Evaluation Task Force

Geological Disposal Facility Evaluation Plan

Evaluation Task Force

August 2025

verian Nuclear Waste Services Evaluation Task Force

Executive Summary

Overview of the GDF programme

A Geological Disposal Facility (GDF) is a highly engineered underground facility designed for the permanent disposal of radioactive waste. The waste is placed in sealed chambers while it decays over time, a process that can last from hundreds to thousands of years. A GDF is expected to be able to provide a permanent and sustainable disposal solution for the Most Hazardous Radioactive Waste (MHRW).

The GDF programme aims to create the UK's first GDF to permanently deal with the current large inventory of MHRW in the UK, avoid the costs associated with overground storage, and support the building of new low-carbon nuclear power stations in line with the new nuclear programme.

Critically, the GDF programme is consent-based, meaning that the host community must actively consent to the construction and presence of the GDF in their area. Potential host communities will receive investment through a Community Investment Fund (CIF), which will be used to realise a number of locally determined benefits. Once a preferred site has been chosen and the community has demonstrated its commitment through a Test of Public Support, the host community will receive further investment opportunities, referred to as Significant Additional Investment (SAI).

Introduction to the GDF evaluation

Verian and Oxford Global Projects were commissioned by the Evaluation Task Force to provide an evaluation plan for the GDF programme. Looking forwards, this evaluation plan should:

- be used to inform the design of an invitation to tender for commissioning an evaluation of the GDF programme
- act as an evidence base for Nuclear Waste Services (NWS) in regard to securing the resources (budget and staffing) to support an evaluation
- be shared with other large infrastructure projects to compare, scrutinise and improve approaches to evaluation scoping for complex projects

The way the GDF programme is designed and will operate poses some challenges for evaluation design. Firstly, there are currently 'known unknowns' in regard to the details of the GDF programme, which will be resolved as the local projects develop. This makes it challenging to scope a detailed plan for precisely how the GDF programme might be evaluated at this early stage of the programme.

In addition, while we have suggested phasing the evaluation to broadly align with the GDF programme implementation activities, many of these activities will occur over a similar timeframe or over overlapping timeframes. This means it may be difficult to find a suitable counterfactual and to disentangle the effects of specific phases of the GDF programme.

Finally, the timeframes involved are very long and there is a huge potential for change in government policy, programme delivery and evaluation techniques, which means that the evaluation approach must also be flexible and be able to change.

Overview of approach

The recommended approach for the GDF programme comprises three evaluation strands:

- implementation and process evaluation (IPE)
- impact evaluation
- value for money (VfM) evaluation

IPE evaluations are designed to understand how the GDF programme has been delivered, whether delivery matches the policy intent, any facilitators and challenges that have been experienced throughout delivery, and any examples of best practice and learning.

Impact evaluations assess observed changes in outcomes throughout the GDF programme and determine whether and how the programme caused these. Impact evaluation will contribute to an understanding of what has worked and what has not worked, and any unintended consequences of the GDF.

Finally, value for money evaluation investigates how well the programme has performed against the National Audit Office's (NAO's) key dimensions of VfM (the 4Es framework), whether the value created justifies the use of resources, and how VfM is likely to evolve over the course of the programme.

Each of these three evaluation strands will feed into one another to create a holistic picture of the GDF 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 VfM calculations.

Given the length of time involved in the GDF programme, key evaluation activities and questions have been split into four evaluation phases that broadly align with programme implementation activities. The evaluation will still explore the overall implementation and impact of the GDF: the phasing is a mechanism that splits the evaluation in order to focus on what can realistically be delivered and achieved within a given timeframe.

The phases are as follows:

1. Community engagement and site evaluation
2. Site characterisation and Test of Public Support
3. GDF technical development, design and construction
4. GDF operation

The first phase includes the establishment of working groups, funding for the initial setting up of community projects, and selection of potential host communities for the GDF site to take forward to intrusive investigations. The second phase involves detailed site and geology investigations, through to the point at which the NWS is confident that a GDF could be designed and constructed safely. Phase 2 would also cover the delivery and emerging impacts of the community projects, and the process of holding a Test of Public Support. The third phase includes the continued delivery and impact of community projects funded through the SAI mechanism after the Test of Public Support, and the decision on the preferred site. It also includes the process and permissions involved in designing and constructing the GDF, up until the first waste emplacement. The final phase will focus

on the extent to which the GDF programme achieves its longer-term outcomes and impacts, particularly in relation to the disposal of MHRW and the ongoing impacts of community projects. This phase will consider – and potentially inform – the waste transfer programme known as the Stores Export Schedule.

Table 1: Overview of evaluation activities mapped against implementation activities

Phase	Implementation activity	IPE and impact evaluation activities
Phase 1	Community engagement and site evaluation	● scoping ● data collection
Phase 2	Site characterisation and the Test of Public Support	● scoping/re-scoping ● data collection ● emerging results/outcomes
Phase 3	GDF Technical Development, Design and Construction	● scoping/re-scoping ● data collection ● emerging results/outcomes
Phase 4	Operation and closure	● scoping/re-scoping ● data collection ● emerging results/outcomes

There are a number of options for the periodicity of assessment for Value for Money. Verian provisionally recommends conducting this at the following critical points as a minimum:

- completion of business case and approvals
- completion of the CIF investment
- completion of siting process
- completion of initial construction
- completion of commissioning and first operations/emplacement
- end of emplacement and significant additional investment
- closure of the GDF

Implementation and process evaluation

Verian recommends that each phase have its own IPE. This would start with a scoping stage, to refine a phase-specific approach to data collection and analysis to ensure best fit and efficiency, followed by data collection and analysis. A summary of IPE activities is provided below in Table 2 and 3.

Table 2: Overview of scoping activities

Scoping activity	Aim
Document review	To gain a holistic and coherent understanding of upcoming implementation
Stakeholder mapping	To identify participants for scoping and data collection activities

Co-production workshops	To revise the phase-specific Theory of Change
Interviews	To identify evidence gaps, feedback on existing data and data collection methods
Design and analysis	To produce a detailed evaluation plan for the relevant phase, including timelines, methods and key aims for data collection, Evaluation questions, and Theories of Change
Validation	To sense-check the evaluation plan and design

Table 2: Overview of data collection and analysis activities

Data collection and analysis	Aim
Design research tools	To tailor the evaluation to the programme
Fieldwork period and analysis	To understand how the programme has been delivered and what stage the programme is at
Fieldwork period and analysis	To understand progress updates, explore barriers, challenges and key areas of learning
Fieldwork period and analysis	Additional fieldwork period to respond to, clarify and triangulate emerging insights from previous fieldwork
Results and learning workshops	Share learnings and validate findings
Revision and full results	Publish findings and implement learnings

Impact evaluation approach

Verian recommends that the IPE and impact evaluation are both situated within a programme-level contribution analysis framework. The recommended approach will require conducting impact evaluation activities throughout the GDF programme to deliver an overarching narrative about the impacts of the programme and, in the long term, an assessment of whether the expected outcomes have materialised along the programme's Theory of Change during its different phases. The recommended methods include:

- contribution analysis
- quasi-experimental impact evaluations
- simulation analysis

The contribution analysis approach would capture evidence from interviews and focus groups with local residents, experts, businesses, voluntary groups and other local stakeholders about their perceptions of the effects of the programme. It would also capture evidence from sentiment and

thematic analysis of media and social media or forums in which the GDF is discussed and would report key changes in outcome indicators and other quantitative evidence.

Quasi-experimental analysis would compare the outcomes of the host community relative to other comparable communities. Alternatively, depending on the types of projects funded through CIF/SAI, it may be possible to compare the outcomes of the programme's beneficiaries to those of non-beneficiaries within the same host community or potential host community. The approach would adopt statistical techniques to account for any existing differences between the host community, potential host community, and other communities (or beneficiaries and non-beneficiaries). It can be applied to outcome data collected from primary surveys – for example, on community trust in the GDF or individuals' wellbeing, or secondary data from public or government sources.

Finally, simulation analysis will be required to update details and assumptions within the GDF programme's business case to evaluate the programme's impacts on outcomes such as carbon savings and employment across the UK (accounting for job multiplier effects). These simulation studies would provide an approximation of the 'realised estimated effects' on such outcomes, which would then be compared with the original economic business case.

It is recommended that the evaluation consider both the GDF programme's local and national-level effects, due to differences in the methods employed in evaluating these different effects. Also, in the long term, while an analysis of the national effects will consider the aggregate impact of the GDF programme, an evaluation of the local effects of the programme will provide a comparison of whether the effects of the GDF programme are more concentrated in the host community or potential host community.

As with the IPE, the impact evaluation will need to be regularly (re-)scoped to confirm the outcomes to be evaluated and which method is to be used. For example, once more detailed plans are made for CIF expenditure the design will need to be refined to ensure it is able to specifically evaluate the intended effects of these activities on the desired local outcomes.

Value for money approach

The VfM assessment should apply the value-for-investment (namely, the King/Oxford Policy Management (OPM)) approach when designing and delivering the VfM assessment. This approach is recommended because it allows for the incorporation of non-monetised benefits which the Greenbook and NAO approaches do not fully consider. The value-for-investment approach is structured around the eight steps set out below, and provides a structured, comprehensive and holistic approach to assessing the NAO's key dimensions that are to be considered when making a VfM judgement: economy, efficiency, effectiveness, and equity (4E's).

Figure 1: Overview of the value-for-investment approach

Source: Julian King & Associates, Value for Investment. Available at: https://www.julianking.co.nz/vfi/ .

At a minimum, the VfM of the GDF programme should be assessed at the end of its life, once all costs and benefits have been realised. Given the long timeframes involved, we would also advise conducting an interim VfM assessment at the start of GDF closure and carrying out VfM assessments of specific activities where there is a discrete set of benefits and associated costs. These could be conducted in the earlier stages (such as the community engagement and site evaluation phase), which would provide an evolving picture of VfM that could be used to periodically assess performance and test the assumptions of the evolving business case estimates. Activity-based assessments could be agreed on a rolling basis at the beginning of each phase of the evaluation, or they could be responsive to significant programme achievements.

Next steps

To put in place an evaluation of the GDF programme, it is expected that the NWS will need to:

- gain senior stakeholder buy-in for the investment of time and resources (staff and budget), with support from the Evaluation Task Force
- confirm the scope of the desired evaluation, in relation to the intended forms of evaluation and phasing/timeframes to be covered
- identify the resource requirement through an assessment of internal team capacity and skills, benchmarking the budget required for an externally commissioned evaluation with other tenders and engaging in market testing
- procure an external evaluation partner/consortium, via a competitive tendering exercise, incorporating a supplier engagement event

Contents

Section 1 Introduction to the GDF programme

This section provides a short introduction to the Geological Disposal Facility (GDF) programme and how it will be implemented.

What is a GDF?

A GDF is a structure designed for the final disposal of radioactive waste. The waste is placed in sealed chambers, where it decays over time, a process that can last from hundreds to thousands of years depending on the radioactivity level of the waste.[1] Since the GDF is constructed at a depth of between 200 and 1,000 metres the waste housed in the facility does not pose a risk to the population. In addition, the depth and construction design of the GDF prevents any potential man-made or natural environmental impacts on the facility.[2] GDFs have been constructed in countries such as France, Canada and Sweden.

Why is the UK developing a GDF?

