Department for Environment Food & Rural Affairs

Complexity Evaluation Framework

Recognising Complexity & Key Considerations for Complexity-Appropriate Evaluation in the Department for Environment, Food and Rural Affairs (Defra)

Version 2.0

Commissioned by the Department for Environment, Food and Rural Affairs (Defra)

Version 1.0 delivered by CECAN Ltd with Risk Solutions and the Tavistock Institute of Human Relations.

Version 2.0 updated by Steer Economic Development.

CECAN Ltd

THE TAVISTOCK INSTITUTE

risksolutions
leave nothing to chance

steer economic development

This publication is available on Defra Science Search: http://randd.defra.gov.uk/

Any enquiries regarding this publication should be sent to us at:
evaluation@defra.gov.uk

www.gov.uk/defra

TABLE OF CONTENTS

0 ABOUT THE COMPLEXITY EVALUATION FRAMEWORK

0.1 HOW TO USE THIS DOCUMENT

This document is divided into two parts:

- *Part 1: What is complexity and what does it mean for Defra?* (Chapter 1) explores what we mean when we talk about complexity, makes the case for why evaluation is crucial when working with complex systems, and introduces the core principles of complexity-appropriate evaluation.
- *Part 2: The Complexity Evaluation Framework* (Chapters 2 to 5) sets out a practical framework of considerations and guidance for those designing, managing and embedding evaluations.

The document is intended to be used as a reference guide. The reader is encouraged to begin with Part 1, and then to concentrate on the sections of Part 2 most relevant to the evaluation element they are working on. A set of user case studies are provided in Annex IV which describe how others have used the CEF to work with complexity-appropriate evaluations, at various stages of development.

Hyperlinks throughout this document give signposts to useful additional resources. To get the most out of the document, reading on-screen/digitally is recommended. A printable 'quick-reference guide' is also available in A3 poster format.

0.2 WHAT IS THE COMPLEXITY EVALUATION FRAMEWORK?

The Complexity Evaluation Framework (CEF) is designed to assist policy makers and analysts to design and deliver effective evaluations under complex circumstances.

A system or process that is *complex* is made up of many diverse components that interact in nonlinear ways and may also adapt or change over time. This can lead to unpredictable behaviour and unexpected outcomes. The domains that Defra deals with are complex and can involve working with complex environmental and social/economic systems, often at the same time.

The purpose of the CEF is to guide the scoping, commissioning, management and delivery of evaluations in the presence of complexity. It provides guidance on complexity characteristics and a framework of considerations to inform conversations between policy leads, commissioners of evaluation, and evaluation contractors. Its aim is to ensure that complexity is recognised and complexity thinking is embedded into evaluation design and delivery to ensure approaches are robust. The CEF is intended to support evaluation across

the policy cycle, from as early in the policy cycle as possible. It could also be used outside of evaluation, including in policy scoping and option appraisal.

This is the second version of the Complexity Evaluation Framework. It was originally developed for Defra by the Centre for Evaluating Complexity Across the Nexus (CECAN), through the engagement of Defra staff and use of academic literature. The first published version was evaluated by an independent evaluator, Steer Economic Development (Steer-ED), who made recommendations to help maximise the value of the CEF and collected case studies of early uses of the CEF (see Annex IV). This second version reflects changes made in response to those recommendations. Full details of Steer-ED's evaluation can be found in their published Evaluation report.

0.3 WHO IS IT FOR?

The CEF has been designed with a number of different users in mind:

- Analysts and policy makers in Defra who are involved in planning, commissioning, managing and/or delivering evaluation;
- Commissioned researchers and evaluators undertaking evaluation for Defra; and
- More broadly, for evaluators in environmental and non-environmental spheres, who are considering the practicalities of evaluating policies and interventions subject to complexity.

0.4 WHY EVALUATE?

Evaluation is an integral part of the government policy making ROAMEF cycle (which stands for Rationale, Objectives, Appraisal, Monitoring, Evaluation, and Feedback), and is formalised in the HM Treasury *Magenta Book* and *Green Book* Evaluation. When performed well, evaluation can give the policy maker valuable, objective, insights as to:

- The **impact or effectiveness** of a policy, including both anticipated and unanticipated effects, providing or enhancing the evidence-base that supports policy decisions;
- An understanding of **how** the intervention worked;
- How successfully it was **delivered;**
- Whether it generated **value for money;**
- Potential **improvements** (in particular if the policy has not performed as expected), and how the effectiveness of the policy could be **maintained over time**;
- An understanding of how **transferable** the evaluation results might be over context, place and time; and
- The **accountability** of the delivering bodies, helping to satisfy requirements for external scrutiny and certain regulatory requirements.

