Evaluation Task Force

One Login Evaluation Plan

Evaluation Task Force

August 2025

PUBLIC

Government Digital Service

Evaluation Task Force

Contents

Introduction

Background

In 2021, His Majesty's Treasury (HMT) allocated the One Login programme funding to enable GDS to build, rollout, run and scale up a live end-to-end authentication and identity verification service that meets the needs of the government's largest service providers as well as those of smaller, less complex services. As of May 2024, much of the build of One Login has been completed, with a detailed rollout plan to 100+ services set in place.

One Login is a single, ubiquitous way for UK citizens to log in, verify, and then 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 externally contracted evaluation partner, PUBLIC, has worked alongside the Government Digital Service to produce a detailed evaluation plan for the One Login programme, alongside one other digital government programme, as part of the Evaluation Task Force's (ETF) ICT Demonstrators project. The project aims to improve and embed evaluation practices across HMG and involves the creation of several demonstrator evaluation plans for five of the programmes within the Government's Major Projects Portfolio (GMPP).

At the time of writing this evaluation plan, One Login is three years into its development, with several services already part or fully onboarded to the One Login programme. As with many projects within the GMPP, One Login has a thorough business case with predefined critical success factors (CSFs) and key performance indicators (KPIs). Alongside the business case, One Login has a detailed rollout plan through March 2025.

This evaluation plan seeks to build on the work conducted by the GDS team to date, including the benefits outlined in the One Login business cases and utilising existing data sources where feasible. Finally, this evaluation plan details several outcomes and impacts that are not defined as KPIs in the business case, with a focus on evaluating the additional qualitative and wider social benefits of One Login.

Following the Machinery of Government change in July 2024, GDS (and by extension the GOV.UK One Login Programme) has joined the Department for Science, Innovation and Technology (along with the Central Digital and Data Office and the Incubator for AI) to create a new digital centre for government. The contents of this paper should not be used in a formal evaluation as there have been several redactions due to the sensitivity of an agile digital programme, which still remains on the GMPP. Although PUBLIC has engaged with the One Login team to develop this evaluation plan at a fixed point in time, any ongoing or future monitoring and evaluation work conducted by the One Login team may deviate from this evaluation plan. The One Login team will be guided by up-to-date programme delivery plans while still following best practice on evaluation and monitoring as outlined by the Magenta Book.

Structure of the report

The evaluation plan is structured in five chapters, outlining a variety of methods and approaches to planning an evaluation. These chapters are organised as follows.

The first section outlines the approach to developing a Theory of Change, articulating the key inputs and activities to date in the creation of the One Login outputs. Against these, the evaluation plan identifies outcomes and impacts of the programme with which to evaluate the success of One Login. The theory of change is accompanied by a database of evaluation metrics, including the individual units, calculations and expected changes related to each outcome. A visual logic model details the causal flow between short-, intermediate- and long-term outcomes and their related impacts.

The second chapter, through a process evaluation outline, sets out a plan to gather lessons learned from the delivery of One Login, from its inputs and activities in the development of the outputs, and the delivery of the outputs to the end-user. This section provides a breakdown for conducting the proposed data collection methods, including interviews, surveys, focus groups and programme documentation.

Subsequently, an impact evaluation outline is used to discuss the opportunities and constraints for evaluation, providing a rationale for the selected evaluation methods. The impact evaluation outline provides detailed explanations of several quantitative evaluation options, supported by targeted use of qualitative methods. The implementation guide presents the recommended plan for the impact evaluation of One Login, through two distinct evaluation phases.

The fourth chapter introduces a Value for Money evaluation outline which lightly builds on One Login's full business case. It explains how insights from the impact evaluation could inform the Value for Money (VfM) assessment. Further, since sustainability is a priority for the One Login team, the VfM evaluation outline contains some proposed methods to account for the environmental costs and benefits to estimate One Login's overall environmental impact.

The final chapter, which concerns implementation and timescales, provides an overview of the wider considerations to be made for the implementation of the full-scale evaluation. It discusses the skills and capabilities required for the evaluation team, as well as a summary of the range of activities to execute this evaluation plan.

Some sections of the evaluation plan have been removed for confidentiality reasons. Please get in touch with the Evaluation Task Force for further information.

Section 1 Theory of Change

This section presents the GDS One Login Theory of Change (TOC). The TOC illustrates the causal flow between the key inputs, their related activities, and the tangible product outputs,and proceeds to identify the relationship between the outputs, objective outcomes and the intended impacts of the service as a whole.

