Merge Ahead: Integrating Process and Impact Evaluations

Patrick Hewlett, ERS
Sue Haselhorst, ERS

ABSTRACT

Traditionally, process and impact evaluations are conducted independently, under separate contracts and in separate time frames. Frequently, there is little opportunity for communication and sharing between the two teams. In this environment recommendations can be vague and/or disconnected from program realities. In the most extreme cases, the two evaluations can yield unrelated and even contradictory recommendations. For example, for a major small-business direct install program in one jurisdiction, the process team recommended increasing outreach to franchises for more efficient sales, while the impact team simultaneously recommended staying away from franchises because of high free ridership.

Merging the process and impact evaluations can provide beneficial synergies and produce avenues for internal collaboration and cross-pollination that should yield more informed and robust recommendations. This approach is not intended to be a shotgun marriage; instead, it should build on the strengths of each discipline: the process team frames the research within the context of the program, while the impact team brings field evidence to the solution. This approach is particularly effective for C&I evaluations based on on-site measurement and verification (M&V).

This paper presents a blueprint for an integrated approach in an effort to develop well rounded and actionable recommendations to improve program implementation. From a review of the methods for integrating the two types of evaluations, to a critical look at the pluses and minuses of integration, this paper outlines the advantages to an integrated approach at the planning, data collection, and reporting phases of evaluation.

1. Introduction

New Englanders have a reputation for bold (but slow) automobile driving. A true New Englander approaches a traffic rotary, accelerates, and makes no eye contact with any other driver. *Usually*, the drivers make it through the circle pretending there are no other drivers on the road. However, rotaries can create bottlenecks and can become a major source of accidents.

As this paper's title suggests, this is an apt metaphor for the traditional evaluation of energy efficiency programs. Evaluations typically involve two separate, often out-of-synch, fast-moving entities—process and impact components. Traditionally, each speeds forward and ignores the other during various phases (planning, data collection, reporting), sometimes creating delays or inefficiencies, sometimes incurring embarrassing "collisions." Instead, a more integrated yield-and-merge can move both sets of activities to a better outcome.

Process and impact evaluations have traditionally been separated due to differences in planning and implementation schedules, level and source of funding, and research objectives. Regulatory reporting requirements drive primary evaluation objectives, such as determining the program gross savings realization rate. However, there is also an expectation that the evaluation findings will result in actionable, practical recommendations for improving the program design from application approval to savings claim to incentive payout.

Traditionally, the process team focuses on the big picture and uses self-reported assessments of program strengths and weaknesses from program staff, partners, and participants. The process scope is often broad and can include assessment of program delivery, marketing and outreach, incentive levels and structures, benchmarking, and customer satisfaction. Assessment of program delivery is focused on whether the design of the program is appropriate. Extensive crosschecks of tracking data, project file review, or on-site inspections are not usually included in a process evaluation; therefore, an assessment of the execution of the design is less robust.

The impact team is usually narrowly focused on producing a single number: the net program savings. To arrive at this number requires a disciplined in-field data collection effort that may involve detailed project file reviews, dozens of site visits with logger deployment, discussions with customers and vendors, and project-specific analysis and subsequent reporting. While the impact team engineers use the products of the program (the applications and project files) to supplement their data collection, they rarely assess how those products conform to program design intent. For example, an impact evaluation rarely reports if measures were correctly classified as prescriptive or custom.

As a result of their independent objectives, the process and impact teams make independent recommendations, which may be contradictory in the worst case or less powerful than they might be with a more holistic approach. For example, a large commercial/industrial efficiency program recently underwent separate, concurrent process and impact evaluations. The process evaluation team assessed the program's tracking system design and quality checks at a variety of stages from application to incentive payout to post inspection. The process team concluded that the design and QC steps were adequate. Concurrently, the impact evaluation team closely tracked savings claims at a site-specific level over the course of on-site M&V of a large number of projects. Contradictory to process team findings, the impact team concluded that the program's tracking system had major errors, such as duplicate records and a high incidence of failure to update records when applications were revised. The impact team recommended that program staff "eliminate duplicate records" and "update savings estimates."

