

# Evaluation of Major Government Projects

*A short guide for Programme Delivery Teams*

*Produced by the Evaluation Task Force, building on five demonstrator studies*

In April 2025 the Evaluation Task Force (ETF) published a review of evaluation in the Government Major Projects Portfolio (GMPP). The review shows that 63% of major projects now have some form of evaluation plan and 34% have a high-quality evaluation in place. Although this represents significant progress from the past, there is still a need for considerable further improvement.

The findings from this review have informed an action plan to overcome these barriers and continue improving major project evaluation over time. One component of the action plan is to improve evaluation skills across GMPP project teams through developing and sharing examples of best practice. To fulfil this commitment the ETF has overseen the development of evaluation plans for five projects on the GMPP. These projects were chosen as they have characteristics identified in the review as challenging for evaluation, such as long delivery timeframes, digital programmes and military programmes.

This short guide provides an introduction to evaluation for major projects and insights from developing the example evaluation plans. Short summaries of each evaluation plan are included as an annex.

## Why Evaluation Matters

Evaluation is an essential part of delivering major government projects. It helps us understand what works, what doesn't, and whether the time, money, and effort we invest lead to the outcomes we expect. For major projects—whether in digital, transformation, military capability, or infrastructure and construction—robust evaluation supports accountability, learning, and better decision-making. Evaluation goes beyond performance or benefits monitoring; good evaluations are able to attribute changes in outcomes of interest to a project, and they make judgments about the effectiveness and value for money of a project and provide recommendations for improvements.

This guide is designed for programme delivery and assurance teams with little or no background in evaluation. It draws on lessons from five demonstrator projects supported by the Evaluation Task Force, and is fully aligned with the HM Treasury

Magenta Book,<sup>1</sup> the government's guidance on evaluation, and the Teal Book,<sup>2</sup> the guidance on project delivery in government. The five demonstrator evaluation plans are published alongside this guide.

## The Purpose of Evaluation

There are three key questions that every evaluation should aim to answer:

First, was the project delivered as intended? This is the focus of **process evaluation**, which helps us understand how the project unfolded in practice and whether it was implemented as planned.

Second, what difference did the project make? This is where **impact evaluation** comes in. It looks at outcomes and tries to determine whether the changes observed can be attributed to the project, rather than to other factors.

Third, was the project good value for money? **Value for money evaluation** assesses whether the benefits achieved justify the resources used, and whether there might have been more efficient ways to achieve similar results.

It is also important that evaluations are designed for the specific circumstances of major projects:

1. **Scale and complexity:** Major projects are large and complex, often being programmes with many components. The scale of major projects means consideration is needed on whether evaluation is conducted at the project, programme or portfolio level - or some combination of these. The complexity means that the evaluation needs careful design, often requiring mixed research methods, as some conventional evaluation approaches (such as randomised control trials) may not be feasible.
2. **Long timeframes:** Many major projects are delivered over several years, and final evaluation results are a long way away. Considering interim evaluation findings is therefore important for course correction and learning during delivery. Given these time frames, impact and value for money evaluation planning needs to be iterative. Earlier iterations should consider the evaluation design and collect baseline data, as well as evidencing whether expected impacts are on track. Later iterations can then assess whether impacts are realised and their value relative to the costs.
3. **Proportionality:** What evaluation types and methods are proportionate depends on the context of the project and varies considerably. Factors to consider include: how contentious the programme is, levels of risk and uncertainty, its cost, and the

---

<sup>1</sup> [www.gov.uk/government/publications/the-magenta-book](http://www.gov.uk/government/publications/the-magenta-book)

<sup>2</sup> [projectdelivery.gov.uk/teal-book/home/](http://projectdelivery.gov.uk/teal-book/home/)

potential for learning to inform decision making and improve the evidence base. The nature of major projects often means detailed process, impact and value for money evaluation is therefore required, but this is not always the case. Proportionality needs to be considered in the specific context of each project.

