenda priority areas.

Furthermore, EPA's understanding of its capacity to address the Learning Agenda's priority questions can facilitate a strategic approach to evaluation and evidence building and prioritize investments in resources and staff. As EPA assesses its capacity to address the priority questions and employ a variety of evidence-building activities, the Agency will consider the coverage, quality, methods, effectiveness, and independence of EPA's evaluation, data, statistics research, analysis, and other evidence-building efforts.

Early in the development of its learning agenda, EPA identified three priority areas: Drinking Water Systems Out of Compliance, Workforce, and Grant Commitments Met. A fourth priority area— Expanding EPA's Toolkit of Air Benefits Assessment Methodologies and Practices—started development after the survey was underway in 2021. Consequently, the findings described in the next section only applies to the first three priority areas. EPA will work closely with the fourth priority area workgroup to assess its capacity to answer priority questions at a later date.

Overview of Findings

Significant Progress to Date

Each workgroup has reported significant progress to date. The Drinking Water Systems Out of Compliance workgroup began answering priority question 1 and is in the preliminary stages of strategizing how to answer questions 2-5 (priority questions for the three learning priority areas are listed in the section that follows). The Workforce workgroup began answering priority questions 1 and 3 and is in the preliminary stages of gathering data to start answering question 4. The Grant Commitments Met workgroup made progress in answering question 1; questions 2 and 3 will be addressed in future years. Key areas in which all three workgroups have made progress include data collection and planning, specifically with regards to identifying staff and contractor needs and requesting resources through EPA's budget and planning processes to advance their work across the next several fiscal years (FY).

Challenges Identified

Data availability within EPA and uncertainty regarding data quality levels, were highlighted as challenges that persist for some priority questions. The data concerns are more significant, especially for the Drinking Water workgroup – the Drinking Water data needed is not owned by EPA, purchasable outside EPA, or required to be shared with EPA by the states. The Workforce and Grant Commitments Met workgroups will have more confidence in the data's sufficiency as they proceed further with data collection and analysis. An additional trend across several workgroups and priority questions is uncertainty regarding evaluation design and staffing – the workgroups had not yet determined their methodological approach to how evaluation will be conducted for some of their priority questions and were not confident they will have sufficient access to qualified internal staff, academics, and/or contractors to support evaluation effectively.

Initial Assessment Summary/Conclusions

Path Forward

EPA's approach was beneficial in offering deeper insight into the successes and challenges of executing the Learning Agenda. EPA will use these findings to develop action plans for the Drinking Water Systems Out of Compliance, Workforce, and Grant Commitments Met priority areas. The action plans will recommend solutions to address gaps in skills, capability, and capacity. The Expanding EPA's Toolkit of Air Benefits Assessment Methodologies and Practices workgroup will engage in a similar process to assess their ability to answer the priority questions and then creating an action plan to assist with execution. By providing the workgroups with an actionable framework to carry out the Learning Agenda, the Agency will make progress in developing a culture based in evidence and evaluation that fosters continuous learning and improvement.

Summary Findings for Each Learning Priority Area

The summary that follows presents the findings for each Learning Priority Area and next steps.

Learning Priority Area: Drinking Water Systems Out of Compliance

Priority Questions

1. To what extent does EPA have ready access to data to measure drinking water compliance reliably and accurately?
2. What factors determine system noncompliance and continuous compliance?
3. How can we determine if a system has the technical, managerial, and financial capacity to provide safe water on a continuous basis to its customers?
4. Does increased use of compliance assurance tools (inspections and enforcement) improve system compliance, and if so, under what circumstances?
5. What EPA oversight activities are effective at assessing and improving state programs' ability to drive compliance?

Overview

Summary:

The Drinking Water workgroup has made progress on priority question 1 but remains in the preliminary stages of strategizing on priority questions 2-5. They are broadly optimistic they can design appropriate evaluations over time but have significant concerns around data access and quality, as well as having access to qualified program evaluation staff. A key priority for this workgroup is engaging with the states to fulfill data needs.

Next Steps:

Moving forward, the Drinking Water workgroup will focus on accessing, analyzing, and generating the necessary data to answer the priority questions and will look to secure additional internal and external support to execute the Learning Agenda.

Learning Priority Area: Workforce

Priority Questions

1. To what extent does EPA have access to the tools and strategies needed to analyze and understand the Agency's near and long-term workforce needs?
2. What are the critical skills needed to support the Agency's mission, now and in the future?
3. What are the leading strategies to attract, recruit, grow, and retain a diverse and talented workforce? What makes people stay in the Agency long-term?
4. How can EPA ensure knowledge is transferred from outgoing to current and incoming staff to support succession planning?

Overview

Summary:

The Workforce workgroup has started answering priority questions 1 and 3 and gathering data for priority question 4. They are broadly optimistic they will have sufficient staffing to continue making progress but are less certain about their ability to access contractor support. A key priority for this workgroup is communications and change management.

Next Steps:

A broad challenge across this initiative is communications and change management, especially given all the high priority communications coming from Human Resources (e.g., return to the workplace). The Workforce workgroup is in consistent contact with the HR community to share information about this Learning Priority Area at the grassroots level. As they move forward, they want to put additional effort into socializing the initiative with staff members at all levels of the organization.

Learning Priority Area: Grant Commitments Met

Priority Questions

1. How do EPA's existing grant award and reporting systems identify and track grant commitments?
2. What EPA practices and tools (1) effectively track grantee progress towards meeting workplan grant commitments including outputs and outcomes and/or (2) support communication of national program level outputs and outcomes?
3. Are the commitments established in EPA's grant agreements achieving the intended environmental and/or human health results, particularly for environmental justice and underserved communities?

Overview

Summary:

The Grant Commitments Met workgroup has made progress answering question 1 by collecting and analyzing relevant data, leaving questions 2 and 3 to be addressed in the longer-term. The workgroup is broadly optimistic they can onboard qualified staff and contractor resources across the next few fiscal years; however, they may have additional human or technological resource needs that they have not yet accounted for since they are still in the early stages of evaluation design. A key priority for this workgroup is ensuring collaboration with various internal and external stakeholders throughout this effort.