Although both recommendations had useful observations, they were flawed in that neither of the recommendations provided program implementers with actionable direction. The process team may have correctly concluded the system design was good; however, without the detailed tracing of data from application to summary report, the assessment was incomplete. Likewise, the impact team's recommendations were actually a diagnosis of a symptom (duplicate records). This absolute, diagnosis-driven recommendation is blind to the actual methods and costs of implementation and does not provide a benchmark against which to measure success, since it is impossible to eliminate all tracking errors.

This paper presents a model for an integrated process/impact evaluation approach drawing on real-world examples of process/impact recommendation successes and misfires. We are unaware of an evaluation where the model has been fully implemented; we hope this paper will inspire efforts to move toward better integration of these two perspectives of a complex process. Section 2 delves into more detail on the concepts and advantages of an integrated approach in theory. Section 3 outlines the blueprint for program administrators and contractors to strategically plan and launch the integrated evaluation concept. Finally, Section 4 discusses some of the barriers and caveats to implementing the integrated model.

2. Integrated Evaluation Approach – Why Merge?

The reason for merging process and impact efforts is to provide more powerful and actionable recommendations for program improvements. For recommendations to be compelling, the implementers must be provided with evidence, an action to undertake, and also a yardstick for assessing the value and the cost of implementation. By joining forces, recommendations to improve the program can better meet these standards. This section outlines the benefits of a merged approach by addressing:

- Value-based recommendations
- Method of implementation
- Harmonized recommendations reflecting the value of the recommendation (in terms of program savings) and the actual mechanics of implementing the recommendation
- Potential for reduced burden on program implementation staff and customers

Valuing Recommendations

Ultimately, recommendations should be made because their implementation will increase savings through expanding markets, improve the savings per measure, improve savings estimates, decrease the cost of delivery, and/or increase customer satisfaction. Yet, recommendations rarely identify the expected impact from these outcomes. Without a value metric, it is difficult for the implementer to prioritize actions, and, frankly, why should an implementer go to great effort to achieve a small change?

One source of valuation is through a savings-based analysis of the contributors to a program-wide realization rate, as conducted in conjunction with the impact evaluation. The impact team, through its field work, can identify sources of error in program estimates of savings and, with input from the process team, can classify where in the process those errors arise. This discrepancy analysis can be used to prioritize process improvements by those most affecting the realization rate. Figure 1 presents one such analysis of a large commercial/industrial program.