The UK has accumulated a significant volume of radioactive waste over time, through various activities that rely on nuclear power for energy.[3] Nuclear power is used to generate approximately one-fifth of the UK's electricity, and is used in the medical, industry, defence, research and other sectors of the country.[4] The use of nuclear energy will be vital to meet current and future demands for energy in the UK, and this will increase the volume of waste and the need for its storage.[5]

Radioactive waste is currently stored in above-ground containers and dedicated storage facilities in sites across the UK. However, continuing to apply this storage method will not only require the construction of additional containers to house future generated waste, it will also require continued maintenance/repackaging of current waste. Additionally, the use of above-ground containers will increase the carbon footprint on the environment, presents a possible security threat, is cost inefficient in the long term, and poses an undue burden on future generations.[6]

To find a sustainable long-term solution for the UK's higher activity waste (MHRW), the UK Government initiated the Managing Radioactive Waste Safely Programme in 2001. Following the development of this programme, the Committee on Radioactive Waste Management recommended geological disposal, along with interim storage, as the best available approach to management of nuclear waste.[7]

---

[1] Nuclear Waste Services (NWS), Introduction to Geological Disposal: Available at: https://assets.publishing.service.gov.uk/media/621794ce8fa8f5490aff8356/Introduction_to_Geological_Disposal.pdf
[2] BEIS, GDF Programme Business Case, 2019
[3] Ibid.
[4] NWS, ibid.
[5] BEIS, GDF Programme Business Case, 2019
[6] Ibid.
[7] GOV.UK, Implementing Geological Disposal, 2014. Available at: https://www.gov.uk/government/publications/implementing-geological-disposal

Therefore, the UK is seeking to develop a GDF with the expectation that this will:

- allow for permanent disposal of the UK's MHRW and the removal of the burden on future generations
- remove the costs associated with the indefinite construction and operation of interim storage facilities
- provide socio-economic benefits to the potential host communities and/or host communities located in proximity to GDF sites [8]
- avoid the risks associated with surface-level storage of MHRW, including vulnerability to environmental impacts, threats or attacks, man-made events or climate change, and inevitable degradation of waste packages and stores [9]
- allow for the continued use of nuclear power for the needs of the country, which depends on the ability to safely store current and future waste[10]

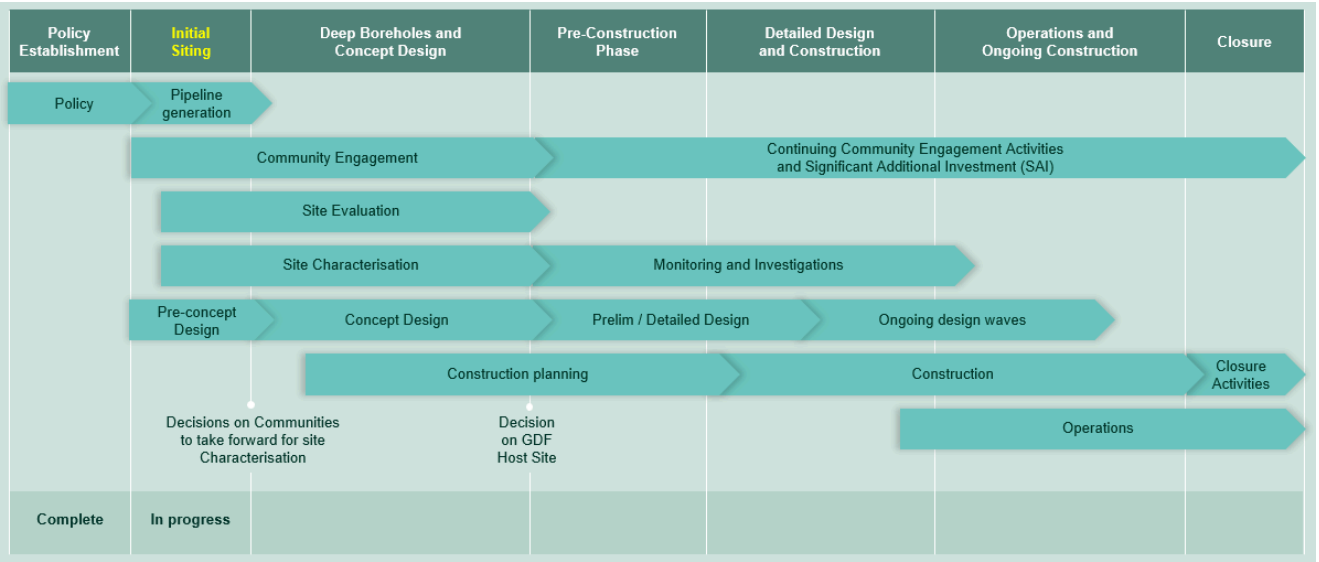
Critically, the GDF programme is consent-based, meaning that the host community must actively consent to the construction and presence of the GDF in their area. Potential host communities will receive investment through a Community Investment Fund (CIF), which will be used to realise locally determined benefits. These benefits include the presence of additional jobs and investment in skills, and community projects aimed at enhancing the wellbeing, environment, or economic landscape of the community (as defined in their own locally determined Community Vision). Once a preferred site has been chosen and the community has demonstrated its commitment through a Test of Public Support, the host community will receive further investment opportunities, referred to as Significant Additional Investment (SAI).

[8] GOV.UK, Implementing geological disposal – working with communities
[9] BEIS (2019), GDF Programme Business Case.
[10] Ibid.

GDF programme implementation

There are a number of key activities which will be conducted as part of the GDF programme, as outlined in Figure 2.

Figure 2: GDF programme implementation plan*



Note:

1. *Dates have been removed to give the graphic more longevity as dates can change over time. The old tranche structure was rescinded so that Nuclear Waste Services (NWS) could work with phasing that is better aligned with the target operating model and business case structure for the GDF programme.

At the time of writing, the GDF policy work had already been completed, and therefore this is outside the scope of the evaluation. Policy establishment involved:

- developing the governance, project management, and framework of the GDF programme
- developing government policy on the community role in the consent-based siting process, including securing approvals for the launch of the siting process

Community engagement and site evaluation will occur over similar timeframes to each other and will broadly consist of the following elements:

- launching the siting process to identify communities that are interested in hosting a GDF
- establishing Community Partnerships to engage with members of the community, including sharing information about the GDF and gathering information on community sentiment towards it
- providing prospective GDF host communities with funding of up to £1 million annually (for communities that progress to form a Community Partnership) to create a community project with the goal of enhancing the wellbeing, environment, or economic landscape of the community

- waste management officials conducting site evaluations in the communities targeted to assess the feasibility of constructing the GDF, with a decision to be made on two sites to take into the next phase

Site characterisation broadly comprises the following elements:

- securing regulatory and planning permissions
- investigations of two sites to confirm feasibility and characterisation of the sites, including borehole drilling
- using this information to confirm whether or not a GDF could be designed and constructed safely in the relevant site
- holding a Test of Public Support to determine whether the community consents to the presence of a GDF
- increasing investment in the community project funding up to £2.5 million annually for those communities that continue to the deep borehole investigation stage (to be used as part of the agreed community project)

GDF design, construction and operation broadly comprise the following elements:

- implementing SAI schemes for community projects designed to benefit the community
- securing all permits and permissions for construction
- creating detailed designs to input into final site selection
- continuing site characterisation into construction
- ongoing construction and emplacement of waste
- decommissioning of above-ground sites

Section 2 Introduction to evaluation

This section provides a short introduction to relevant evaluation concepts, approaches and methods.

Why evaluate?

An evaluation is a systematic assessment of the design, implementation and outcomes of an intervention. It involves understanding how an intervention is being, or has been, implemented, and what effects it has, for whom and why. It identifies what can be improved and estimates the overall impacts and cost effectiveness of the intervention.[11]

In the world of project management for the major infrastructure sector, learning usually takes place at an individual level through gaining experience and learning by doing. However, this kind of learning is not sufficient to allow complex organisations to improve performance. These issues were identified in the National Audit Office's (NAO's) review of the Government Major Projects Portfolio (GMPP)[12] and the Department for Levelling Up, Housing, and Communities (DLUHC) investments.[13] Despite representing billions of pounds of government funding, the NAO found that projects frequently failed to deliver the benefits promised, or, worse still, the NAO was unable to assess the realisation of benefits because the benefits had not been baselined or evaluated. Accordingly, it was not possible to determine if these investments represented good value for money (VfM), to make evidence-based investment decisions, or to learn how performance could be improved.

For effective organisational learning, structured and formal evaluation is required. This kind of evaluation has reached a mature level in the fields of medicine and international development but is still nascent in the world of major project management.

The GDF programme is a complex, substantial and multi-decade investment. The programme expects to deliver a wide range of environmental, economic and social benefits at the community, regional and national level. As such, evaluation will be important to capture learning that can guide ongoing implementation, inform other types of large-infrastructure programmes, and provide accountability for delivering the objectives of the programme.

Forms of evaluation

An implementation and process evaluation (IPE) is an evaluation that is focused on understanding how a particular programme has been delivered, why it has been designed that way, and what experiences of delivery have been like. Monitoring activities are often integrated within IPE and are focused on collecting data to understand the nature and progress of interventions and delivery.

---

[11] HM Treasury, The Magenta Book, 2020. Available at: https://www.gov.uk/government/publications/the-magenta-book

[12] NAO, Projects leaving the Government Major Projects Portfolio, 2018. Available at: https://www.nao.org.uk/reports/projects-leaving-the-government-major-projects-portfolio/

[13] NAO, Levelling up funding to local government, Department of Levelling Up, Housing, and Communities (DLUHC), Session 2023–2024, 17 November 2023. Available at: https://www.nao.org.uk/wp-content/uploads/2023/11/levelling-up-funding-to-local-government.pdf

An impact evaluation is an evaluation that is focused on understanding what effect a programme has had. This often means assessing to what extent programmes have caused the intended outcomes and impacts, and also identifying whether unintended consequences have occurred.

A VfM assessment is focused on making a judgement about whether a particular programme represented value. This is often done by assessing the costs and inputs of a programme, and balancing these against the impacts, benefits and savings achieved.

Evaluation scoping

Verian and Oxford Global Projects were commissioned by the Evaluation Task Force to scope this evaluation plan for the GDF programme. Several steps were taken to develop the plan, as outlined below. Throughout scoping, weekly meetings were held with the Evaluation Task Force and NWS, and additional meetings were conducted/written updates provided when necessary and appropriate. The contract was active from January to June 2024.

Inception meeting

The purpose of the inception meeting was to introduce key team members, to discuss scoping activities and the scoping timeline, and to agree how Verian would work together with the Evaluation Task Force and NWS. This included discussing policy context and programme implementation, as well as identifying key stakeholders for workshops.

Document review

A number of GDF programme documents were shared with Verian, to familiarise the evaluation team with the programme and support the development of the programme Theory of Change. These documents included:

- the business case
- frameworks for the community engagement campaign
- benefit management strategies
- programme execution plans
- programme information packs

Theory of change and evaluation questions

The documents were reviewed and input to a pro-forma analysis framework, structured around a Theory of Change (including inputs, activities, outputs and impacts) to enable an initial draft Theory of Change to be developed. The draft Theory of Change was then refined following a first round of comments by both the Evaluation Task Force and NWS.

A Theory of Change workshop lasting 1.5 hours was held with five key NWS stakeholders to review the Theory of Change components, to determine if these were correctly worded and also whether any components were missing. Key dependencies between different activities were discussed, and draft evaluation questions were posited. Following the workshop, the Theory of Change was further revised and reviewed by the Evaluation Task Force and NWS. The long timeline of the programme

meant that the developed Theory of Change was rather complex to read and so at this point phasing was introduced and the Theory of Change was split into sub-theories by phase.

A second workshop was held with the same participants, to explore the causal mechanisms linking activities, outcomes and impacts, to identify key programme risks and underlying assumptions. Following a further round of revisions, the Theory of Change and refined evaluation questions were reviewed by the Evaluation Task Force and NWS, before being confirmed as suitable for sign-off. However, the Theory of Change is a live document and should be reviewed and updated as the evaluation progresses.

Evaluation planning

Based on the Theory of Change and evaluation questions, an options paper was produced. This considered:

- the periodicity of different forms of evaluation
- key priorities for different stages of implementation
- possible outcomes and indicators

The options paper also sought to address several challenges for evaluation design. Firstly, there are currently 'known unknowns' in regard to the details of the GDF programme at the time for writing, which will be resolved as the local projects develop. For example, we do not yet know the characteristics of chosen sites and the CIF/SAI opportunities have not been fully defined. This makes it challenging to scope a detailed plan for precisely how the GDF programme might be evaluated at this point.

There is also the challenge of complexity and overlap in programme activities. While we have suggested phasing for the evaluation to broadly align with GDF programme implementation activities, many of these activities will occur over similar or overlapping timeframes. This means it may be difficult to find a suitable counterfactual and to disentangle the effects of specific phases of the GDF programme. For example, many activities will occur sequentially during the construction and (partial) operation of the GDF, and some long-term effects of one phase on certain overlapping (shared) outcomes will also materialise in a later phase.

Finally, the timeframes involved are very long – much longer than most evaluations of public policies and programmes. There is thus a huge potential for change in government policy, programme delivery and evaluation techniques, which means that the evaluation approach must be flexible and able to change alongside such changes.

The content of the options paper was reviewed by the Evaluation Task Force and NWS, before being presented for consideration by NWS stakeholders in a 1.5-hour-long workshop. Based on the workshop feedback this evaluation plan was produced, to lay out the recommended approach for an IPE, impact evaluation, and VfM assessment of the GDF programme.

Section 3 GDF evaluation plan

This section sets out the overall approach for the evaluation of the GDF programme, including the Theory of Change, key evaluation questions, and the suggested timeline.

