Department for Transport

Best Practice Guidance for planning the Fuller Evaluations of Local Authority Major Schemes

DRAFT

GSR GOVERNMENT SOCIAL RESEARCH
Analysis for Policy

While the Department for Transport (DfT) has made every effort to ensure the information in this document is accurate, DfT does not guarantee the accuracy, completeness or usefulness of that information and it cannot accept liability for any loss or damages of any kind resulting from reliance on the information or guidance this document contains.

The DfT has actively considered the needs of blind and partially sighted people in accessing this document. The text will be made available in full on DfT's website in accordance with the W3C's Web Content Accessibility Guidelines. The text may be freely downloaded and translated by individuals or organisations for conversion into other accessible formats. If you have other needs in this regard please contact DfT.

Contents

Acknowledgements

In developing this guidance, DfT has drawn on a number of sources which are referred to throughout. The previous guidance by NERA Economic Consulting, MVA and David Simmonds Consultancy was used as a foundation and this has been updated and developed using the framework of evaluation design designed by the Tavistock Institute for DfT and the HM Treasury Magenta Book. These key sources are:

The evaluation of major local authority transport schemes: a guide by NERA Economic Consulting, MVA and David Simmonds Consultancy (2006)
http://webarchive.nationalarchives.gov.uk/+/http://www.dft.gov.uk/pgr/regional/ltp/major/pdfevalmajlocautranpro.pdf

Guidance for transport impact evaluations: choosing an evaluation approach to achieve better attribution by the Tavistock Institute in consultation with AECOM (2010)

http://webarchive.nationalarchives.gov.uk/20111005180109/http://www.dft.gov.uk/publications/guidance-for-transport-impact-evaluations

The Magenta Book, Guidance for evaluation: by HM Treasury, (2011)
http://www.hm-treasury.gov.uk/magentabook

1. Introduction

Summary

The Department for Transport (DfT) has set out its framework for the monitoring and evaluation of Local Authority Major Schemes[1]. This best practice guide is designed to support Local Authorities to develop and implement plans for undertaking a fuller evaluation, although some aspects of this guidance might also be useful for Local Authorities designing standard and enhanced monitoring plans.

This guide is designed to:

- Bring together elements of existing best practice guidance for evaluation in one place for Local Authority Major Schemes;
- Encourage best practice without being prescriptive; and,
- Provide readers with a summary of the broad spectrum of evaluation approaches and techniques available to enable informed decision making regarding production of an evaluation plan.

This introductory section will help the evaluation planner to understand:

- The scope of the guidance and intended audience;
- The benefits to be gained from good quality evaluation; and,
- How the guidance is structured.

---

[1] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/9154/la-major-schemes-monitoring-evaluation.pdf

1.1 Scope and purpose of the guidance

This best practice guidance document aims to support Local Authorities in designing and developing plans for fuller evaluation. It introduces a systematic approach to identifying the core evidence requirements and designing a suitable and proportionate scheme evaluation. This guidance supplements the Framework for the Monitoring and Evaluation of Local Authority Major Schemes, which sets out DfT's approach to standard monitoring and evaluation, and identifies the schemes within the 'Supported Pool' in 2010 and the 'Development Pool' in 2012 which are required to undertake a fuller evaluation. Local Authorities who are required to develop a fuller evaluation plan are recommended to use this guidance to assist them in this process. However, Local Authorities undertaking standard and/or enhanced monitoring may also find aspects of the guidance to be useful.

The guidance draws upon and updates the previous guidance on Major Scheme evaluation[2], incorporating the latest thinking on designing and conducting transport evaluations. In particular, this document reflects the framework outlined in the DfT publication Guidance for transport impact evaluations[3], which readers can refer to for further detail on the steps involved for relevant evaluations, and best practice guidance from the HM Treasury, the Magenta Book[4].

The guidance is not intended to be prescriptive, since each scheme and its context are unique and the evidence requirements from evaluations vary widely. The guidance aims to provide an introduction to the broad spectrum of evaluation approaches and techniques available in order to assist evaluation planners in making an informed choice regarding the most appropriate, and proportionate, approach to evaluate their particular scheme.

---

[2] NERA/MVA (2006) *The evaluation of major local authority transport schemes: a guide* http://webarchive.nationalarchives.gov.uk/+/http://www.dft.gov.uk/pgr/regional/ltp/major/pdfeval majlocautranpro.pdf

[3] Hills, D., & Junge, K. (2010), *Guidance for transport impact evaluations: choosing an evaluation approach to achieve better attribution,* DfT http://webarchive.nationalarchives.gov.uk/20111005180109/http://www.dft.gov.uk/publications/g uidance-for-transport-impact-evaluations

[4] HM Treasury, (2011) *The Magenta Book, Guidance for evaluation* http://www.hm-treasury.gov.uk/magentabook

1.2 What is evaluation and what benefits can it bring?

> **Some definitions…**
>
> **Evaluation** is the assessment of the effectiveness and efficiency of the scheme during and after implementation. It seeks to measure the causal effect of the scheme on planned outcomes and impacts. It is used to assess whether the anticipated benefits and value for money have been realised and whether any unanticipated impacts have occurred.
>
> **Monitoring** is the collection of data to check progress against planned targets and benefits, and can be defined as the formal reporting and evidencing that spend and outputs are successfully delivered and milestones met. Monitoring data can play a key part in evaluation by providing valuable evidence throughout the life of the scheme.

Evaluation is an objective process which examines how effectively a scheme was designed and implemented, and with what results. It focuses on the practice and experience of the initiative, based on observation about what actually happened during and following implementation - rather than what was expected to happen.

Importantly, evaluation is a tool that can offer Local Authorities, transport analysts, policy makers and other individuals responsible for assessing transport schemes rich quantitative and qualitative data on a scheme's outcomes and impacts and the reasons for why and how they came about.

Without proportionate evaluation there is a risk of some, or all, of the following:

- not knowing how to improve this or future schemes, in terms of efficiency and effectiveness;
- continuing to fund inefficient and ineffective types of schemes;
- lack of evidence to respond to criticism from external stakeholders;
- not being able to convince others that your scheme is successful and offers good value for money;
- not being able to demonstrate that the benefits observed have been generated by your scheme, rather than caused by other factors;
- not understanding why the scheme worked or didn't work.

1.3 Guidance structure

The remainder of this guidance is split into three sections:

Section A: Establishing the evaluation requirements

Section B: Selecting an appropriate evaluation

Section C: Planning evaluation delivery and use of evidence

These sections are based on the key steps in effective evaluation design outlined in the Magenta Book[5]. The sections and their content are listed in Figure 1.1.

The structure of the guidance complements the Monitoring and Evaluation Plan section headings identified in 'Appendix 5 – Monitoring and Evaluation Plan' of the Framework for the Monitoring and Evaluation of Local Authority Major Schemes. Figure 1.1 highlights where the guidance maps on to the relevant section headings identified in the Appendix 5 of the DfT Monitoring and Evaluation Framework for Local Authority Major Schemes (the Monitoring and Evaluation Plan). All sections of the guidance will be relevant for evaluation planners in producing a fuller evaluation plan.

---

[5] HM Treasury, (2011) *The Magenta Book, Guidance for evaluation*
http://www.hm-treasury.gov.uk/magentabook

Figure 1.1: Structure and Content of the Guidance

Guidance Section	Step	Corresponding monitoring and evaluation plan headings

A: Establishing requirements

1. Defining scheme objectives and intended outcomes

2. Defining the audience for the evaluation

3. Identifying the evaluation objectives and research questions

1. Scheme background and context

2. Scheme objective and outcomes

3. Evaluation objectives and research questions

B: selecting appropriate type of evaluation

4. Selecting the evaluation approach

5. Identifying the data requirements

4. Outline evaluation approach

5. Data requirements

6. Data collection methods

C: Planning delivery and use of the evaluation

6. Identifying resources and governance arrangements

7. Conducting the evaluation

8. Using and disseminating the findings

7. Resourcing and governance

8. Delivery plan

9. Dissemination plan

2. Section A - Establishing the evaluation requirements

Introduction

Making explicit the underpinning assumptions about the outcomes the scheme is being designed to deliver, and how, is key in terms of planning an evaluation. Furthermore, defining which stakeholders will be interested in the evaluation findings and the type of evidence they are looking for the evaluation to generate (provide accountability, generate knowledge etc.) is an important step in starting to shape the evaluation questions.

This section will help the evaluation planner to work through the first three steps:

1. **Define the scheme's objectives and outcomes** and the underpinning theory about how the scheme will achieve these;

2. **Consider the needs of the audience for the evaluation** and understand how this frames the evaluation design; and,

3. **Establish the research questions** which the evaluation should seek to answer.

Working through these steps will help you develop sections 1-3 of the monitoring and evaluation plan required and deliver the following outputs:

- A description of the scheme inputs and outputs;
- A logic map (or maps) for the scheme;
- A list of objectives with associated measures; and,
- A set of research questions.

2.1 Defining the scheme's objectives and outcomes

> **Key questions to consider:**
>
> - What does the specific background and context of the scheme mean for the evaluation design?
> - What is the scheme logic (i.e. the causal theory) underpinning how the scheme will deliver the intended objectives?
> - In practice, how will the inputs lead to outputs, outcomes and impacts in the particular scheme context?

2.1.1 Scheme background and context

This is a short preliminary step designed to ensure that the background circumstances for the planned evaluation have been fully considered. The evaluation plan should seek to provide the following type of information about the scheme:[6]

Scope: What is the scheme's purpose? Who are the intended beneficiaries and how will they be reached / effected by the scheme[7]? What evidence already exists about the effectiveness of similar schemes?

The proposed scheme will have been designed with specific aims and objectives in mind, tailored to reflect the local context and the transport challenges faced. The evaluation design will therefore need to consider how the unique contribution of the scheme to achieving the intended aims and objectives can be established.

Context: How complex is the scheme setting? Could other factors be influencing the intended outcomes and impacts?

The scheme will be implemented alongside existing transport infrastructure and service provision, as well as, wider social and economic circumstances. It is important, for understanding causality and the transferability of the evaluation findings, that this context is clearly articulated within the evaluation plan. For instance, schemes which are designed to release capacity to accommodate future growth through new development and / or regeneration, there will be a number of wider influences on the successful delivery of the new development which will in turn affect whether the scheme has achieved its aims and objectives, so understanding and tracking these will be important for the evaluation.

---

[6] This should draw on the evidence provided in the Strategic Case and plans for benefits management from the Delivery Case.
[7] The DfT Behavioural Insights Toolkit will help with this: https://www.gov.uk/government/publications/behavioural-insights-toolkit.

Scale: What is the scale of investment and anticipated impact?

Schemes selected for fuller evaluation vary in terms of scale of investment, size of anticipated effect / returns on investment and the geographical spread of the scheme (e.g. from a site specific to area-wide measures). The evaluation will therefore need to be designed so that approaches and methods are commensurate with the scheme's scale.

Implementation: How will it be implemented and where? Which stakeholders will be critical to the successful implementation of the scheme?

The implementation model will be an important factor in the overall success of the scheme which will need to be clearly articulated. It is likely that third party suppliers will perform key roles and involving these stakeholders in the evaluation will be important both to hear their views and to gain access to any data collected locally.

Timeframe: What timeframe will it be delivered in and when are the outcomes and impacts expected to be realised?

Schemes are often implemented within a limited time period, though their effect can be permanent (e.g. some infrastructure schemes). The longer a completed scheme has been in operation, the more important it will be for the evaluation approach to be able to identify what factors other than the scheme could be responsible for the impacts observed. As well as assessing their short-term outcomes, it will also be particularly important to make some kind of assessment, or forecast, of likely longer- term impact and confirmation that the scheme is on track to deliver these.

Nature of anticipated objectives: What are the objectives of the scheme?
What changes will the scheme deliver?
What benefits will it generate?

These will have been a core part of the business case for the scheme and its benefits realisation plan, but it is useful to ensure that they are presented in a way which is clear**,** specific**,** up to date, measurable with a clear line of sight linking them to the scheme. This should also be presented in the form of Logic Map which is discussed further in 2.1.2 below.

> **Tip Box 1:** When setting out the scheme objectives, it can be useful to differentiate between objectives which the scheme is expected to deliver over a short time period (1-3 years, referred to as outcomes) and those which might take longer to be achieved (i.e. impacts). You should also consider what the primary and secondary goals of the scheme are.

2.1.2 Logic Mapping

Logic mapping is a systematic and visual way of presenting the key steps required in order to turn a set of resources or inputs into activities and outputs that are designed to lead to a specific set of changes or outcomes / impacts. The aim is to articulate the underlying causal theory based on the assumptions and evidence underpinning the rationale for the scheme.

Causality is central to logic maps, as they order events in such a way that the presence of one event or action leads to, or causes, a subsequent event or action.

Logic maps should seek to:

- Articulate what needs to happen in order for the anticipated outcomes and impacts to be achieved;

- Provide a clear line of sight between the inputs and the anticipated impacts;

- Visualise unintended effects;

- Highlight gaps in the evidence base and therefore help to focus evaluation effort accordingly;

- Outline the stages between the inputs and the desired impacts, which provides a transparent assessment framework within which existing evidence and evaluation results can be combined to provide answers to the evaluation questions; and

- Point to where the links between the inputs, outputs, outcomes and impacts are unclear, which aids delivery as well as evaluation design.

Figure 2.1 below illustrates the main components of a Logic Map. For further information please refer to *A Logic Mapping: Hints and Tips Guide* available from DfT's website https://www.gov.uk/government/publications/logic-mapping-hints-and-tips-guide.