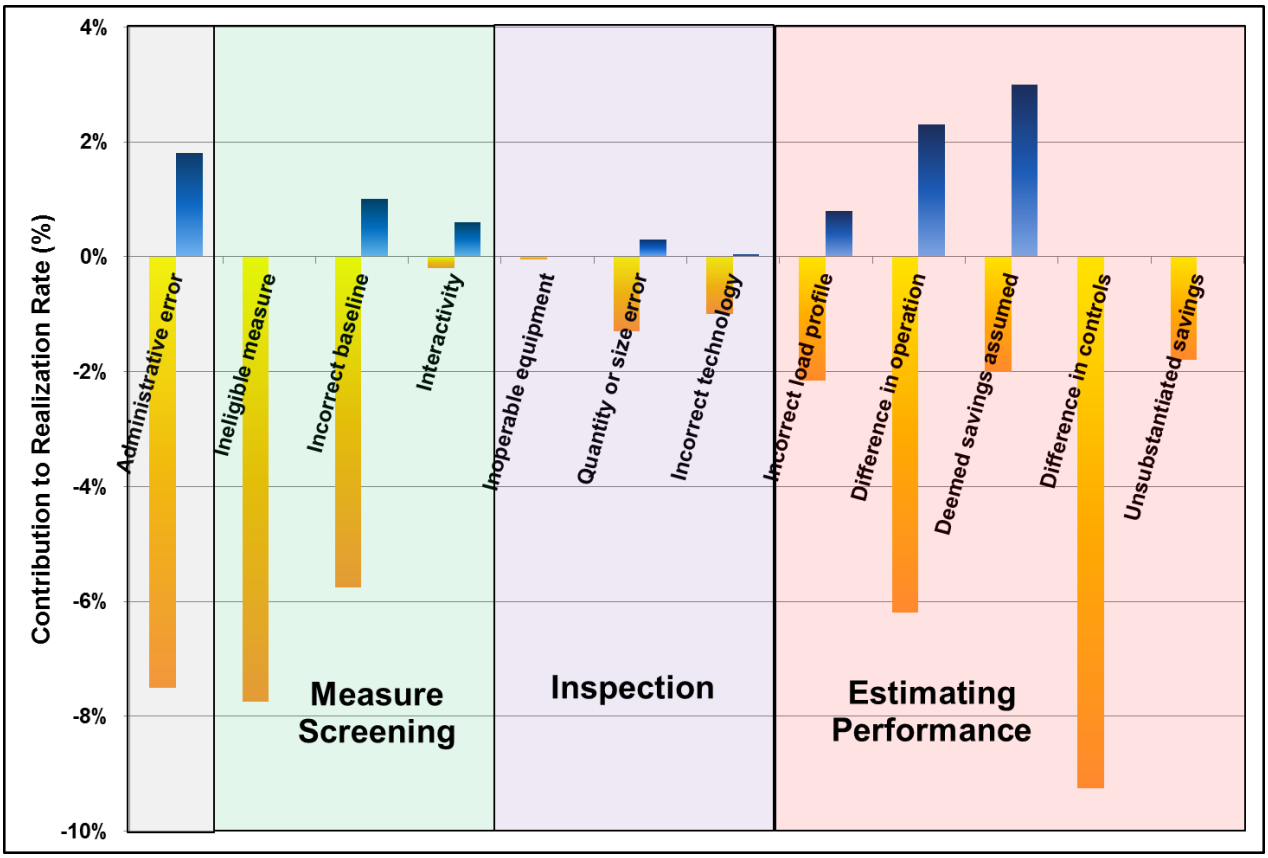

Figure 1. Example of Program-Wide Realization Rate Analysis

In the example evaluation above, the field engineers analyzed the sources of errors for each of the individual project savings estimates. For example, if the applicant used the existing equipment to define the baseline for an end-of-life replacement, the engineer would calculate the impact of that baseline error on the savings estimate and classify that discrepancy as an "incorrect baseline." The discrepancies across projects can be aggregated on a program basis and, in turn, can then be used to disaggregate the program savings realization rate into process-driven components. The program administrators are therefore able to deduce not only the contributors to the evaluation's realization rate but also the magnitude of each discrepancy.

Referring to the same figure, the evaluators concluded that discrepancies related to equipment inspection are relatively minor, indicating that the program's inspection process is working. Even discrepancy categories with relatively high plus/minus magnitude, such as deemed savings assumptions, somewhat "cancel out" and therefore would not warrant a high-priority recommendation. Tracking and measure screening errors, however, were significant, and in this case, led to the recommendation to "eliminate duplicate records." The benefits of valued recommendations should expand the traditional scope of impact evaluation— evaluators should be charged not only with assessing the net savings of a program, but also with prioritizing those improvements to the program process that could have the greatest impact on net savings.

Implementation Method

In a similar manner, the cost and feasibility of a recommendation is rarely included in an impact evaluation report, although high cost and impracticality can invalidate a recommendation.

For example, an impact evaluation of a large, state-wide commercial/industrial efficiency program led to a recommendation that all multi-site customers should be assigned a unique identification number to more thoroughly track program outreach and project progress. While simple in concept, many entities, including utility companies and other program administrators, have struggled in vain to establish reliable customer ID systems. The impact team did not have the intimate knowledge of the program's data system capabilities or the manpower required for such a task—information typically gathered extensively by a process evaluation team—and therefore made a somewhat uninformed recommendation in a vacuum.

While it may not be possible to quantify actual dollars, a recommendation should identify a tangible path to implementation, which might require an intermediate step, such as a feasibility study.

Harmonized Recommendations

The examples noted so far in this paper have cataloged flawed recommendations. How is an integrated recommendation different? Table 1 compares traditional evaluation report recommendations, from past process and impact efforts, with collaborative, robust recommendations achievable with a more integrated evaluation approach. These recommendation examples, while drawing on actual experience, are illustrative.