Next Steps:

Now that the workgroup has collected the necessary data to answer question 1, they will focus on further analyzing the data and understanding the path forward for question 2. Moving forward, they plan to reach out to EPA offices to observe how different programs are conducted and learn more about which best practices the Agency should pursue. As they work through priority questions 2 and 3, they foresee challenges shifting from an output-orientation to an outcome-orientation, as well as determining how to demonstrate the environmental results EPA has achieved.

AGENCYWIDE ASSESSMENT:

Assessing EPA's Skills, Capability, and Capacity Based on a Maturity Model

Maturity Model Overview

Context

Maturity models assess a current state or level of effectiveness along with criteria for achieving the next desired level of performance. For EPA, a maturity model will serve as a roadmap to help establish an evidence-based culture where Agency decisions are informed by evidence, and performance is routinely evaluated for potential improvements. Stakeholder feedback and management buy-in are critical to ensuring that the maturity model will be actionable and can drive EPA towards achieving its desired state. Implementation of the maturity model will enable the Agency to take stock and chart a path forward to ensure it makes progress in critically important areas to EPA. Looking forward, EPA will pursue a holistic approach that integrates the requirements of the Evidence Act with strategic planning and budgeting, regulatory development, program management, scientific research, and continuous improvement efforts. This integration will reinforce the importance of each initiative and foster Agencywide long-term culture change.

EPA 's maturity model addresses five domains: Data Use, Evaluation, Research, Statistics, and Lean Management. For each domain, the maturity model considers dimensions such as coverage, quality, methods, independence, and effectiveness. The final maturity model is included in **Appendix A** of this document.

FY 2022 Actions

In FY 2022, EPA is piloting the maturity model in order to improve applicability and gather initial information about Agencywide organizational capacity. A key feature of this pilot is developing tools to supporting accurate assessment of maturity. Once the pilot is complete, EPA will analyze the results and determine how to proceed with broader implementation.

The Role of the Data Skills Assessment Survey

Data Skills Assessment

Since data and analytics are increasingly becoming part of everyday business, with different jobs requiring different types of data skills, scientists and data specialists may require advanced technical skills to

support data gathering, conversion, cleansing, and analysis; while non-specialists often need to interpret data, communicate its importance, and use it to make decisions.

In support of the Evidence Act and Federal Data Strategy requirements, EPA launched an Agencywide Data Skills Survey in April 2021 to gain input on staff use of data to perform their work. The survey was designed to identify strengths and gaps related to critical data skills, assess staff capacity for those skills, and take actions to ensure its workforce is prepared to support evidence-building activities. In addition to skills questions, questions regarding attitudes and perceptions of EPA's overall culture with respect to data were included in the survey. The survey consisted of the following six categories:

- Respondent Office, Role, and Data Responsibilities
- Awareness of Options to Access, Share, and Manage Data
- Skills to Interpret Data and/or Analysis
- Skills to Visualize Data
- Skills to Communicate
- Organization Value/Use of Evidence

A total of 2,665 EPA staff completed the survey. Of this number 2,015 answered all the questions while 650 completed a portion (one percent – 98 percent). The current analysis only includes completed surveys.

Survey Results

Approximately 97% of survey responders (1952 out of 2015) responders use data or data and information in their work. After comparing data-focused responses to the skill level criteria, about 22% (434 out of 1952) of data-focused responders met the criteria for Level 1 – Novice (can complete simple, well-defined tasks with instruction or guidance) or above. The remaining 1518 responders did not meet the criteria to attain the Novice skill level.

Approximately 10% of data-focused responders (193 out of 1952) met only the criteria for Level 1 – Novice. Of the more highly skilled groups, approximately 10% of data-focused responders (193 out of 1952) met the criteria for Level 2 – Savvy and approximately 2% of data-focused responders (48 out of 1952) met the criteria for Level 3 – Advanced. Savvy responders can complete and assist others to complete tasks/problems on their own. Advanced responders can complete or assist others to complete complex problems and tasks. The finding that only 22% of data-focused responders were able to meet the Novice or better skill level may indicate that additional training or communication on data analysis concepts and evidence-based decision making may be useful for a wide range of EPA data-focused staff.

EPA recognizes that defining expectations for data skill attainment is a significant effort that goes beyond the scope of this initial survey. Refinement of Agency-thinking on data skill expectations will continue to evolve and mature over time. The data skill levels used in this survey are an initial step to begin discussions in FY 2022 but are not intended to provide a definitive analysis of EPA staffs' capacity for data analysis.

Summary Observations

The implementation of the initial phase of the capacity assessment was instrumental in highlighting common and unique challenges across the learning priority area workgroups. The process served to raise internal awareness regarding available capacity, skills, and expertise. The resulting action plans will provide the workgroups with an actionable framework to execute the Learning Agenda.

EPA has made steady progress in implementing the second phase of our capacity assessment. The maturity model has been developed and piloting the maturity model in FY 2022 is the next critical step in the process.

Last, the results of the Data Skills Survey show that data-focused responders at all skill levels recognize the importance of data skills to their work. For example, 75% of Novice responders, 97% of Data Savvy responders, and 100% of Advanced responders agree data skills are important. However, differences in responses between answers to survey questions by skill level provide EPA with an initial understanding of staff capacity to perform different types of data analysis skills and highlight opportunities that the Agency may consider taking to increase attainment of these skills. The results of the Data Skills survey will provide valuable information in helping to provide a baseline against which to measure progress in assessing the Data Use maturity model.

Collectively, these experiences have helped to raise agency awareness about the Evidence Act to internal audiences. It has also helped the Agency gain much needed experience assessing our capabilities and understanding the effort, time and cost required to implement the Evidence Act. Equally important is the value in helping internal audiences see the connection between how evidence building activities can help achieve the Agency's mission.