Figure 2.1: Main components of a Logic Map

Source: Hills, D., & Junge, K., (2010) *Guidance for transport impact evaluations*
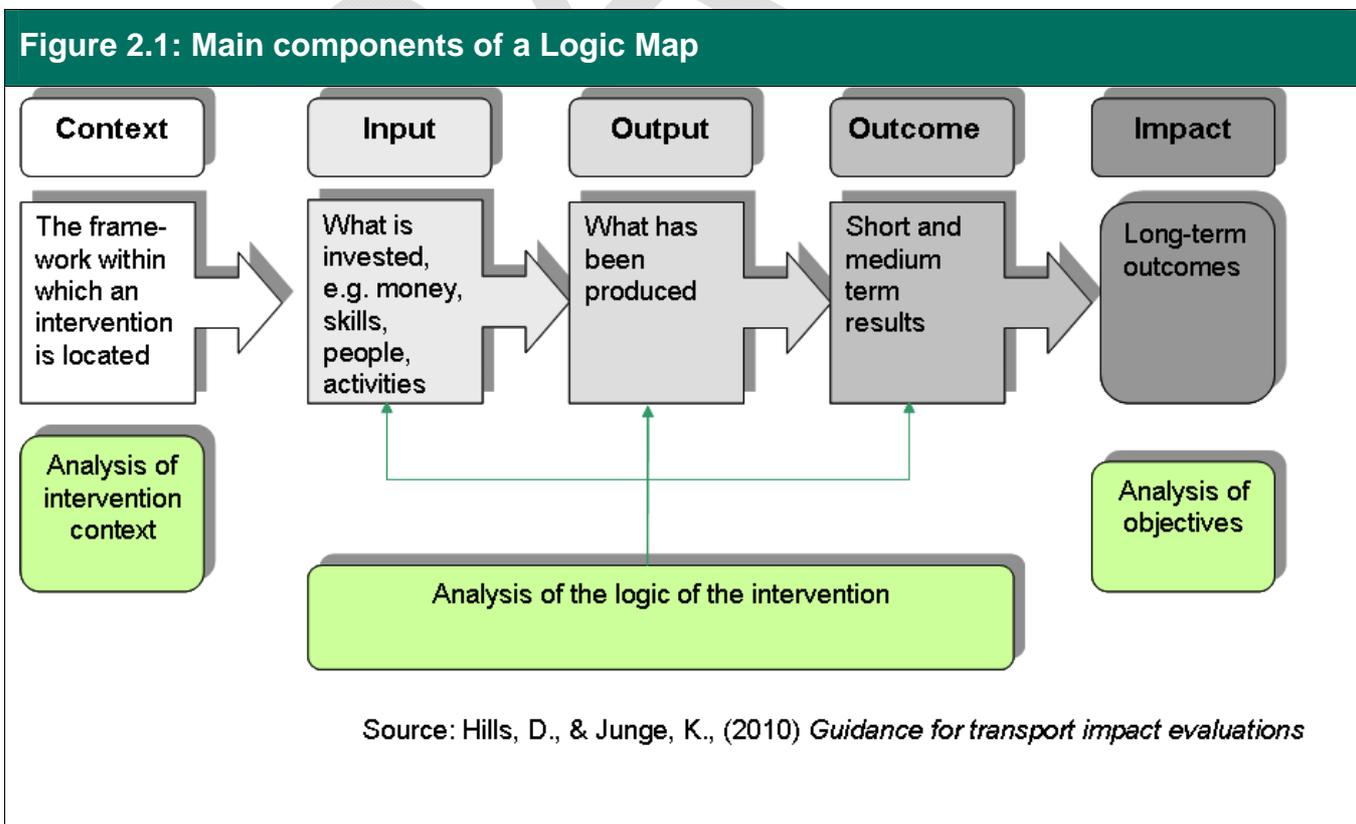
2.2 Defining the audience for the evaluation

Key questions to consider:

- Who will be the main users of the findings?
- What types of evidence do they need?
- How do the needs of different stakeholders vary?
- How will they be engaged?

To ensure the evaluation evidence delivers optimum benefits it is important that the requirements of the anticipated users of the findings are considered at an early stage.

When starting to design the evaluation, it is therefore important to understand:

Who the main users of the evidence will be and how will they be engaged? For instance, Local Authority officers, key stakeholders in local decision-making such as, local community members, transport users, Councillors, emergency services, local transport operators / service providers, DfT analysts and policy makers, other Government Departments, Public Accounts Committee, Local Government Association etc.

Are there different expectations for the results? For instance, what are the expected outcomes and impacts and how do they relate to wider strategic objectives? Are there specific areas of local learning the evaluation should address?

What would allow them to make effective use of the evaluation findings? This includes consideration about the quality of evidence which is required to generate transferable findings and transparent reporting and dissemination of evaluation findings.

Tip Box 2: Independent Peer Review

For evaluation results to be their most useful, stakeholders will need to have confidence in their robustness, relevance and clarity. Peer review can be a valuable process in achieving this, particularly for complex, large-scale, innovative or contentious scheme evaluations. Having external specialists review the evaluation will add independence and credibility to the findings and, in the case of ongoing evaluations, will deliver insights into how an evaluation could be improved.

External assessment can come at different stages, either at evaluation completion or periodically throughout the delivery of the evaluation. It can be sourced from different areas including developing an expert panel (which might consist of a range of evaluation practitioners and front line delivery personnel), or through an academic institution to provide methodological review.

2.3 Identifying the evaluation objectives and setting research questions

> **Key questions to consider:**
>
> - How do these respond to the needs of key stakeholders?
>
> - How will the evaluation build on the existing evidence base?
>
> - What questions will the evaluation answer?
>
> - Is prioritisation necessary (for example due to time or resource constraints), if so what are the priority questions?

2.3.1 Setting the evaluation objectives

The third step in planning an evaluation is to set the objectives for the evaluation which will in turn inform the development of the specific research questions to be addressed. Framing and presenting the objectives of the evaluation is important for creating a common understanding of, and consensus around them and the subsequent research questions. This is a separate activity from defining the scheme's objectives, but the information gathered during that step will be relevant here.

The evaluation objectives for Local Authority Major Schemes need to build on the aims for fuller evaluations set out in the monitoring and evaluation framework for Local Authority Major Schemes[8]:

- Whether the scheme was delivered effectively and efficiently;

- The causal effect of the scheme on the anticipated outcomes and whether these have contributed to the intended impacts; and,

- Whether the scheme had any unintended adverse or positive effects.

Underpinning these aims, the objectives of the fuller evaluations could:

- Provide accountability for the investment in the scheme;

- Provide transferable evidence to inform decision making about future spending decisions on similar schemes;

- Generate knowledge about the success and cost effectiveness of the scheme in achieving its objectives;

- Assess the sustainability of the scheme;

- Learn about barriers to success and how they can be overcome; and,

- Improve the efficiency and effectiveness in the delivery of future schemes based on learning from this scheme.

---

[8] Monitoring and Evaluation Framework for Local Authority Major Schemes, DfT, 2012, page 8

The evaluation objectives for your scheme should respond to the DfT's requirements and also consider the needs of other key stakeholders identified in step two (section 2.2). The evaluation objectives will be tailored to the scheme and what it is trying to achieve, so the logic mapping and other activities undertaken in the previous step will need to be drawn on here.

Tip Box 3: Setting evaluation objectives

Avoid a lengthy list of objectives: The focus should be limited to no more than three to five objectives. It is generally preferable to explore a few issues in depth rather than to examine a broad set superficially.

Use clear outcome-focused language: Do not state the objectives in technical or process terms.

Source: "Writing Terms of Reference for Evaluation: A how-to guide", Independent Evaluation Group (2011) http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/ecd_writing _TORs.pdf

---

2.3.2 Designing the research questions

Once the evaluation objectives have been established, more detailed research questions can be prepared.

The following considerations should be made when designing the research questions[9]:

- There should be a logical progression between the purpose of the evaluation, its specific objectives, and the questions posed in relation to each objective;

- Questions should be specific and focused on the scheme being evaluated;

- Depending on the type and purpose of the evaluation, such questions are likely to address specific demands for information related to the following broad areas of inquiry (See Annex A for some examples):

  1. Have the right things been done? (relevance, effectiveness)
  2. Have things been done well? (efficiency, effectiveness)
  3. What results have been achieved? (effectiveness, impact, cost/ effectiveness)
  4. How do the results compare with an alternative intervention to achieve the same objective? (relative effectiveness, impact, cost/ effectiveness)
  5. How could things be done better in the future?
  6. Are the results sustainable?

---

[9] Source: "Writing Terms of Reference for Evaluation: A how-to guide", Independent Evaluation Group (2011) http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/ecd_writing_TORs.pdf

In constructing the list of evaluation questions, it is important to prioritise these areas of inquiry according to the information needs of stakeholders and the overall rationale driving the evaluation. More information on formulating research questions can be found in Annex A.

Summary of Section A

Section A sets out what evaluation planners should consider for the completion of part 1 – 3 of their evaluation plans (see Figure 2.2). Through this process it should become clear what evaluation planners intend to deliver through evaluation (e.g. evaluation objectives and research questions) and how this interlinks with the context in which the evaluation takes place (e.g. scheme background and objectives/outcomes). However, it does not address how the evaluation will be delivered. Decisions regarding and prioritisation of evaluation objectives and research questions only address what will be evaluated, not how it will be evaluated.

Figure 2.2

Guidance Section	Step	Corresponding monitoring and evaluation plan headings
A: Establishing requirements	1. Defining scheme objectives and intended outcomes	1. Scheme background and context
	2. Defining the audience for the evaluation	2. Scheme objective and outcomes
	3. Identifying the evaluation objectives and research questions	3. Evaluation objectives and research questions

3. Section B - Selecting the appropriate type of evaluation

Introduction

There are a variety of approaches to evaluation, which can produce different types of evidence. Fuller evaluations will generate evidence on the process, impacts and cost effectiveness of the scheme (economic evaluation). Evaluation planners should consider how the evaluation will be undertaken to incorporate these three aspects and there will be specific considerations to make about how to design the evaluation approach which for each aspect.

This section provides an overview of the broad range of evaluation approaches set out in the Magenta Book. It is not possible to go into detail about each approach in this document but by providing insight into the range of approaches available, evaluation planners can start to make informed decisions about the relative strengths and weakness of each approach when applied in their scheme context.

This section will help the evaluation planner to work through the key steps in selecting the right type of evaluation:

1. **Selecting the evaluation approach**; and,

2. **Identifying the data requirements**

Working through these steps will help you develop sections 4-6 of the monitoring and evaluation plan required for fuller evaluations and can help you deliver the following outputs:

- Outline evaluation approach;

- Plans for the data requirements, methodological and analytical approaches which can be used to develop the measures template (see Appendix 5 of the DfT Monitoring and Evaluation Framework for Local Authority Major Schemes.

3.1 Selecting the evaluation approach

Key questions to ask:

- How will the evaluation design combine requirements for process, impact and economic evaluation?
- Will a scoping / feasibility study be required?
- How extensive is the evaluation likely to be?
- What level of robustness is required?
- Can proportionate steps be taken to increase the robustness of the evaluation, if so what are these?

This step will build on the scope for the evaluation which is discussed in section 2.1 and articulated through the evaluation objectives.

3.1.1 Types of evaluation

This section starts with a brief overview of the different types of evaluation. The Monitoring and Evaluation Framework for Local Authority Major Schemes sets out that fuller evaluations should seek to answer the following high level questions and those stated in Annex A:

- How was the scheme delivered?
- What difference did the scheme make? and,
- Did the benefits justify the cost?

The Magenta Book explains that these questions can be answered through three different types of evaluation, which are defined as[10]:

Process evaluations seek to answer the question of how the scheme was delivered by analysing the processes by which the scheme was implemented (see section 3.1.2). This is important for understanding how and why a scheme was successful (or not) in delivering the intended benefits and generates valuable lessons about how to improve the management and implementation of current and future schemes;

---

[10] Taken from Chapter 2 of the HM Treasury, (2011) *The Magenta Book, Guidance for evaluation* http://www.hm-treasury.gov.uk/magentabook

Impact evaluations focus on demonstrating and quantifying the difference made by the scheme (see section 3.1.3). Impact evaluations provide robust, objective evidence about whether the scheme caused the intended outcomes and impacts and for whom. It is also used to assess any wider or unintended effects of the scheme; and,

Economic evaluations use the evidence generated through impact and process evaluations to assess whether the outcomes and impacts generated by the scheme justify the investment. [This section of the guidance is still under development but further information can be found in WebTAG Unit 3.5.]

The design of the fuller evaluation will need to consider the data which is required for combining these three types of evaluation and the approaches to data collection which are most suitable given the context of the scheme. If possible, building in a scoping / feasibility study to test out / develop methods for delivering the selected approach will be a valuable exercise.

All three types of evaluation are very complementary and should be delivered as part of the scheme lifecycle. Figure 3.1 below shows how they interlink and when they should be undertaken.