Table 1. Comparison of traditional and integrated evaluation recommendations

Traditional process recommendations	Traditional impact recommendations	Integrated recommendations
Database is fully populated and includes appropriate fields for tracking applications.	Eliminate duplicate records in the tracking dataset and always update tracking system with the most current estimates of savings.	Database design conforms to best practices. Apparently, QC processes are inadequate because tracking errors degraded the realization rate by 6%. Convene a team of database administrators, program implementers, and the contractor to map out the tracking database and identify specific checkpoints in the system for reducing these errors.
Sufficient application review processes are in place.	Measures should always be assigned the correct baseline. In particular, older capital equipment replacements are typically at end of life and should reflect a code-driven baseline.	Incorrect baseline assignments eroded the savings by 8%. Provide the application reviewer with specific training regarding equipment age, effective useful life (EUL), and baseline.

Aggressively recruit vendors for selling technologies to multiple facilities.	Savings estimates were particularly poor for vendors addressing multiple facilities.	The savings estimates used by multi-site vendors degraded realization rates by about 12% because they relied on over-optimistic deemed savings values. Identify multi-site vendors and proactively assist them in identifying acceptable savings values and algorithms.
Customers prove eligibility through extensive screening of project applications, which drives down participation.	Ineligible customers have been approved due to inaccurate account information as determined through field data collection. Each project should require an on-site pre-inspection to ensure eligibility.	Develop a benchmarking system to screen customer-claimed usage by square footage and facility type. Require field inspection for those customers that do not meet the benchmark, which is expected to occur about 15% of the time.

Reduced Burden on Program Staff and Customers

Additional reasons to merge are to conserve program staff time and to foster customer goodwill. By planning joint program staff interviews and joint process/impact customer surveys, the interview burden on both program administrators and customers may be reduced.

3. A Blueprint for Program Administrators

Hiring a single contractor for both impact and process efforts or initiating a process and impact efforts at the same time will not guarantee an integrated approach. Integration requires planning and a commitment from several parties to take the steps necessary to produce well-rounded recommendations. Figure 2 illustrates how process and impact evaluations might come together to form a more cohesive evaluation product, while the following subsections address the approach for integration at each phase.