## **Embedding Evaluation in Delivery**

The most successful evaluations start early—ideally at the same time as the project itself is being designed. Planning ahead means you can build in the right data collection processes from the start, avoiding the need to retrofit evaluation methods later, which can be costly and less reliable. Early planning also encourages project teams to clarify their objectives and set out a clear “theory of change”—a description of how the project is expected to achieve its outcomes.<sup>3</sup>

Evaluations should be proportionate to the scale, complexity, and risk of the project. Not every evaluation needs to be large or complex. What matters most is that the evaluation is useful, realistic, and designed to provide insights that can improve current and future delivery. Projects should engage evaluation experts, either internal government analysts or external evaluation suppliers, to jointly determine what evaluation is useful, proportionate and feasible in the context of the project.

An evaluation plan is an essential component of any business case or other funding application a department wants to make for a project. For projects within scope of the HM Treasury Approvals Process, the guidance states that “Business cases are unlikely to receive approval without a robust and proportionate evaluation plan, properly aligned with the delivery context of the specific project or programme.”<sup>4</sup> Evaluation plans for major projects should be registered on the Government Evaluation Registry.<sup>5</sup>

Evaluation should be considered iteratively throughout the business case cycle. In early stages the aims and questions for the evaluation should be established, along with an initial consideration of the evaluation design and what resources are likely to be needed. In later stages, the evaluation design should be scoped out in further detail, with baseline data and initial process evaluation evidence starting to be collected. The evidence requirements for evaluation should be considered alongside broader evidence needs of the project, especially benefits management, to avoid any duplication of effort.

Once an outline evaluation plan is established, it can be used as a basis for more detailed evaluation planning. Table 1 summarises the key components of an evaluation plan.

---

<sup>3</sup>

<https://assets.publishing.service.gov.uk/media/67adca3663d7f4ccee9e11e/2-etf-evaluation-academy-20-developing-a-theory-of-change-slides.pdf>

<sup>4</sup> [www.gov.uk/government/publications/treasury-approvals-process-for-programmes-and-projects](http://www.gov.uk/government/publications/treasury-approvals-process-for-programmes-and-projects)

<sup>5</sup> [www.gov.uk/guidance/guidance-on-using-the-evaluation-registry](http://www.gov.uk/guidance/guidance-on-using-the-evaluation-registry)

*Table 1. Developing an evaluation plan*

Evaluation aims and objectives	Include information on what the evaluation is trying to achieve, what questions need to be answered and how they align to the project. This should include whether the aim is to understand how well the project is being implemented, what outcomes or impacts are being delivered, or whether the project represents good value for money.
Theory of Change	Set out a Theory of Change, describing what the programme/intervention aims to achieve and how.
Key outcomes and how to measure them	Describe what the key outcomes are for the evaluation, and what you need to measure to meet the evaluation aims.
Evaluation approach, design, methods and rationale	Decide on whether the evaluation will include an impact, process and/or value for money evaluation. Will the approach to the evaluation be experimental, quasi-experimental or theory-based? Will the methods be qualitative, quantitative or mixed?
Data requirements	Set out what data is needed, how it will be accessed or collected.
Key milestones	Note down key decision points for confirming the evaluation approach and methods. What are the expected timelines for delivering the evaluation, and how does this align with key decision points for the programme/intervention?
Resources	Agree on an evaluation budget and what other resources are needed including specialist evaluation expertise.
Governance	Include information on the governance structure for the evaluation, which should include senior oversight.
Use of findings	Clarify how evaluation results will be used to inform decision-making and how learnings will be shared with other stakeholders.

## Lessons from the Demonstrator Projects

The Evaluation Task Force supported five government major projects across different sectors to develop example evaluation plans. Several consistent lessons emerged that are relevant to delivery teams across government.