Figure 3.1: How evaluation interlinks with the scheme's lifecycle
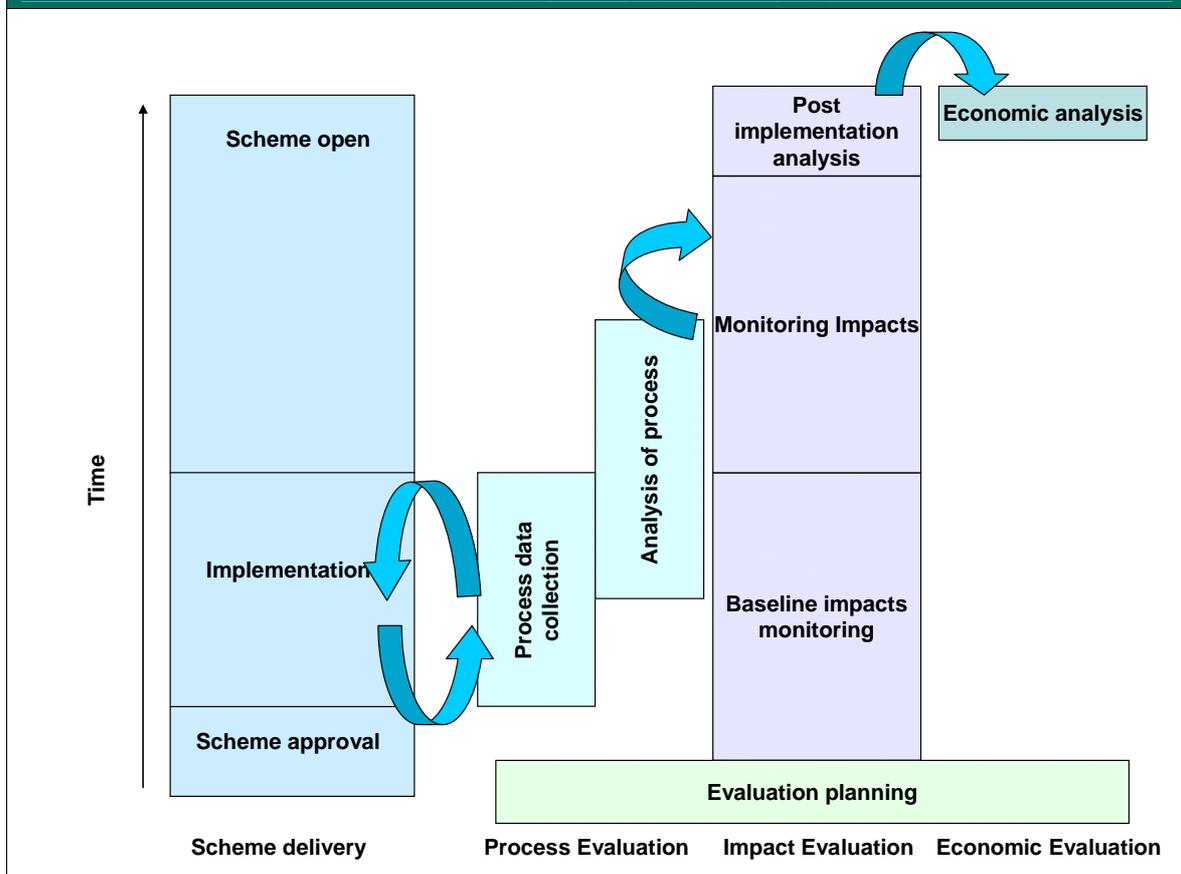
The next sections provide more detail on process and impact evaluations and some of the key considerations for evaluation design.

3.1.2 Process evaluation

Process evaluations examine how a scheme is delivered in practice and provide understanding about how and why a scheme results in certain outcomes (or fails to). They are particularly valuable when delivered alongside impact evaluations. It offers complementary evidence by providing crucial insights into why or why not the scheme worked and to test the logic map on which the scheme is based which helps attribute the evaluation findings to the scheme. On their own, they cannot tell you what a scheme has achieved or whether it has worked.

Process evaluation can be used throughout the life of a scheme to provide timely feedback about whether a scheme is being implemented as expected, whether important outputs are being delivered and if there are any parts of the scheme which are not working or which are working particularly well.

Adapted from the Magenta Book 2011

To get the most out of a process evaluation, data should be collected and analysed during the implementation stages. It offers real-time feedback which can contribute to continuous improvement in delivery. It also gathers evidence which contributes to the analysis and interpretation of the impact and economic evaluations (see tips for getting the most out of a process evaluation in Table 3.1 below).

Designing a process evaluation

Process evaluations tend to use a range of qualitative and quantitative research methods. They draw on performance and financial management data, feedback from the project management team, delivery team and wider stakeholders, they can also build in evidence from the end users' perspectives about whether the delivery process and / or outputs have met their needs and expectations.

When planning the scheme's process evaluation, you will need to consider how it will build on and utilise the evidence collected and analysed from the standard monitoring and the evidence required for scheme project management. Table 3.1 below provides an illustration of the types of evidence collected for scheme project management and assurance, standard monitoring and process evaluation within a fuller evaluation project. This helps to set out the distinction between the sets of activities but also highlights how they can complement each other and the importance of developing an integrated plan for data collection.

The evaluation planner will need to refer back to the research questions for the evaluation and may even need to define a further level of questions specific to the scheme process. Annex A provides some examples of types of process evaluation questions.

Table 3.1: Types of evidence collected on scheme implementation process			
Type of process information[11]	**Project management**	**Standard Monitoring**	**Fuller evaluation**
Whether the scheme was implemented as planned			
Scheme planning	Check point assessment that the scheme plan in on track to deliver. This will identify any issues or delays as part of the management process.	Retrospective assessment of scheme delivery against originally planned timeframes. Using real-time monitoring of implementation.	This should look to draw on these two activities plus also explore why delays / changes to the plan were encountered, whether they were, or could have been, foreseen / mitigated, the knock-on effects to the overall delivery of the scheme and cost, and the lessons to be learnt for future delivery.
Delivery context			Exploring whether factors external to the scheme have impacted on implementation and, if so, in what way.
Costs	Ongoing financial monitoring and contingency planning.	Outturn investment costs, the use of contingency budgets, any cost savings or overruns, outturn operating, maintenance or other capital costs.	This could provide a wider analysis of the inputs to the scheme beyond the financial. For instance, staff, expertise and specific tools required to deliver the scheme. It would also generate more detailed understanding of the reasons behind any divergences from the planned scheme budget.
What has worked more or less well			
Stakeholder management	Ongoing review that the stakeholder management plan is appropriate, up to date and being implemented.	Reporting what stakeholder management approaches were actually adopted and identifying lessons learnt from these approaches.	This should build on the standard monitoring and may involve a wider range evidence gathering than standard monitoring. For instance, obtaining evidence directly from the stakeholders about their experiences.
Risk management	Ongoing management of risks which are escalated when required. Consistent assessment of risks and clear	Review of effectiveness of risk management processes. This might	This may involve more in-depth data collection to learn lessons what factors were critical to successful risk management,

---

[11] These are based on the Monitoring and Evaluation Framework for Local Authority Major Schemes, there may be other aspects which a process evaluation could also assess. These are not mutually exclusive categories.

	ownership.	include an assessment of which risks were realised and with what impact.	how challenges were overcome and the implications of any unforeseen risks. It may also be useful to undertake case study analyses of specific risks which have been escalated and how these are managed.
Mitigation Measures	Ongoing review of delivery of mitigation measures and their effects on reducing potential impacts to tolerable levels.	Description of how proposed mitigation measures have changed during / following implementation and the reasons for the changes.	This may also explore how effective mitigation measures were built into the scheme and whether these delivered the intended effects and whether any unintended issues arose which required additional mitigation.
Whether it is delivering the expected outputs and outcomes			
Scheme outputs	Ongoing assurance that outputs comply with agreed quality criteria and delivered within a defined quality assurance process.	Describe actual scheme outputs and identify any changes to these since funding approval and any changes to ongoing operational assumptions (e.g. fare levels and provision of services).	Assessing whether the outputs have been delivered to the required quality standard, whether they are bringing about the changes anticipated in the logic map to deliver the outcomes and whether any unanticipated changes have occurred as a result of the outputs.
Benefits	Ensuring that the benefits are defined and measurable and the plan in place is being actioned and monitored, ahead of checking whether benefits have been realised.	An assessment of which benefits have been realised during implementation and whether the scheme on track to deliver all anticipated benefits. Also an assessment of whether the scheme has reached the intended beneficiaries.	This may explore in more depth the experience of service users, wider members of the target population and delivery partners during and following implementation. This is in order to test whether the anticipated benefits were realistic, to assess whether external factors have affected the achievement of the benefit (and, if so, how), explore any disbenefits or unintended consequences and to draw out lessons about the effectiveness of the scheme in delivering the anticipated benefits

Tip Box 4: Tips for getting the most out of a process evaluation

- Design in real-time data collection and analysis to enable opportunity to feedback into delivery. Don't underestimate the importance of good records management. Implementation diaries can provide a useful resource for documenting the delivery stages.

- Monitor whether the scheme remains on track to deliver the anticipated outputs and outcomes. Process evaluations explore in depth what lessons have been learnt and why the scheme is on track or not. This evidence can be used to feed into project management and assurance activities.

- Capture lessons learnt during delivery, leaving this to after implementation risks loss of corporate memory.

- Process evaluations can offer the most value when undertaken by independent evaluators who can learn from the delivery team and also provide an objective assessment of the documentary evidence and the feedback provided by stakeholders and the target population.

- Process evaluations should not just focus on internal processes but also on the experience and perceptions of other delivery partners, stakeholders and end users / target population, who are all critical to the success of the scheme.

- Consider which evidence is best collected via a process evaluation about input and output measures for cost-benefit analysis. If the specific data requirements for economic analysis are not considered when designing the process evaluation, a meaningful economic evaluation might be effectively precluded, as it will not be possible to collect the information retrospectively.

- Some schemes may have elements which are particularly innovative or novel which may warrant a more detailed exploration than other elements of scheme delivery.

When designing a process evaluation it is important to think about the quality of the data collected. It is likely to need to draw on both objective data (such as administrative data) and subjective data (based on people's experiences and perceptions) which means that triangulating these sources of data is important to build robust analysis and inform interpretation. Section 3.3.2 discusses data collection approaches. For more information about delivering a process evaluation see Chapter 8 of the HMT Magenta Book.

3.1.3 Impact Evaluation

Assessing the difference a scheme has made involves a focus on the outcomes of the scheme and should seek to answer the following:

- Did the scheme achieve its stated objectives and realise its benefits?

- Were there observable and statistically significant changes in the outcomes? If so, how big was the change? How much can be said to have been caused by the scheme as opposed to other factors?

- Did any outcomes occur which were not originally intended, and if so, what and how significant were they?

In planning for the impact evaluation it is important to consider which type of impact evaluation (these are set out below) is most appropriate for the scheme. It is important to invest time at this stage in understanding the different types of approaches and the benefits and limitations they generate as this will have implications for the conclusions which can be confidently taken from the evaluation. Therefore, the needs of the evaluation audience should also be taken into consideration at this stage. It may be appropriate to undertake a scoping or feasibility study before deciding which approach is most suitable.

Factors which will influence this consideration include the:

- Objectives of the evaluation and the needs of the audience;

- Quality and reliability of different evaluation approaches to meet the above;

- The feasibility of the evaluation approach given the scheme design;

- Time and resources available for the evaluation;

- Availability and reliability of the existing evidence base and the contribution this evaluation will make to it;

- Complexity of the scheme design and the relationship to the outcomes;

- Degree of influence of confounding factors on the outcomes and impacts; and,

- Quality of existing data sources and measurability of outcomes.

This section provides an overview of the range of different types of impact evaluation approaches and the degree of rigour they generate. However, to start with there are some underlying concepts which should be considered as they affect the quality of an impact evaluation:

Attribution

Evaluating impact involves assessing:

- whether something has happened using descriptive statistics; and,
- whether the scheme was responsible.

Separating out the effects of the scheme from the other influences is the defining feature of an impact evaluation, but is also one of the biggest challenges.

A key concept is the **counterfactual** – what would have occurred had the scheme not taken place. By definition it cannot be observed directly, because the scheme has taken place. Good quality impact evaluations seek to obtain a good estimate of the counterfactual, usually by comparison to areas / populations not exposed to the scheme. For more information please refer to Chapter 9 of the Magenta Book.

For some schemes, it just may not be possible to isolate an effect because the scheme is being delivered as part of a wider package. In these cases, **Contribution Analysis** may be appropriate because it does not attempt to prove that one factor 'caused' the desired outcome, but rather to explore the contribution an initiative is making to observed results. It seeks to answer the question "in light of the multiple factors influencing a result, has the intervention made a noticeable contribution to an observed result and in what way?" (Mayne, 2012[12]). It does this by developing a 'theory of change' (discussed more below) showing the links between the activities, outcomes and contexts of the policy and collecting evidence from various sources to test this theory, the aim is to build a credible 'performance story'. This can demonstrate whether interventions were indeed an important influencing factor in driving change, perhaps along with other factors.

For further information please refer to guidance prepared by the Scottish Government http://www.scotland.gov.uk/Resource/Doc/175356/0116687.pdf

Strength of evaluation design

Internal Validity and **External Validity** are key tests for assessing the strength of the evaluation design.

Internal Validity is used to assess the extent to which the design will provide results which are a true reflection of the impact on the target population. This means that assumptions underpinning the evaluation design and analysis will need to be formally tested. Some impact evaluation and data collection approaches provide stronger assurances to internal validity than others (for instance Randomised Control Trials (see Table 3.2 below) are considered to be the strongest approaches for delivering internal validity), and this should be a consideration when deciding on approaches.

External Validity is important for deciding whether the impact estimated for the scheme can be extrapolated / generalised to others areas. Threats to external validity include:
- Producing evaluation findings which are not representative of the wider population;

---

[12] Mayne. J, (2012) 'Contribution analysis: Coming of age?' Evaluation 18(3) 270-280

- Not accounting for the influence of other contextual factors in the analysis of scheme impact; and,
- Not assessing any substitution or displacement effects (for instance whether a reduction in congestion in one area has led to increased congestion in another area).

Experimental impact evaluations

Experimental (also referred to as empirical) impact evaluations seek to find out whether the scheme caused the anticipated outcomes to occur. They require both a measure of the outcome and a means of estimating what would have happened without the scheme, usually using a comparison group. Types of experimental impact evaluations are given in Table 3.2.

Experimental impact evaluation is considered to be the most robust impact evaluation approach for internal validity and attribution as it seeks to isolate the effect of the scheme from the other factors affecting the outcomes and impacts of interest.

Experimental evaluations are *most suited* to schemes when:

There is a means of estimating the counterfactual;

Data of adequate quality and quantity will be available to support the analysis and estimation of the counterfactual; and,

The level of scheme effect is large enough to be detected amongst the random fluctuations (or noise) in the outcome measures.