GDF Theory of Change

The Theory of Change has been split into four phases, over which the evaluation will take place and which are broadly aligned with the GDF programme implementation activities – the key difference being that the evaluation will focus on the phasing of when outcomes are expected to materialise. The evaluation will still explore the overall implementation and impact of the GDF programme: the phasing is a mechanism to manage the extended evaluation timelines, in order to focus on what can realistically be delivered and achieved within a given timeframe. The phases are detailed in Table 3.

Table 3: IPE and impact evaluation activities in ToC phases

Phase	Implementation activity	IPE and impact evaluation activities
Phase 1	Community engagement and site evaluation	● establishment of working groups ● funding for the initial setting up of community projects ● selection of the final two potential host communities for the GDF site to take forward to intrusive investigations
Phase 2	Site characterisation and Test of Public Support	● detailed site and geology investigations, through to the point at which the NWS is confident that a GDF can be designed and constructed safely ● delivery and emerging impacts of the community projects, and the process of holding a Test of Public Support
Phase 3	GDF technical development, design and construction	● continued delivery and impact of community projects funded by the SAI after the Test of Public Support, the decision on the preferred site, as well as the transition of the Community Partnerships to liaison groups ● process and permissions involved in designing and constructing the GDF, up until the first waste emplacement
Phase 4	GDF operation	● analyse the extent to which the GDF programme achieves its longer-term outcomes and impacts, particularly in relation to the disposal of MHRW and the ongoing impacts of community projects

The Theory of Change reflects the current plans for the GDF programme at the time of writing, with further scoping expected to be necessary as each phase approaches. For example, the expected benefits from community projects are not yet fully articulated and this may change the Theory of Change and/or require sub-theories to be developed at the community level. The full Theory of Change is available in the Technical Appendices.

Evaluation questions

There are three overarching evaluation questions:

- Is delivery of the GDF programme on track, and what challenges, risks and facilitating factors have been experienced throughout delivery?
- To what extent has the GDF programme contributed towards achieving its intended outcomes and impacts, and how / why?
- Is the GDF programme delivering VfM, and how / why? To what extent is it likely to deliver VfM in the future?

Three forms of evaluation are proposed to respond to these evaluation questions:

- IPE
- impact evaluation
- VfM assessment

Each evaluation strand will have a number of additional sub-questions, as set out below. It is expected that at each scoping stage further phase-specific evaluation questions will be identified to respond to emerging information needs as the programme continues.

IPE

IPE evaluations are designed to understand how the GDF programme has been delivered, whether delivery matches the policy intent, any facilitators and challenges that have been experienced throughout delivery, and any examples of best practice and learning. There are nine key evaluation sub-questions for the IPE assessment, as follows:

- EQ1. Has the GDF programme delivered its activities as intended, and is it on track to do so going forwards?
- EQ2. Is the GDF programme currently at the expected stage of delivery?
- EQ3. Is the GDF programme on track to enter the next phase of delivery?
- EQ4. What are the critical risks and facilitating factors that have been experienced, and which risks may occur in the next stages of delivery? What actions could be taken to avoid these?
- EQ5. Are there any examples of facilitating factors or best practices that have been experienced during delivery? What actions could be taken to capitalise on these or otherwise improve experiences of delivery?
- EQ6. What are the outcomes and impacts that the GDF programme intends to achieve, and how is it intended to achieve these?
- EQ7. To what extent have the outcomes and impacts that the GDF programme set out to achieve occurred, and to what extent can they be attributed to the GDF programme?

- EQ8. To what extent does the evidence suggest that the GDF programme will go on to contribute towards the realisation of outcomes and impacts it intended to achieve?
- EQ9. Have there been any unintended consequences or impacts of the GDF programme?

Impact evaluation

This will assess observed changes in outcomes throughout the GDF programme and determine whether and how the programme caused these. The impact evaluation will contribute to an understanding of what has worked and what has not, and any unintended consequences of the GDF programme. The evaluation sub-questions for the impact evaluation are:

- EQ7. To what extent have the outcomes and impacts that the GDF programme set out to achieve occurred, and to what extent can they be attributed to the GDF programme?
- EQ8. To what extent does the evidence suggest that the GDF programme will go on to contribute towards the realisation of outcomes and impacts it intended to achieve?
- EQ9. Have there been any unintended consequences or impacts of the GDF programme?

VfM assessment

The purpose of the VfM assessment is to investigate how well the programme has performed against the NAO's key dimensions of VfM (4Es framework), whether the value created has justified the use of resources, and how VfM is likely to evolve over the course of the programme. The evaluation sub-questions which the VfM assessment seeks to answer are as follows:

- EQ10. How well has the programme performed against the NAO's key dimensions of VfM: economy, efficiency, effectiveness, and equity (the 4 Es)?
- EQ11. For completed phases, does the value created justify the resource use? For future phases, how is VfM likely to change and evolve over the course of the programme?
- EQ12. What is the financial cost of the GDF programme phases and the programme as a whole, including expenditure on CIF, SAI, business case development and inputs into evaluation, characterisation, design, construction, operation and closure?
- EQ13. For future phases, what are the projected whole-life costs, how have these changed since the baseline, and how confident can we be that they will materialise?
- EQ14. What cost savings or benefits have been realised as a result of the GDF programme phases and the programme as a whole?
- EQ15. For future phases, what are the projected cost savings or benefits, how have these changed since the baseline, and how confident can we be that these will materialise? For completed phases, does the value created justify the resource use?

Evaluation timeline

In line with the evaluation phasing, a high-level evaluation timeline is proposed, comprising the following elements:

- a dedicated IPE for each evaluation phase
- an ongoing impact evaluation
- at least one VfM assessment at the end of the GDF programme

The first element, which requires an IPE dedicated for each evaluation phase, includes a (re-)scoping phase before implementation begins, a data collection period during and after implementation, and an analysis and reflection period.

Furthermore, the ongoing impact evaluation with regular re-scoping and review points based on critical points of implementation and/or when impact estimates are required by stakeholders. These are to be defined by the NWS but they might include business cases and Infrastructure and Project Authority (IPA) gates.

At least one VfM assessment should be conducted at the end of the GDF programme, once all costs and benefits have been realised. Given the long timeframes involved, we would also advise conducting an interim VfM assessment at the start of GDF programme closure and carrying out VfM assessments of specific activities where there is a discrete set of benefits and associated costs. These could be conducted in the earlier stages (such as the community engagement and site evaluation phase), which would provide an evolving picture of VfM that could be used to periodically assess performance and test the assumptions of the evolving business case estimates. Activity-based assessments could be agreed on a rolling basis at the beginning of each phase of the evaluation (during scoping) or could be responsive to significant programme achievements.

Specific dates and years have not been mapped against this plan, given the level of uncertainty in regard to how long some implementation activities are likely to take. During re-scoping, it may be decided that sub-phase evaluations would be useful for those phases with particularly long timeframes (such as construction and operation).

A visual depicting the evaluation timeline is available in the Technical Appendices. A number of considerations should be taken into account when further developing this into a more detailed evaluation timeline. Namely, evaluators should:

- plan data collection and/or analysis so as to deliver outputs (emerging, interim and final findings) to contribute where possible to business cases and infrastructure and project authority gates, and to feed into the next phase of implementation
- ensure the IPE and impact evaluation findings feed into scoping for subsequent phases
- scope any data collection across evaluation strands to enable efficiencies to be identified between the two – for example, asking some impact-related questions within IPE interviews, or adding implementation-related questions to an impact-focused survey

Analysis and reporting

A cross-cutting approach to analysis and reporting is recommended (i.e. encompassing process, impact and VfM as relevant) to facilitate an evaluation narrative that captures the programme as a whole. This would be an explicit consideration at the relevant scoping stages, including developing or reviewing the contribution analysis framework to inform all evaluation strands and facilitate cross-pollination of findings.

There are other considerations with regard to reporting that will need to be taken into account at each scoping stage. These include the following:

- When is it most useful for results to be shared?
- What audiences is it useful to share results with?
- How detailed/concise do the results need to be?
- How should a balance be struck between standalone and synthesis reports required to cover in-phase and across-phase findings, respectively?
- How should the results be validated and scrutinised, for example by academic or expert peer groups such as the DESNZ peer review group, and how often should this be built in?
- Are multiple reports required? What format should these be in (i.e. PowerPoint, Word, HTML)?
- Should the results be published, and what processes and requirements relating to government publication (for example, accessibility criteria, Environmental Implementation Review requirements) would need to be followed in order to ensure this is done sensitively, ethically and efficiently?

Section 4 Implementation and process evaluation

This section outlines a recommended approach to evaluating the implementation of the GDF programme.

An IPE of the GDF programme would answer the following evaluation questions:

1. What are the outcomes and impacts that the GDF programme intends to achieve, and how is it intended to achieve these?
2. Has the GDF programme delivered its activities and outputs as intended, and is it on track to do so going forwards?
3. How was the GDF programme delivered and why was it delivered in this way?
4. What barriers or challenges have been encountered in delivering the GDF programme, and what examples are there of best practice? Why have these occurred?
5. What learnings about the delivery of the GDF programme could be applicable and relevant for improving the future delivery of the GDF programme and / or other mega-infrastructure projects?

Overall, it is recommended that each phase of the evaluation includes an IPE, comprising the following elements:

- A scoping phase to review and refine the theory behind the programme, to refine approaches to collecting data, and to design the data collection approach
- Mixed-methods (i.e. qualitative and quantitative) data collection to holistically answer the IPE questions
- Analysis and reporting focused on learnings for the future of the GDF programme and other major infrastructure projects, to be aligned with the contribution analysis framework (see impact section below)

Scoping

Given the length of the GDF programme it will be critical to build in explicit and regular periods for reflection and revision in the evaluation design. (Re-)scoping will be important for the following purposes:

- co-producing a detailed implementation plan that maps the key activities and their timing for the forthcoming evaluation phase
- revising and adapting the Theory of Change
- identifying what, if any, data is already being collected for the upcoming business case(s)
- assessing whether existing approaches to monitoring, data collection or indicators can or should be revised to reflect current progress, changes in circumstance, or to improve robustness or accuracy, and, if so, how they should be revised
- assessing whether additional data is needed in order to give indications of progress, and, if so, what form that data collection should take

- identifying any efficiencies that could be achieved to reduce the time and cost of the evaluation, or to reduce participant burden
- assessing whether any other changes need to be made in order to ensure alignment with the impact evaluation or VfM assessment, including consistency with a contribution analysis framework

Overall, the following steps are recommended:

1. Document review of existing data and upcoming implementation activities, including business cases
2. Mapping of relevant stakeholders for upcoming scoping activities
3. Co-production workshop(s) with stakeholders across various disciplines and responsibilities, for the purpose of creating and revising a holistic and realistic implementation framework with clear targets, and revising the Theory of Change. It is recommended that members of the Community Partnerships and / or community project teams are involved in this process where possible and relevant for the phase.
4. Interviews with stakeholders involved in business cases and / or relevant upcoming implementation activities to identify evidence gaps and clarify any questions from co-production workshops
5. Interviews with stakeholders involved in data collection to understand limitations of or challenges with existing data and to test any new approaches to data collection
6. A design and analysis period, which draws on insights from the scoping activities
7. Validation workshops to present and revise the updated approach with critical stakeholders to ensure the revised scope is fit for purpose, including evaluation questions, the Theory of Change, and proposed data collection methods.

Data collection and analysis

Approaches to data collection within an IPE may change for each phase, and must necessarily be matched closely with the phase-specific evaluation questions, the Theory of Change, and the co-produced implementation plan.

Table 4: Proposed data collection approach for IPE evaluation questions

Evaluation focus	Key evaluation question	Suggested data collection method
Theory of change	What are the outcomes and impacts that the GDF programme intends to achieve, and how is it intended to achieve these?	Revising and adapting the Theory of Change through document review, stakeholder interviews and stakeholder workshops, specifically to ensure they reflect any changes and clearly articulate causal pathways.

Evaluation focus	Key evaluation question	Suggested data collection method
Monitoring	Has the GDF programme delivered its activities and outputs as intended, and is it on track to do so going forwards?	Document review and progress or milestone reporting (see table below). Interviews and / or surveys with key staff involved in delivery. Analysis of financial data and cashflow curves – comparison between elapsed time and expenditure. Secondary data analysis, including of data collected and submitted by NWS. Analysis of the data against key targets set in business cases.
Process	How was the GDF programme delivered and why was it delivered in this way?	Interviews and organisational assessment surveys with key staff involved in design and / or delivery.
	What barriers or challenges have been encountered in delivering the GDF programme, and what examples are there of best practice? Why have these occurred?	Interviews with key staff involved in GDF programme delivery. Interviews with or site visits to other audiences as applicable based on insights gathered – for example interviewing members of the potential host community/house community, construction workers, and those receiving training.
	What learnings about delivery of the GDF programme so far could be applicable and relevant for improving the future delivery of the GDF programme and / or for other mega-infrastructure projects?	Interviews with key staff involved in GDF programme delivery. Validation and learning workshops with GDF programme staff, presenting updates on progress and key findings / learnings. Learning workshops with representatives of other major infrastructure projects.