About the Theory of Change

This TOC acknowledges that One Login is in ongoing development, with multiple previous, current and future planned stages of delivery for various product functionalities. Although this is unlikely to alter the overall One Login impacts, a non-static product output has potential implications for future additional outcomes and/or currently inconceivable impacts. This TOC is therefore only an analysis of One Login in its current state at the time of publishing this report, March 2024.

Rationale

The rationale underpinning the One Login project is to develop a way for UK citizens to log in, verify and reuse identity when accessing online government services. One Login intends to replace offline identity-proofing methods with a single, secure and accessible login system. The project aims to achieve efficiency gains, improve user experience and accessibility, increase trust in digital services, and generally improve digital maturity across government.

Inputs

The key inputs to the One Login project can be broadly categorised into two types of input - financial and human capital. There is one financial input, which is the funding received by GDS. The other three inputs concern the skills, knowledge and expertise of staff involved in the development and delivery of the programme. Because the One Login service is intended to deliver cross-government digital transformation, the input of GDS staff and resources is supported by two other inputs - namely, the staff and resources employed by other government departments, as well as the staff and resources employed by partners in the One Login project.

Activities

Several activities must be undertaken as part of the One Login project. There are six overarching activities, which can be categorised as follows:

- strategy, policy and programme delivery
- verification of retirement and service migration
- engagement and onboarding of government services
- monitoring and anti-fraud development
- identity provision
- end-user support

Outputs

The above activities are intended to lead to four key outputs, namely:

- the One Login service
- migration and onboarding processes
- anti-fraud tooling
- identity verification tool

Outcomes

The outputs produced by the One Login project are expected to lead to a variety of outcomes. These outcomes are expected to materialise at different stages, and some of the early outcomes may lead to further outcomes as the project progresses.

The early intended outcomes are:

- large-scale migration to One Login
- large-scale adoption of One Login
- a decrease in user time spent verifying identity
- a decrease in user travel to verify identity
- a decrease in the energy spent to verify identity verification

The above outcomes are, in turn, expected to lead to a series of intermediate outcomes, including:

- a fall in the whole-life costs of verifying identity
- a centralisation of government identity verification systems
- an increase in the control and scalability of identity verification improvements
- enhanced security
- an improvement in the inclusivity and accessibility of government identity verification systems
- an improvement in user experience (UX)

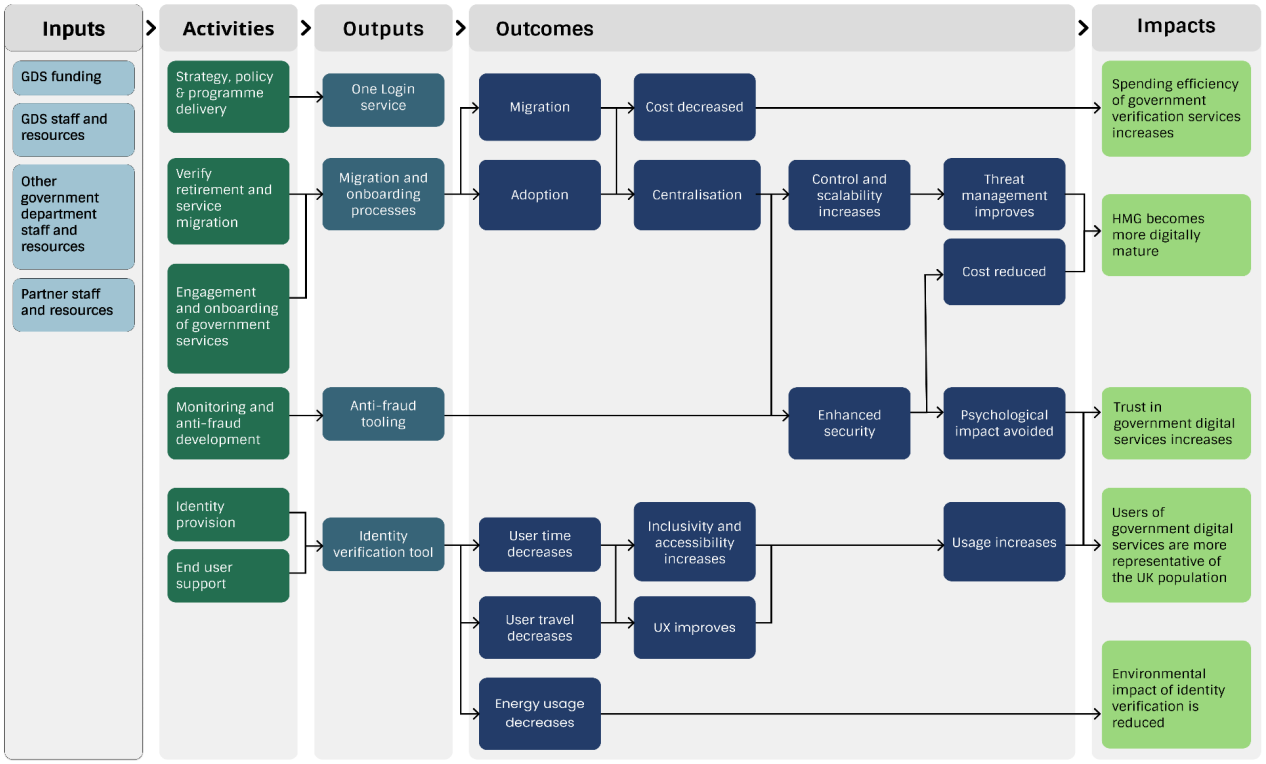
In later stages of the One Login project, the intermediate outcomes will materialise in the following long-term outcomes:

- an improvement in government agility to deal with threat vectors
- a reduction in the costs of identity fraud through government services
- prevention of the psychological impacts caused by identity fraud
- increased usage of government digital services

Impacts

Ultimately, the long-term impact of One Login is to improve the digital maturity of government and the efficiency of government verification services, through an increase in spending efficiency, as well as a reduction in the environmental impact of identity verification. One Login is also expected to have positive impacts on the users of government digital services by increasing users' trust in government digital services and ultimately making users of government digital services more representative of the UK population.

Figure 1: One Login Theory of Change Visual Logic Model



Note:

1. Some detail has been redacted as noted above.

Section 2 Process Evaluation Outline

This chapter lays out the considerations for a process evaluation of One Login in the two strands of *development* and *delivery*. This approach concerns the One Login programme in its current state, which is three years into development with continued development efforts planned, and at a stage where many government services and end-users are part and fully onboarded to its verification service, with a planned roadmap for continued onboarding of users.

Following Magenta Book[1] best practices, this process evaluation gathers lessons learned from the delivery of One Login, from its inputs and activities in building the outputs. With respect to the programme's TOC, this step of the evaluation will focus on 1) the development of the outputs and 2) the delivery of the outputs to the end-user. Broadly, the questions that are asked seek to identify:

1. What worked well and less well, and why?
2. What could be improved?
3. How has the context influenced delivery?

More specifically, this process evaluation outline seeks to evaluate One Login's transition between the programme's focus on development, build and roll out to steady state operations; running a high-performing service that meets the needs of services and their users; and ensuring the programme's resilience and effectiveness with regard to the onboarding of new services.

Detailed instructions for carrying out each of the methods described in this process evaluation outline, including in-depth interviews, surveys and questionnaires, focus groups and programme documentation review, can be found in *Annex A: Process Evaluation Methods*.

[1] HMT (2020), The Magenta Book (PDF, 2.4MB)

Section 3 Impact Evaluation Outline

This section discusses the opportunities for, and constraints to, evaluation that inform the selection of our proposed evaluation methods.

For those outcomes of the database that are assigned to be evaluated, the experimental and quasi-experimental methods outlined in this section are designed to guard against the problem of *selection*. Selection refers to how the set of services that choose to onboard to One Login are different from other services. These differences rule out simple comparisons of outcomes between these two groups as a valid evaluation technique. The proposed methods provide strategies for identifying groups of services that can be validly used to represent the outcomes of the onboarded services if they had not started using One Login. This comparison group is referred to as the 'control'.

In addition to applying an appropriate research design, an exercise known as 'matching' improves the identification of a control group. 'Matching' allows us to find statistical twins for services that have been onboarded to One Login among the services that have not. This exercise relies on those characteristics that are observable, and so are captured in available data. Nonetheless, any characteristic that may influence a service's performance that is missed by the available data can create misleading results.

Opportunities for evaluation

The impact evaluation plan outlined in this section takes advantage of several features of large-scale, government-led, digital projects. These features present unique opportunities to reduce the burden on the evaluator, and ultimately, minimise the cost of conducting the evaluation. Where possible, it is recommended that evaluators use the One Login performance data, embedded feedback collection, and existing datasets.

Performance data in digital projects refers to the data that may be passively and continuously collected via the digital systems employed, or naturally collected through the normal running of the service. This typically includes metrics related to user engagement statistics, website/service traffic, load times/technical performance, and error rates. Embedded feedback collection leverages the digital format to integrate user feedback mechanisms directly into the system interface. This approach enables the collection of immediate and context-specific insights from users while they interact with the digital environment. Methods such as pop-up surveys, feedback buttons, and interactive chatbots can be employed to gather qualitative and quantitative feedback. This real-time data reflects user experiences and satisfaction levels without the typical delay associated with traditional feedback methods. Additionally, embedded feedback collection negates the requirement for outreach and communication with users post-experience with the system, reducing the time taken and cost to collect the data.