Table 3.2: Evaluation Designs
Randomised Control Trials (RCTs)
Usually regarded as the most robust means of evaluating policy because it has strong internal validity as the treatment and control groups are randomly assigned and thus are equivalent in characteristics. Therefore, they only differ in their exposure to the scheme and any observed changes will be due to the intervention. Refer to Cabinet Office paper on RCTs (also note that could answer some questions if not suitable to the whole scheme?)
Quasi Experimental Designs (QED)
Used where the allocation cannot be random, for instance if the intervention is to be given to a whole group or area and it is impossible to control who has access to the scheme - often true in the field of transport. In these cases one of two approaches (or a combination of both) will be used: A comparator group is selected and matched in key characteristics to ensure it is as similar as possible to the treatment group; Acknowledging that the comparison group is non-equivalent, but obtaining it in a way that allows selection bias to be modelled.

For more information see the Magenta Book Chapter 9 and Guidance for Transport Impact Evaluations[13].

In a transport context it may not be feasible to control who benefits from the scheme at an individual level and even if comparing geographical areas there are limitations in the extent to which areas can be 'matched' especially at higher spatial levels. It might be more feasible at a local level but this depends on:

- whether the boundaries of the scheme are contained within a distinct geographical area or cut across a number of areas; and,

- whether the scheme impact across the spatial level is large enough to be detected.

Tip Box 5 – Evaluating Voluntary Participation

"Voluntary participation in an intervention is an example of non-randomness that is a particular problem for the evaluator" (HMT Magenta Book, 2011, pp 108). However, this is a common occurrence for transport schemes. Evaluations which only assess the impact on scheme users are at risk of not being able to generalise the findings to the entire target population as scheme users are likely to have different transport needs and motivations to non-users. Similarly, evaluations which seek to compare users with non-users to estimate the counterfactual will risk generating biased findings because the two groups are non-equivalent.

The Magenta Book recommends that an approach to this is to evaluate the scheme on the basis of intention to treat (ITT). The 'treatment group' consists of all those within the target population, even those who do not use the transport scheme. The comparison group is drawn from areas that do not have access to the scheme.

Impacts estimated on an ITT basis tend to be smaller than those based on the actual treatment group, since it includes non-users. However this approach has stronger internal validity since it measures the reach of the scheme investment and the effect per person in the treatment area.

---

[13] The Tavistock Institute for DfT, 2010 http://assets.dft.gov.uk/publications/pgr-evaluation-evaluationguidance-transportimpact-doc/guidance.doc

Theory-based impact evaluations

Theory-based evaluation approaches provide an overarching framework for understanding, systematically testing and refining the assumptions held about how the scheme will deliver the anticipated outcomes and impacts. They seek to not only understand whether a scheme has worked, but also why and under what conditions the scheme will generate the changes necessary. Theory-based evaluations draw heavily on the logic map to assess whether the underlying theories of change have been delivered. They attempt to make credible causal claims in the absence of experimental approaches (Mayne 2012)[14]

Specifically they are *appropriate* for schemes which are complex - for instance in terms of having a range of activities or delivery mechanisms; work in a variety of settings where the routes to achieving the impact are not straightforward / direct, or where the outcomes and impacts are also likely to be influenced by the wider delivery context.

These evaluations can therefore be particularly relevant where relatively little is known about how the scheme impacts occur (causality), for instance in the case of innovative schemes. Theory-based evaluations also provide the opportunity for interim evaluations to track progress towards longer-term impacts, even if these have not yet been achieved.

Importantly, theory-based methods have the flexibility to mix different evaluation methods. For instance, a theory-based evaluation provides a framework for drawing on experimental methods to measure the impact of the scheme. Incorporating a number of techniques within their design allows such evaluations to triangulate evidence[15].

Although theory-based methods are flexible, their *limitations* mean they will not be suitable in some cases. For instance, they:

Are less able than experimental evaluations to provide robust and quantifiable evidence of links between the scheme and its outcomes;

Can be challenging and resource intensive, where the scheme is large or particularly complex, because of the increased number of factors, mechanisms, groups and circumstances that have to be considered; and

Require the evaluator to work closely with scheme design and delivery teams in mapping the intervention logic which, combined with the incorporation of a wide range of different data sources, can be seen as threatening the objectivity of the evaluator.

For more information see the Magenta Book Chapter 6 and Guidance for Transport Impact Evaluations[16]

---

[14] Mayne. J, (2012) 'Contribution analysis: Coming of age?' Evaluation 18(3) 270-280
[15] Triangulation, or the integration and mixing of evidence, from different sources is a technique to generate robust conclusions.
[16] The Tavistock Institute for DfT, 2010 http://assets.dft.gov.uk/publications/pgr-evaluation-evaluationguidance-transportimpact-doc/guidance.doc

Before and After Studies

The data collected for the standard monitoring of all schemes focuses on observing changes to key metrics before and after the scheme has been implemented. However, as the Magenta Book explains, "sometimes the level of evidence available [from before and after studies] falls far short of what would generally be regarded as a true impact evaluation"[17]. Therefore, evaluation planners should consider alternative approaches first. If this is considered to be the only suitable approach for the scheme the evaluation planner needs to provide a coherent justification in the evaluation plan noting that this. approach may be appropriate in just a few cases when the causal logic is able to confirm that that the scheme is the only factor which could influence the outcomes being measured. They also need to be aware of the limitations to how the evidence can be used (presented below) and consider how they may mitigate these risks in their evaluation design.

Before and after approaches do have some significant *limitations*, however, which should be borne in mind. For instance, they:

- Are good for assessing change where the level of change is large and the link between the scheme and the outcome is relatively straightforward, but they are weak when the level of change is quite small and the link between scheme and outcome is less direct or more complex;

- Are poor at providing clear evidence of a causal relationship between the intervention and any outcomes and impacts, where there are other external factors that may have had an influence on the observed change;

- Predict causal relationships between the scheme and its outcomes (e.g. via the logic maps) but never fully test these assumptions – again other factors may be influencing the outcomes which are not investigated;

- Do not provide answers if the observed outcomes are a significant departure from what was anticipated; and

- Are unlikely to capture what any unintended consequences / benefits or displacement effects of the scheme were.

These factors would affect the confidence with which the evaluation findings were interpreted and would reduce the opportunity to apply experiences into future scheme design. Inappropriate use of outcome evaluations can produce false negatives (where the findings suggest little/no change where the scheme was in fact successful) or false positives (the findings suggest change where in fact there was none), resulting in misdirected implications.

---

[17] HMT Magenta Book 2011:122

3.2 Identifying the data requirements

Key questions to ask:

- What data are required?
- What is already being collected / available?
- What additional data need to be collected?
- What are the key timeframes for data collection?
- How and when will the baseline be established?
- Who will be responsible for data collection and what processes need to be set up?
- How will the data be gathered, transferred and stored?
- What data transfer, data protection and ethical requirements need to be taken into account when planning the data collection?
- How will the data be verified to ensure it is accurate and consistent with the relevant requirements?
- How can data quality and robustness be maximised?

A good evaluation relies on good quality data. Establishing data requirements when planning an evaluation is important to ensure time to identify relevant existing data sources and set up any bespoke data collection.

The most pressing driver for this will be the requirement to establish a baseline for measuring change as part of an impact evaluation. What data is collected will depend on the type of evaluation proposed and the research questions to be answered which have been set out in earlier planning steps. The logic map can also help to guide planning for data collection. Table 3.3 summarises some of the key questions for evaluation planners to ask to help them consider their data requirements.

Table 3.3: Key questions and considerations for assessing data requirements	
Key question	**Consideration**
What type of data will be required to answer each research question?	- Is numerical data required? - Is documentary data required? - Is observational data required? - Is data to describe people's experience, opinions, and views required? - Will a combination of these types of data be required?
Who or what can provide this data?	- Which end users, service providers, stakeholders, databases etc. would have this data and / or need to be consulted? - Who will have responsibility for gathering the data? - How will the data be accessed? What format will the

	data be provided in? How will the data be gathered, transferred and stored?
	• Will the data be gathered in a reliable and consistent fashion?
	• Are there any potential sensitivities / data protection / ethical issues in collecting data from these groups, areas, databases etc?
What section of the population of interest should data be collected from?	• Who is the population of interest?
	• Will the research be a census of all available data / population of interest or will a sample of the population studied?
	• For qualitative sampling – what range of people, experience, organizations, contexts etc. need to be covered?
	• For quantitative data, what types of estimate will the data need to provide and how precise? Which sub-populations need to be included? What impact does this have on the sample size required?
	• For quantitative and qualitative data – what sampling frames are available or will need to be created?
What are the key timeframes for data collection?	• When will data collection start for each measure?
	• How frequently will the data be collected?
	• What are the strengths and limitations to this approach?
How will the data be analysed?	• Does the method of analysis that will be used require a particular sample size or type of data to have been collected?
	• How will the data be verified to ensure it is accurate and consistent with the relevant requirements?

Types of data for fuller evaluations

The specific data required for all of the aspects of a fuller evaluation will relate to the inputs, outputs, outcomes, and impacts of the scheme (which will have been articulated through the logic map). The Magenta Book[18] explains there are four main types of data which should be considered:

Existing administrative data (that is not being collected specifically for the evaluation) e.g. Reported road accident statistics (STATS19), bus punctuality / reliability;

Long-term, large-scale, often longitudinal national survey datasets (e.g. ONS Labour Force Survey, Census, National Travel Survey);

---

[18] The Magenta Book, Chapter 7, HM Treasury, 2011

Monitoring data or performance management data that are already being collected to support the implementation of the scheme (e.g. delivery data, financial data, quality assurance information); and,

New / bespoke primary data collected specifically to support the evaluation information needs (e.g. specially commissioned travel surveys, traffic counts, qualitative interviews / focus groups with delivery teams, stakeholders and end users).

Timing of data collection

Planning for data collection should be considered during the evaluation design stage to ensure that[19]:

- the data is collected at the point in time and is available when it is required;

- the data requirements can be designed into the scheme as far as possible, so they become part of a routine process;

- baselines and counterfactual data can be collected; and,

- where third party organisations need to provide some of the data, the requirement to do this can be built into their contract (or Service Level Agreement or Memorandum of Understanding etc) from the outset.

The timing of the data collection needs to be considered carefully in relation to the timeframe of the scheme's expected delivery, outcomes and impacts and the plans for reporting back the findings (see Section 2.1. regarding how to define these in the evaluation context and Figure 3.1 for more detail about how data collection fits with the overall evaluation life cycle).

Establishing a baseline is a very important exercise in order to assess change over time and should be measured before any effects of the scheme are felt. The timing of this will depend on how the scheme is being implemented. For instance, the implementation of some schemes will have an immediate effect on the local transport network e.g. if an existing route / service is closed / redirected during the implementation of the scheme. This may mean that any baseline measurement during this period may pick up these effects rather than providing an accurate measure of the situation before the scheme. In these instances, the baseline would be more reliable if it was measured before work on the scheme had started. On the other hand, to minimise the risk that the length of time between the baseline measurement and scheme opening could mean that the baseline measure becomes out of date, schemes which are implemented in isolation from the existing transport network should seek to minimise the time lag between baseline measurement and scheme opening as far as possible.

---

[19] The Magenta Book, Chapter 7 section 7.12, HM Treasury, 2011

The Monitoring and Evaluation Framework[20] sets out an expectation that two reports will be generated following the delivery of the scheme. The first is an initial report based on the data collected within the first year following scheme opening and the second is a final report based on the data collected within the first five years following scheme opening. This is to ensure the monitoring and evaluation reports capture both short-term and longer-term outcomes and impacts.

Data collection will in reality span a number of years and evaluation planners should set out the time points for different aspects of the data collection which will be used to inform the process and the impact evaluation. Appendix 3 of the Monitoring and Evaluation Framework for Local Authority Major Schemes sets out some considerations for the timing for fuller evaluation data collection and reporting which should be reflected in the evaluation plan.

Tip Box 6 : Timing of data collection to measure change over time

- Consider the number of observation points during the data collection period. In order to track changes over time, the evaluation plan should consider whether it is feasible to collect time series data with multiple data points collected in a consistent and comparable fashion.
- Use the logic map to consider whether the frequency of data collection is suitable for each measure.
- When planning data sources consider whether how frequently the data is collected, whether it is continuous or intermittent and the strengths and limitations of how the data can be used to track changes over time.
- For primary data collection consider whether it is possible to design a longitudinal approach.

Tip Box 7: Cross Sectional and Longitudinal Data Collection	
Cross- Sectional	**Longitudinal**
Data is collected from a sample at a given point in time - providing a 'snapshot' in the particular sample studied.	Data is collected from the same sample at two or more points in time.
Pros: Generally less expensive, useful for exploratory research.	**Pros:** Greater power in assessing change over time.
Cons: Less robust for assessing change over time as may not be comparing like with like	**Cons:** More expensive, participants may drop out, generally need larger sample size.

---

[20] The Monitoring and Evaluation Framework for Local Authority Major Schemes, paragraph 7.6, DfT, 2012

Monitoring data

Monitoring data can play a key role in fuller evaluations and can include data relating to each component of the logic map. The Monitoring and Evaluation Framework for Local Authority Major Schemes sets out the level of standard and enhanced (where applicable) monitoring data which schemes undertaking fuller evaluations should be preparing to collect.

Evaluation planners should consider how the fuller evaluation will draw on the data collected for standard and enhanced monitoring, whether any further monitoring data could also be relevant and then consider how any remaining evidence gaps can be best filled[21]. It is important to note, however, that an evaluation may not be able to rely solely on monitoring or local indicator data (secondary data) and it will often be necessary to collect primary data. Care should also be taken to assess the quality of the monitoring data being collected and ensure that any limitations of the data are fully understood.

Tip Box 8: Data quality

- Explore the availability and suitability of existing monitoring data, consider what data may be available from other sources e.g. bus operators, datasets for secondary data analysis or any existing vehicles of primary data collection such as local travel surveys or omnibus surveys to help reduce evaluation costs (see also Section 4.1.1).

- However, in making a decision about the use of these data sources, considerations need to be made about:

  - Their suitability and relevance in answering the research questions. For instance, do they directly measure the metrics under investigation or will they only provide a proxy measure? What assumptions will underpin the way proxy measures can be analysed and interpreted and what risks will this have for the evaluation conclusions? Are there any key aspects of the causal theory which will go untested?