Some potential indicators for monitoring delivery of the GDF programme are summarised in Table 5. We would expect this list to be reviewed and refined at each scoping stage to align it with the implementation plan and business cases.

Table 5: Proposed data collection approach for monitoring indicators

Monitoring indicators	Method
Number of communities engaged in the process	Document review / interviews
Stage of communities engaged in the process	Interviews
Maturity of upcoming business case	Interviews / business case maturity checklist
Community sentiment towards and knowledge and awareness of the GDF programme	Secondary data analysis / surveys / interviews with members of the host community or Community Partnership
Funding that has been spent and secured for the next stage of delivery	Document review / interviews
Procurement capacity	Interviews / organisational readiness assessments
Confidence of Community Partnerships in regard to achieving the next stage of delivery	Interviews / surveys
Confidence of NWS in regard to achieving the next stage of delivery / permissions	Interviews / surveys
Number of delivery and permissions milestones achieved	Secondary data analysis
Identified barriers, risks and challenges	Interviews / risk registers / board reports
Facilitating factors and best practice	Interviews
Number of direct jobs related to construction and design of the GDF	Surveys / secondary data analysis / programme documentation
Number of direct jobs related to community projects	Surveys / secondary data analysis programme documentation

Monitoring indicators	Method
Number of indirect jobs related to the presence of the GDF and / or community projects	Surveys / secondary data analysis
Extent of waste emplaced and amount of MHRW in surface storage	Secondary data analysis / programme documentation
Quality and availability of training	Secondary data analysis / interviews
Safety events, injury frequency, risks or 'near-miss' safety incidents	Secondary data analysis, safety records
Potential host community / host community population statistics	Secondary data analysis
Costs incurred	Secondary data analysis for financial records
Costs avoided, including change in cost associated with surface storage of MHRW	Secondary data analysis
Progress against commercial strategy (i.e. contractor boards)	Secondary data analysis
Site evaluation data	Secondary data analysis
Evidence of the GDF programme enabling new low-carbon nuclear power stations	Policy documentation / interviews
Reduction in carbon emissions associated with new low-carbon nuclear power stations	Official analyses
Evidence of the avoidance of environmental/safety/security risks due to reduction of surface storage of MHRW	Risk registers / interviews
Sharing of specialist knowledge related to the GDF	Learning legacies / knowledge sharing activities

The list below provides a template for the data collection approach to be applied in each IPE phase. The iterative nature of data collection is intended to align with the contribution analysis approach (discussed later in relation to the impact evaluation):

1. Designing research tools and mapping relevant stakeholders or participants.
2. Fieldwork period A, which is focused on how the programme has been delivered and what stage of delivery the programme is at, using stakeholder interviews, secondary data analysis and document review.
3. Interim analysis period.
4. Fieldwork period B, which is focused on any updates to progress, and identifying and exploring barriers, challenges and learning. Based on stakeholder interviews, secondary data analysis and document review.
5. Interim analysis and reflection period.
6. Fieldwork period C, which is an optional additional period for further questions, clarifications or insights based on fieldwork periods A and B: for example, including additional interviews with members of the host community.
7. Final analysis period.
8. Provisional results, validation and learning workshops with stakeholders, and learning workshops with representatives of other large major infrastructure projects.
9. Revision and full report on results.

Considerations, risks and resourcing

Ultimately, the methods which it will be most appropriate to employ, and the intensity with which they are employed, should be considered at re-scoping. It is likely that efficiencies could be achieved in the recommended process for scoping and/or data collection, for example by:

- combining fieldwork periods
- choosing to omit additional periods of fieldwork or learning workshops
- conducting fewer interviews or workshops

However, applying such efficiencies may affect the level of depth that an evaluation is able to answer the relevant evaluation question. For each phase, the scoping stage will be critical in order to strike the appropriate ‘balance’ between conducting the evaluation efficiently and also allowing for sufficient depth to be captured.[14]

Decisions regarding the intensity and periodicity of the evaluation are likely to be made on a phase-by-phase basis, and outlined in the scoping or re-scoping phases. These will make recommendations on the evaluation design for the relevant phase based on:

- the current stage of implementation, the type of activities that are occurring, and the length of time before the next stage of implementation

[14] In particular, within an IPE scoping stage it may be useful to identify an expected ‘saturation point’ (the period at which no new information is likely to be garnered from further data collection) and to design the evaluation in such a way that does not exceed this point. Saturation points will differ depending on the nature of the questioning and data collection method, but an experienced evaluation partner would be able to advise on this.

- the extent of data that has already been collected and its level of suitability
- the level of resources available among relevant stakeholders, and engagement with the evaluation
- the available budget and access to relevant stakeholder information
- the nature, intensity and complexity of the upcoming activities

The IPE would need to consider the following issues:

- confirmation bias in interviews and surveys, whereby stakeholders have a desire to present the programme as successful and therefore omit or downplay slow progress or barriers experienced – this can be offset by collecting multiple different forms of evidence to triangulate insights
- inconsistencies in secondary data, particularly where this is self-submitted or amalgamated from multiple sources
- out of date documents or not enough documents being provided
- whether the timing of the fieldwork periods is aligned with with implementation activities
- ways of encouraging engagement with the evaluation process and reducing participant burden
- planning in advance to ensure insights are shared in a timely fashion and acted upon during delivery to maximise value for the stakeholders and communities involved

Section 5 Impact evaluation

This section outlines a recommended approach for evaluating the impacts of the GDF programme on outcomes at a local and national level.

The impact evaluation of the GDF programme would answer the following questions:

1. What are the outcomes and impacts that the GDF programme intends to achieve, and how is it intended to achieve these?
2. To what extent have the outcomes and impacts that the GDF programme set out to achieve occurred, and to what extent can they be attributed to the GDF programme?
3. To what extent does the evidence suggest that the GDF programme will go on to contribute towards the outcomes and impacts it intended to achieve?
4. Have there been any unintended consequences or impacts of the GDF programme?

Recommendations in regard to assessing the local and national-level impacts are reported separately here, due to differences in the methods employed in each case. Furthermore, in the long term, while an analysis of the national effects will consider the aggregate impact of the GDF programme, an evaluation of the local effects of the programme will provide a comparison of whether the effects of the GDF programme are more concentrated in the host community.

Overall, we recommend:

- designing a theory-based evaluation and using a contribution analysis approach at the programme level
- conducting contribution analysis, quasi-experimental impact evaluations, and simulation analysis of individual aspects of the programme

A theory-based evaluation at the programme level, using contribution analysis, would allow evaluators to explore the extent to which the GDF programme has contributed towards achieving its intended aims. This will entail considering different types of evidence that are relevant to different outcomes along the programme's Theory of Change. It will also require consideration of evidence collected during different phases and activities involved in developing the GDF.

Furthermore, within the programme-level contribution analysis framework, we recommend undertaking focused contribution analysis, quasi-experimental impact evaluations, and simulation analysis of individual aspects of the programme and its outcomes. For example, a more focused analysis could also look at the impacts of the CIF, or even individual investments within that fund (as a type of case study). Each of these separate pieces of evidence would contribute to the broader overarching contribution narrative for the programme and, in the long term, the assessment of whether the expected outcomes have materialised along the programme's Theory of Change during its different phases.

Scoping

The impact evaluation is intended to occur continuously over the duration of evaluation – as the evaluation moves into each subsequent phase there will be an opportunity to re-scope evaluation

activities as required. The precise evaluation questions the impact evaluation can reasonably answer will also depend on the method or approach selected to evaluate the programme, as well as the specific outcomes included in the evaluation.

Consideration should be given at each scoping stage to evaluating which method will be used. This should reflect the updated details, evaluation budget and specificities of the programme. For example, the design will need to consider the specific activities and projects arising from the CIF, to properly evaluate the effects of these activities on local outcomes. There may also be a need for prioritisation of the outcomes selected for impact analysis, based on a revised Theory of Change.

Evaluating local-level effects

It is expected that the GDF's programme activities will have a large and significant impact on the local host community. The expected impacts on the local community include the following:

- local development
- wellbeing and the local economy
- trust in and support for the GDF

During the community engagement and site evaluation and characterisation activities (Phases 1 and 2), the CIF is expected to impact local development outcomes agreed by the Community Partnership, which could help build support for the GDF programme before the Test of Public Support. The planned CIF strategies intend to impact outcomes that include local economic development, enhancing the natural or built environment, and community wellbeing.

During evaluation Phase 3 (GDF technical development, design and construction) and Phase 4 (operation), the development, construction and operation of the GDF is expected to provide new jobs and infrastructure in the local economy. This could also encourage business and migration to the local area. Overall, the development of a GDF in a host community is expected to increase the wellbeing of the host community.

Throughout the community engagement and site evaluation, site characterisation, and GDF technical development, design and construction activities (Phases 1–4), the heightened local engagement and benefits that accrue in the host community due to the GDF are expected to increase trust in and support for the facility.

To evaluate the impacts of the GDF programme on local communities, we propose two recommendations. Firstly, we recommend applying a theory-based evaluation approach, including using a contribution analysis to develop a plausible narrative regarding the programme's contribution to the local community and the realisation of relevant outcomes. This can be performed at a programme level and for different phases/benefits of the programme.

Our second recommendation is to embed a quasi-experimental impact evaluation in a contribution analysis, which would provide a strong source of evidence on specific outcomes to inform a broader narrative about the GDF programme's impact on the local community. This option will additionally require that data to measure outcomes are available and in a suitable unit of analysis (e.g. individual-, address-, or street-level) or that the evaluation's budget allows for large-scale surveys in the host

community and a comparable comparison community. We consider this a vital element of any successful GDF programme evaluation.

This approach would require different types of data to be collected and analysed. To this end, we recommend conducting the following:

- interviews and focus groups with local residents, businesses, voluntary groups and other local stakeholders in order to gather evidence about perceptions of the effects of the programme
- a local community survey of people living in the GDF host area (and other communities where a CIF is established), to collect evidence about how key outcomes for residents in the community have changed
- a comparison survey of people living in other similar areas, which can be used to estimate what the change in key outcomes would have been in the GDF host area without the programme
- analysis of existing datasets, which could be used to track certain relevant outcomes in a host area, and potentially in comparison areas, such as Department for Work and Pensions (DWP) and HM Revenue & Customs (HMRC) records of benefits claims and tax payments, business records from the Inter-Departmental Business Register, and data from the Annual Population Survey, property prices and the Census.

It is likely that there will be opportunities to bring the data collection for the IPE and impact evaluation together for efficiency and to minimise burden. If we conduct the scoping for these strands together then, for example, the resident surveys could be used to provide insights into the effectiveness of community engagement activities, and the IPE data can help explain why some outcomes may be taking longer to materialise than expected.

Theory-based evaluation approach using contribution analysis

The five key steps involved in a contribution analysis are shown in the figure below. They require that a detailed Theory of Change is created, which details the expected mechanisms through which the programme will have an impact. Evidence is then generated and evaluated to assess the plausibility of the logic and whether the expected outcomes have materialised along the Theory of Change. In some cases, this evidence-generating process is planned in iterative steps, whereby evidence is collected to develop and then confirm the contribution narrative (or ‘contribution claims’).

Contribution analyses can be flexible in regard to the forms of evidence used to assess the plausibility of the causal links and claims made about the programme’s impacts. For the evaluation of the GDF programme, this means different forms of evidence, such as quantitative or qualitative and data from different sources or stakeholders, can be used to inform the overall assessment of the plausibility of the contribution of the GDF programme to the local community and their intended benefits.