Existing data sets offer another avenue for evaluators of One Login to extract valuable insights without the need for extensive net-new data collection efforts. As discussed later in this section, institutions including The Office of National Statistics (ONS), the Organisation for Economic Cooperation and Development (OECD) and the World Bank (WB) conduct several, large-scale surveys on digital maturity and trust in government. The published results of these surveys could provide control group data for several metrics in the One Login evaluation plan. It is recommended that, where available, the

future evaluation team utilise external data sources from reputable sources.

Selected evaluation methods

This subsection discusses the selected approaches for the impact evaluation of One Login and the implications, risks and mitigation strategies of each approach.

Based on the opportunities and constraints discussed above, the current rollout plan of One Login and the corresponding methodological limitations, it is proposed that an *event study* - a specialised form of difference-in-difference (DiD) analysis - be used to measure a prioritised set of outcomes defined by the GDS team. This will form the basis for the majority of the impact evaluation.

Where only aggregated data can be collected, conducting a standard DiD evaluation is recommended. If sufficient data for a DiD approach cannot be collected, as a third-best option, evaluators can use *qualitative* and *non-experimental* methods as a final opportunity to gather limited evidence.

Finally, it is useful to explore an alternative form of a *randomised controlled trial* (RCT) in a secondary phase of evaluation, once the One Login implementation roadmap of the current business case has been realised, to evaluate a wider number of outcomes.

Qualitative methods

Strictly experimental or quasi-experimental methods may not be feasible or appropriate for every outcome or impact outlined in the TOC. Metrics relating to these outcomes have been highlighted to utilise qualitative approaches to evaluation. Qualitative methods can be particularly useful to capture how One Login has impacted end users from their own perspective and is, therefore, our suggested evaluation approach for many user-centred metrics. However, a qualitative approach can be taken for the evaluation of a range of metrics in the TOC if an experimental or quasi-experimental method is chosen not to be implemented.

Qualitative methods do not employ treatment and control groups that are typical of experimental or quasi-experimental designs, instead gathering data directly from users, with a focus on experiences, perceptions, and behaviours in a natural setting. Qualitative methods, therefore, prioritise deep insights into context and meaning, rather than determining causal relationships. While qualitative methods can infer theoretical causality by identifying patterns and relationships that suggest possible causal links, they cannot be used to definitively attribute scientific causality to an intervention in the absence of controlled comparisons.

Qualitative methods, therefore, provide a lower level of evidence for the impact of One Login than the experimental and quasi-experimental methods discussed in this section. However, due to the relative ease of their implementation, qualitative methods can be effective in gaining rapid insights and should form the minimum level of evaluation employed by a future evaluation team. Additionally, the GDS already conducts user research alongside One Login development and delivery, therefore future evaluators should seek to align qualitative evaluation efforts with existing user research engagements.

To assess the impact of a qualitative nature, focus groups or in-depth interviews can be employed. For larger-scale qualitative evaluation, surveys and embedded feedback forms can be effective.

Non-experimental methods

This subsection outlines the method of *calibration* that may be used to investigate the relationship between One Login and the long-term impacts it aims to bring about.

To assess the effect of One Login on the impacts outlined in the TOC, it is particularly useful to exploit third-party data. The Office of National Statistics (ONS), the Organisation for Economic Cooperation and Development (OECD) and the World Bank (WB) are examples of institutions that may provide relevant data. Other multilateral organisations, universities, research centres and NGOs may also have similar offerings.

Exploiting such sources has many benefits. These data often offer comprehensive global coverage enabling the setting of benchmarks. Such surveys are often designed to be longitudinal or cover large periods, and so allow for investigations tracking changes over time. Finally, as these data are free to access, they are also highly cost-effective for any future evaluation.

The following subsection demonstrates how a method of calibration could be used to assess to what extent One Login contributes to increasing trust in digital public services.

Measuring trust in digital public services

The OECD has designed a 'Trust in Government' survey[2] [3] that asks respondents a range of questions that indicate their level of trust in the government. This survey is conducted biennially by the ONS on behalf of the OECD. Some survey questions explicitly refer to a respondent's view on how their personal data is used by public entities. Others provide proxies for a service's ease of use and level of fairness/accessibility.

This data is nationally representative of the UK's adult population. These surveys are also carried out in the 37 other member countries, enabling the framing of changes with global trends. Currently, two waves of the survey have been conducted, and the OECD expects to have this flagship 'Trust in Government' survey continuously administered over the long-term.