Agency Evaluation and Evidence Building Activities

As part of the capacity assessment, the Evidence Act requires federal agencies to include a list of activities and operations currently being evaluated and analyzed. EPA's list of planned evaluations for FY 2022 is included in **Appendix B** of this document. This list was compiled in conjunction with the development of EPA's FY 2022 Congressional Justification (CJ). In support of the process, EPA issued a data call to all EPA offices from April – May 2021 requesting National Program Managers (NPMs) and Regional Offices identify all significant planned evaluations and other evidence-building activities that will be initiated in FY 2022.

Evaluations and evidence building activities were defined consistent with definitions included in the Evidence Act and OMB A-11. EPA defined significant as activities: supporting an Administration or other Agency priority; mandated by Congress; and/or highlighted as a priority for resource allocation.

Appendix A : Maturity Model by Domain

Data Use

Data Use ensures the right people are aware of, have appropriate access to, and have the necessary tools and skills to analyze and interpret the data they need to inform policy or programmatic decisions.

Level 1	Level 2	Level 3	Level 4	Level 5
Appropriate data **RARELY** (0-10%)[1] reflect programmatic and policy priorities and are **RARELY** (0-10%) available to support programmatic and policy decisions. Data are **RARELY** (0-10%) developed to a consistent standard. Appropriate staff **RARELY** (0-10%) have the necessary tools and skills to analyze and interpret data in a way that can inform programmatic or policy decisions.	Appropriate data **INFREQUENTLY** (11-50%) reflect programmatic and policy priorities and are **INFREQUENTLY** (11-50%) available to support programmatic and policy decisions. Data are **INFREQUENTLY** (11-50%) developed to a consistent standard. Appropriate staff **INFREQUENTLY** (11-50%) have the necessary tools and skills to analyze and interpret data in a way that can inform programmatic or policy decisions.	Appropriate data **FREQUENTLY** (51-75%) reflect programmatic and policy priorities and are **FREQUENTLY** (51-75%) available to support programmatic and policy decisions. Data are **FREQUENTLY** (51-75%) developed to a consistent standard. Appropriate staff **FREQUENTLY** (51-75%) have the necessary tools and skills to analyze and interpret data in a way that can inform programmatic or policy decisions.	Appropriate data **ROUTINELY** (76-90%) reflect programmatic and policy priorities and are **ROUTINELY** (76-90%) available to support programmatic and policy decisions. Data are **ROUTINELY** (76-90%) developed to a consistent standard. Appropriate staff **ROUTINELY** (76-90%) have the necessary tools and skills to analyze and interpret data to inform programmatic or policy decisions.	Appropriate data **ALMOST ALWAYS** (91-100%) reflect programmatic and policy priorities and are **ALMOST ALWAYS** (91-100%) available to support programmatic and policy decisions. Data are **ALMOST ALWAYS** (91-100%) developed to a consistent standard. Appropriate staff **ALMOST ALWAYS** (91-100%) have the necessary tools and skills to analyze and interpret data to inform programmatic or policy decisions.

[1] For the purposes of the maturity model, frequency is to be measured in terms of the rate at which a given attribute's criteria point is performed/realized across an office's *total* workload portfolio (rather than, for example, a specific subset of that total workload).

Evaluation

Evaluation[2] is an assessment using systematic data collection and analysis of one or more programs, policies, and organizations. The purpose of evaluation is to make recommendations to improve, advance, or modify existing programs, policies, projects, or operations.[3]

	Level 1	Level 2	Level 3	Level 4	Level 5
Coverage	Evaluation activities **RARELY** (0-10%) reflect policy and programmatic priorities of the agency and have the potential to impact those priorities. Activities **RARELY** (0-10%) address high priority questions and serve information needs of EPA and EPA's stakeholders.	Evaluation activities **INFREQUENTLY** (11-50%) reflect policy and programmatic priorities of the agency and have the potential to impact those priorities. Activities **INFREQUENTLY** (11-50%) address high priority questions and serve information needs of EPA and EPA's stakeholders.	Evaluation activities **FREQUENTLY** (51-75%) reflect policy and programmatic priorities of the agency and have the potential to impact those priorities. Activities **FREQUENTLY** (51-75%) address high priority questions and serve information needs of EPA and EPA's stakeholders.	Evaluation activities **ROUTINELY** (76-90%) reflect policy and programmatic priorities of the agency and have the potential to impact those priorities. Activities **ROUTINELY** (76-90%) address high priority questions and serve information needs of EPA and EPA's stakeholders.	Evaluation activities **ALMOST ALWAYS** (91-100%) reflect policy and programmatic priorities of the agency and have the potential to impact those priorities. Activities **ALMOST ALWAYS** (91-100%) address high priority questions and serve information needs of EPA and EPA's stakeholders.
Type	Little to no formal evaluation work. Ad hoc descriptive studies[4]	Ad hoc Process or Implementation Evaluation[5]	Regular investments in Process or Implementation Evaluation; Rare/Ad hoc Outcome Evaluation[6]	Regular investments in Process or Implementation, Outcome Evaluation; Rare/Ad hoc Impact Evaluation[7]	Regular investments in Process or Implementation, Outcome Evaluation, and Impact Evaluation

[2] For the purposes of this work, "program evaluation" and "evaluation" are synonymous. Evaluations may address questions related to the implementation or institution of a program, policy, or organization; the effectiveness or impact of specific strategies related to or used by a program, policy, or organization; and/or factors that relate to variability in the effectiveness of a program, policy, or organization or strategies of these. Evaluations can also examine questions related to understanding the contextual factors surrounding a program, as well as how to effectively target specific populations or groups for a particular intervention.
[3] Program evaluation standards, and associated definitions, can be found in OMB M-20-12.
[4] *Descriptive Studies* can be quantitative or qualitative, and seek to describe a program, policy, organization, or population without inferring causality or measuring effectiveness.
[5] *Process or Implementation Evaluation* assesses how the program or service is delivered relative to its intended theory of change, and often includes information on content, quantity, quality, and structure of services provided.
[6] *Outcome Evaluation* measures the extent to which a program, policy, or organization has achieved its intended outcome(s) and focuses on outputs and outcomes to assess effectiveness. Unlike impact evaluation, it cannot discern causal attribution but is complementary to performance measurement.
[7] *Impact Evaluation* assesses the causal impact of a program, policy, or organization, or aspect of them on outcomes, relative to a counterfactual. In other words, this type of evaluation estimates and compares outcomes with and without the program, policy, or organization, or aspect thereof. Impact evaluations include both experimental (i.e., randomized controlled trials) and quasi-experimental designs.