Table 6: Steps of contribution analysis[15]

Step	Activities
Programme Logic	● clarify the logic behind how the intervention is intended to affect the outcome ● include external factors, assumptions, risks and other influences on the outcomes ● use these factors to form 'links' in the Theory of Change which represent the path by which impact is generated
Existing evidence	● gather and assess evidence which exists on the links and outcomes in the programme logic ● assess the plausibility of the intervention influencing the outcomes in light of the evidence ● develop a 'contribution story'
Assess contribution story	● assess the credibility of the contribution story ● ensure there is agreement from relevant and knowledgeable stakeholders ● assess whether the theory is validated by the observed results ● identify the weaknesses and challenges to the story
Generate Evidence	● gather evidence to address gaps or weaknesses in the contribution story
Assess and iterate	● draw on the evidence gathered to assess the contribution story once more and potentially iterate the ToC

Different approaches to collating evidence to inform a contribution analysis also exist. For example, evidence can be based on approaches that directly ask stakeholders about their perceptions of the programme, its value and importance in regard to an outcome or output occurring, and the extent to which it has contributed to outcomes. Alternatively, more indirect approaches also exist, such as approaches based on the Qualitative Impact Assessment Protocol (QuIP).

Of course, regardless of the method or source of data used, the quality of the evidence used should still be as high as possible, as the strength of the assessment rests on the strength and relevance of the evidence. This, therefore, also requires that evaluators consider the caveats and limitations of different sources of evidence when considering the strength of different claims about the programme's contributions and impacts. Furthermore, a limitation of many contribution analysis studies is that they are informed by purely qualitative evidence, which provides little indication of the scale or magnitude of the impacts of a programme. Using quantitative data impact evaluation methods (such as quasi-experimental evaluations, discussed below) to evidence aspects of the Theory of Change and key outcomes can help to inform our understanding of the potential magnitude of effects that can be attributed to the programme.

[15] Mayne, J. 'Contribution Analysis: Addressing Cause and Effect', in *Evaluating the Complex*, K. Forss, M. Marra and R. Schwartz (Eds.), Transaction Publishers; Piscataway, New Jersey, 2011

Quasi-experimental evaluation approach

A quasi-experimental design is used to infer the effects of a programme by comparing the outcomes realised by those affected by a programme with those of a comparable group that has not been affected by it. In this case, we can envisage a quasi-experiment with two different types of comparisons:

- a comparison of beneficiaries and non-beneficiaries in the same community
- a comparison of residents in the host community with those in another comparable community

The comparison of beneficiaries and non-beneficiaries in the same community involves the comparison of the outcomes realised or not realised by the beneficiaries of the activities funded by the CIF with those of other non-beneficiaries from the same community. However, the feasibility of applying this approach will depend on the specific projects the CIF invests in, and whether this has spillover effects on others in the community or impacts the entire community.

It may also be possible to compare the overall outcomes realised by the host community (or other communities with a CIF) with those of other communities. Communities would be selected based on observable characteristics which suggest they are similar to the host community to use as a comparison group. Since the GDF programme is expected to have national impacts that may also impact other communities, in the long term this comparison would make it possible to examine whether the benefits of the programme are more concentrated in the host community than in other comparable areas in the UK.

The precise quasi-experimental approach used will depend on the availability of outcome data and the specifics of the implementation of particular aspects of the programme (such as the CIF). In the Technical Appendices we elaborate further on ways we may incorporate specific quasi-experimental designs, such as difference-in-differences and regression discontinuity designs, into the evaluation of specific phases/benefits of the programme.

Overall, we recommend considering a quasi-experimental approach as a source of evidence within a broader contribution analysis, and we recommend that the evidence identified from this approach should also be triangulated with other sources of evidence to support and add nuance and breadth to the evaluation's findings. This recommendation is made due to the following caveats about applying this approach in this context. Firstly, it may not be possible to account for all relevant outcomes and impacts using a quasi-experimental approach. For some outcomes data might not be available or accurately quantifiable, and only a limited number of outcomes will feature in a quasi-experiment in practice as increasing the number of outcomes included in an evaluation reduces the ability to detect effects. In addition, a quasi-experiment will require strong assumptions about the comparability of the comparison group. A comparison group should be as similar as possible to the group affected by the programme (whether the local community as a whole or specific sub-groups of beneficiaries within the community). However, it is generally impossible to be sure that any quasi-experimental approach has fully accounted for all relevant differences between groups. These methods therefore rely on the assumption that any remaining differences are not driving any difference in outcomes between the groups. Finally, it is often challenging to robustly identify localised and sub-group effects if they exist using a quasi-experimental analysis, since the approach does not use randomisation, which means it cannot be completely ruled out that other factors are impacting the outcomes.

Evaluating national-level effects

While the direct impacts of the GDF programme will be concentrated in the local host community, there will also be overarching benefits that accrue in the wider economy during the GDF's construction and operation. Understanding the impacts of the GDF programme on the broader UK economy will be crucial to understanding the overall scale and social value of the programme's outcomes.

The expected broader impacts of the GDF programme are as follows:

- long-term safe disposal of MHRW and a reduction in the inventory of radioactive waste in surface storage
- new low-carbon nuclear power stations, by providing a permanent waste disposal route for their waste and spent fuel
- reduction in the environmental impact of surface storage
- reduction in the costs associated with indefinite construction and operation of interim storage facilities
- removal of the security and hazardous waste risks of above-ground storage of nuclear waste materials
- increase in economic activity and potential creation of a large number of jobs and supplier contracts, some which may materialise in the host community, but nonetheless are likely to spill over into the broader national economy, especially in cases where contractors are located in other parts of the country
- new export of specialist knowledge as the specialist knowledge, skills and experience developed during the implementation of the GDF programme create new specialist expertise that its contractors can then export

Some sub-evaluation questions specific to the national-level effects may include:

- How much MHRW has been stored in the GDF? Has there been an associated reduction in surface storage? How does this compare to the expectations set out in the business case?
- To what extent has there been a reduction in the financial cost of surface storage and how does this compare to the expectations set out in the business case?
- Has the GDF programme enabled the development of new low-carbon nuclear power stations?
- What is the impact of the GDF programme on the UK's carbon footprint relative to maintaining indefinite construction and operation of interim storage facilities?
- What is the impact of the GDF programme on employment in the UK?
- What is the impact of the GDF programme on the UKs international trade and investment?
- To what extent has implementing the GDF programme reduced national security risks and hazards in the UK?

Evaluating the realised national impacts of a programme is highly complex and can be difficult since it is often challenging to measure many kinds of outcomes (such as carbon savings) or to understand

what would have been the national-level outcome had the programme not been implemented (i.e. to account for a counterfactual).

Common approaches to evaluating such national impacts of a programme include using simulated models and theory-based evaluation approaches.[16]

Simulated models, which are often based on the expected implementation of a programme and its activities, make assumptions about the relationships between factors, and about savings and benefits, to infer the impacts of a programme on an outcome. For example, for a programme that reduces energy consumption, we can model the associated impacts of the programme on carbon savings (without collecting data on carbon emissions) based on assumptions about the relationship between energy consumption and carbon emissions. To assess the realised benefits of a programme, these models can often be updated during or after the programme is completed to reflect details about its actual implementation and activities. This provides an approximation of the 'realised estimated effects', which can be compared with the original economic business case provided. Where possible, these quantitative estimates are also regularly monetised to inform cost–benefit analysis assessments, which can contribute evidence to broader assessments of a programme's VfM.

It is also common that programmes and reforms with nation-wide impacts use theory-based approaches – for example, to assess the contributions of an intervention to outcomes. This is more common when outcomes are not easily quantified or simulated, or cannot be monetised.

Below we provide further details of possible approaches that could be used to evaluate the impact of the GDF programme's different national impacts. We categorise them according to likely quantifiable and non-quantifiable impacts for the purposes of a cost–benefit analysis.

Quantified (monetised) benefits

Long-term disposal of MHRW and associated cost savings associated with surface storage

An evaluation of the long-term disposal of MHRW would involve determining the amount of MHRW waste stored in the GDF and the associated reduction in the use of surface storage. The financial cost of the GDF and cost savings associated with reduced surface storage would then be calculated. These figures can then be compared to the expectations set out in the economic case to determine if the forecast cost–benefit ratio has been achieved. This approach would provide evidence on the extent to which the GDF programme fulfilled the assumptions of the business case and would provide evidence for the broader assessment of VfM. Data on radioactive waste storage and expenditure could come from programme documentation. However, it would be preferable to leverage findings from future official reviews of the GDF programme, such as the NAO's Major Project Reviews.

[16] Modern econometric approaches, such as the synthetic control method, have also been applied to evaluate the impacts of national interventions. However, we do not cover these here since we do not consider them relevant to the GDF programme. In particular, it is unlikely that we will be able to attribute effects using such an approach, due to changes in other contextual and policy factors.

Environmental impacts

An evaluation of the carbon savings and other environmental benefits of the GDF programme will require a simulated model that estimates the magnitude of, and forecasts the timing of, carbon savings attributable to the programme's activities.[17] This should also account for any expansion of the use of nuclear power and carbon implications of a change in the UK's energy mix. The carbon output of the GDF programme would need to be compared to an estimated counterfactual policy scenario, such as one assuming indefinite construction and operation of interim storage facilities. The social (monetised) value of carbon savings can then be easily calculated using techniques described by the supplementary guidance to HM Treasury's Green Book on the valuation of energy use and greenhouse gas emissions.

Since such models are usually highly sensitive to the assumptions about how a programme will be implemented, the values used as conversion factors, and the timing of the programme's key milestones, the evaluation's design should plan for periodic updates of the model to assess the estimated 'realised' benefits. These are the estimated benefits given the actual implementation of the programme and any updates to key inputs or assumptions (e.g. the monetised value of carbon).

Job creation

The employment the GDF creates will be an important part of the benefits realised by the programme. This may include jobs directly created by the GDF and its suppliers, as well as the indirect jobs related to this economic activity. Capturing the total employment effects of the GDF across the economy will require applying an estimated job multiplier. This is a factor that reflects how many additional jobs result from the creation of one new job, which will be used to estimate the number of new jobs created by the GDF. Evaluating this aspect of the programme would involve using data collected for the implementation assessment within the IPE discussed above (e.g. on the realised number of new jobs created by the GDF and its suppliers across the UK) and combining this with a model of the job multiplier effects of the GDF. The DWP Social Cost–Benefit Analysis framework discusses principles to be applied in estimating the (monetised) social value of changes in employment. The evaluation of this aspect of the programme will also need to account for any jobs created or lost during decommissioning of facilities in a counterfactual scenario, such as a scenario assuming indefinite construction and operation of interim storage facilities.

New export of specialist knowledge

The evaluation of the GDF programme could also consider trying to capture information about the number of export opportunities or, preferably, the total financial value of such opportunities that arise among its suppliers related to the work and experience generated by the GDF programme. Alternatively, it could capture information about whether suppliers are investing in, or receiving investment in, export services related to the GDF project. This would require planning to collect confidential information from a supplier survey, or another similar data collection method, which could also be aligned with data collection efforts with suppliers for the IPE discussed above.

---

[17] Similar models are available for other GMPP projects, such as the UK's smart meters rollout programme. Available at: https://assets.publishing.service.gov.uk/media/5d7f54c4e5274a27c2c6d53a/smart-meter-roll-out-cost-benefit-analysis-2019.pdf

Non-quantified (non-monetised) benefits

It is commonly the case with large infrastructure projects that impact outcomes are difficult to quantify and/or that it is difficult to monetise the social value of these benefits.[18] Nevertheless, understanding whether these benefits have been realised should also be considered in efforts to understand the broader social value and contribution of the GDF programme.

Enabling the development of new low-carbon nuclear power stations

Evaluation of the carbon savings and environmental benefits associated with the development of new nuclear power stations could be achieved by applying a quantified approach (see environmental benefits section above). However, care must be taken when claiming these benefits in relation to the GDF programme because it is the new nuclear power stations that are delivering carbon savings, not the GDF. As such, the GDF can only be considered an enabler to the new nuclear power stations. If both the GDF and the new nuclear power stations claimed the carbon reduction, twice as many carbon savings would be claimed as really existed. Therefore, a non-quantified approach needs to be applied to understand the contribution that the GDF makes to quantified carbon savings associated with the new nuclear power stations. This would involve determining the extent to which the GDF enabled the development of new low-carbon nuclear power stations – or at least a caveated statement about the magnitude of carbon savings the GDF has enabled indirectly via other nuclear programmes.

The most conclusive evidence would be provided by official policy documents or regulations that stipulate that new nuclear power stations will not be permitted without a viable long-term disposal solution for nuclear waste. By implication, if there was no GDF (or a credible alternative) new nuclear power stations would not be permitted and there would be no associated carbon savings. However, expert opinion on this matter would be valuable because governments frequently change policies and regulations, and it is not unreasonable to consider that new nuclear power stations may have been approved in the absence of a GDF.