Measuring government maturity

A well-designed login service can significantly enhance a government's digital maturity. It allows users to access multiple applications with one set of login credentials, improving user experience by reducing password fatigue and support costs associated with password resets. It can increase security by minimising the number of attack vectors and enabling better management of user access through centralised controls. By simplifying the authentication process, One Login encourages higher usage of digital government services, leading to increased digital engagement and efficiency. As governments become increasingly reliant on technology, their digital maturity directly impacts their ability to respond effectively to the demands of citizens. As such, this dimension of public service is researched by public sector bodies like the World Bank and the United Nations, which have each produced useful indices for the measurement of digital maturity:

- World Bank's GovTech Maturity Index

[2] OECD (2023). Trust in Government

[3] ONS (2023). Trust in government, UK

- UN E-Government Index

The World Bank's GovTech Maturity Index evaluates governments based on their technology and data infrastructure, digital services, and institutional and regulatory frameworks. By providing a clear framework for comparison, the Index helps governments identify areas of strength and opportunities for improvement. For countries like the UK, it offers insights into how well they are leveraging technology to serve their citizens and maintain competitiveness on the global stage. This data is comprehensive and publicly available. However, this index is relatively new and is available for 2020 and 2022 only.

The United Nations E-Government Index is another key metric for assessing digital maturity in the public sector. It measures how digital technologies are being used by governments to deliver public services and engage with citizens. The index considers a variety of factors including online service delivery, telecommunication infrastructure, and human capacity. The most relevant component of this index to One Login is the Online Service Index (OSI). The OSI is a key subset of the UN E-Government Index that measures the quality, relevance, and availability of online services provided by a government. It specifically evaluates how public services are delivered electronically through national government portals and websites. This data is publicly available and has been gathered since 2001.

Section 4 Value for Money Evaluation Outline

This chapter outlines the proposed approaches for the economic evaluation design of One Login. It is preferable that the value for money (VfM) assessment be aligned with One Login's comprehensive business case. Therefore, the recommendations in this chapter do not provide additional guidance to factors already well-documented in the business case. Instead, they offer some additional considerations and valuation strategies, especially relating to evaluating programme sustainability. This chapter also provides general advice on evidential requirements for impacts to be treated as properly monetisable.

Time horizon

According to HMT Green Book, an economic appraisal should attempt to capture holistic costs and benefits.[4] Currently, the business case uses time horizons of three years and five years, which it notes represents a conservative account of the realisation of the expected benefits from One Login.

The proposed impact evaluation provides a detailed method to observe benefits as they are realised through an event study. Further, the programme's TOC presents outcomes as short-, medium- and long-term. Therefore, an evaluator will be able to determine when the full spate of specified benefits has been accrued. This means that some economic benefits will be realisable before the stated three-year and five-year time horizons, while others may persist far beyond this scale. The impact analysis can, therefore, inform the decision regarding the appropriate timing for conducting a VfM assessment.

Costs

Programme costs

Evaluators should consider the economic case in the business case for One Login, as this is where the breakdown of the costs associated with the service is shown.

Environmental costs

We also suggest a future evaluation to scrutinise the environmental cost of the administration of One Login. This inclusion would provide a more holistic picture of the cost of this centralised service. Additionally, as the environmental benefit of the channel shifts to online from offline verifications is taken into account, to understand the net environmental impact of the service the following costs need to be addressed. These costs may be broadly categorised under hardware, software, and people costs.

[4] HMT (2022). The Green Book

Table 1: Summary of environmental costs

Environmental Cost	Description
Hardware	Hardware refers to the hardware components used in the delivery of the programme, which would fall under Build and Rollout costs and Operational costs.
Energy efficiency	Rating the energy efficiency of the hardware components used in the build and operations of One Login can indicate their environmental costs. Assessing the carbon cost of the average laptop hour, and scaling by an estimate of laptop hours used can give this element of the cost.
Lifecycle management	Lifecycle management involves including the costs of upgrading components of hardware to elongate its life. Effective lifecycle management should reduce the yearly investment in hardware as the total cost is distributed over a longer period.
Software	Software refers to the carbon footprint of software employed in the delivery of the programme, which will mostly fall under Operational costs.
Cloud computing	This requires understanding the total energy consumption from cloud computing associated with the operations of One Login.
Data centre design	This requires assessing the total energy consumption at the data centre's sites being employed for the operations of One Login.
People	This refers to the carbon footprint of the individuals employed in the delivery of the programme.
Telecommuting	The primary source of an individual's environmental cost comes from travel costs. Accounting for this can involve determining the prevalence of hybrid/remote working among the workforce associated with One Login.
Environmental cost of single account login	Given that this assessment primarily focuses on the ongoing costs of the administration of One Login, it is important to isolate the energy costs associated with an individual account login. This accounting will allow an evaluator to disentangle the marginal environmental effect of using One Login from the total environmental effect of using One Login.