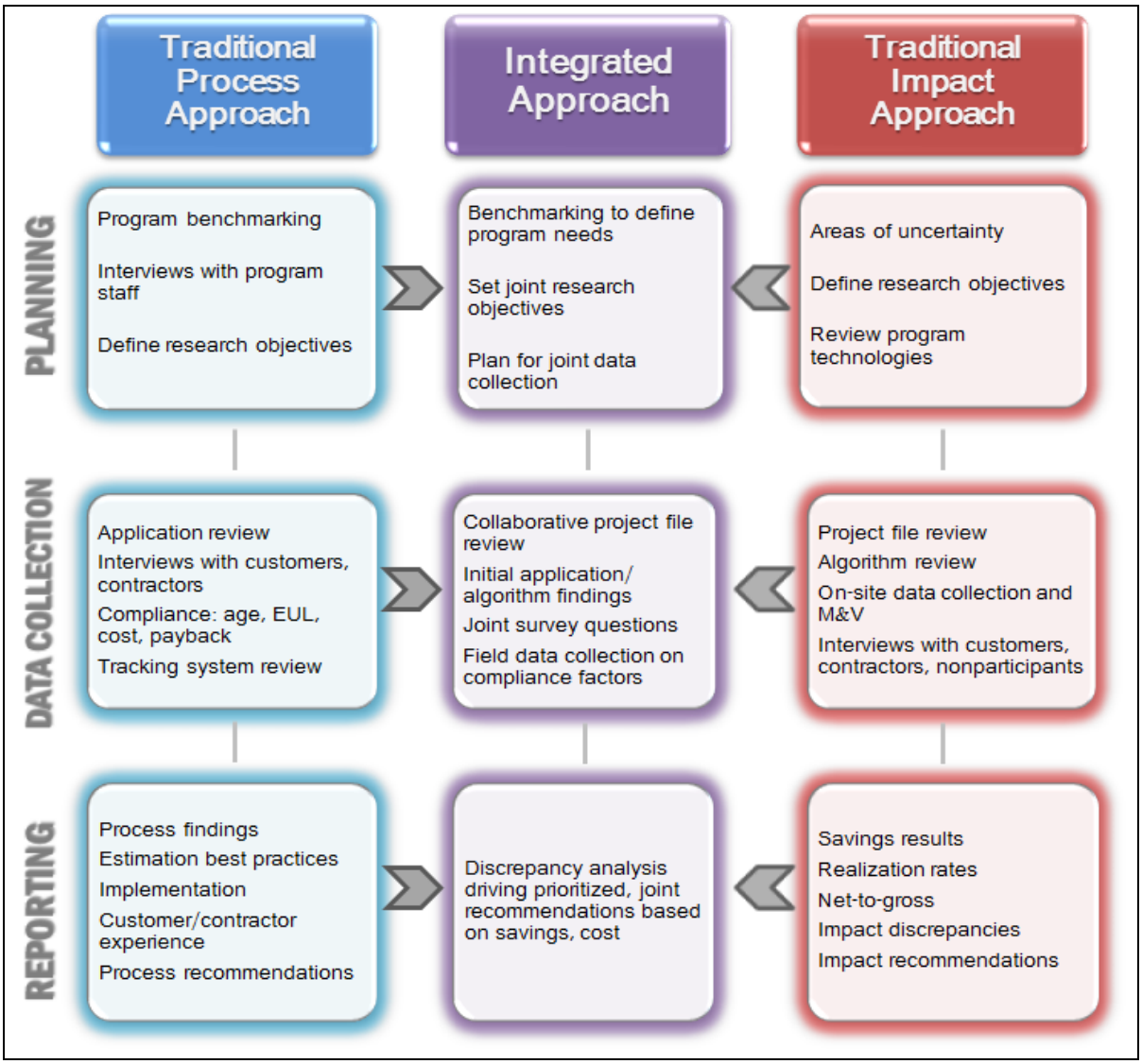

Figure 2. Benefits of an Integrated Process/Impact Approach at Various Evaluation Phases

Planning

An integrated master plan identifies joint areas of investigation, synergistic data collection, and strategies for developing joint findings and recommendations. This enables comprehensive and complementary research. Exposure to each other's planning will enable each plan to be better grounded and focused on the important issues facing the programs.

Since a number of data collection activities, including discrepancy analyses, application file desk reviews, and savings algorithm reviews, will play important roles in both evaluations, coordinating the nature and timing of those activities during planning is important.

Strategic integrated planning should first and foremost lead to the development of joint objectives and should include a framework for valuing recommendations. As noted in the introduction, the strength of the process team is framing the questions in the context of the program while the strength of the impact team is collecting extensive project-specific data. Table

2 presents examples of research questions and corresponding data collection that might lead to joint recommendations.

Table 2. Example research questions and data collection plans

Research question	Data collection plan
Is project screening effective? Do measures meet program eligibility requirements or is there evidence of gaming?	Confirm for the on-site sample whether the site paperwork conforms to eligibility requirements. Collect and report additional parameters, including benefit- cost ratio (BCR) screening result, measure cost, and quality of the cost data.
Are on-site inspection protocols sufficient to ensure quantity and do technologies in the field match what was claimed by the applicant? Is there too much inspection, costing the program resources and customer goodwill?	Quantify the impact on savings due to observed differences in quantity installed or technology.
Are savings left on the table at the customer facility?	Identify and communicate additional feasible measures at the facility.
Could simplified energy estimation approaches (using calculators or deemed savings) provide reasonable savings estimates at lower cost with faster approval turnaround?	Examine the variance in savings that would have occurred using deemed savings approaches.
Have earlier recommendations been implemented?	Compare present-day applications to an earlier benchmark.

A great starting point for the development of overarching objectives is a joint interview of program staff. The final research focus and ensuing data collection plan are likely to be developed through several iterations, based on program staff notions about program weaknesses, the process team's other objectives, and the impact team's capacity for additional data gathering.

Once the objectives are established, the two teams should agree to the method of valuing the impact of the recommendation. Typically, it will be in terms of savings– but it could also be in terms of dollars (for example, incentive dollars awarded for ineligible measures) or customer satisfaction.

Finally, the two teams need to coordinate delivery schedules to ensure that each party has an opportunity to review and provide real feedback on survey instruments and interim results. Additionally, the schedules should allow for regular check-ins to ensure that the mechanism for reporting joint recommendations is working. This may be one of the more challenging parts of the integration effort.

Data Collection and Analysis

After planning, the process and impact evaluation teams' core activities will be implemented independently. The evaluations will feed each other information at specific scheduled intervals as well as on an as-needed basis. Integrated data collection enhances the quality of both evaluation branches in a way that is not possible with their traditional separation.