0.5 EVALUATION AS A TOOL FOR NAVIGATING COMPLEXITY

When seeking to influence or manage complex systems, evaluation is crucial, helping to understand and navigate this complexity. In the presence of complexity, a good evaluation, carefully planned and managed, can help policy makers to:

- Understand and navigate the systems which they are seeking to influence;
- Understand the challenges posed by complexity;
- Anticipate and take steps to manage these challenges.

Evaluation in the presence of complexity can be difficult, given the multiple considerations, layers and unexpected possibilities. The purpose of this document is to assist policy makers to successfully navigate these challenges.

0.6 ADDITIONAL RESOURCES AVAILABLE

The framework is intended to be used in conjunction with other relevant guidance, including the HM Treasury *Magenta Book*, the key government guidance document on best practice evaluation, and its supplementary guide *Handling Complexity in Policy Evaluation*. The CEF also suggests a range of additional tools and resources that the reader may refer to for further or technical information.

1 PART 1: WHAT IS COMPLEXITY AND WHAT DOES IT MEAN FOR DEFRA?

> ***"The world is complex. Despite our best efforts, we cannot always accurately predict what will happen when we implement policies. As individuals and organisations interact with one another and the policy, unpredictable things can happen. So we will ensure we learn, genuinely and openly, about the effects we have had, and adapt our programmes accordingly."***
>
> Defra and the Environment Agency (2018) *Our waste, our resources: a strategy for England*

1.1 WHAT IS COMPLEXITY?

The terms *complexity* and *complex* are used to describe certain properties and behaviours of the world around us. A system or process that is complex is made up of many diverse components (e.g. people or organisms) that interact with each other in nonlinear[1] ways (i.e. where changes in outputs are not proportional to changes in inputs). Their behaviour may also adapt or change over time. This can lead to unpredictable behaviour and unexpected outcomes.

This differs from how the word *complex* is used in everyday conversation, where it is often used to mean 'difficult' or 'complicated' instead.

Example

Evaluating policy for air pollution is complex: not only are there multiple causal factors to consider, but there are many actors and interventions delivering the policy.

"[it] consists of a lot of moving parts: different areas are taking forward several different types of measures; that makes it quite difficult to compare like with like across different areas."[2] – Interviewee

[1] For a more detailed description see *Non-linearity* on page 43.

[2] For specific considerations and resources to help manage this complexity in evaluation, see Chapters 2-5.

1.2 SOURCES OF COMPLEXITY

In terms of policy-making and policy evaluation, complexity can arise at any one or more of multiple different levels. These include: the system in which the policy takes place (e.g. a complex socio-ecological system); and the policy itself (multiple actors, multiple actions).

For the evaluation, complexity can arise from any combination of the above sources and/or from the multiple and diverse stakeholder aims and perspectives involved.

When designing evaluations, and seeking to determine whether complexity-appropriate approaches should be applied, it is useful to identify potential sources of complexity in both the system and the intervention:

- A **simple intervention** could have a single objective, a clearly defined target group and relatively few delivery activities, delivered by one agent, within a relatively short timescale. A **simple system** will likely be highly predictable, easily controlled, and with few actors affecting the system.
- A **complicated intervention** could be an intervention with multiple objectives, a programme of activities over different time periods, delivered by a number of different agents or require agents with specific expertise. A **complicated system** can have a large number of different actors, forces or relationships affecting how the system operates, but will be possible to predict with some confidence.
- A **complex intervention** and **complex system** may demonstrate complexity characteristics, such as non-linearity, adaptation and emergence. These characteristics are described in more detail below.

In practice, few policy evaluations may be categorised as ‘simple’, even where there is a single ‘simple’ intervention, because the environment in which the intervention operates usually forms part of a wider complex system.

1.3 RECOGNISING COMPLEXITY IN DEFRA

The domains that Defra deals with are complex. As the UK government department responsible for safeguarding our natural environment, supporting our food and farming industry, and sustaining a thriving rural economy, Defra’s remit involves working with complex environmental *and* social/economic systems, usually together at the same time.

The content and types of policies Defra delivers mean policy design, implementation and evaluation can be challenging.

Figure 1: Challenges facing evaluating complex policy interventions in Defra (Boyd, 2015).