	Level 1	Level 2	Level 3	Level 4	Level 5
Methods/ Appropriateness	Do not have process or have processes which are **RARELY** (0-10%) followed to reduce risks associated with the adoption of inappropriate methods or selective reporting of findings, and instead promote accountability for reporting methods and findings.	Have processes which are **INFREQUENTLY** (11-50%) followed to reduce risks associated with the adoption of inappropriate methods or selective reporting of findings, and instead promote accountability for reporting methods and findings.	Have processes which are **FREQUENTLY** (51-75%) followed to reduce risks associated with the adoption of inappropriate methods or selective reporting of findings, and instead promote accountability for reporting methods and findings.	Have processes which are **ROUTINELY** (76-90%) followed to reduce risks associated with the adoption of inappropriate methods or selective reporting of findings, and instead promote accountability for reporting methods and findings.	Have processes which are **ALMOST ALWAYS** (91-100%) followed to reduce risks associated with the adoption of inappropriate methods or selective reporting of findings, and instead promote accountability for reporting methods and findings.
Documentation	Evaluation activities **RARELY** (0-10%) ensure evaluation's design and methods are available in sufficient detail to achieve rigor, transparency, and credibility	Evaluation activities **INFREQUENTLY** (11-50%) ensure evaluation's design and methods are available in sufficient detail to achieve rigor, transparency, and credibility.	Evaluation activities **FREQUENTLY** (51-75%) ensure evaluation's design and methods are available in sufficient detail to achieve rigor, transparency, and credibility.	Evaluation activities **ROUTINELY** (76-90%) ensure evaluation's design and methods are available in sufficient detail to achieve rigor, transparency, and credibility.	Evaluation activities **ALMOST ALWAYS** (91-100%) ensure evaluation's design and methods are available in sufficient detail to achieve rigor, transparency, and credibility.
Quality	Evaluation activities **RARELY** (0-10%) meet the standards of Relevance and Utility, Rigor, Objectivity, Transparency, and Ethics.	Evaluation activities **INFREQUENTLY** (11-50%) meet the standards of Relevance and Utility, Rigor, Objectivity, Transparency, and Ethics.	Evaluation activities **FREQUENTLY** (51-75%) meet the standards of Relevance and Utility, Rigor, Objectivity, Transparency, and Ethics.	Evaluation activities **ROUTINELY** (76-90%) meet the standards of Relevance and Utility, Rigor, Objectivity, Transparency, and Ethics.	Evaluation activities **ALMOST ALWAYS** (91%-100%) meet the standards of Relevance and Utility, Rigor, Objectivity, Transparency, and Ethics.
Practicality	**RARELY** (0-10%) ensures that findings from evaluations and other evidence-building activities can be to be acted upon or implemented in a timely fashion.	**INFREQUENTLY** (11-50%) ensures that findings from evaluations and other evidence-building activities can be acted upon or implemented in a timely fashion.	**FREQUENTLY** (51-75%) ensures that findings from evaluations and other evidence-building activities can be acted upon or implemented in a timely fashion.	**ROUTINELY** (76-90%) ensures that findings from evaluations and other evidence-building activities can be acted upon or implemented in a timely fashion.	**ALMOST ALWAYS** (91-100%) ensures that findings from evaluations and other evidence-building activities can be acted upon and implemented in a timely fashion.

	Level 1	Level 2	Level 3	Level 4	Level 5
Use	Information **RARELY** (0-10%) informs agency decision making in key areas such as budgeting, program improvement, accountability, management, rulemaking, and policy development.	Information **INFREQUENTLY** (11-50%) informs agency decision making in key areas such as budgeting, program improvement, accountability, management, rulemaking, and policy development.	Information **FREQUENTLY** (51-75%) informs agency decision making in key areas such as budgeting, program improvement, accountability, management, rulemaking, and policy development.	Information **ROUTINELY** (76-90%) informs agency decision making in key areas such as budgeting, program improvement, accountability, management, rulemaking, and policy development.	Information **ALMOST ALWAYS** (91-100%) informs agency decision making in key areas such as budgeting, program improvement, accountability, management, rulemaking, and policy development.
Independence	**RARELY** (0-10%) ensures the independence and objectivity of personnel conducting and managing evaluations and other evidence-building activities.	**INFREQUENTLY** (11-50%) ensures the independence and objectivity of personnel conducting and managing evaluations and other evidence-building activities.	**FREQUENTLY** (51-75%) ensures the independence and objectivity of personnel conducting and managing evaluations and other evidence-building activities.	**ROUTINELY** (76-90%) ensures the independence and objectivity of personnel conducting and managing evaluations and other evidence-building activities.	**ALMOST ALWAYS** (91-100%) ensures the independence of personnel conducting and managing evaluations and other evidence-building activities.
Objectivity[8]	Staff tasked with evaluation activities **RARELY** (0-10%) strive for objectivity in the planning and conduct of evaluations and evidence-building activities, and in the interpretation and dissemination of findings.	Staff tasked with evaluation activities **INFREQUENTLY** (11-50%) strive for objectivity in the planning and conduct of evaluations and evidence-building activities, and in the interpretation and dissemination of findings.	Staff tasked with evaluation activities **FREQUENTLY** (51-75%) strive for objectivity in the planning and conduct of evaluations and evidence-building activities, and in the interpretation and dissemination of findings.	Staff tasked with evaluation activities **ROUTINELY** (76-90%) strive for objectivity in the planning and conduct of evaluations and evidence-building activities, and in the interpretation and dissemination of findings.	Staff tasked with evaluation activities **ALMOST ALWAYS** (91-100%) strive for objectivity in the planning and conduct of evaluations and evidence-building activities, and in the interpretation and dissemination of findings.