Benefits

It is recommended that the benefits of a VfM assessment are evaluated according to the structure outlined in the business case, considering its level of detail. By adhering to the framework established within the business case, evaluators can ensure a thorough examination of the potential advantages and align the VfM assessment closely with the established objectives and criteria.

Translating impacts to value for money

The benefits case currently outlines how monitoring data can be converted into monetisable benefits. This business case works on the assumption of making a direct point-in-time comparison between an initial baseline figure, and later monitoring data. It does not factor in evidential requirements for establishing whether a change can be attributed to the One Login service via an impact assessment, or statistical method. Ideally, conclusions about monetisable benefits should be drawn from the value for money factors in the impact evaluation findings rather than simple before-and-after benchmarking. This also will allow the One Login service to potentially prove monetizable benefits when monitoring data shows underperformance against benefits case objectives - in the case where the impact evaluation shows that the One Login service has had a positive impact, compared with a counterfactual group, but where external factors have led to underperformance against benefit targets.

Evaluation criteria: the 4Es model

The "4 Es" framework is a holistic approach used in VfM evaluations to ensure that projects, programmes, or policies are assessed comprehensively beyond just financial metrics. Each of the four Es - Economy, Efficiency, Effectiveness and Equity - addresses a different aspect of performance, allowing evaluators to capture a broad spectrum of impacts.

One Login is a complex programme where benefits are expected to be realised in several outcomes. To understand the cost-effectiveness across the full suite of potential benefits, one suggestion is ‘normalising’ units to determine the programme’s effectiveness in bringing about one standard deviation of change. This kind of analysis helps avoid the problem of non-monetisable benefits by including all quantifiable benefits in the assessment. However, not all expected benefits are necessarily quantifiable. Therefore the cost-effectiveness measure should be treated as a lower bound of the programme’s true cost-effectiveness. Cataloguing qualitative evidence of additional benefits not included in the cost-effectiveness analysis would help frame interpretation.

By engaging with key stakeholders, such as the delivery team of One Login, the relevant members of staff of the services that have been onboarded to the platform and the users themselves can help inform the degree of weight attached to each of the Es.

Economy

‘Economy’ concerns the cost-minimization aspect of resource utilisation. It assesses whether resources (such as financial, human, and material) have been obtained at the lowest possible cost

while meeting required quality standards. In a VfM evaluation, the Economy aspect checks if spending is done wisely and sustainably, ensuring that the programme does not overspend unnecessarily and that it obtains resources at optimal prices.

In relation to the TOC, this component generally looks to answer whether the cost of the ‘inputs’ can be justified by their quality.

Efficiency

Efficiency refers to how well resources are converted into outputs, or the amount of output per unit of input. It reflects the relationship between the resources employed and the results achieved. In VfM evaluations, the Efficiency principle is crucial because it determines whether the programme achieves maximum productivity with minimum wasted effort or expense, thus providing more services or better quality services without increasing the expenditure.

In relation to the TOC, this component generally looks to answer how well the inputs were converted to outputs.

Effectiveness

Effectiveness evaluates the extent to which the intended outcomes or objectives of a programme are achieved. It focuses on the impact and relevance of the outcomes in addressing the identified needs or problems. In a VfM context, effectiveness is about ensuring that the programme achieves its goals and delivers tangible benefits to stakeholders, which justifies the investment made into the programme.

In relation to the TOC, this component generally looks to answer how effectively the outputs brought about the outcomes. The impact evaluation most directly addresses this question.

Equity

Equity considers fairness and justice in the distribution of benefits and costs among various stakeholder groups, including vulnerable and marginalised populations. It aims at ensuring that all groups have access to the benefits provided by the programme and that no particular group bears an undue share of the costs. VfM evaluations use the Equity principle to assess whether the programme contributes to social inclusion and equality.

The 4 Es and VfM evaluations

Integrating the "4 Es" into VfM evaluations enables a more comprehensive assessment of a program’s performance and worth. Rather than focusing solely on cost reduction, this framework allows evaluators to consider how effectively and efficiently a programme operates while ensuring that it is fair and equitable. This holistic approach not only underscores the financial aspects but also highlights the social impacts, governance, and sustainability of the programmes. By addressing each of these dimensions, VfM evaluations can provide decision-makers with detailed insights into both the tangible and intangible benefits of a programme, thereby facilitating better-informed policy decisions and improvements in public sector management.