In addition, many other important factors are likely to support the development of new nuclear power stations. This means that the development of new nuclear power stations cannot solely be attributed to the GDF programme. As such, policy documentation will be used to validate the notion that the GDF programme contributed to the development of new nuclear power stations, while expert opinion will be used to provide context relating to the importance of the GDF programme in enabling the nuclear power stations. It should be noted that it is possible that new nuclear power stations may be removed from the UK Government's future energy strategy and roadmap to net zero, which would undermine the benefit of enabling new nuclear power stations.

National security and hazard reduction:

The impacts of developing a GDF on national security is a key example of a non-quantifiable outcome, the social value of which is also difficult to assess. It may be possible to quantify the social cost of a major incident or hazard occurring in the UK if the GDF did not exist, using a simulated

[18] For example, another GMPP infrastructure project, the smart meter rollout programme, also highlights various non-quantified benefits of the project, such as helping improve competition in the domestic energy market and the value of increasing the provision of data to consumers and businesses, that it is qualitatively important to consider when evaluating the merits of the programme.

scenario analysis, but there is no guarantee that an event or incident might occur in a (counterfactual) world where the UK did not pursue developing a GDF, and so it is difficult to attribute this as a realised benefit of the programme.

Alternatively, the contributions of the GDF programme to the UK's national security could be evaluated by periodic expert assessments that compare the security threats associated with different policy scenarios (e.g. building the GDF versus maintaining above-ground storage facilities). To understand whether the expected security benefits of the programme are being realised, periodic assessments would need to update understanding about the risks of incidents or hazards occurring relative to current (ex-ante) assessments about the programme. These assessments would also need to reflect the risks caused by the security levels at the GDF and the condition of the GDF at the time of the assessment. For example, if current assessments suggest that the GDF will prevent the risk or prevalence of material discharge that may impact local communities, but following the GDF's construction incidents occur where this is not the case, then this should be reflected in the evaluation of the (un)realised benefits.

Overall, this evidence would contribute to the broader assessment and narrative about the contributions of the GDF programme to outcomes, and its importance to and value for the UK. This evidence could also be triangulated with evidence from a broader range of expert opinions and perspectives on the issue (for example, collected through in-depth qualitative interviews).

Section 6 Value for money evaluation

This section outlines the key challenges in conducting a VfM assessment for the GDF and discusses methods and approaches that are relevant for answering the evaluation questions.

A VfM assessment of the GDF would answer the following evaluation questions:

1. How well has the programme performed against the NAO's key dimensions of VfM: economy, efficiency, effectiveness, and equity (the 4 Es)?
2. For completed activities, does the value created justify the resource use? For future activities, how is VfM likely to change and evolve over the course of the programme?

Assessing the costs would require information to answer the following questions:

1. What is the financial cost of the GDF activities and the programme as a whole, including expenditure on CIF, SAI, business case development and inputs into evaluation, characterisation, design, construction, operation and closure?

2. For future activities, what are the projected whole-life costs? How have these changed since the baseline? How confident can we be that the estimates are accurate?

Key questions for assessing benefits include:

1. What cost savings or benefits have been realised as a result of the GDF activities and the programme as a whole?
2. For completed activities, does the value created justify the resource use? For future activities, what are the projected cost savings or benefits? How have these changed since the baseline? How confident can we be that these will materialise?

The recommended approach involves using the value-for-investment (i.e. King/Oxford Policy Management (OPM))[19] approach to design and deliver the assessment. This approach has been selected over the standard Greenbook and NAO approaches because it allows for the incorporation of non-monetised benefits, which the Greenbook and NAO approaches do not fully consider. The value-for-investment approach is structured around eight steps (see figure 6 below) and provides a structured, comprehensive and holistic approach to assessing the NAO's key dimensions to be considered when making a VfM judgement: economy, efficiency, effectiveness, and equity (4 Es).

As a minimum, the VfM of the GDF should be assessed at the end of the life of the facility, once all costs and benefits have been realised. Given the long timelines for all costs to be accrued and benefits realised, we would also advise conducting an interim VfM assessment at the start of GDF closure. This would capture all the costs and benefits realised so far and make forecasts for the remaining costs and benefits yet to be realised.

Given the long timeframes involved, we would also suggest carrying out VfM assessments of specific activities where there is a discrete set of benefits and associated costs. These could be conducted in the earlier stages (such as during the community engagement and site evaluation phase), to provide an evolving picture of VfM that could be used to periodically assess performance and test the assumptions of the evolving business case estimates – thus providing an early warning system and

[19] See: VFI | Julian King & Associates. Available at: https://www.julianking.co.nz/vfi/

learning opportunity. These activity-based assessments could be agreed on a rolling basis at the beginning of each phase of the evaluation, or they could be responsive to significant programme achievements.

The ability to assess the VfM of a distinct set of activities will be dependent on identifying the specific costs and benefits associated with the activities. This requires financial reporting that enables activity-based accounting. Therefore, we strongly recommend reviewing financial reporting capabilities to check that they can provide this level of disaggregation.

Challenges in assessing value for money

The GDF programme is expected to bring about a wide array of benefits alongside the main objectives of providing a permanent solution for the disposal of MHRW. These include the following:

- socio-economic benefits, such as the creation of jobs (local and national), local skills and training, improved local infrastructure, retention of young people in the host community, and overall improved socio-economic wellbeing of the host community
- financial benefits, such as reducing the need for ongoing costs associated with the storage of waste above ground, and supporting long-term economic growth
- environmental benefits, such as enabling the development of new low-carbon nuclear power stations and associated reduction in carbon emissions, and reducing energy and fuel usage associated with above-ground storage
- security benefits associated with enabling the decommissioning of nuclear assets and safer storage of radioactive waste underground

Determining the amount of money spent on delivering these objectives is a relatively simple exercise, assuming there is good financial record-keeping and activity-based accounting. Of course, whole-life costs can only be calculated at the end of the programme (around 2200, at the time of the closure of the GDF), but the full and actual costs of different activities of the project can be calculated at the end of each activity, potentially with the addition of an updated forecast for whole-life costs.

However, determining the value of the benefits produced through the financial investment is a more complex endeavour. This is because not all benefits are immediately amenable to being monetised, as would be required for a traditional cost–benefit analysis (per the HM Treasury Greenbook guidance).[20] As seen in the economic case of the GDF business case, a traditional cost–benefit analysis can only include standardised and easily monetised benefits, such a job creation and reduction in operating expenditure costs. This means that important but non-monetised benefits are excluded from the analysis, resulting in an underestimation of the total value of the programme. For instance, the value of the GDF programme's security benefits are not included in the benefit–cost ratios.

[20] HM Treasury, 2014. Supporting public service transformation: cost benefit analysis guidance for local partnerships. Available at: https://assets.publishing.service.gov.uk/media/5a7dbd4340f0b65d8b4e3357/cost_benefit_analysis_guidance_for_local_partnerships.pdf

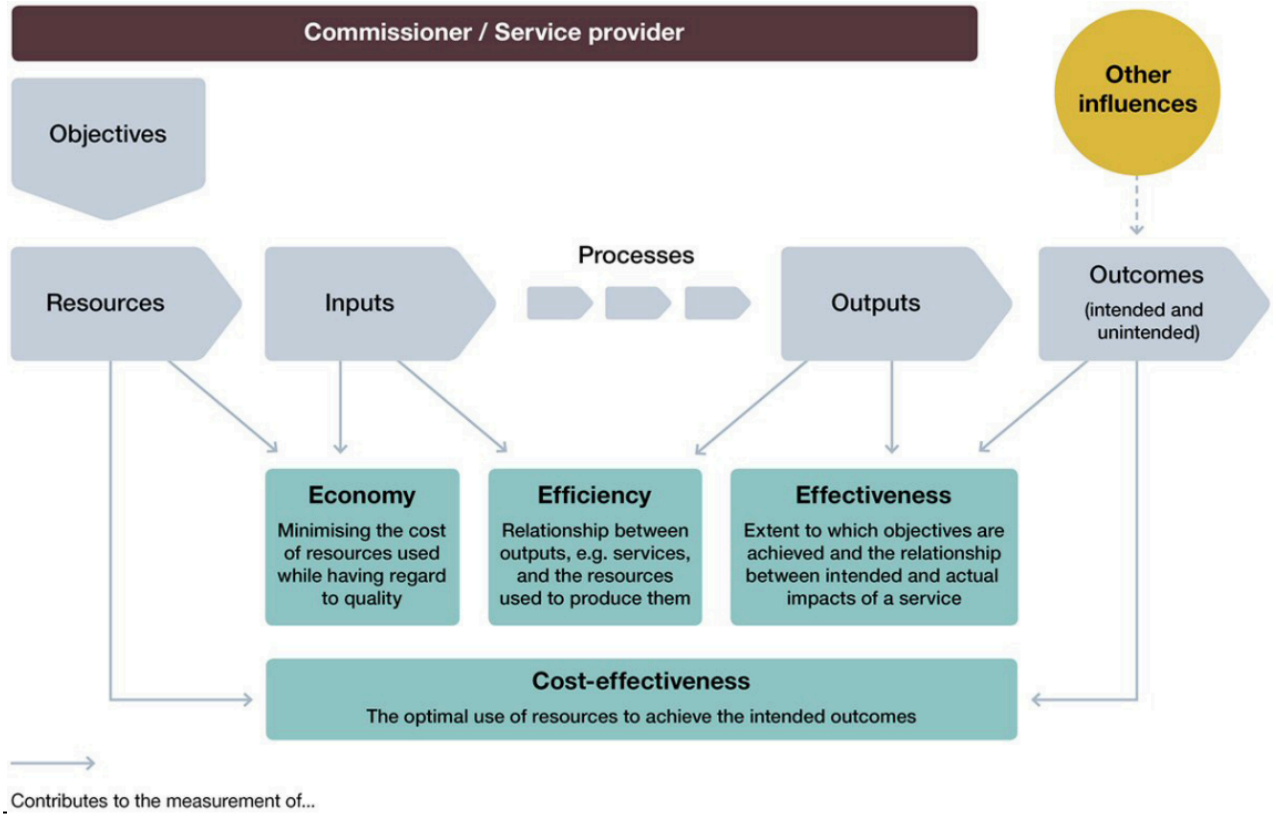
Options for assessing value for money

The Green Book[21] acknowledges this limitation of assessing VfM, and the need to consider more than just benefit–cost ratios in a judgement on VfM. However, the document does not set out a process or methodology for systematically and rigorously incorporating wider considerations into a VfM judgement.

The NAO's approach for assessing VfM provides a more encompassing framework on the dimensions of VfM. Defined as "the optimal use of resources to achieve the intended outcomes", the NAO identifies economy, efficiency, effectiveness, and equity (the 4 E's) as the key dimensions to consider when making a VfM judgement. The interaction between these dimensions of VfM are shown in Figure 3. While the dimensions provide a helpful framework for assessing VfM, this framework does not fully reflect the interactions between the different Es, which are important in order to capture the value of a complex and long programme, such as the GDF programme. The NAO approach also does not provide detailed guidance on how non-monetised impacts can be valued in a judgement of VfM.

The economy aspect is concerned with minimising the cost of resources used while having regard to quality, while the efficiency principle considers the relationship between outputs and the resources used to produce them. Effectiveness concerns the extent to which objectives are achieved, and the relationship between the intended and actual impacts of the service. Finally, the equity aspect concerns the extent to which outcomes reached all intended people. These are summarised further in below Figure 3.

Figure 3: Value for money flow diagram



[21]HM Treasury, 2022. The Green Book. Available at: https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government

Instead, there is a need to consider all the steps in the Theory of Change in a more holistic manner. A practical evaluation of VfM also requires a broader and more strategic definition of the 4 Es, including considering if and how alternative approaches compare to each other, what else could be done to improve VfM, and how VfM may evolve over time.