Specific examples of areas where there are opportunities to increase quality through integrated data collection include the following:

- The impact team will examine the programs' savings algorithms and data collection techniques to identify opportunities for improvement and will feed proposals to the process team to determine alignment with program processes.
- The impact team will collect additional information through desk or on-site reviews tailored to support process objectives such as measure ineligibility and improper baseline selection. The results will then be examined for patterns, which will feed back to the process team, so they can investigate the root causes of these problems.
- Integrating data collection by incorporating two surveys into one will save money by eliminating the one-time costs that are inevitably associated with each standalone survey, such as questionnaire approval cycles and computer-assisted telephone interviewing (CATI) center setup.
- Site-specific data collected by the impact team will lead to identifying and quantifying sources of discrepancy on a program basis. These quantitative findings will be offered to the process team to ground their observations and recommendations in actual program outcomes.

It is useful to note that there are limits to data integration. While it is often possible to combine process and impact questions into a single survey, there is a practical limit to the length of the survey. If the question batteries are extensive, it may be more optimal to conduct two separate surveys. Likewise, the impact team can collect additional site-specific information at a small marginal cost, but the desired data needs to be identified very clearly upfront and specified in a format that permits aggregation. This requires some thought as to how the engineers would, for example, characterize whether a measure was correctly categorized as a prescriptive or custom measure. In addition, some specific joint data collection – such as determining broadly lost opportunities at each facility – could add significant scope to on-site work.

Reporting

The benefits of an integrated evaluation approach are fully realized during the reporting phase. An effective impact evaluation report offers concrete, prioritized recommendations to improve the program's attributed savings per incentive dollar spent. To accomplish that objective, the impact team typically sifts through a number of site-specific project files and updates project savings claims based on field inspection and/or M&V. The most effective impact evaluations identify reasons for discrepancy in realization rate and associate those reasons with savings magnitudes. Aggregating a large number of site-specific findings into one set of program-wide recommendations can be challenging. However, this program-level discrepancy analysis is vital to deciding which recommendations are most impactful and cost-effective for a program with limited budgets and manpower.

4. Conclusion

The benefits of an integrated evaluation approach have been examined in this paper from both conceptual and applied perspectives. Nonetheless, process and impact evaluations traditionally are contracted and executed separately. Both program administrators and

contractors face several hurdles before the status quo is improved. This section examines three primary reasons why process and impact evaluations are typically separate and presents the integrated planning required to address each. Finally, a program that would most benefit from an integrated approach is conceptually described.

Timing – Process evaluations typically occur before impact evaluations for a number of reasons. Often the evaluated program is relatively new or redesigned, and program administrators are interested in the new program's effectiveness and ease-of-use for customers and contractors. Process evaluation typically can provide actionable feedback to PAs much faster than impact evaluation. An integrated approach would require strategic planning to ensure that the impact data collection does not disrupt the process evaluation's fast feedback, but that there is a reconvening of the teams at the conclusion of the impact data collection to formulate joint recommendations.

Budget – It is likely that additional budget would be required for both teams to allow for augmented planning and the formulation of joint recommendations at the conclusion of both team's efforts. Additional on-site data collection costs may also be incurred for more complicated data collection, such as an investigation of lost opportunities. This additional cost may be offset somewhat by reductions in telephone survey costs where joint surveys are possible.

Objective – Process and impact evaluations have remained separate because their research objectives are traditionally not thought to intersect. As this paper has examined, there is common ground between process and impact sides. Cost savings can be incurred in the data collection phase if planned holistically, but more importantly recommendations become more effective when prioritized with savings and grounded within program capability.

Not every efficiency program would benefit from an integrated evaluation approach. The best candidates for integration offer a level of flexibility in schedule, budget, and deliverable constraints. For example, a program conducive to integration would typically feature process and impact evaluation cycles with concurrent planning periods and/or overlapping schedules. Programs open to the idea of integrated recommendations—and the slight increase in cost for joint planning and data collection—would be fitting candidates for this proposed approach. From a contractual standpoint, programs with process and impact evaluation funding shared from a common source could be more smoothly streamlined using the integration concepts introduced in this paper. Most importantly, program staff must be receptive to the additional logistical planning and joint check-ins necessary to properly monitor and execute an integrated approach during each phase.