Section 5 Implementation and Timescales

Details for the execution of each evaluation type can be found at the end of the corresponding outline, however this section focuses on the wider considerations to be made for the implementation of the full-scale evaluation. This includes the skills and capabilities required and further recommendations for the procurement of the evaluation if conducted by an independent evaluation team.

Whether conducted internally or by an external organisation, several skills and capabilities are required within the evaluation team to execute the delivery plan. These skills have been categorised by the recommended project role (see Table 2).

Table 2: Required skills and capabilities by recommended project role

Recommended project role	Required skills and capabilities
Project and delivery management	• overseeing major research projects • managing multidisciplinary teams • risk and timeline management • external stakeholder engagement and problem-solving
Statistical and data analysis	• using different data types (e.g. geospatial data) and impact evaluation approaches • conducting statistical matching • using statistics to derive causality
Qualitative and social research methods	• designing surveys and using data collection tools • conducting interviews, focus groups and workshops for user research • coding qualitative responses
Economic analysis	• conducting CBA and SCEA approaches • conducting NPV assessments • cost accounting • evaluating the potential risks associated with

	economic decisions
Technical advisory	● sufficient technical expertise for appraising technical development and architecture decisions ● assessing associated value for money

See *Annex C: Minimum and maximum evaluation approach* for a high-level summary of the range of activities to execute this evaluation plan, based on a ‘minimum-vs-maximum’ approach.

Annex A: Process Evaluation Methods

This section summarises the individual methods of data collection described in the process evaluation outline of this evaluation plan. Each method is supplemented by a stepwise guide for executing the process evaluation.

We recommend that the future process evaluators of the One Login programme employ a mixed-methods approach, incorporating each of the methods described in this section where identified as most appropriate in the process evaluation outline. Additionally, as previously mentioned, we understand that the GDS already conducts user research alongside One Login development and delivery, and we therefore encourage future evaluators to align process evaluation efforts with existing user research engagements.

Further best practices for the collection of data via all methods mentioned in the process evaluation outline can be found in *Chapter 4: Data collection*, data access and data linking of The Magenta Book.[5]

In-depth interviews

In-depth interviews gather detailed data through one-on-one conversations. They can provide insights into interviewees' experiences, attitudes and perceptions. Interviews hold a comparative advantage in their ability to explore topics deeply, allowing for a nuanced understanding of complex phenomena, and surpassing the surface-level data obtained through other methods like surveys or focus groups. Interviews are the preferred method of collecting feedback, as ensured anonymity allows for participants to speak freely.

Surveys and questionnaires

Surveys and questionnaires are used when collecting data from a large number of respondents, where the goal is typically to obtain information that is representative of the general population. These generate quantitative data through standardised response options and have the benefit of providing respondents with anonymity.

When conducting surveys, a sampling method must be chosen so that results can be systematically framed. For responses that are being collected from the delivery or development teams themselves, we suggest stratified random sampling where the strata are different teams or subunits within the broader delivery and development functions.

For responses that are being gathered from One Login’s end-users, we suggest random sampling to ensure that the responses are representative of the total population. It is important, however, to note that in some cases the end-users are other government services and in some they are civilians. For the former, a stratified sampling approach may be more appropriate where the strata are government departments.

[5] HMT (2020). The Magenta Book (PDF, 2.4MB)

Focus groups

Focus groups collect data from groups of participants who engage in a structured conversation around a central topic, guided by a moderator. This method allows for the recovery of consensus or identification of conflict lines around a topic and is best suited for answering solutions-oriented questions.

Programme documentation review

Programme documentation review as a research method involves systematically recording and analysing information about an output. It includes various documents such as programme manuals, reports, policies, and other written materials. The process typically involves collecting, organising, and analysing these documents to understand the programme's goals, implementation strategies, and output performance data.

Annex B: Impact Evaluation Methods

In this section, we summarise the steps to take for each of the methods described in the impact evaluation outline of this evaluation plan. For the quasi-experimental methods, we discuss considerations for constructing a control group, utilising the baseline questionnaire and KPIs. For the experimental methods, we discuss considerations for sample size, the duration of phases and unit of analysis, utilising the baseline questionnaire and KPIs, and analysing the treatment effect. For all methods discussed, we provide a stepwise guide to executing the impact evaluation.

Conducting an Event Study

Constructing a control group

We recommend drawing the control group from the population of eligible services, yet to request integration with One Login, that have already been identified in the business case. To determine which eligible services are similar to onboarded services, matching can be used. Matching would look for closeness between two services on characteristics for which data is available. The degree of closeness is assessed using standard statistical techniques like coarsened exact matching or propensity score matching.