- **Developing the theory of change:** an investment of time upfront was needed to develop a mutual understanding of the aims of the project and of the evaluation between project teams and evaluators. This is especially important for major projects given their scale and complexity. Developing the theory of change together was a very valuable way to do this. Each project benefitted from setting out the logic behind their intervention—what they planned to do, how it was expected to work, and what outcomes they aimed to achieve. This helped teams identify where to focus their evaluation efforts and what data they needed to collect.
- **Evaluation questions:** Another key learning was the importance of focusing on a small number of priority evaluation questions. Rather than trying to evaluate everything, the most useful plans targeted specific areas of interest—such as user experience, implementation challenges, or whether expected benefits were actually realised.
- **Senior leadership:** support from projects' senior leadership was critical. Where senior responsible owners and programme boards championed evaluation from the outset, it was easier to secure the necessary time, resources and buy-in across delivery teams.
- **Collaboration:** evaluation planning worked best when policy teams, analysts, operational leads and finance teams worked together from the beginning. This helped ensure the evaluation design was practical, grounded in delivery reality, and aligned with broader departmental priorities.
- **Actionable:** all five projects reinforced the importance of using findings—not just producing reports. Evaluation is not just about accountability; it should support continuous learning. The most successful plans built in mechanisms to feed insights back into project delivery and share learning across teams and departments.

## Next Steps for Delivery Teams

If you are working on a major project, it is never too early to start thinking about evaluation. Speak to your departmental evaluation team as early as possible. Use the Teal Book and Magenta Book as your guide for technical standards and best practice, and don't hesitate to contact the Evaluation Task Force for advice or support.

Evaluation should be seen as an integral part of successful delivery—not a box-ticking exercise. It helps demonstrate whether public money is being well spent and provides vital evidence to improve outcomes for citizens.

By planning early, focusing on what matters, and working collaboratively, delivery teams can ensure their projects generate the insight needed to support better government.

## Annex: Demonstrators

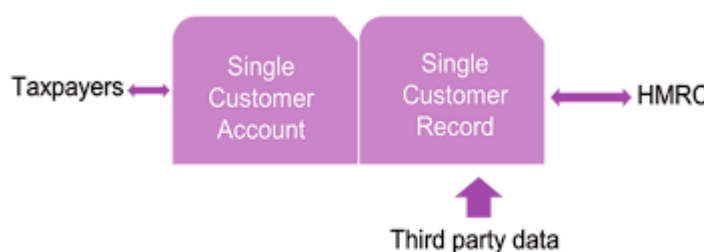
### HM Revenue & Customs: Single Customer Account



HM Revenue  
& Customs

#### Project Overview

The Single Customer Account (SCA) programme is a large-scale, digital project, developing and delivering a single destination for individuals and businesses to manage their tax affairs with HMRC. The existing HMRC online experience is highly fragmented, with access to basic information and critical services divided across multiple systems and accounts. The SCA will deliver an omni-channel user experience that reduces the burden on customers while easing demand for non-digital contact channels. The programme seeks to achieve efficiency and compliance gains and improve customer and staff experience.



Source: HMRC

#### Evaluation Design

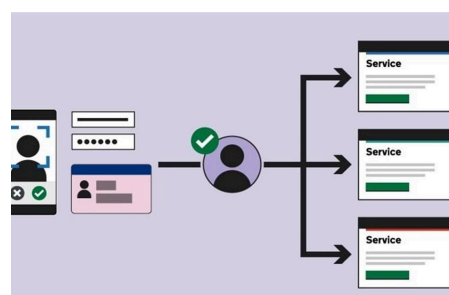
Many early evaluation activities have already taken place. The existing work, including previous Theories of Change, a high-level evaluation plan and specified evaluation questions, considerably aided the design of the SCA evaluation plan. A process evaluation will investigate whether the development of the outputs and delivery of outputs to the end-user was successful and whether there were any unintended challenges or barriers. An impact evaluation will assess what difference the programme made through a mix of experimental and quasi-experimental methods as well as qualitative assessment of user experiences. This might involve A/B testing to compare the performance of two versions of the service or specific features. Or it could involve using historical data to compare performance over time between taxpayers who benefit from the SCA and those whose accounts had not yet been moved to the new system. A value for money evaluation will build on existing cost-benefit analysis to identify additional cost and benefit types, including unintended or indirect consequences, with a particular focus on accounting for non-monetised benefits.