[8] See OMB guidance M-20-12 Program Evaluation Standards and Practices for federal standards of objectivity in program evaluation practices, e.g.: "…Evaluators should strive for objectivity in the planning and conduct of evaluations and in the interpretation and dissemination of findings, avoiding conflicts of interest, bias, and other partiality."

Research

Research and development activities are defined as creative and systematic work undertaken to develop new data, information, and technologies to support credible decision-making to safeguard human health and ecosystems from environmental pollutants and to enable implementation of programs and policies designed for this purpose. These activities involve both environmental and public health research to better understand and characterize the risks associated with exposure to environmental pollutants; sources, fate, and transport of pollutants in the environment; and solutions to monitor, prevent or mitigate environmental pollutant exposures. Further, agency decision making also include social science and economic research and analysis regarding policy options and decision making.

	Level 1	Level 2	Level 3	Level 4	Level 5
Coverage	Research and development planning not informed by internal or external to the Agency stakeholders and has no external scientific expert review; therefore, research is not ensured to support the agency's strategic goals and objectives.	Research and development planning informed by internal Agency stakeholders but not external stakeholders and has no external scientific expert review; therefore, research is not ensured to support the agency's strategic goals and objectives.	Research and development planning informed by internal and external to the Agency stakeholders but has no external scientific expert review; therefore, research is not ensured to support the agency's strategic goals and objectives.	Research and development planning informed by internal and external to the Agency stakeholders but is only informally reviewed by external scientific experts; therefore, research should support the agency's strategic goals and objectives but lack rigor.	Planning of research and development activities informed by internal and external to the Agency stakeholders with formal external scientific expert review; therefore, research should support the agency's strategic goals and objectives.
Quality	Research and development activities are planned and conducted such that they **RARELY** (0-10%) meet Agency quality policy requirements (CIO 2105.1) to ensure that Agency work products are accurate, traceable, reproducible, and defensible.[9]	Research and development activities are planned and conducted such that they **INFREQUENTLY** (11-50%) meet Agency quality policy requirements (CIO 2105.1) to ensure that Agency work products are accurate, traceable, reproducible, and defensible.	Research and development activities are conducted such that they **FREQUENTLY** (51-75%) meet Agency quality policy requirements (CIO 2105.1) to ensure that Agency work products are accurate, traceable, reproducible, and defensible.	Research and development activities are conducted such that they **ROUTINELY** (76-90%) meet Agency quality policy requirements (CIO 2105.1) to ensure that Agency work products are accurate, traceable, reproducible, and defensible.	Research and development activities are conducted such that they **ALMOST ALWAYS** (91-100%) meet Agency quality policy requirements (CIO 2105.1) to ensure that Agency work products are accurate, traceable, reproducible, and defensible.

[9] Term definitions provided in the glossary below.

	Level 1	Level 2	Level 3	Level 4	Level 5
Methods	Prior to public release, research and development products do not require Quality Assurance review, line management approval, internal (to the Agency) scientific peer review, or external (to the Agency) scientific peer review.	Prior to public release, research and development products require internal (to the Agency) scientific peer review; but Quality Assurance review, line management approval, and external (to the Agency) scientific peer review are not required.	Prior to public release, research and development products require line management approval and internal (to the Agency) scientific peer review, but Quality Assurance review and external (to the Agency) peer review are not required.	Prior to public release, research and development products require Quality Assurance review, line management approval, and internal (to the Agency) scientific peer review, but external (to the Agency) scientific peer review is not required.	Prior to public release, research and development products require Quality Assurance review, line management approval, internal (to the Agency) review, and external (to the Agency) scientific peer review.
Effectiveness	Research and development products **RARELY** (0-10%) meet the needs of identified stakeholders (i.e., Partner Agencies, Program Offices, States, Tribes, Communities, NGOs)	Research and development products **INFREQUENTLY** (11-50%) meet the needs of identified stakeholders (i.e., Partner Agencies, Program Offices, States, Tribes, Communities, NGOs)	Research and development products **FREQUENTLY** (51-75%) meet the needs of identified stakeholders (i.e., Partner Agencies, Program Offices, States, Tribes, Communities, NGOs)	Research and development products **ROUTINELY** (76-90%) meet the needs of identified stakeholders (i.e., Partner Agencies, Program Offices, States, Tribes, Communities, NGOs)	Research and development products **ALMOST ALWAYS** (91-100%) meet the needs of identified stakeholders (i.e., Partner Agencies, Program Offices, States, Tribes, Communities, NGOs)
Independence	Research and development activities and results **RARELY** (0-10%): adhere to human subject research standards; are free from inappropriate influence; follow Scientific Integrity policy; and **RARELY** (0-10%) have appropriate levels of internal and external review and clearance.	Research and development activities and results **INFREQUENTLY** (11-50%): adhere to human subject research standards; are free from inappropriate influence; follow Scientific Integrity policy; and **INFREQUENTLY** (10-50%) have appropriate levels of internal and external review and clearance.	Research and development activities and results **FREQUENTLY** (51-75%): adhere to human subject research standards; are free from inappropriate influence; follow Scientific Integrity policy; and **FREQUENTLY** (50-75%) have appropriate levels of internal and external review and clearance.	Research and development activities and results **ROUTINELY** (76-90%): adhere to human subject research standards, are free from inappropriate influence, follow Scientific Integrity policy; and **ROUTINELY** (75-90%) have appropriate levels of internal and external review and clearance.	Research and development activities and results **ALMOST ALWAYS** (91-100%): adhere to human subject research standards; are free from inappropriate influence; follow Scientific Integrity policy; and **ALMOST ALWAYS** (90-100%) have appropriate levels of internal and external review and clearance.