  - The level of rigour and robustness they provide. Have / will the data been / be produced using rigorous methods? How have risks to data error been minimised? Are there risks of systematic bias in the data? Is the quality of the data suitable for the planned level of data analysis?

  - When the data will be available for analysis, the time period that will be covered and the frequency of data collection (i.e. will it provide a time series or snapshot?) and,

  - The extent to which evidence gaps will remain, whether these are critical to the evaluation and whether these can be filled through bespoke data collection?

---

[21] The Magenta Book, Box 7.C 'Designing an effective monitoring system' offers a useful tool to evaluation planners, HM Treasury, 2011

Tip Box 9: Categorisation of metrics/indicators

In addition to being categorised according to scheme component, indicators can also be classed in other ways.

Contextual indicators are important in providing a clearer picture of the situation in which a scheme is implemented and changes occur. Contextual indicators can reveal how an outcome may have been influenced by something other than the scheme itself (such as the wider environment of a scheme or other changes taking place) and can show how transferable a scheme's context is for similar implementations in the future, e.g. fuel prices, cost of living, characteristics of the local population, geographical characteristics.

If an indicator is unable to be measured directly, then a **proxy indicator** can be identified to be used as a 'stand in'. The proxy indicator will refer indirectly to the subject of interest.

3.3.2 Primary data collection

Primary data collection techniques are generally categorised as either quantitative or qualitative. Key methods for gathering quantitative and qualitative data, as well as an illustration of the type of questions they can help to address, are given in Figures 3.2 and 3.3 below.

Qualitative data typically provides evidence for the 'how' and 'why' questions of an evaluation. Qualitative techniques can be used for one aspect of an evaluation (e.g. to develop an intervention logic) and is often a key source of data for process evaluations. Data collection techniques are discussed in Chapter 7 of the HMT Magenta Book.

Tip Box 10: Triangulation of data

It can be useful to collect data using a number of different methods, e.g. monitoring data, bespoke surveys, qualitative interviews, to answer particular research questions. Combining data from more than one source can increase confidence in the results, corroborate findings and deepen understanding.

When planning for data collection using a range of data collection techniques it is important to consider how the evidence will be synthesised and analysed so robust conclusions can be drawn across the entire evidence base.

Figure 3.2: Quantitative tools and techniques

Question type

What are current travel behaviours? How do they differ by groups?

What level of change has been produced? What change could have occurred even without the scheme?

Were there unintended or displacement effects?

Who is/was affected by the scheme?

How are the scheme impacts distributed?

Were the scheme assumptions correct?

What was the impact on congestion/ emissions/ noise?

Tools

Surveys
E.g. Residents/ households Businesses Employees Schools Traveller intercept surveys

Travel diaries

National datasets

Transport models

Performance data

Scheme budget accounts

Monitoring data
E.g. Modal counts Emissions Noise traffic surveys

Patronage data

Appraisal models

Figure 3.3: Qualitative tools and techniques

Question type

Why do different groups behave as they do?

What is the intervention logic (context-input-output-outcome-impact)?

What are the barriers to change?

How were delivery issues overcome?

What were the factors that contributed to success?

How can this scheme/ similar schemes be improved?

How effective were risk management strategies?

What caused the unintended outcomes?

Tools

Interviews
E.g. with the delivery team, Key recipient groups and non recipients, Transport Providers,
Other stakeholders

Workshops

Focus groups

Evidence reviews
E.g. budget accounts Implementation diaries Delivery timetables Interim reports meeting minutes

Consultations

Logic mapping

3.3.3 Collecting reliable data

Care should be taken to establish the quality of all data collected, and collecting robust data must be addressed in the evaluation design - one of the key limiting factors in how strong evaluation evidence can be is how the scheme is delivered in the first place. Issues of sampling strategy, sample size, questionnaire design, and fieldwork approach are just some of the key factors that will affect how reliable and credible data will be (see Boxes below).

Evaluation planners will need to have the necessary skills within the evaluation team to ensure data collected reaches the required standard.

Tip Box 11: Sampling

Sampling describes the methods by which "units" are chosen to participate in the research. "Units" is a technical term used to describe the finest level of detail which your research will go down to. A unit could be an individual person, a household, a business, an area, etc. Sampling can affect the results of your research in important ways. Choosing the wrong sampling approach, or not paying close attention to how the sampling is conducted, can lead to inaccurate findings.

Random sampling will provide the most accurate reflection of the characteristics of the population, especially where it is combined with strategies for maximising the response rate amongst all subgroups (e.g. men and women; different ethnic groups; different age groups). Where it is possible to use a random sample, it should always be preferred for reliability reasons. There are many variations on random sampling which are used in different circumstances (e.g. cluster sampling, stratified sampling, etc), but all share the same key principle of equal chance of selection. This helps to minimise systematic bias in the sample.
Pros: less prone to bias; more likely to generate reliable results.
Cons: more expensive; more complicated to administer; survey providers may be reluctant to use this approach.

Quota sampling is more likely to introduce bias into the findings. Sophisticated forms of quota sampling may be able to limit the extent of that bias, but will be more complex and expensive to administer and thus offer lower savings compared to random sampling. Where quota sampling is the only option, ensure that bias is minimised by designing the quotas carefully and not limiting recruitment in ways which might affect the response e.g. to particular places or times of day.
Pros: cheaper; simpler; quicker.
Cons: less reliable due to risk of bias.

Tip Box 12: Sample Size

Although increasing sample size can be costly, having a sample which is too small can undermine the value of the research, particularly if you are interested in particular groups. The smaller the sample, the wider the degree of uncertainty there will be in the findings. An appropriate balance between robustness and cost can be achieved through careful sample design. Consult a statistician to find out the minimum sample size required to produce data which you can rely on.

3.3.4 Analysis of data

Analysis of data can take various forms. There are many issues to consider on the issue of quantitative analysis which this guidance does not cover, but evaluations planners can find more information in chapters 9 and 10 of the Magenta Book In many cases, evaluation planners will benefit from engaging with experienced practitioners to design and conduct such analysis to ensure the conclusions are robust and avoid the many potential pitfalls.

Because the rigorous collection and analysis of qualitative data can be just as challenging as quantitative data, there are significant benefits to bringing in external specialists to assist. In the case of process evaluations, it is important for the findings to be gathered and interpreted by an independent group who are distanced from the delivery and implications of success or failure.

The evaluation planner should set out plans for analysing and synthesising the evaluation findings within the broad evaluation approach which has been selected for the scheme.

3.3.5 Ethical and data protection considerations

In planning the data collection and evaluation design, it is necessary to consider not only practical and resource implications but also some key ethical issues. The Ethics Checklist provided in Annex C may be useful in working through each potential methodology and ensuring it is appropriate for the situation and audience.

Summary of Section B

Section B set out what evaluation planners should consider for the final parts of their evaluation plans parts 4 - 6, see Fig 3.4.

Figure 3.4: Considerations for Evaluation Planners
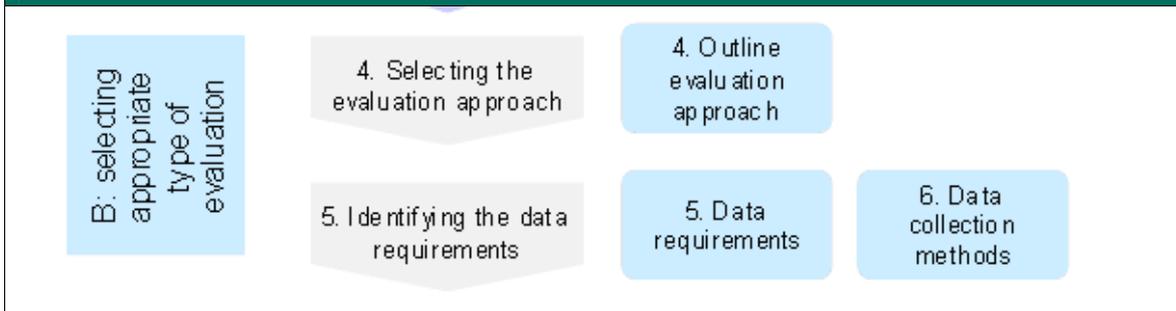
4. Section C – Evaluation Delivery and Use of Evidence

Introduction

This section helps evaluation planners to make decisions about the how the evaluation will be delivered, resourced and managed. It also considers how to plan for using the evidence produced by the evaluation and to ensure the messages are disseminated effectively.

This section will help the evaluation planner to work through the final three steps:

1. **Identifying necessary resources and governance arrangements** for the evaluation;
2. **Quality Assurance** including an assessment of delivery risk; and,
3. **Using and disseminating the evaluation findings.**

Working through these steps will help you develop sections 7-9 of the monitoring and evaluation plan required for fuller evaluations and deliver the following outputs:

- An evaluation budget;
- A governance structure and management arrangements;
- A clear assessment of the capabilities and skills required to deliver the evaluation and a strategy for ensuring these are resourced; and
- A reporting and dissemination plan.

4.1 Identifying the necessary resources and governance arrangements

> **Key questions to consider:**
>
> - What is a proportionate level of resource for the evaluation?
> - What is the best governance structure to have in place?
> - What budget is to be used for the evaluation? Has sufficient allowance been built in?
> - Who will be the project manager, provide analytical support, and be on the project board?

4.1.1 Resources

An evaluation's resource requirements can vary according to the type of scheme being implemented, the evaluation's objectives, its context and the level of data collection required. Consideration therefore needs to be made as early as possible on the appropriate levels of resource to commit to the evaluation (such as staff time and research budgets etc.) and which elements of work should be conducted in-house or contracted out. It will be necessary to take an individual assessment of the scheme in question, taking key factors into account.

Proportionality is an important consideration - ensuring the resources allocated reflect the scale, risks and profile of the scheme. The scheme's *complexity* may also affect the resources needed. Chapter four of the Magenta Book[22] sets out some factors to consider when setting the resources for an evaluation (summarised below). These principles were applied by DfT when selecting schemes for fuller evaluation (see Monitoring and Evaluation Framework for Local Authority Major Schemes[23] Table A5). They will also be valuable to use by the evaluation planners when deciding on the budgets for their scheme evaluation.

---

[22] HM Treasury, (2011) *The Magenta Book, Guidance for evaluation* http://www.hm-treasury.gov.uk/magentabook
[23] Monitoring and evaluation framework for Local Authority Major Schemes, DfT, 2012, p30

Tip Box 13: Factors affecting appropriate resourcing of an evaluation

- Scheme nature: For schemes which *are large scale, high-profile, innovative or high risk* then a thorough and robust evaluation is needed to assess whether the scheme has been successful;
- Use of evidence: More resource may be needed for evaluations to produce *generalisable evidence* which will *influence* future decision making or fill important gaps in the *evidence base*
- Variability of impact: The effects of schemes with highly uncertain outcomes or with significant behavioural effects are likely to be more difficult to isolate, and there is likely to be a greater case for conducting a more extensive evaluation.

For more information please refer to table 4.C in chapter 4 of the Magenta Book (2011)

The types of resources likely to be needed for an evaluation are given in Figure 4.1. Sections A and B of the evaluation plan will have provided a clear understanding of the scope of the evaluation which needs to be resourced. At this stage it is important to be clear about the resources available and how far they will deliver against the objectives for the evaluation which have been set.

Tip Box 14: Balancing costs and quality (also see Tip Box 8)

- Pressures on the evaluation project timescales and resources may risk compromises in quality and therefore the robustness and value of the evaluation. Any trade-offs should be first assessed in terms of the risks they pose to the evaluation and whether these can be maintained at a tolerable level and ensuring stakeholders agree with this. It may for example, be preferable to reduce the scope rather than the quality of the evaluation.

- Balancing in-house and external resources – A judgement needs to be made about the scale and type of evaluation that is required or possible and the trade-offs that this would require, including whether it should be commissioned externally or conducted (either partly or wholly) in-house (this should include the consideration of the risks to the perceived independence of the evaluation if it is wholly conducted in-house).

- Maximising opportunities for economies of scale by exploring ways in which methodological approaches can be shared across schemes.

- In some circumstances a scoping or feasibility study may be conducted to support this decision making process. This can foster greater understanding of what can and cannot be evaluated, and therefore what level of investment is required.

Figure 4.1: Types of resources employed in evaluation (adapted from Table 4:B of the Magenta Book)

It is important to establish a budget for the evaluation during the planning stages

The timeframe for the evaluation data collection, analysis and reporting will have an impact on the budget planning as evaluations will continue to collect data years after the scheme has delivered

Making the best use of good quality monitoring data or existing data can help manage evaluation costs as long as the data is suitable for the metric being measured

Financial resources

Delivery Bodies

A successful evaluation depends on the early and continued engagement and cooperation of the organisations and individuals involved in delivering the scheme. It is important to communicate what the evaluation seeks to address, what input will be required from them and how they might benefit from the findings

Peer Review

In order to ensure quality it is valuable to have aspects of the evaluation independently peer reviewed. These might include the evaluation plan, methodology, evaluation tools or findings

Evaluations are multidisciplinary and a range of analytical specialisms (social research, economics, statistics) can be drawn upon to help design, steer, quality assure or deliver aspects of the evaluation

Analytical Support

The balance between using in-house or external organisations depends on the capabilities and capacity of the in-house teams and the need for independent evidence. It may be suitable to develop an approach which draws on both

Both internal and external evaluations require a dedicated project manager (with the specialist expertise to assure quality) who is responsible for: procurement, day-to-day management, quality assurance, budget management and directing the evaluation team

Management Resources

Wider Stakeholders

Key stakeholders may need to be involved in the evaluation but the scale of this involvement needs to be considered on an individual basis but could include representation on advisory or steering groups.

Internal Resources   *External Resources*

4.1.2 Evaluation Governance

The evaluation plan should set out the governance structure which will be put in place for the delivery of the evaluation. It should specify who will be responsible for delivering the plan and managing quality assurance and risks.