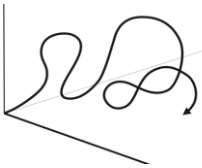
Defra's policies are multi-faceted	Defra's contribution is part of an interconnected system of impacts and governance	Defra's policy domains are associated with long timescales and unpredictability

• Defra's policies include:
  • multiple interventions
  • delivering multiple outputs
  • to produce multiple impacts
  • for multiple beneficiaries
• Impacts are diverse: environmental, social, and economic
• Multi-disciplinary knowledge and skills are involved
	• The following make it particularly challenging to isolate Defra's contribution:
	  • localised policy delivery
	  • multiple interventions and interveners
	  • global systems
		• Timescales (e.g. for certain ecological changes, such as wetland restoration) are very long, in contrast to political and economic cycles
		• Working with complex systems involves unpredictable, non-linear and emergent impacts

To help review the sources of complexity which may be present in the interventions or systems being evaluated, the section below provides a selection of characteristics and behaviours that complex systems in Defra's remit might exhibit. These apply to both environmental and social/economic systems – and the systems that Defra encounters in setting policies are often a combination of the two.

Annex II: A visual guide to understanding complexity for Defra presents these seven characteristics in more detail, plus an additional six that may also be of interest[3].

Note that in some cases there may well be overlap between the definitions, and/or a system could be described by multiple terms presented. The aim of this guide is to support readers in considering complexity within policy areas; and to provide suitable language for describing the complexity of the system they are working with.



Non-linearity: when the effect of inputs on outputs are not proportional. Outputs may change exponentially, or even change direction (e.g. after increasing for some time, they may begin decreasing), despite small or consistent changes in inputs.

E.g.: increasing payment rates for land management does not translate into a corresponding increase in their uptake. Land managers do not behave as the rational agents of traditional economic theory – there are other factors at play.

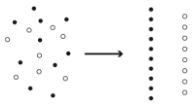
[3] These additional six characteristics are: *domains of stability*, an *open system*, *distributed control*, *levers and hubs*, *nested systems*, and *multiple scales and levels*.

E.g.: a new product may be slow to take-off, but after a certain point sales accelerate, before slowing again as the market is saturated.

Feedback: when a result or output of a process influences the input either directly or indirectly. Feedback can accelerate or suppress change.

E.g.: positive feedback leads to accelerating change. For example, as the climate changes, permafrost melts and releases more greenhouse gases, contributing further to climate change.

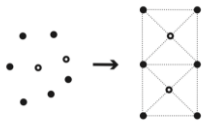
E.g.: negative feedback creates stability in systems. For example a thermostat, or the human body shivering/sweating, maintains a constant temperature.



Self-organisation: higher-level patterns can arise from the local interaction of autonomous lower-level components.

E.g.: sheep paths – these informal paths across land have no architect; they are formed by erosion caused by the footfall of individuals over time. Patterns of paths develop as each individual chooses their own route.

E.g.: multiple individuals locally clearing non-crop species leading to large-scale habitat fragmentation.



Emergence: new, unexpected higher-level properties can arise from the interaction of components. These properties are said to be emergent if they cannot easily be described, explained, or predicted from the properties of the lower level components.

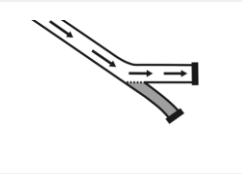
E.g.: community resilience – a community's capacity to function in and respond to shocks and extreme events – is an example of emergence; it is shaped by and arises from interactions between human and environmental components.

Tipping points: the point beyond which system outcomes change dramatically. The threshold is the point beyond which system behaviour changes; from where it may be difficult to return to the previous system state.

E.g.: a species' population reducing in numbers to such an extent that it cannot re-establish itself in the wild.

E.g.: building new relationships between industry partners can lead to sudden step-changes in collaboration or knowledge sharing, developing grounds for innovation.

Path dependency: Current and future states, actions, or decisions depend on the sequence of states, actions, or decisions that preceded them – namely their (typically temporal) path.



E.g.: the organisation chosen to lead a new policy initiative influences which other organisations also become involved; similarly, species which colonise a habitat first have 'founder effects', determining the ultimate composition of the community.

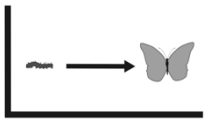
Adaptation: Components or actors within the system are capable of learning or evolving, changing how the system behaves in response to interventions as they are applied. So, for example, in social systems people may communicate, interpret and behave strategically to anticipate future situations. In biological systems, species will evolve in response to change.

E.g.: when bacteria evolve to become resistant to antibiotics

E.g.: when an individual or organisation finds a way to circumvent a new tax or regulation. For example in response to chemical safety regulations, an organisation might look for 'loopholes' by designing new chemical compositions that are not covered by the regulatory requirements.