Glossary

Accuracy: The degree of agreement between an observed value and an accepted reference value. Accuracy includes random error (precision) and systemic error (bias or recovery) that are caused by sampling and analysis. A data quality indicator.

Defensible: The ability to withstand any reasonable challenge related to the veracity or integrity of laboratory documents and derived data.

Reproducibility: Obtaining consistent results using the same input data, computation steps, methods, and code, and conditions of analysis.

Traceable: When a measurement result can be related to appropriate standards, generally national or international standards, through an unbroken chain of comparisons.

Statistics

Statistics and statistical activities are the collection, compilation, processing, or analysis of data from a sample of a population for the purpose of describing or making estimates concerning that population. This includes the development of methods or resources that support those activities. Statistical evidence is the information produced from statistical activities.

		Level 1	Level 2	Level 3	Level 4	Level 5
Coverage	Strategic Goals	Statistical activities and the development of statistical evidence **DO NOT or RARELY** support the agency's strategic goals and objectives.	Statistical activities and the development of statistical evidence **INFREQUENTLY** support the agency's strategic goals and objectives.	Statistical activities and the development of statistical evidence **FREQUENTLY** support the agency's strategic goals and objectives.	Statistical activities and the development of statistical evidence **ROUTINELY** support the agency's strategic goals and objectives.	Statistical activities and the development of statistical evidence **ALMOST ALWAYS** support the agency's strategic goals and objectives.
	Decision-Making	Statistical activities and the development of statistical evidence **ARE NOT or ARE RARELY** available to use for operational, management, and policy decision-making.	Statistical activities and the development of statistical evidence are **INFREQUENTLY** available to use for operational, management, and policy decision-making.	Statistical activities and the development of statistical evidence are **FREQUENTLY** available to use for operational, management, and policy decision-making.	Statistical activities and the development of statistical evidence are **ROUTINELY** available to use for operational, management, and policy decision-making.	Statistical activities and the development of statistical evidence are **ALMOST ALWAYS** available to use for operational, management, and policy decision-making.
Quality	Data Quality	Statistical activities **DO NOT or RARELY** meet data quality standards (relevant, accurate, timely, and credible).	Statistical activities **INFREQUENTLY** meet data quality standards (relevant, accurate, timely, and credible).	Statistical activities **FREQUENTLY** meet data quality standards (relevant, accurate, timely, and credible).	Statistical activities **ROUTINELY** meet data quality standards (relevant, accurate, timely, and credible).	Statistical activities **ALMOST ALWAYS** meet data quality standards (relevant, accurate, timely, and credible).
	Transparency	Statistical activities and the development of statistical evidence **ARE NOT** or **ARE RARELY** transparent, including with respect to methods and data quality.	Statistical activities and the development of statistical evidence are **INFREQUENTLY** transparent, including with respect to methods and data quality.	Statistical activities and the development of statistical evidence are **FREQUENTLY** transparent, including with respect to methods and data quality.	Statistical activities and the development of statistical evidence are **ROUTINELY** transparent, including with respect to methods and data quality.	Statistical activities and the development of statistical evidence are **ALMOST ALWAYS** transparent, including with respect to methods and data quality.

		Level 1	Level 2	Level 3	Level 4	Level 5
Methods		Statistical activities and the development of statistical evidence **DO NOT or RARELY** employ appropriate AND rigorous methodological approaches.	Statistical activities and the development of statistical evidence **INFREQUENTLY** employ appropriate AND rigorous methodological approaches.	Statistical activities and the development of statistical evidence **FREQUENTLY** employ appropriate AND rigorous methodological approaches.	Statistical activities and the development of statistical evidence **ROUTINELY** employ appropriate AND rigorous methodological approaches.	Statistical activities and the development of statistical evidence **ALMOST ALWAYS** employ appropriate AND rigorous methodological approaches.
Effectiveness	Program Output	Statistical activities and the development of statistical evidence **DO NOT or RARELY** support the agency's program outcomes.	Statistical activities and the development of statistical evidence **INFREQUENTLY** support the agency's program outcomes.	Statistical activities and the development of statistical evidence **FREQUENTLY** support the agency's program outcomes.	Statistical activities and the development of statistical evidence **ROUTINELY** support the agency's program outcomes.	Statistical activities and the development of statistical evidence **ALMOST ALWAYS** support the agency's program outcomes.
	Intended Use	Statistical activities and the development of statistical evidence **DO NOT or RARELY** meet their intended outcomes, including serving the needs of stakeholders and being disseminated publicly.	Statistical activities and the development of statistical evidence **INFREQUENTLY** meet their intended outcomes, including serving the needs of stakeholders and being disseminated publicly.	Statistical activities and the development of statistical evidence **FREQUENTLY** meet their intended outcomes, including serving the needs of stakeholders and being disseminated publicly.	Statistical activities and the development of statistical evidence **ROUTINELY** meet their intended outcomes, including serving the needs of stakeholders and being disseminated publicly.	Statistical activities and the development of statistical evidence **ALMOST ALWAYS** meet their intended outcomes, including serving the needs of stakeholders and being disseminated publicly.
Independence	Bias & Oversight	Statistical activities **DO NOT or RARELY** have the appropriate levels of internal and external oversight.	Statistical activities **INFREQUENTLY** have appropriate levels of internal and external oversight.	Statistical activities **FREQUENTLY** have appropriate levels of internal and external oversight.	Statistical activities **ROUTINELY** have appropriate levels of internal and external oversight.	Statistical activities **ALMOST ALWAYS** have appropriate levels of internal and external oversight.
	Accountability & Controls	Science Integrity and Data policies **DO NOT or RARELY** identify accountabilities and controls for maintaining independence and objectivity in statistical activities and statistical evidence.	Science Integrity and Data policies **INFREQUENTLY** identify accountabilities and controls for maintaining independence and objectivity in statistical activities and statistical evidence.	Science Integrity and Data policies **FREQUENTLY** identify accountabilities and controls for maintaining independence and objectivity in statistical activities and statistical evidence.	Science Integrity and Data policies **ROUTINELY** identify accountabilities and controls for maintaining independence and objectivity in statistical activities and statistical evidence.	Science Integrity and Data policies **ALMOST ALWAYS** identify accountabilities and controls for maintaining independence and objectivity in statistical activities and statistical evidence.