Evaluations are likely to be delivered over a long timeframe and therefore it may be appropriate to manage the project over a number of stages to ensure that quality is maintained throughout and the evaluation keeps on track. Figure 4.2 below presents some considerations for evaluation planning. This suggests that a minimum of four stages may be appropriate but the evaluation planner may wish to break this down further to better reflect different activities especially as some of the aspects are iterative rather than linear such as data collection, analysis and reporting.

The evaluation plan provides a good opportunity to consider and explain how the evaluation will dovetail with the other forms of project assurance being applied to the scheme.

Figure 4.2: Evaluation Planning Stages

Assurance reviews at key delivery points
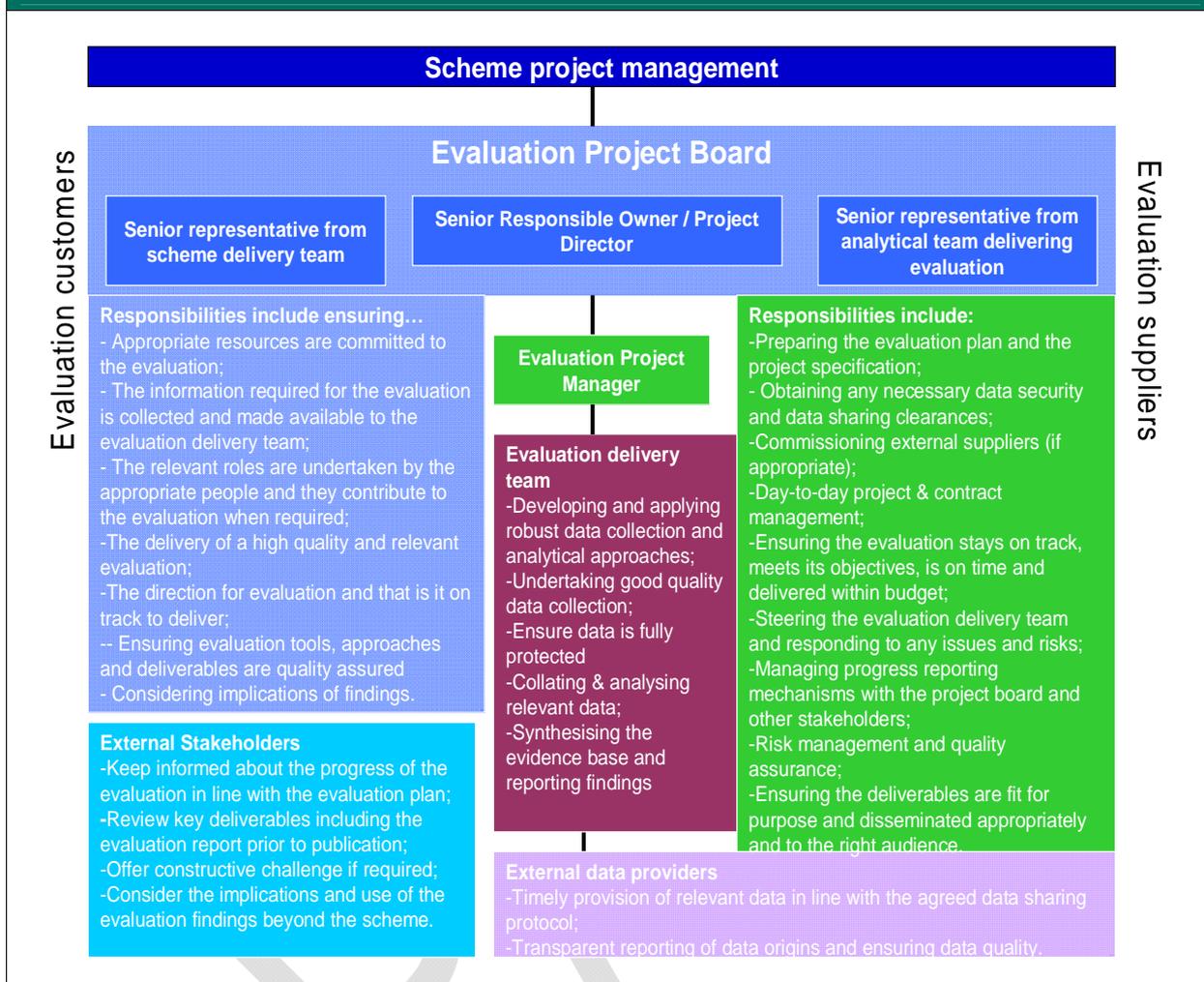
Develop idea

Submit Business case

Agree evaluation plan

Scheme business case development

SCHEME DELIVERY

Benefits

Benefits Realisation

Evaluation

Develop

Baseline

Process Evaluation data collection and analysis

Post scheme data collection

Analyse impacts

Economic analysis

Report

Data collection

Evaluation planning, Methodological development, feasibility testing, securing resources, commissioning specialist skills / peer review

Development of data collection / collation, testing and piloting data collection tools, quality assuring data collection tools, primary research fieldwork, data management / processing / cleaning, feedback any lessons learnt from process evaluation if it can help improve delivery

Applying, testing & QA of analysis approaches & model development

Evidence synthesis, stakeholder & peer review, publication and dissemination

Establishing a project management team is important during the evaluation planning stages. Roles and responsibilities of the team should be clearly articulated and it should be adequately resourced. A typical evaluation project management team structure may look similar to the illustrative Evaluation Project Management Team Figure 4.3.

Figure 4.3: Evaluation Project Management Team[24]

Scheme project management

Evaluation customers

Evaluation suppliers

Evaluation Project Board

| Senior representative from scheme delivery team | Senior Responsible Owner / Project Director | Senior representative from analytical team delivering evaluation |

Responsibilities include ensuring…
- Appropriate resources are committed to the evaluation;
- The information required for the evaluation is collected and made available to the evaluation delivery team;
- The relevant roles are undertaken by the appropriate people and they contribute to the evaluation when required;
- The delivery of a high quality and relevant evaluation;
- The direction for evaluation and that is it on track to deliver;
-- Ensuring evaluation tools, approaches and deliverables are quality assured
- Considering implications of findings.

External Stakeholders
- Keep informed about the progress of the evaluation in line with the evaluation plan;
- Review key deliverables including the evaluation report prior to publication;
- Offer constructive challenge if required;
- Consider the implications and use of the evaluation findings beyond the scheme.

Evaluation Project Manager

Evaluation delivery team
- Developing and applying robust data collection and analytical approaches;
- Undertaking good quality data collection;
- Ensure data is fully protected
- Collating & analysing relevant data;
- Synthesising the evidence base and reporting findings

Responsibilities include:
- Preparing the evaluation plan and the project specification;
- Obtaining any necessary data security and data sharing clearances;
- Commissioning external suppliers (if appropriate);
- Day-to-day project & contract management;
- Ensuring the evaluation stays on track, meets its objectives, is on time and delivered within budget;
- Steering the evaluation delivery team and responding to any issues and risks;
- Managing progress reporting mechanisms with the project board and other stakeholders;
- Risk management and quality assurance;
- Ensuring the deliverables are fit for purpose and disseminated appropriately and to the right audience.

External data providers
- Timely provision of relevant data in line with the agreed data sharing protocol;
- Transparent reporting of data origins and ensuring data quality.

---

[24] Adapted from Table 5.C HM Treasury, (2011) *The Magenta Book, Guidance for evaluation*
http://www.hm-treasury.gov.uk/magentabook

4.2 Conducting the Evaluation

> **Key questions to consider:**
>
> - Have the roles and responsibilities been clearly allocated?
> - Who will be responsible for specification development, tendering, project management and quality assurance?
> - What quality assurance processes should be put in place?
> - How will risks be managed?
> - Is there a clear plan setting out how the evaluation will be delivered and the critical paths to achieving this?

At this step of the evaluation plan, the scope, scale and methodological approach for the evaluation should have been established. However, there are a large number of project management decisions and tasks to be resolved and articulated to set out a clear delivery plan for the evaluation.

Ideally the delivery plan will provide sufficient granularity for each of the evaluation strands being undertaken (process, impact and economic), it should be sufficiently clear who is responsible for the activities and when they will be undertaken.

Relevant activities include[25]:

Drafting a specification of requirements

What data will be collated and by whom (see Section 3.1.3); what data needs to be collected, when and in what format (see Section 3.2);

Commissioning (if appropriate)

What is it that needs to be commissioned, data collection, analysis etc; what needs to be taken into account when commissioning e.g. internal process, best practice, and how will this impact on the delivery timeline?

Day to day project management

What structures, meetings and communications are needed to ensure the smooth running of the day to day management; considering the planning stages (Fig 4.2). What are the most important activities in each stage of the evaluation project?

Setting up the project board and developing the terms of reference

---

[25] A more detailed list is given in Chapter 5 of HM Treasury, (2011) *The Magenta Book, Guidance for evaluation* http://www.hm-treasury.gov.uk/magentabook.

Is there high level buy in to ensure everyone who needs to be on the board and take decisions will commit the time; is there enough time planned to ensure all members of the board are in place to make the decisions and undertake the activities for evaluation assigned to them?

Defining when the evaluation will start and end

Based on the decision made about data collection timeframes (Section 3.2), it should be possible to set out a timeline covering the start and completion of the evaluation. This should consider the relevant evaluation stages, building on figure 4.2 and the critical paths for achieving the plan.

Allocating roles and responsibilities

Clearly allocate who will be responsible for which activities and ensure that this is clearly communicated with them, along with any expectations about their roles and standards for any deliverables. Figure 4.3 provides a useful basis for considering different roles.

Identifying project risks and mitigating actions

The evaluation plan should include a risk management strategy and risk register, setting out who is responsible for owning these risks. See section 4.2.2 for further information.

Setting up quality assurance processes

Section 4.2.1 below provides more information of the consideration which need to made when designing quality assurance processes. It is important that a suitable strategy for assuring the quality of the evaluation is reflected in the evaluation plan and the governance structure.

Budget management

Articulate how the available budget will be allocated across activities covering the duration of the evaluation timeframe (based on considerations identified in section 4.1). How will plans be adapted if aspects of the project become delayed or altered?

4.2.1 Quality Assurance

The Magenta Book clearly explains that implementing good quality control and quality assurance mechanisms are crucial for any evaluation, they are:

> "Without [quality control and assurance], the methods and results from the evaluation cannot be guaranteed to be of sufficiently high standard or fit for purpose. This means the resulting evidence is not robust enough to provide answers to the questions the evaluation was designed to resolve or to reliably inform the decision making process."
> HM Treasury, (2011) *The Magenta Book* (4.3)

- Ensuring that the evaluation design, planning and delivery are properly conducted, conform to professional standards (such as ethical assurance), and that minimum analytical standards are adhered to;

- Being managed within the governance process with all project members having responsibility for ensuring quality. The evaluation manager within the commissioning body should take responsibility for applying quality control criteria. The use of external assessors and / or peer reviewers can also be useful in providing independent assurance.

- Ensuring consistency in data collection, methodology, reporting and interpreting of findings. Therefore, it should be undertaken throughout the evaluation to identify and assess the implications of weaknesses in methodology, design, data collection etc. enabling early response and mitigation measures to be put in place which can help to avoid or reduce any adverse effects on the reliability of the findings.

The Magenta Book sets out four principles as being critical to managing an evaluation in a way which will satisfy quality assurance criteria. These are presented in Figure 4.4. below.

| Figure 4.4: Principles underpinning evaluation quality assurance[26]	
INDEPENDENCE **Researcher independence and objectivity are essential for any evaluation...** **However, this does not automatically necessitate the use of external contractors or keeping the evaluation team at arm's length. A suitable balance should be reached to enable interaction between the research team and the policy and delivery leads, while retaining independence and objectivity.**	**INCLUSIVITY** **The inclusion of recipients, delivery bodies or stakeholders within the evaluation enhances the potential learning from the evaluation and acceptance of its results...** **However, this should be actively managed as a continuous process of communication and engagement. This is likely to involve improving the awareness of the evaluation, obtaining feedback on the evaluation design and communicating scoping, interim and final findings and conclusions.**
ROBUSTNESS **Ensuring that evaluation plans and deliverables are assessed against required analytical standards so that there is an assessment of whether:** **The planned evaluation is likely to provide robust evidence to answer the research questions;** **The evaluation findings and conclusions are presented, reported and interpreted by users accurately and clearly.**	**TRANSPARENCY** **The transparent reporting of the deliverables must be a feature of any evaluation, especially at key points of the evaluation lifecycle:** **The evaluation planning;** **Evaluation governance;** **Evaluation reporting.**

4.2.2 Risk Management

Assessing the possible risks that an evaluation may face is a key component of designing an evaluation and is particularly useful to do once the approach and data collection tools are decided. Evaluation projects should have procedures for risk management, typically in evaluation this includes an assessment of limitations to the proposed evaluation approach, the focus is on risks associated with the evaluation, rather than the scheme. Figure 4.5 shows some of the questions that can be asked to help identify any risks.

Tip Box 15: Risk

Evaluation managers should take as wide a view as possible when identifying risks, for instance considering delivery, methodological, political, resourcing, economic and social factors. Taking early action to reduce the likelihood and impact of risks will be important. Detailed risk assessments should be drawn up which take account not only of the potential impacts, but also their likelihood and possible mitigation actions. Efforts should also be made throughout the evaluation to re-assess risks and determine what, if anything, has changed over time.