These characteristics can lead to unpredictable behaviour and unexpected outcomes in response to planned policy and delivery. In particular, when dealing with complex systems, there will be:

Unknowns: Because of a complex system's nonlinear causal structure and the number of interactions between its components as well as with the system's wider context, there are likely to be many factors which influence (or have the potential to influence) a system of which we are not aware. The inevitable existence of such unknowns mean we often see unexpected indirect effects of our interventions.



Change over time: Complex systems inevitably develop and change their behaviour over time due to their interconnectedness and adaption. For example, ecosystems undergo succession over time, i.e. the types of plants that occupy a given area change over time (e.g. from annual plants, to scrub, to woodland). Similarly, social norms evolve over time.

Unpredictability: For all practical purposes, complex systems are fundamentally unpredictable. The number and interaction of inputs, causes, mechanisms and feedbacks mean it is not possible to accurately forecast complex system behaviour with precision. Random 'noise' can have a large effect.

1.4 COMPLEXITY-APPROPRIATE EVALUATION IN DEFRA: TOWARDS A NEW FRAMEWORK FOR WORKING WITH COMPLEXITY

> ***"Like any other experimental system, policy uses trial and error; or learning by experience; or convergence of the truth about 'what works'; or adaptive management as the process by which we move forward, and evaluation is right at the centre of this. Evaluation also has parallels with business processes — like 'lean' and continuous improvement. Ideally, policy should be operationalised using the scientific method: — (theory), hypothesis, experiment, evaluate outputs against expectation, re-formulate hypothesis.***
>
> ***"However, with complex, multidimensional, non-linear problems, like we find in real life, evaluation becomes much more difficult. Not only are outcomes very uncertain, but the definition of what we are trying to achieve can be uncertain: policies are more than a linear relationship between a problem and its solution — instead they are often part of the continuous management of the intractable."***
>
> Ian Boyd, Defra Chief Scientific Advisor, 2018.[4]

Defra recognises the importance of policy evaluation to monitor the impact of policy on people, industry, other organisations, and the environment. Defra's commitment to evaluation is identified in the *Supplementary evidence report* of the *25 Year Environment Plan* which highlights the need for robust evidence to accurately estimate the impact of policies and sets out proposals for a new monitoring and evaluation framework for the Plan.

The need for and importance of evaluation in Defra is further amplified by the complexity of the systems that Defra works with – as described above – and further still where rapid policy generation is required, for example in delivering new policies following EU Exit.

However, evaluation also needs to be planned and managed in a way that is appropriate for this complexity. Because complex systems are particularly susceptible to unpredictable change, policy teams and other decision makers and stakeholders may find it beneficial to embed evaluative activity into the policy cycle at regular and more frequent intervals. As such, policy teams may need to involve evaluators more throughout the lifecycle of a given policy or policies.

---

[4] Boyd, I., (2018). Policy, evaluation and implementation, in *The Evaluator*, Autumn 2018, pp 6-7, UK Evaluation Society.

Complexity-appropriate evaluation also requires an iterative approach. The core elements of an evaluation are typically described as a set of 'stages', conventionally expressed in a linear way, for example: defining what is to be evaluated; scoping, designing and conducting the evaluation; and synthesising and disseminating findings. When working with complexity, since the subject of the evaluation is susceptible to unpredictable change, it is important to revisit and update both one's understanding of the system and the design of the evaluation regularly.

Key point

Complexity-appropriate evaluation is iterative and embedded throughout the policy cycle.