#### Next Steps

The SCA evaluation plan will allow HMRC to put in place a robust and comprehensive evaluation, to build on their benefits management work and assess the success of the programme across process, impact and value for money.

### Project Overview

One Login is a cross-government, digital transformation project, developing and delivering a single, ubiquitous way for UK citizens to log in, verify and reuse identity, and access government services online. With more than 340 services on GOV.UK, citizens can hold up to 191 accounts, accessed via 44 different sign-in methods and a multitude of ways to prove their identity. One Login aims to replace siloed and offline identity-proofing methods, with a single, secure and accessible login system. The programme seeks to achieve efficiency gains, improve user experience and accessibility, increase trust in digital services and improve digital maturity across HMG.



Source: GDS

### Evaluation Design

The evaluation plan proposes taking forward a process, impact and value for money evaluation. The aim of the process evaluation is to understand how users engage with One Login and how it can be expanded to other services in the most efficient way. For the impact evaluation, a two-stage approach is proposed. This would first involve conducting a longitudinal comparison of outcomes of the services that use One Login to those that do not, followed by a phase-in randomised controlled trial in stage two to produce more robust evidence for the remaining One Login outcomes. A value for money evaluation will focus in particular on estimating non-monetised benefits, including improved sustainability or trust in government services. Evidence from a value for money evaluation will influence decisions about the roll out of One Login.

### Next Steps

The evaluation plan will enable GDS to allow GDS to implement an evaluation within the practical constraints of current delivery, and build further robust, comprehensive evidence for the impact of One Login in the future. The evaluation plan will be useful to other digital projects, as it places a strong focus on assessing channel shift benefits, moving from non-digital forms of communication to digital; improvement in trust and confidence in online services; and reduced environmental impact.



# Ministry of Defence: Armed Forces Recruitment Programme



## Project Overview

The UK Armed Forces are undergoing a process of substantial transformation in how they recruit their people. The rationale underpinning this transformation is two-fold: firstly, Defence has recognised that existing recruitment approaches are failing to attract and recruit the required quality and quantity of recruits due to various challenges faced by existing single Service (sS) recruitment approaches. Additionally, the Armed Forces are navigating a challenging recruiting environment within an increasingly competitive labour market. This reinforces the need to maximise effectiveness in the Armed Forces recruitment model.

## Evaluation Design

The evaluation plan recommends undertaking a process, impact and value for money evaluation of the Armed Forces Recruitment Programme (AFRP) and tri-service recruitment approach. The process evaluation should collect programme documentation and data, and carry out consultations with a wide range of stakeholders to generate evidence on the effectiveness of the programme's delivery and learn about barriers to implementation. An impact evaluation should use a mixture of pre-post comparisons and Contribution Analysis to explore how the tri-service recruitment approach has contributed to the observed outcomes, and what other factors may have led to changes in these outcomes. The value for money evaluation will focus on four areas of the programme: costs, efficiency, effectiveness, and equity (e.g. whether benefits are distributed equitably). The evaluation will generate learnings about how effectively the programme has been implemented, whether it has met its aims, and whether it represents good value for money.

## Next Steps

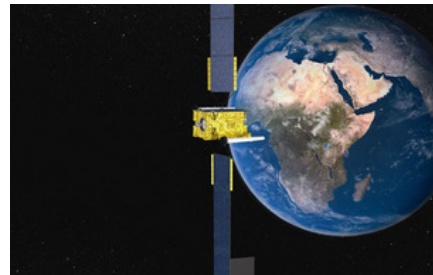
The evaluation plan recommends taking forward a multi-stage evaluation, including an early process evaluation to inform the design and demand planning of the procurement, an interim evaluation to identify key learning opportunities for ongoing delivery of the tri-service recruitment approach, and a final evaluation to feed into the design of a new contract or approach.