Lean Management

Lean management is an approach to managing an organization that supports continuous improvement by using Lean principles and tools paired with routine measurement, visual management and regular engagement between management and staff to identify and solve problems, realize, and sustain process improvements, and more effectively achieve agency priorities.

		Level 1	Level 2	Level 3	Level 4	Level 5
Leader Behaviors	Senior Leader	Senior leaders (AA/RA) do not perform, or infrequently engage in leader behaviors such as operation site visits (e.g., Gemba walks) and business reviews.	Senior leaders engage in leader behaviors. For example, they perform regular operations site visits (e.g., Gemba walks) and participate in business reviews at least monthly.	Senior leaders lead regular business reviews to discuss organizational performance measures, developing metrics and targets to assess performance as appropriate.	Senior leaders conduct regular deep dives on specific topics at business review meetings to focus on and help improve organizational performance.	Senior leaders use data and evidence to support continuous improvement and to achieve organizational priorities and mission critical goals.
	Mid-Level Manager	Mid-level managers do not perform, or infrequently perform, operations site visits to assess or review the process (e.g., Gemba walks).	Mid-level managers perform regular operations site visits (e.g., Gemba walks) at least once a week.	In addition to performing regular site visits, mid-level managers and subordinate staff (e.g., branch chiefs) meet regularly around a leader performance board which tracks each process's goals, metrics, and performance and covers all visual management tools (e.g., flow and performance boards) within their division/unit/office.	Mid-level managers lead advanced process reviews with their subordinate staff (e.g., branch chiefs) to commit to accomplishing a priority goal.	Mid-level managers use operations site visits, visual management, process reviews, and data and evidence from their leader performance board to prioritize and attain goals for processes that support EPA's mission.

		Level 1	Level 2	Level 3	Level 4	Level 5
	Front-Line Manager	Managers do not attend, or infrequently attend, weekly huddles with their teams and have not identified a process for improvement.	Managers attend huddles with their team and work to improve at least one process for which the team is responsible.	Managers regularly attend huddles and choose lean management system (LMS) tools to address opportunities for improvement. (E.g., convene problem-solving by team and proving-solving guides when performance targets are missed.)	Managers regularly attend huddles with team to review LMS tools used (e.g., action registry, countermeasure form, etc.,) and monitor status of countermeasure implementation.	Leaders and managers at every level of EPA regularly attend huddles / business reviews around schedule (see what processes are on/off track) -better (engage in efforts to improve one thing) action -(use problem solving and data driven solution) tied to their core work, problem solving, and A3 projects.
Visual Management		No visual management (VM) (or other appropriate tools) exists; data are not captured.	Visual management (or other appropriate tools) exists; data are captured and used.	Visual management (or other appropriate tools) includes all the necessary components to facilitate its utility; data are captured consistently.	Visual management (or other appropriate tools) is used consistently; data are captured routinely and used to identify issues, engage in problem solving, and process improvement.	Visual management and captured data engage all levels of management, increase transparency with agency stakeholders, and connect the agency's mission, organizational goals, and priorities.
Standard Process		No or very little documentation of the process and its steps exist.	The Team has documented the key milestones which includes timeframes for example in the form of a standard operating procedure (SOP).	Visual management and documented instructions (e.g., SOP) cover process steps thoroughly and describe what success looks like (e.g., target levels of performance, timeframes). Process steps are completed using the documented approach.	Team uses standard work consistently and gathers data to improve and standardize additional processes. Desired levels of performance maintained using standard work.	Team regularly uses and revises standard work, incorporating best practices from across the agency and in industry, to achieve priority agency goals and mission. Conducts data driven analysis of process performance and makes improvements to effectively achieve mission.

		Level 1	Level 2	Level 3	Level 4	Level 5
Cascading Performance Measures	Senior Leader	Senior leaders do not leverage, or infrequently leverage, Bowling Charts during Monthly Business Reviews (MBRs).	Senior leaders leverage Bowling Charts – which capture the organization's key performance metrics.	Senior leaders have articulated a priority measure and have measured organizational performance against that measure using Bowling Charts.	Senior leaders have facilitated improvement on Bowling Chart measures. Senior leaders frequently review these measures to drive improvement. Cascading performance measures demonstrate a clear connection between measures and metrics at some levels of the organization.	Senior leaders ensure that agency measures and leadership priorities – including LTPGs, APGs, Enterprise Risk, and Administrator Priority – are tracked using visual management (e.g., Bowling Charts) at the right levels of the organization.
	Mid-Level Manager	Managers are not aware, or are only vaguely aware, of what teams are measuring.	Managers are consistently aware of what teams are measuring.	Managers ensure that teams' lead/lag goals and targets align with priorities.	Managers have facilitated documented improvement on teams' lead/lag goals and targets. Managers frequently review measures to drive improvement.	Managers ensure teams' lead/lag goals and targets are tied to Bowling Chart measures with associated actions, problem-solving, and performance trends.
	Front-Line Manager	Teams do not measure, or infrequently measure, performance.	Teams consistently measure performance. Teams have identified performance metric targets and priority lead/lag goals, e.g., outcomes and outputs.	Teams use graphs and/or charts to track identified performance metric targets and priority lead/lag goals.	Teams have contributed to improvement on the performance metric targets and priority lead/lag goals identified.	Teams ensure performance and priority lead/lag goals are directly tied to Bowling Chart measures. Teams frequently review measures to drive improvement.