Key Performance Indicators

The One Login team has defined KPIs that it aims to monitor in conjunction with its services. This practice will enable the causal impact to be estimated once a control group provides corresponding data. Frequent collection of KPIs will allow for event studies to be conducted, as we have outlined in the subsection above.

Conducting a phase-in RCT evaluation

Sample size

Sample size affects statistical power. Statistical power is the likelihood of an impact being detected, given that it occurred, by a statistical test. For any given sample size, there is a minimum detectable effect it can recover. This minimum detectable effect (MDE) also varies by outcome as it depends on characteristics specific to data, such as its mean and standard deviation.

This allows us to determine the MODE when we randomise the remaining services to establish the upper limit of what's feasible for this test. If the MDE turns out to be impractically large—for instance, a 50% change—then this would suggest that a phase-in design is not suitable for evaluating this metric.

This exercise can be replicated for other outcomes where sufficient data exists to perform similar calculations. Once we identify which outcomes have a feasible MDE under full participation, we can further refine our approach by experimenting with reduced sample sizes to see how this affects the MDE.

This iterative method helps in pinpointing the smallest sample size capable of detecting

a reasonably sized effect, optimising our evaluation strategy.

Duration of phases

Determining the duration of phases is a key factor in determining which outcomes can be reasonably evaluated. If, for example, the impact of moving to One Login on a certain outcome takes longer to materialise than the designated duration of the phases then that outcome should not be evaluated using this method.

In the case where a phase-in design is adopted following an event study, results from the latter can be used to either (i) calibrate the phases such that impact may be observed on a maximum number of outcomes or (ii) determine which outcomes can be evaluated for a given duration that is determined to be feasible under institutional constraints regarding the onboarding of services.

Unit of analysis

In this setting, it is not feasible to randomly assign users to benefit from One Login as this would require each onboarded service to split their beneficiaries into treatment and control groups. In effect, this would require each onboarded service to conduct an RCT. The simplest unit of analysis is therefore individual services, which is the lowest level at which randomisation is feasible.

Balance of characteristics

Following the assignment of eligible services to either the treatment or control group, a balance of characteristics check must be conducted. This verifies the efficacy of randomisation. As services are the unit of analysis, the primary concern of this check is their characteristics.

Treatment effect

In the setting of One Login, an RCT would draw its participants from services that request to be onboarded. Therefore the interpretation of any effect recovered through this design would be a treatment-on-treated (ToT) effect. A ToT is the expected effect on the kind of services that would request onboarding to One Login. Our capability to generalise results to services that did not request integration with One Login is therefore limited.

Annex C: Minimum and maximum evaluation approach

Stage 1: Interim process and impact evaluation

Table C.1: Interim Process Evaluation Minimum and Maximum Approach

Process evaluation method	Minimum	Maximum
Interviews	Interviews primarily with the GDS internal team, a small number of service teams (3-5)	Interviews with the GDS internal team and a larger number of service teams (c.20)
Surveys	No surveys used	A survey is issued to all participating service teams
Feedback forms	No feedback forms	Simple feedback embedded into GOV.UK / One Login pages
Data collection	Data collected from a small number of control services to establish event study counterfactual (5+)	Data collected from a larger number of control services to establish event study counterfactual (10+)
Qualitative research with service teams	A small number of interviews with service teams to measure the impact of wider outcomes (3-5)	A larger number of interviews with service teams to measure the impact of wider outcomes (c.20)
Qualitative research with end users	A small number of interviews with end users, normally embedded into existing GDS user research processes	A larger number of interviews with end users, with a dedicated research strand

Survey with end users	No survey of end users	Commissioning of a survey sample of end users

Stage 2 - Full process, impact, and VfM evaluation

Table C.2: Full Process Evaluation Minimum and Maximum Approach

Process evaluation method	Minimum	Maximum
Interviews	Interviews primarily with the GDS internal team, a small number of service teams (5-10)	Interviews with the GDS internal team and a larger survey to all participating service teams
Surveys	No surveys used	A survey is issued to all participating service teams
Feedback forms	No feedback forms	Simple feedback embedded into GOV.UK / One Login pages
Data collection	Fewer RCT participants (15 treatment + 15 control)	Fewer RCT participants (30 treatment + 30 control)
Qualitative research with service teams	A small number of interviews with service teams to measure the impact of wider outcomes (3-5)	A larger number of interviews with service teams to measure the impact of wider outcomes (c.50)
Qualitative research with end users	A small number of interviews with end users, normally embedded into existing GDS	A larger number of interviews with end users, with a dedicated

Process evaluation method	Minimum	Maximum
	user research processes	research strand
Survey with end users	No survey of end users	Commissioning of a survey with a representative sample of end users