# Ministry of Defence: Skynet Integrated Enterprise Solution



## Project Overview

SKYNET is a family of military communications satellites that provides strategic communication services to the UK Armed Forces and its allies. The SKYNET Integrated Enterprise Solution (SKIES) programme is developing and assessing options for the future commercial/delivery model for the SKYNET programme, whilst exploring the implementation of a new operating model to enable more collaborative working between Industry and the Ministry of Defence. This will lead to more effective decision making in areas such as route to market, programme and system integration and the utilisation of constrained resources, while exploiting future technology and innovation. At the same time SKIES will need to align and be consistent with broader UK Government and defence policies and strategies.



## Evaluation Design

Source: MoD

The evaluation plan consists of a process and impact evaluation which will follow a realist approach, alongside continuous monitoring, to assess whether the SKIES programme is meeting its strategic objectives, is being delivered to plan, and is improving routes to market and yielding wider benefits. A realist approach—a theory-based approach to evaluation which considers the specific context in which a project operates—was recommended instead of an experimental or quasi-experimental design because no comparison case could be identified to act as a counterfactual. The evaluations will be based primarily on programme data and documentation as well as stakeholder interviews.

Building on the process and impact evaluation, the economic evaluation will assess whether the approach to procurement and managing risk and reward ensured value for money, and how an agile supply chain schedule could identify efficiencies or savings. It will also look at how the skills and workforce plan achieved value for money through efficient recruitment and retention processes.

## Next Steps

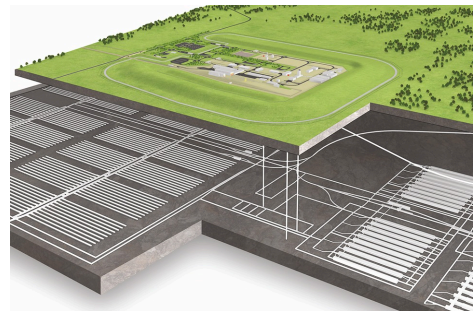
The evaluation plan recommends evaluation activities and data collection to take place in several phases, starting before appointing a supplier, so that the evaluation can inform the delivery and design of the programme at key decision points.

## Nuclear Waste Services: Geological Disposal Facility



### Project overview

The Geological Disposal Facility (GDF) programme refers to a suite of activities to be carried out to construct a GDF in the UK. Geological disposal involves isolating radioactive waste deep underground, inside a suitable rock volume to ensure that no harmful quantities of radioactivity ever reach the surface environment. A GDF will be a highly engineered structure consisting of multiple barriers that will provide protection over hundreds of thousands of years. The presence and operation of a GDF will allow for the development of new low-carbon nuclear power stations, and to provide socio-economic benefits for communities involved in the process and the UK more generally.



Source: GDF

### Evaluation design

The evaluation plan benefited from early involvement of key stakeholders. This highlighted the importance of co-design for evaluation in the context of large and complex programmes, given the need to build mutual understanding of the programme and what types of evaluation may be appropriate.

The plan recommends carrying out a process, impact and value for money evaluation. Each of these evaluations is expected to feed into one another to create a holistic picture of the programme's implementation and impact overall. The process evaluation will provide critical context for the impact evaluation, in terms of why impact may or may not have been observed. The results of the impact evaluation will feed directly into the value for money calculations. Each evaluation should be conducted in alignment with a theory-based contribution analysis approach, collecting multiple different forms of evidence and using an iterative approach to build a contribution story.

### Next steps

To proceed with an evaluation, senior leadership buy-in and support will naturally be a key facilitator. To achieve this support, it is important to find opportunities to meet and brief senior colleagues on the purpose of the evaluation, as well as how this investment will be beneficial to the organisation and to the GDF programme. Given the length of the programme, the evaluation plan has been designed to be iterative, with phases of evaluation activity aligned to phases of the wider programme.