		Level 1	Level 2	Level 3	Level 4	Level 5
Problem Solving		Problem solving does not occur formally or consistently; or problem solving occurs reactively.	Teams engage in problem-solving to identify, analyze and solve problems. Lean tools (e.g., tick sheets) are leveraged to support process improvements realized from problem solving.	Teams engage in proactive problem-solving to identify and solve problems through the use of visual management (e.g., flow boards) and Lean tools (e.g., advanced problem solving such as use of a 4-Square and/or root cause analysis tools). Systematic problem-solving tools and expertise (e.g., Lean facilitator, coach, or problem-solving guide) are utilized to determine root causes and devise countermeasures for issues, resulting in improved and sustained performance.	Teams at every level of an organization rely on data and evidence to proactively identify, analyze, and solve problems through the effective use of available visual management and other continuous improvement tools and techniques.	Mission-critical problems are routinely and proactively identified and addressed. Associated processes are improved and sustained through the use of problem-solving activities. Organization uses an internal system to raise and respond to problems raised at the right level. Problems that cannot be solved by teams are up-leveled effectively and efficiently to the level of leadership that can help remove barriers and get performance back on track.
Business Reviews & Huddles	Monthly Business Reviews	Business reviews do not occur at all or occur infrequently.	Senior leaders use business reviews to evaluate organizational performance of key metrics on the Bowling Chart.	Business reviews – held by senior leaders on a monthly basis – include a standard agenda that ensures review/presentation of key documents (e.g., countermeasure worksheets, action registry, Bowling Chart).	Throughout the business review, senior leaders engage in: more detailed conversations on organizational performance; recognition of processes improved, and employee ideas implemented; and discussion of up-leveled problems as appropriate. Business reviews influence actions taken by senior executives to better support teams in their organization.	Monthly business reviews, held by senior leaders, are used to advance progress towards priority agency goals and EPA's mission.

		Level 1	Level 2	Level 3	Level 4	Level 5
	Huddles Reviews	Teams do not conduct, or rarely conduct, process reviews around visual management.	Teams conduct regular huddles around visual management.	Teams use regular huddles – facilitated around visual management – to document process-related problems and the actions being taken to fix them.	Teams use regular huddles – facilitated around visual management – to document process-related problems and the actions being taken to fix them, up-leveling to management where necessary. Front-line teams leverage process reviews to inform and influence mid-level manager decision-making and actions.	Teams routinely engage in huddles around visual management to: proactively identify problems; leverage data and evidence to inform and influence decision-making; and ensure front-line objectives and metrics tracked achieve organizational objectives and agency mission goals and priorities.
	Advanced Huddles	Advanced huddles do not occur or rarely occur.	Advanced (4DX style) huddles occur weekly at the mid-level management level.	Regularly held advanced huddles include documentation of weekly commitments and leverage the lag/lead goals established by team.	Regularly held advanced huddles lead to follow-through on commitments and result in progress towards the team's lead measure.	Regularly held advanced huddles are used to connect frontline lead measures with agency goals and to achieve mission critical outcomes.
Employee Engagement	Managers	Managers do not encourage, or rarely encourage, the generation of ideas among staff.	Managers regularly encourage the generation of ideas among staff.	Managers create structured idea discussion opportunities where ideas can be presented, discussed, and prioritized.	Managers actively work to implement ideas that result in improvements across the organization.	Managers actively work to promote and champion implemented employee ideas.
	Staff	Staff do not, or rarely, generate or share ideas.	Staff feel comfortable sharing ideas with their team and manager.	Staff are empowered and supported by managers to implement or pilot ideas.	Staff create a positive feedback loop of innovation by continually discussing, promoting, and implementing ideas among themselves and with their managers.	Staff engagement/idea sharing efforts lead to improved employee morale, processes, and delivery of mission.

Appendix B : List of Programs being Evaluated or Analyzed

List of Programs Being Evaluated or Analyzed

The Foundations for Evidence-Based Policymaking Act (Evidence Act) provides a framework to promote a culture of evaluation and continuous learning to ensure Agency decisions are made using the best available evidence. Below is a list of programs being evaluated or analyzed by the Agency. This list was developed from EPA's FY 2022 Evaluation Plan and other Evidence-Building Activities.

EPA's FY 2022 Evaluation Plan and other Evidence-Building Activities describes significant program evaluations and other significant evidence-building activities the Agency plans to undertake in FY 2022. Significant evaluations and other evidence-building activities include those that support EPA's ability to meet an Administrator Priority, is mandated by Congress, or being highlighted as a program priority.

FY 2022 Planned Evaluations

- IT Modernization of EPA Pesticide Tracking Systems - Office of Chemical Safety and Pollution Prevention
- Evaluate Impact of Pre-Deadline E-reminders on Discharge Monitoring Report (DMR) Non-Receipt - Office of Enforcement and Compliance Assurance

FY 2022 Additional Planned Activities to Support EPA's Portfolio of Evidence

- Office of Air and Radiation:
    - Title V Permitting Program Reviews
    - Our Nation's Air: Status and Trends Through 2021
- Office of Chemical Safety and Pollution Prevention & Office of Research and Development:
    - Reducing Use of Animals in Chemical Testing
- Office of Land and Emergency Management:
    - Population Analysis
    - Annual Evidence Literature Search
    - Redevelopment Economics at Remedial Sites (non-federal facility)
    - Redevelopment Economics at Federal Facilities
- Office of Mission Support:
    - EPA Space Reduction – Annual Review
    - Strategic Sourcing
- Office of Research and Development:
    - Research Area: Assessment and Management of Harmful Algal Blooms
    - Research Area: Waste Recovery and Beneficial Use

- Office of Water:
    - Drinking Water Infrastructure Revolving Fund State Reviews
    - Public Water System Supervision (PWSS) Program Reviews
    - Safe Drinking Water Information System (SDWIS) National Regulation Non-Compliance Review