---

[26] Adapted from Section 4.5 of the Magenta Book 2011

Figure 4.5: Identifying Risks[27]
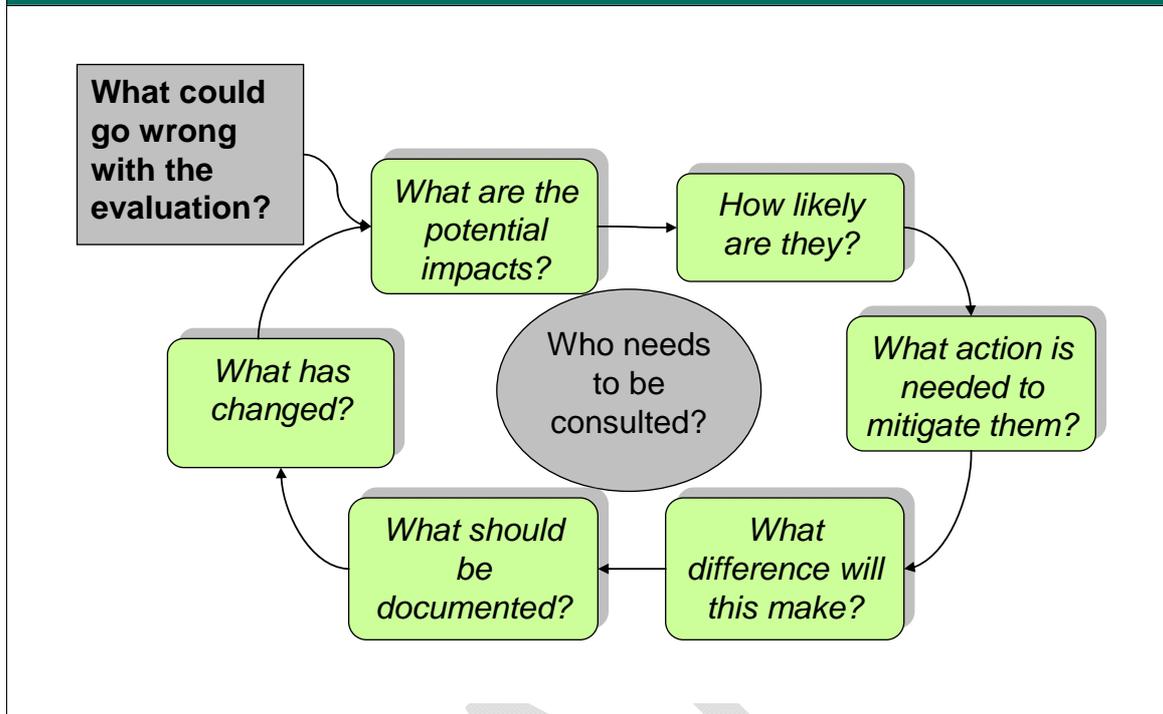
Adapted from NERA/ MVA, (2006), *The evaluation of major local authority transport schemes: a guide*

Some common issues which scheme evaluations can face, and which should therefore be considered, are:

- Failure to understand or address the *objectives* of the scheme

- Failure to agree the purpose or purposes of the evaluation (e.g. between evaluation manager and evaluation contractor, or with wider stakeholders)

- Starting data collection after the intervention has started having an effect (and therefore failing to gather robust 'baseline' data)

- Completing the evaluation or trying to collect outcome and impact data too early (before outcomes or impacts can be observed)

- Failure to understand limitations of the data sources for answering the questions posed for the evaluation

- Having an evaluation design which fails to provide robust data

- Failure to foresee future analytical or data requirements at the design stage

---

[27] Adapted from *The evaluation of major local authority transport schemes: a guide* by NERA Economic Consulting, MVA and David Simmonds Consultancy (2006)

- Failure to gather sufficient data or good quality data (e.g. a low response rate in a survey leading to results that cannot be validated statistically)

- Using a methodology which cannot attribute the outcomes to the intervention

- Producing evaluation findings which are not actionable or do not have clear implications

- Poor or disrupted planning, because of insufficient time and resources, too low a management priority, or inadequate response to unforeseen events

4.3 Using and disseminating the evaluation findings

Key questions to consider:

- What will the findings be used for, when will they be needed, and what decisions will they feed into?
- How will the findings be fed back into decision making and how can they have maximum impact?
- How will the findings be shared and disseminated? How will this need to vary by stakeholder groups?
- What are the timeframes for reporting the evaluation findings?

For evaluations to be useful and influential, communication of the findings must be appropriately tailored and targeted at their intended audience. The information gathered from steps 1 and 2 should be used to consider how and when they findings will be used and the evidence and communication needs of the different evaluation user groups.

4.3.1 Developing a dissemination plan

Dissemination is the practice of sharing knowledge about a scheme, its evaluation and resulting evidence with a range of stakeholders. This can take many forms and should be considered a key part of the scheme, to be done throughout the delivery and scheme lifecycle as appropriate.

Provisions should be made to share emerging knowledge, results and best practice with key partners. This could come in the form of meetings, workshops, websites, leaflets or briefings. Such dissemination and discussion can be as valuable for the scheme promoter as it is for other interested stakeholders in providing an effective opportunity for feedback which could be fed into scheme delivery improvements as in the ROAMEF cycle below[28] (Figure 4.6).

Dissemination should therefore not be confused with the requirement for written interim and final reports - these are just one important aspect of dissemination.

Dissemination plans should be set out in section 9 of the evaluation plan, and implemented over the course of the evaluation. Having a thorough dissemination plan is a key part of generating credible evaluation findings which can be acted upon and used to influence future decision making.

Building on this, evaluation planners must also consider exactly how the evaluation findings can be used, both internally and by other groups. This is important for a number of reasons; for instance, doing this shows that the evaluation has provided useful evidence and that the local authority is committed to taking relevant and timely action to address the findings.

---

[28] HM Treasury, (2011) *The Magenta Book, Guidance for evaluation* http://www.hm-treasury.gov.uk/magentabook

Reporting on lessons learnt is also extremely valuable for other stakeholders, such as DfT and other evaluation planners, who can use this insight in the development of future schemes.

Figure 4.6: Evaluation and the ROAMEF cycle
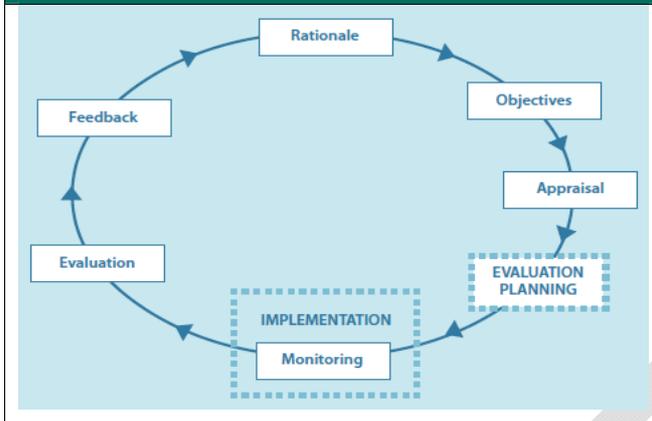
Understanding the relevant stakeholders and their requirements early on is an important part of planning the dissemination, see Section A defining the evaluation audience. Below is a list of some groups which are likely to be relevant:

- Evaluation sponsors
- Scheme management
- Delivery bodies
- Sponsors
- Policy makers
- Beneficiaries
- Government Departments
- Interest groups
- Academia
- Local transport users

4.3.2 Evaluation reports

It is a crucial role for the evaluation managers to design and deliver reporting styles and procedures that are tailored to the evaluation's specific purpose(s) and to its different audiences. Evaluation reports should aim to be accessible, relevant and useful documents which can be referred to by the scheme stakeholders and wider audiences to make meaningful conclusions from the scheme's implementation.

Evaluation reports are also an important opportunity to describe the methodological and data characteristics upon which the findings are based. Providing this level of detail in reports is key to ensuring the findings and conclusions are seen as credible and useful. Care should always be taken to report findings in an ethical way, for instance ensuring data handling, non-disclosure and other ethics requirements are met (see Annex C).

Rather than just provide the data outputs, evaluation reports have a key role in synthesising evidence and drawing overall conclusions. Providing accountability is a key principle of evaluation and therefore evaluation reports should make clear reference to the targets stated in the scheme's benefits realisation strategy as set out in Section, 2.1 and should provide evidence surrounding to what extent these have been met.

An evaluation report[29] should have coverage of the:

- Scheme and its context: including the main elements, intended objectives, rationale and surrounding circumstances;

- Evaluation scope: exactly what was evaluated, the objectives, which approach was taken and why (particularly explaining decisions to prioritise certain objectives);

- Evaluation design: Details about how the evaluation was designed and conducted;

- Methodology: Description of the methods used to measure the evaluation questions or indicators (including an indication of how robust the data is - clear articulation of any limitations to the data and caveats, sampling techniques/design, sample sizes, response rate, how the data was analysed, significance etc.) and analytical techniques;

- Evaluation results: Presentation and interpretation of the key findings with the appropriate indications of robustness, level of confidence in the findings and ability to generalise from the findings. Possible contextual factors of importance/ issues that arose. Provision of defensible evidence and appropriate management of different sources of information;

- Conclusions: Weighing up the evaluation evidence to produce a judgement on the value of the scheme and the lessons for future schemes;

- Recommendations (if appropriate): Could include ways of improving the delivery of the scheme, feedback for future scheme designs / appraisal, what will be done with the evaluation findings, identification of remaining gaps in the evidence base, and / or ways to improve evaluation designs in the future.

---

[29] The European Commission guide on programme evaluation provides a checklist of items for an evaluation report to include, which is a useful starting point for thinking about reporting requirements. European Commission, (1997), *Evaluating EU Expenditure Programmes: A Guide* http://ec.europa.eu/dgs/information_society/evaluation/data/pdf/lib_master/eur_budg_guide_ex_post_and_intermediate.pdf

4.3.3 Best practice reporting

While each evaluation report will vary depending on the particular scheme, evaluation approach and audience, there are a number of reporting principles which evaluations should always aim to follow (see Annex D). Section 10.33 of the Magenta Book also provides further details about best practice in reporting and disseminating evaluation findings.

Section 4.2 described the value in having a level of external review in evaluation, and the reporting stage is one important element where this may be used. Peer review (whether by academics or other experts) will add credibility to the evaluation findings and can sometimes provide an alternative, useful perspective for the evaluation planners.

4.3.4 Publication

A wide range of stakeholders can have an interest in the outcome of a scheme and the expectation set out in section 7 of the DfT Monitoring and Evaluation Framework is that "all Evaluation reports will be published by the scheme promoter on an appropriate website with the Department providing a link from its own website". It should be made clear in the dissemination plan, to both the evaluation team and externally, what reporting and dissemination approaches the evaluation will use, when, and who the audience will be.

There are, of course, other methods for dissemination aside from written research reports. Online portals and resources can be very useful - they are an opportunity to share results, progress and methodological issues with other professionals in the field and can be valuable for other Local Authorities and organisations developing transport schemes in the future. Road Safety Time Bank[30] and UK Morse[31] are two examples in the road safety field, for instance.

---

[30] See www.roadsafetyhub.co.uk
[31] See www.uk-morse.com

Summary for Section C

Section C set out what evaluation planners should consider for the final parts of their evaluation plans parts 7 – 9, see Figure 4.7. Evaluation planners should now have a clear plan of how they will resource the evaluation, what governance structure they will have in place and how to ensure quality and minimise risks associated with evaluation delivery. A clear evaluation delivery plan should be in place which links the evaluation project management process to the technical evaluation approach, methods and requirements set out in Section B. A dissemination plan for the evaluation in line with the requirements of schemes for publishing should also be part of the overall plan.

Figure 4.7: Considerations for Evaluation Planners Parts 7-9
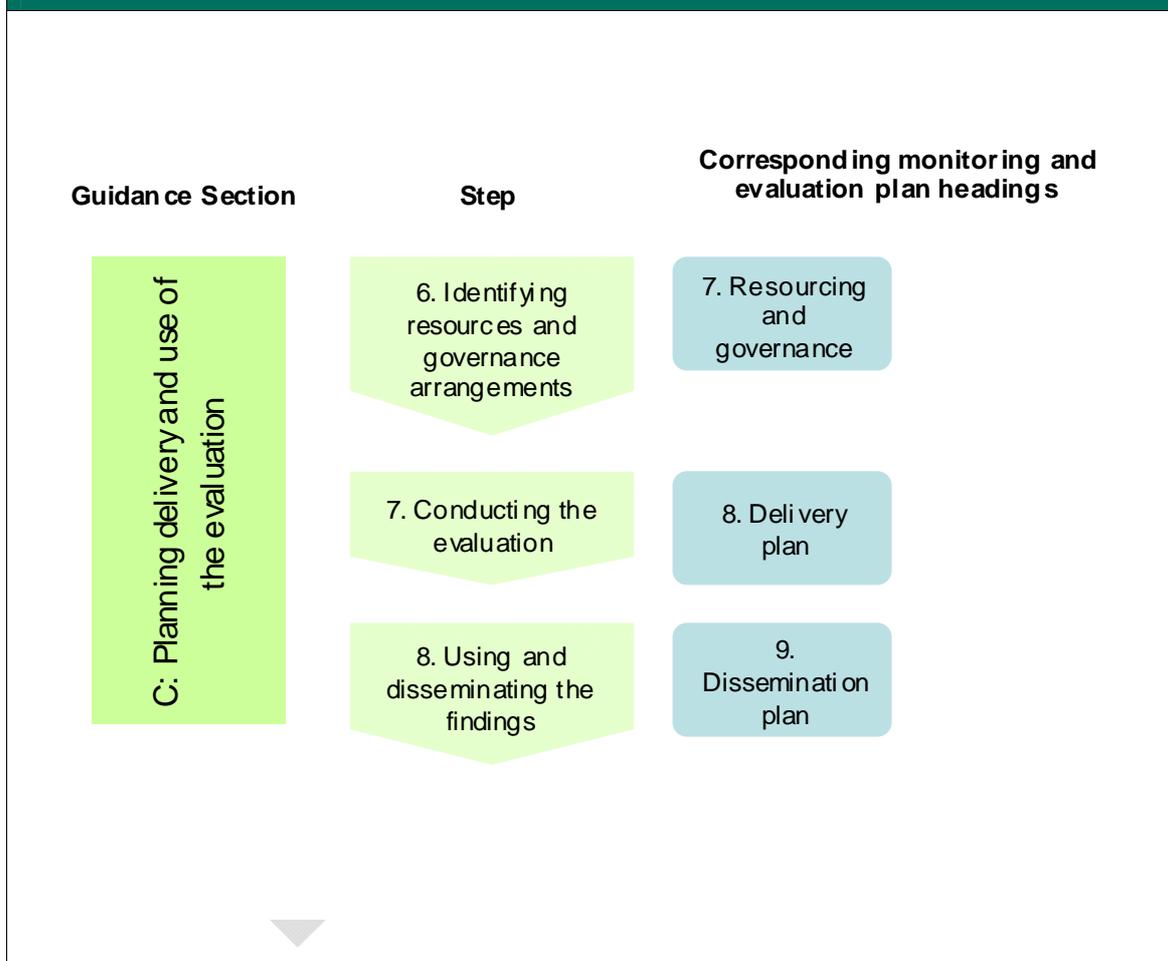
5. Overall Concluding Remarks

This guidance is produced to provide more details and best practice references to Local Authority Major Scheme's evaluation planners selected for fuller evaluation in the Monitoring and Evaluation Framework. It can also be used to support other Local Authority Major Schemes plan for their standard and enhanced monitoring requirements. The guidance provided in each section here is linked back to the expected content of Monitoring and Evaluation plans that need to be provided by evaluation planners and is based on the best practice set out in the Magenta Book and other sources.