These limitations are overcome by King and OPM's value-for-investment approach to assessing VfM.[22] The King/OPM approach provides a structured, systematic and transparent approach to assessing the 4 Es in a comprehensive and holistic manner. It then draws together the findings for each of the 4 Es to provide pragmatic and strategic evaluative judgements on the following questions:

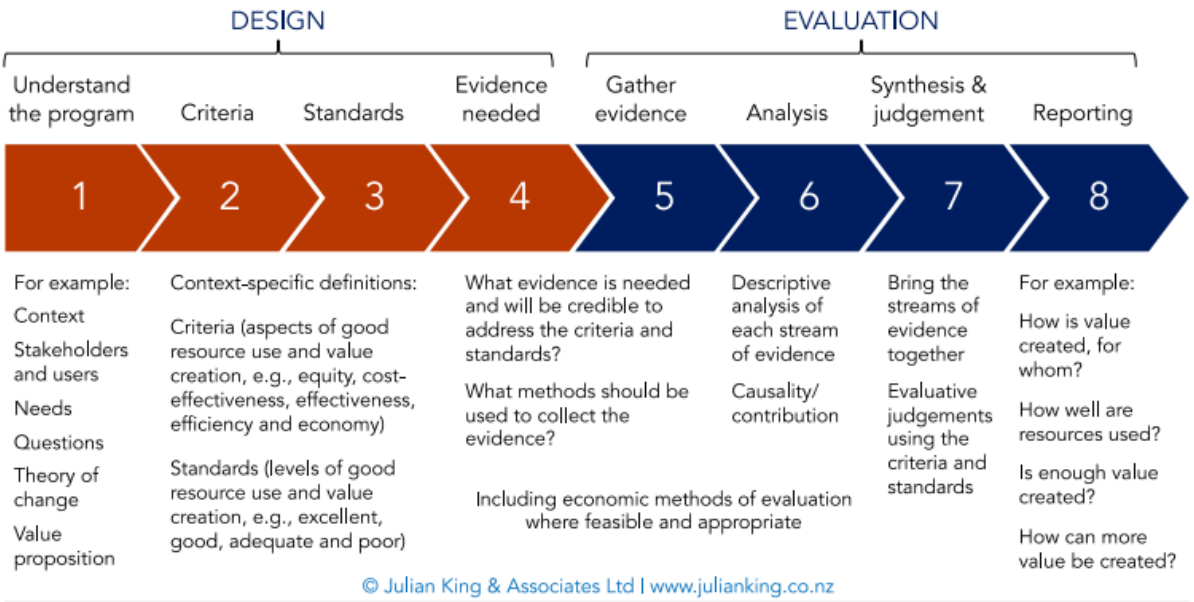
1. How well have resources been used?
2. What value has been created?
3. Does the value created justify the resource use?
4. How can the resource use be improved?

It is therefore recommended that the VfM assessment for the GDF programme adopts the King/OPM approach.

The value-for-investment approach

Figure 4 presents the value-for-investment (i.e. King/OPM) approach in designing and delivering the VfM assessment, as well as key considerations for assessing the 4 E's and ultimately coming to judgements on the VfM of the GDF programme. The King/OPM approach is structured around eight steps, as summarised below.

Figure 4: Value-for-investment approach to assessing VfM



[22] Oxford Policy Management, Assessing Value for Money: the Oxford Policy Management Approach. Available at: https://www.opml.co.uk/sites/default/files/migrated_bolt_files/opm-vfm-approach-2.pdf

Step 1: Understand the programme

This involves understanding the context of the programme, consulting with stakeholders, understanding the programme's objectives and value proposition (the benefits the programme is intended to deliver), and developing a Theory of Change. For the GDF programme, this work has already been completed through the GDF business case activities and documentation, and the Verian-led Theory of Change workshops that have been held to support the wider evaluation objectives.

Steps 2 and 3: Develop VfM criteria and standards

Criteria and standards provide an explicit and transparent basis for interpreting the evidence and making the evaluative judgements (Step 7). These criteria and standards should reflect the key elements of the Theory of Change and value proposition, and should be agreed with key stakeholders before the evaluation begins. For the GDF programme, this work should take place during the evaluation design. The value-for-investment approach refers to criteria and standards as follows:[23]

- "criteria of merit or worth are selected dimensions of performance that are relevant to a particular programme and context. They describe, at a broad level, the aspects of performance that need to be evidenced to support an evaluative judgement about VfM. When using the 4Es as the basis for VfM assessment, context-specific definitions of economy, efficiency, effectiveness, and equity need to be developed."

- "performance standards provide defined levels of VfM for each of the criteria. In our VfM assessment, we typically label these levels as excellent, good, adequate, and poor. However, in some cases, an alternative framing (e.g. stages of growth such as emerging, evolving, embedding, and excelling) may be preferred."

Criteria do not need to be specific and measurable like indicators, but they need to clearly communicate the different types – and the nature – of value that the programme will deliver. This means the VfM assessment of the GDF programme will have to develop definitions of value for each of the benefits that will be included in the assessment. While this has been done for the monetised benefits in the business case, clearer expectations will be required for the value that the GDF programme will deliver in relation to the non-monetised benefits. The CIF strategy goes some way to achieving this for the CIF activities, but further definition of the expected benefits would be helpful, and other goals/areas of the programme will require similar consideration. We would recommend this be done through a participatory workshop exercise with key stakeholders.

Once the overarching criteria have been developed, sub-criteria based on the 4E's can then be developed to aid in the assessment of specific aspects of performance.

For Economy, some example sub-criteria may include benchmarking fee rates or unit costs; ensuring all inputs were utilised and necessary; promoting competitive procurement practices; and leveraging

[23] ibid.

economies of scale or partnerships, such as with other government departments, following due diligence with regard to financial management.

For Efficiency, it may be useful to consider proportional spend on overheads and time and money spent on specific activities (e.g. rounds of submissions as a result of rejected applications). Other sub-criteria may include assessing if the programme is delivering the expected work on time, to budget, and with the required quality (technical efficiency); considering if a different mix of inputs could have been more productive (allocative efficiency); and checking if lessons have been learned and processes/practices optimised (dynamic efficiency).

To assess Effectiveness, it may be helpful to determine if the outcomes identified in the Theory of Change are being delivered according to plan. In the long term, this will be supported by evidence from the impact evaluation, which will compare the intended outcomes specified in the business case and benefits/value criteria with the results achieved. However, the process evaluation will also provide interim information on the effectiveness of early activities, such as business case and planning approvals, the siting process, and the market engagement and procurement strategy.

Measuring Equity will involve assessing whether the benefits of the GDF programme are fairly distributed and reaching the intended target groups. For example, is the siting process benefitting less well-off communities? Were all communities given the opportunity to apply and make their voice heard? Is the intended mix of local and national job creation being delivered? Is the GDF programme benefitting all members of the local community equally or do some groups benefit or lose more than others? Given the long time horizons of the GDF programme, consideration of the equity of inter-generational costs and benefits may be pertinent.

Finally, as mentioned previously, benefit–cost ratios will be inadequate to assess the cost effectiveness of the GDF programme because not all of its core objectives can be monetised (which could result in a negative return on investment). For this reason, the business case takes the position that a GDF is necessary and that a return on investment is not a primary consideration: the goal is to provide a permanent solution for the disposal of MHRW. This pivots the cost effectiveness analysis towards focusing on whether the GDF that is delivered compares favourably in terms of costs and benefits to comparators or alternative forms and options that are not applied (such as the scenarios in the economic case). Such scenario analysis relies on multiple assumptions and is fraught with uncertainties. Therefore, we suggest that the primary objective of the cost effectiveness analysis should be to determine if enough value has or will be created to justify the resource use.

Steps 4 and 5: Identify the evidence required, select methods, and gather evidence

This step involves determining what evidence is needed to address the criteria and standards, what evidence will be credible for this purpose, and what methods should be used to gather the evidence. This should be done in a structured manner for each criterion and standard to be assessed. For this reason, a VfM framework is advised.

Data would then be collected according to the VfM framework and an evaluation design developed that can establish causality or contribution (to determine if the programme caused or contributed to the observed results) and adjust for additionality concepts, such as dead weight (outcomes that would have happened without the programme), shared effects (other programmes contributed to the

results), gains though positive externalities (programme provided wider benefits than were anticipated), losses through displacement or negative externalities (results that cannot be claimed by the programme of negative effects), and sustainability (will results increase, stay the same, or reduce over time).

For the GDF programme, these data collection and design considerations will be dealt with as part of the wider evaluation framework because the VfM assessment will use the same methods and data collection activities as are employed by the other evaluation strands. Any additional information required, such as calculations of expenditure or monetisation of benefits, will ideally be available from programme and government sources, and thus will only need to be validated.

Benchmarking is an important criterion for value for money

Benchmarking of inputs provides compelling evidence on the economy of the programme (*did similar programmes spend more or less on the same resource inputs?*), while benchmarking of outputs provides important evidence on the effectiveness of the programme (*did similar programmes produce more or less benefits?*). Benchmarking of duration or time spent on activities is also an important indicator of efficiency. Combined, this information can be a powerful indicator of cost effectiveness and can inform the overall judgement as to whether the value created justifies the resource use. This is because it provides insight into whether alternative approaches or implementation would have offered better or worse VfM.

Given the unusual characteristics of the GDF programme, direct comparators may be difficult to identify. However, GDFs in other countries are also being developed and following a similar course to the UK GDF: for example, the Cigéo deep GDF to be built in France. These projects could provide useful comparators as the UK GDF progresses in parallel with others.

Additionally, when broken down by activity, suitable comparators will become evident. For instance, there are likely to be many similar investment programmes to the CIF, which could be used as comparators (for example, levelling up funds provided by DLUHC). Elements of the construction of the GDF, such as tunnelling or boring, are also likely to be comparable to other projects. And costs and outcomes for business case development, design, consulting, and approvals could also be compared: for example, in relation to time spent, fee rates, and submission acceptance rates. Some of these comparisons may need to be cautious because inputs and outputs will not all be standardised so it may not always be possible to compare ‘apples with apples’. but a similar, if not identical, comparator will provide important context for a VfM judgement. From an organisational learning perspective, benchmarking analysis can also be used to improve or test the quality of estimates which are used to forecast future costs, schedule and benefits (see below).

Cost and schedule overruns, and benefits shortfall are also important criteria for value for money

Cost and schedule overruns are important indicators of efficiency. This is because these overrun indicators provide evidence of cost and schedule control and delivery performance in terms of meeting budget and schedule expectations.

Cost and schedule overruns, and benefits shortfall, can also be used to predict the likelihood of the programme delivering to plan because they provide a leading indicator of forecasting accuracy. If a

programme systematically underestimates its cost and duration (leading to a cost or schedule overrun) or overestimates its benefits (leading to a benefits shortfall) it is unlikely to deliver the

outcomes and VfM set out in the business case. Therefore, cost and schedule overruns and benefits shortfalls can also be important indicators of effectiveness, cost effectiveness, and VfM.

Cost and schedule overruns are calculated as the actual project cost or duration divided by the estimated cost or duration. Benefits shortfalls are calculated in the same way (actual project benefits divided by the estimated project benefits) but because the cost and schedule are usually underestimated, and benefits overestimated, cost and schedule are typically presented as *overruns* and benefits are typically presented as *shortfalls*. They can be expressed as a ratio or as a percentage overrun of the estimate.

Since calculation of overruns/shortfalls relies on comparison between estimates and actuals at completion, they could only be calculated for the GDF programme once the programme is complete. However, overruns/shortfalls for earlier stages of work could be calculated as activities and phases are completed: for example, the CIF or siting process. Prior to, or alongside, this, the evolution of estimates for future activities could be used as a leading indicator for overruns/shortfall: for example, comparing a re-forecast for a future activity with its prior or original estimate to see if costs or durations have increased or benefits have decreased (for instance, due to changes in scope or unforeseen challenges).

However, when using overruns/shortfalls as indicators, it is important to consider the influence of estimating maturity. Due to the higher levels of uncertainty at early design stages, early estimates will be inherently less accurate than more mature, later stage estimates. Therefore, some level of inaccuracy at early design stages is to be expected, and it is usual for the accuracy of estimates to improve as the project design matures. For this reason, it is important to keep records of early estimates (including recording the maturity/version of the estimate) and to regularly recalculate overruns/shortfalls as the programme evolves and activities are completed. In this way, the accuracy of forecasting can be tracked over time and monitored to check if there has been improvement. If there appears to be a systematic bias towards overruns/shortfalls (which is a common pattern due to psychological factors collectively known as the planning fallacy), some level of adjustment or recalibration of estimates is likely to be necessary.

In this case, we recommend using a methodology known as reference class forecasting. This uses the actual performance of past completed projects to forecast the likelihood and severity of an overrun/shortfall for a planned project, and then establishes uplifts to estimates that correct for any systematic bias observed in the dataset. Since the GDF programme is unlikely to have completed enough projects to form the basis of a reference class, past similar projects completed elsewhere can be used for this purpose.