The guidance:

- Brings together elements of existing best practice guidance for evaluation in one place for Local Authority Major Schemes;

- Promotes best practice without being prescriptive (requirements are set out in the Monitoring and Evaluation Framework which this document supports); and,

- Provides the readers with a summary of the broad spectrum of evaluation approaches and techniques available to enable informed decision making regarding production of an evaluation plan.

The guidance demonstrates the values of evaluation beyond planning to implementation and sharing of findings with external audiences to improve delivery of transport schemes and data used in transport modelling and other local decision making processes. Different methods are described and the value of using and combining a variety of approaches and methods for data collection to get a more rounded picture of how the scheme worked and what impact it has had on different audiences.

The guidance should enable the scheme promoter to design an evaluation which will help to:

- Improve this or future schemes, in terms of efficiency and effectiveness;

- Develop the evidence base for future funding decisions;

- Show robust evidence in response to criticism from external stakeholders;

- Demonstrate to others that your scheme is successful and offers good value for money;

- Demonstrate that the benefits observed have been generated by your scheme, rather than to other factors;

- Understand why the scheme worked or didn't work.

Annex

Annex A: Evaluation Questions

The Framework[32] sets out three types of questions which fuller evaluations should be designed to answer within these each scheme will have to define more specific questions. Tables A.1. A.2 and A.3 set out some guidance for schemes when developing evaluation questions[33] within each of these three types of questions.

Table A.1: Issues to consider when developing evaluation questions
How was the scheme delivered?
How will you understand why the scheme does or does not achieve anticipated outcomes?
Which aspects of the delivery process are innovative or untested?
Is it important to learn about scheme reach, usage, user experience, attitudes?
What contextual factors might affect delivery (e.g. economic climate, development in the geographic area, other transport schemes etc.)?
What delivery process information would be necessary, or useful, for assessing scheme impacts?
What were the experiences of service users, delivery partners, service providers, local businesses, and other stakeholders?
How complete are current data collection processes? Are the issues to be considered likely to need tailored data collection?
What lessons need to be learnt to improve future delivery?

Table A.2: Issues to consider when developing evaluation questions
What difference did the scheme make?
How will you know if the scheme is a success? Which of the outcomes will be

---

important to assess?
Do you need to quantify impacts, as well as describe them? How measurable are the various outcomes which might describe the schemes impacts?
How complex is the impact logic map? How important will it be to control for confounding factors?
What are the impacts for the target group? Do you need data on average or marginal impacts?
Were there different impacts for different groups or on different groups?[34]
How developed is the existing evidence base? Could it enable the scope of the evaluation to be restricted to those areas, impact or processes where knowledge is most uncertain?
What longer-term or wider knock-on effects should be considered? How will you know whether there were any unintended effects?

Table A.3: Issues to consider when developing evaluation questions
Did the benefits justify the costs?
How should costs and benefits of the scheme be assessed and can they be quantified?
How do the outcomes contribute to social wellbeing[35], and how do they generate costs?
How can the evaluation collect data which can be used to test assumptions in the appraisal model?
How will the evaluation assess displacement, optimism bias, additionality etc?
How can the benefits be attributed to the scheme investment?

---

[34] Summary Guidance on Social and Distributional Impacts of Transport Interventions, WebTAG Unit 2.13, http://www.dft.gov.uk/webtag/documents/project-manager/unit2.13.php
[35] Social wellbeing would include journey ambience, security, accessibility, safety etc which are part of full appraisal. Some of these can be monetised.

Annex B: Glossary

Term	Definition
Appraisal	The process of defining objectives, examining options and weighing up the costs benefits, risks and uncertainties of those options before a decision is made.
Attribution	The ascription of a causal link between observed (or expected to be observed) changes and a specific intervention. This takes into account other interventions, anticipated or unanticipated confounding factors or external influences.
Audit	A control function, which is primarily concerned with verifying the legality and regularity of the implementation of resources in a programme. Audit has traditionally covered areas such as the verification of financial records (financial audit) and is therefore less wide-ranging in scope than an evaluation.
Benefits management	The identification and agreement of a programme, project or scheme's expected outcomes and how they will be realised, measured and monitored.
Counterfactual	An estimate of what would have happened if the scheme had not taken place. The status quo or other baseline option used in the original appraisal should normally inform the counterfactual. However, viewing events from a post hoc position, evaluators may judge that the counterfactual would actually have been quite different from what was envisaged at the time of the appraisal, due to, for example, alternative states of the world and/or alternative management decisions. In such circumstances it may be helpful to consider other counterfactuals in addition to the original baseline option.
DfT	Department for Transport
Experimental methods	A theoretical way of deriving the counterfactual situation, and hence the net impact of an intervention. It involves comparing two groups which are identical in all respects except one: exposure to the intervention. Differences between the groups which have been exposed (the programme group) and the group which has not (the control group) are then attributable to the intervention. Quasi-experimental designs are a class of causal evaluation designs which take a more practical approach than is the case with true experimental designs. Control groups can still be used, but these have to be assigned through some non-random process. Alternatively, one can examine beneficiaries before and after exposure to the intervention.
Impact evaluation	The evaluation of, usually, the effectiveness of economic, social and environmental impacts of a programme or project. An impact

	evaluation may or may not also consider value for the money spent on the programme or project.
Intervention	Collective noun used to cover transport policies, programmes, schemes, projects and packages.
Intervention logic	The conceptual link from an intervention's inputs to the production of its outputs and, subsequently, to its impacts on society in terms of results and outcomes. The examination of the programme's intervention logic will be of central importance in most evaluations. The evaluator needs to ask how the programme achieves its specific objectives, and how do the specific objectives contribute to the attainment of the general objectives? The terms "theory of action", "programme logic" and "programme theory" are sometimes used to mean more or less the same thing.
Monitoring	The continuous process of examining the delivery of outputs and trends, usually with the intention of immediately correcting any deviation from operational objectives. Monitoring often generates data which can be used in evaluations.
Outcome	The likely or achieved short-term and medium-term effects of an intervention's outputs.
Outputs	The goods and services produced by an intervention (e.g. training courses for the long-term unemployed). See also intervention, intervention logic, operational objectives. These may also be referred to as activities.
Quasi-experimental method (non-equivalent group design)	The non-equivalent comparison group (NECG) design involves the evaluator selecting a group of units similar to those receiving the new policy or programme that is being tested. Such a group is called a comparison group (similar to a control group in a social experiment) and acts as a counterfactual.
Package	A combination of measures introduced to address common or shared objectives. These measures are not normally part of a co-ordinated set of activities at a national level (unlike a programme), but may still be used to work towards achieving one (national) goal. Package evaluations therefore tend to generate knowledge about the impact of the package as a whole.

Term	Definition
Process evaluation	An evaluation of the planning and management of a project as a whole, or of some specific components of it, such as capital procurement or concession negotiation and management.
Realist Evaluation	Realist evaluation assumes that most interventions have varying impacts under different sets of circumstances. Therefore, the context in which a programme is implemented is an important determinant for the outcomes. addresses the psychological and motivational responses that lead to behaviour change. Realist evaluation starts from the premise that causal outcomes follow from mechanisms acting in context, or put differently: Outcomes are explained by the action of particular mechanisms in particular contexts. Realistic evaluation is therefore concerned with "understanding causal mechanisms and the conditions under which they are activated to produce specific outcomes." (Tilley 2000, 5).
Scheme	A single endeavour undertaken to create a unique result or change in response to a specific problem. Scheme evaluations generally aim to justify the investment into a particular endeavour and to assess whether the anticipated impacts have been achieved and how.
Theory of Change	Theory of change involves a systematic and cumulative study of the links between activities, outcomes and context of an initiative. It involves the specification of an explicit theory of how and why a programme or project might cause or have caused an effect and the use of this theory to guide the evaluation. The focus of the theory of change approach is therefore on causal pathways.

Further sources of information and definitions:

http://www.worldbank.org/oed/ecd/docs/annex_e.pdf

Annex C: Government Social Research (GSR) Ethics Checklist

This summary checklist outlines the five key principles to adhere to when conducting and using primary research. GSR's full checklist can be found on the GSR website: http://resources.civilservice.gov.uk/wp-content/uploads/2011/09/gsr_ethics_checklist_tcm6-7326.pdf.

Sound application and conduct of methods and appropriate dissemination
Scope out existing research – e.g. is the research duplicating work already done? Will it add a burden to over-researched groups?
Confirm methodology - e.g. is the method appropriate for the groups/ issue involved?
External scrutiny - will this be required/ helpful?
Dissemination strategy - what level of dissemination is appropriate?

Participation based on valid informed consent
Consent to take part in primary research
Consent via gatekeepers or proxy - is this required? How can you maintain representativeness?
Children and young people (aged 15 and under)
Vulnerable adults
Access protocols - e.g. for schools, courts, police, prisons

Enabling participation
Reducing the barriers to participation
Ensuring hard-to-reach groups are included

Avoidance of personal harm
For research participants
For interviewers/ researchers

Non-disclosure of identity and personal information
Data protection
Research findings

Annex D: Evaluation reporting principles

Structure	
Ensure that…	**Avoid…**
-The structure of the report meets the needs of the sponsors and the principal stakeholders -The report includes an executive summary which can be circulated as a separate document -The report includes a copy of the terms of reference -Authorship and ownership of the report is clear and contact details are provided	-Moving between unrelated issues -Including too many complex tables, figures and technical detail in the main report. Detail is best kept in an appendix

Clarity	
Ensure that…	**Avoid…**
-The purpose and objectives of the evaluation are made clear -It is clear what was evaluated -There is a description of the full evaluation methodology as well as the rationale behind it -The reporting style is clear, concise and accessible -The findings are clearly communicated in an objective way, with all conclusions linked directly to detailed evidence -Relevant and appropriate conclusions and implications are drawn out -Data is disaggregated where appropriate (e.g. by area/ population group/ over time) and differences shown	-Lengthy or unclear executive summaries -Providing insufficient detail on evaluation approach or scheme implementation and objectives -Citing information without providing sources -Drawing conclusions which are not firmly based on evidence

Dissemination	
Ensure that…	**Avoid…**
-Findings are timely and relevant, e.g. feeding into strategy/ able to feed into upcoming reviews/ policy -Evaluation findings are tailored to the needs of relevant stakeholder groups -The dissemination strategy uses a range of appropriate routes (e.g. report, presentations, info sheets) -Findings are distributed to all possible interest groups -Findings, conclusions and recommendations are clear	-Drawing only high level conclusions which provide little input for future scheme delivery

Annex E: Bibliography and useful links

Aecom, (2009), *Evaluation of Better Use Interventions*
http://assets.dft.gov.uk/publications/cycling-city-and-towns-evaluation-approach/frameworkreport.pdf

European Commission, Evalsed, *Defining evaluation questions and criteria*
www.ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/designing_implementing/designing_planning/questions_en.htm

European Commission, (1997), *Evaluating EU Expenditure Programmes: A Guide*
http://ec.europa.eu/dgs/information_society/evaluation/data/pdf/lib_master/eur_budg_guide_ex_post_and_intermediate.pdf

HM Treasury (2011) *The Magenta Book*
http://www.hm-treasury.gov.uk/data_magentabook_index.htm

Hills, D. (2010) *Logic mapping: hints and tips*, DfT
http://www.dft.gov.uk/publications/logic-mapping-advice-guide/

Hills, D., and Junge, K. (2010), *Guidance for transport impact evaluations: Choosing an evaluation approach to achieve better attribution*, DfT
http://www.dft.gov.uk/publications/guidance-for-transport-impact-evaluations/

HM Treasury, (2003), *The Green Book*
www.hm-treasury.gov.uk/data_greenbook_index.htm

Lyons, G., Musselwhite, C., Dudley, G., Goodwin, P. and Wiltshire, P. (2009), *Business Attitudes to Transport: knowledge review of existing evidence*
http://webarchive.nationalarchives.gov.uk/20100413203302/www.dft.gov.uk/pgr/scienceresearch/social/businessattitudes/

NERA/ MVA, (2006), *The evaluation of major local authority transport schemes: a guide*
http://webarchive.nationalarchives.gov.uk/+/http://www.dft.gov.uk/pgr/regional/ltp/major/pdfevalmajlocautranpro.pdf

Useful links:

European Commission's evaluation guidelines:
http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/guide/designing_implementing/use_en.htm

Information Commissioner's Office - Freedom of Information Act 2000:
http://www.ico.gov.uk/for_organisations/freedom_of_information.aspx

DfT Transport Analysis Guidance (WebTAG)
http://www.dft.gov.uk/webtag/