The advantage of this technique is that it enables the recalibration of estimates before a pattern in an organisation’s specific forecasting accuracy can be established. It also permits comparison with similar projects so that the GDF programme’s performance can be understood in the context of the significant challenge that is estimating. Estimating is notoriously difficult for any major project, and especially so at early design stages, when uncertainty is high and there are such long timelines to completion. Through comparison with projects that have faced similar challenges we can establish what constitutes a reasonable expectation for performance. Given the influence of estimate maturity

on forecasting accuracy, it is important to compare projects at a similar baselined level of maturity as the planned project: for example, full business case (higher maturity), outline business case (medium maturity), or strategic outline business case (low maturity). For instance, if GDF programme activities experience or forecast a cost overrun of 20% compared to early design stage estimates, we may find that this level of overrun actually represents good performance in comparison with past projects because the median overrun in past projects was 40%.

Steps 6, 7 and 8: Analysis, synthesis, and judgements, and reporting

This step involves analysing each stream of evidence to identify findings and results. These are then triangulated and synthesised to produce findings for each of the 4 Es. Finally, the totality of the findings is considered to reach a judgement on the VfM questions:

1. How well have resources been used?
2. What value has been created?
3. Does the value created justify the resource use?
4. How can the resource use be improved?

Judgement is based on the criteria and standards developed during the design stage, but this may continue to be a participatory process involving key stakeholders, in terms of transparently grading the findings against the criteria and standards. Reporting should be structured according to the VfM criteria, with each criterion addressed systematically in turn. The report should provide a compelling performance story that considers both monetised and non-monetised benefits and presents them within the relevant context – ideally with examples, benchmarks, or comparators to justify judgements and contextualise performance.

For the GDF programme, this will involve assembling the results according to the evaluation framework and presenting findings for each of the 4 Es. We then propose that a workshop be held to explain and justify the findings and then map them against the criteria and standards and reach conclusions that have the buy-in of all stakeholders. Although this process will involve parties that are not independent, the process will be overseen by the independent evaluators and graded against previously agreed criteria. Therefore, it can be considered robust and transparent, which guards against subjectivity and bias.

The report will have to consider both monetised and non-monetised benefits in relation to the costs, including whole-life costs, and will need to be creative in its comparisons with other programmes, given the unusual nature of the GDF programme.

Options for the periodicity of assessment

As mentioned previously, the overarching purpose of the VfM assessment is to provide answers to the four key VfM evaluation questions, and to draw an overall conclusion as to whether the GDF programme has provided acceptable VfM.

At a minimum, the VfM of the GDF programme should be assessed at the end of its life, once all costs and benefits have been realised. Since some benefits may not be realised until after the GDF closes, it may be advisable to delay the VfM assessment until the measurement of all benefits is possible. However, this could significantly delay the results of the VfM assessment and make the

findings less useful. Therefore, we would advise conducting an additional interim VfM assessment at the start of GDF closure. This would capture all the realised costs and benefits so far, and would provide forecasts for the remaining costs and benefits yet to be realised.

Given the long timeframes involved, we would also suggest carrying out VfM assessments of specific activities where there is a discrete set of benefits and associated costs. These could be conducted in the earlier stages (such as the community engagement and site evaluation phase), which would provide an evolving picture of VfM that could be used to periodically assess performance and test the assumptions of the evolving business case estimates – thus providing an early warning system and learning opportunity.

It is advised that the assessment of a specific set of activities takes place at the end of each phase of work, because some activities and their costs and benefits will roll across phases and not be realised until after the phase is complete. These activity-based assessments could be agreed on a rolling basis at the beginning of each phase of the evaluation, or they could be responsive to significant programme achievements. However, the conclusion of the following activities is likely to be a useful point to conduct VfM assessments once their benefits have been realised:

- completion of business case and approvals
- completion of the CIF investment
- completion of siting process
- completion of initial construction
- completion of commissioning and first operations/emplacement
- end of emplacement and significant additional investment
- closure of the GDF

The ability to assess the VfM of a distinct set of activities will be dependent on identifying the specific costs and benefits associated with the activity: for instance, isolating the costs and benefits associated with delivering the CIF/SAI, while excluding the costs and benefits associated with other areas, such as GDF construction. This means that sound financial systems will be needed to capture and isolate the specific costs associated with a set of activities (rather than using total current programme spend). As such, we strongly recommend reviewing financial reporting capabilities to check that they can provide this level of disaggregation.

Section 7 Reflections and next steps

This section provides some reflections on the process of scoping an evaluation for a mega-infrastructure project, and sets out the next steps for an evaluation of the GDF programme.

Scoping a mega-infrastructure project evaluation

Firstly, we present some learning about the scoping process itself, in terms of how to facilitate a streamlined approach and a quality output:

- ensure all evaluation partner organisations are covered within any non-disclosure agreement, to facilitate knowledge sharing
- commit time early on to map key stakeholders, identify their role in project governance, ensure their attendance at workshops, and agree a clear timetable for getting their feedback on outputs, including, most critically, the options paper/workshop
- plan how to engage policy and operations stakeholders (who often have limited understanding of evaluation and its value) in scoping activities, to make the most of their time and expertise and to foster a sense of collaboration and ownership in the evaluation design
- ensure comprehensive access to programme documentation from the outset, as delayed access to key documents (e.g. local delivery plans, internal data collection processes) will have an impact on the level of detail that can be achieved in scoping the evaluation methods and integrating the proposed approach with existing programme activities

The process of scoping an evaluation for the GDF programme has revealed a number of key considerations that may be useful for other major infrastructure projects. These include:

- phasing an evaluation in a manner consistent with implementation phases or critical decision points, in order to manage longer-term evaluations
- using theory-based approaches that integrate quantitative impact methods in order to understand the level of, and reason for, the impact of projects, particularly those being run in complex environments.
- considering how best to specify monitoring data requirements in order to satisfy the needs of both evaluation activities, business cases, and benefits realisation processes
- ensuring that policies and processes for data sharing (for example, data sharing agreements, privacy policies etc.) are built in from the outset
- involving evaluation partners as early as possible to ensure that measures are in place to ensure robust learning is gathered from the earliest opportunity (for example, collecting appropriate baseline data, conducting detailed stakeholder mapping, and identifying existing data sources and seeking opportunities to reduce stakeholder burden)

Next steps

The next steps in the GDF programme evaluation for NWS are to gain senior buy-in for the evaluation, to confirm the scope of the intended evaluation, to identify the resource requirement, and to secure an evaluation partner.

Gaining senior stakeholder buy-in

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 the programme. It is expected that the Evaluation Task Force can offer reassurance to such stakeholders on the significance and value of robust evaluation for major projects.

Confirming the scope

The GDF programme timeline is long, and it is unlikely that any evaluation could/would cover the full 100+ year period. As such, NWS needs to consider which / how many of the recommended phases and forms of evaluation it wishes to take forward.

From a timeline perspective, it seems sensible to conduct evaluation Phases 1 and 2 together (i.e. site evaluation and community engagement, and site characterisation and Test of Public Support), given that (a) these are the earliest phases of activity, and (b) so much of the implementation activity covers the same period. If this is to be contracted out to an evaluation partner, there should be a proviso that continuation into Phase 2 is contingent on the successful meeting of agreed standards/key performance indicators during Phase 1.

This evaluation plan suggests conducting three forms of evaluation together – IPE, impact evaluation, and VfM assessment – to gain a holistic picture of GDF programme delivery, impact and value, to maximise efficiencies, and ensure triangulation between evaluation strands. However, NWS may wish to start with the IPE strand, to work on specification of programme monitoring data, combined with learning and good practice in community-engagement techniques (for example). Once the detailed plans for CIF projects are made, the impact component could come online, with some further scoping to pin down expected outcomes and metrics, and the expected counterfactual approach.

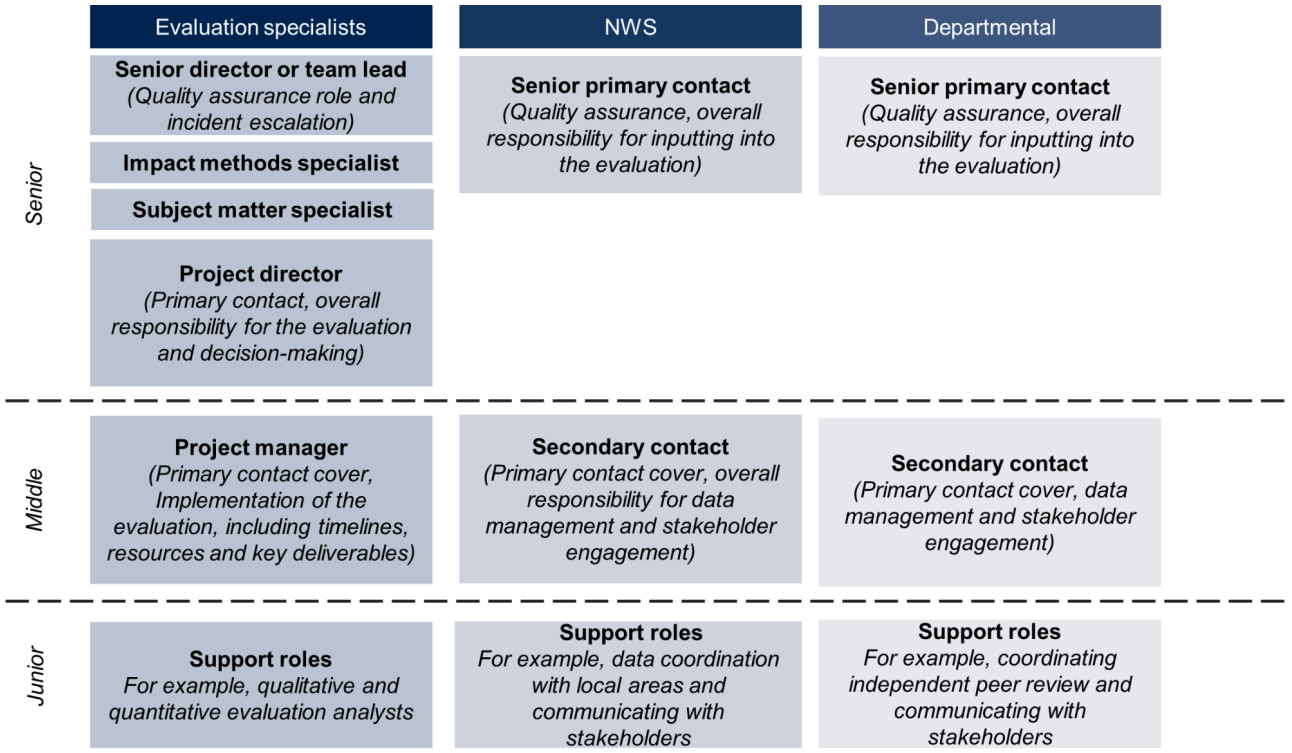
Identifying the resource requirement

An evaluation will require budgetary and staff resources. In terms of staff resources, NWS should assess their internal skills and capacity for evaluation, to identify which / if any of the recommended evaluation activities may be conducted in-house and which staff could do these activities and / or manage an external evaluation partner.

It is also recommended that NWS seek out evaluation expertise within the sponsor department or Evaluation Task Force, to support in-house evaluation work, with commissioning an external evaluation, and providing ongoing evaluation support and quality assurance. Figure 7 sets out the kind of staffing required to manage a large-scale, complex evaluation.

In relation to the budget requirement for an external evaluation, this will depend on the scope of the commission (as discussed above). Once the timeline and evaluation scale are established, it will be sensible to benchmark the potential budget against published tenders of other infrastructure projects. Few projects will be directly comparable in scale and scope overall but there will be similarities in scale and scope for individual GDF programme evaluation phases.

Figure 5: Example staffing structure for an evaluation



It would also be sensible to conduct market testing to refine the requirement and gain an understanding of what resources (i.e. budget) would be needed to fulfil a contract. It is standard practice to send out an expression of interest form to organisations using the Government's Dynamic Purchasing System framework, exploring the feasibility of the suggested evaluation approach, the timelines, and the budget envelope, as well as testing the level of interest and expertise amongst responding organisations.

Commissioning an evaluation partner

Given the expected scale of the GDF programme evaluation, and given that the programme does not have an in-house evaluation team, an external expert evaluation partner should be procured, via a competitive tendering process. It is recommended to conduct a supplier engagement event well in advance (around six months ahead) of an invitation to tender being released, to stimulate supplier interest and refine the statement of requirements.

The NWS may want to encourage responses from consortia to ensure that the evaluation partner is suitably equipped with methodological and sector-specific knowledge. Supplier engagement events are a useful opportunity for suppliers to identify potential partners. Contact details can be shared (with permission from attendees) for that purpose.

This evaluation plan can be used to shape the statement of requirements or specification that accompanies the invitation to tender and the plan/or relevant excerpts can be shared as part of the tender process. The response deadline should be at least four weeks after the tender publication to enable suppliers to fully understand the brief and account for its complexity in their response.