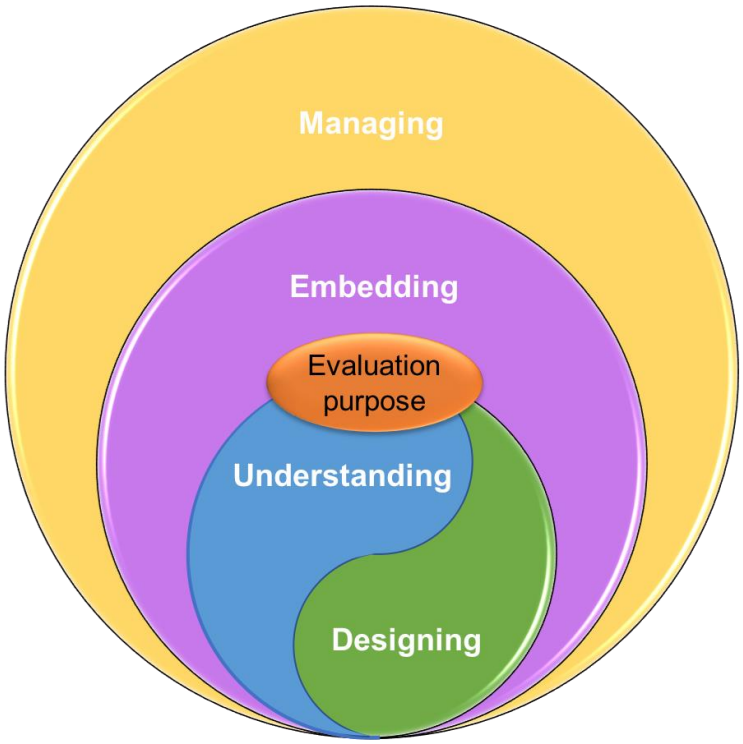


Figure 2: Nested components of a complexity-appropriate evaluation

In this framework, evaluation is described instead as a process of nested components (see Figure 2). The evaluation is centred around and defined by the evaluation purpose[5]. This

[5] For example: learning (how do I make this work better?); accountability (how well did it work?); and improving the knowledge base (how can I make similar policies work better in future?)

itself may be subject to some degree of change over the course of the evaluation, for example as understanding increases or as stakeholders and their objectives change. The evaluation purpose informs an iterative process of understanding the system and intervention (**Understanding**) and adapting the evaluation design (**Designing**). These will both continue to develop and be updated throughout the evaluation (for example, as an intervention is evaluated, more will be understood about the intervention and any new changes in its context, and therefore how best it can be evaluated). All of these activities are conducted with the ongoing engagement of stakeholders, and understanding and learning are fed back and embedded into relevant processes both inside and outside of the evaluation (**Embedding**). Finally, all of these interacting components of an evaluation are led and managed by one or more individuals (**Managing**).

1.5 THE STRUCTURE OF THE REST OF THIS FRAMEWORK

In the following sections, this framework focuses on two sources of complexity which cause challenges for evaluation in Defra:

1. **Complexity in Defra's policy domains**: this complexity may manifest itself as a policy target (system, process or outcome) that is hard to control or manage, such as biodiversity or water quality. It may also arise from or be exacerbated by the characteristics of the policy itself (e.g. multiple actors; multiple actions) and the context in which it takes place (e.g. in a large scale complex social and/or ecological system, nationally and internationally, and/or one of several interventions). Even a simple intervention *may* require a complexity-appropriate approach to evaluation due to the complexity of the system within which it is being introduced.
2. **Complexity arising from the involvement of multiple and diverse stakeholders in the evaluation**, including from their multiple and diverse aims and perspectives.

We explore the implications of these complexity issues for the inner three nested elements of an evaluation set out above in Figure 2 (page 12):

A. **Understanding**
B. **Designing**
C. **Embedding**

Considerations about how each of these relate to the **evaluation purpose** are integrated within each of these chapters.

This framework is written for commissioners of evaluations. Therefore, the final element of evaluation – **Managing** – is the overarching perspective from which the issues in this framework are explored. As such it has no chapter of its own; considerations for managing

complexity-appropriate evaluations are embedded throughout each of the **Understanding**, **Designing** and **Embedding** chapters.

In addition, Annex IV contains a collection of **case studies** which show examples of how early users of the CEF have integrated the recommendations from this document into management of an evaluation.

2 PART 2: THE COMPLEXITY EVALUATION FRAMEWORK: AN OVERVIEW

The following chapters set out a framework of considerations to ensure that complexity thinking is embedded into evaluation design, with examples, suggestions, and pointers to useful resources. Together, these chapters are for those who recognise or anticipate aspects of complexity in their evaluation work and want a formal framework of considerations for how to approach the management and evaluation of complex systems going forward.

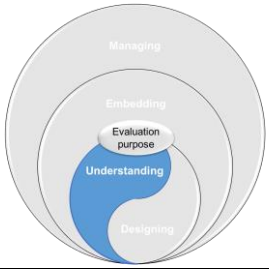
The table below is an abridged summary highlighting some of these key issues and questions. It provides users with a rapid overview of the Complexity Evaluation Framework and can be used to point the reader to specific chapters for further information and resources.

These questions are also included on the A3 summary version of the CEF (available as a separate file) and in *Annex I: Key Issues and Questions*, which can be printed and used as a quick reference guide.

MANAGING		
	This framework is written predominantly for those commissioning or managing evaluations. Therefore, the **Managing** element of an evaluation is the overarching perspective from which the issues in this framework are explored. As such it has no chapter of its own; considerations for managing complexity-appropriate evaluations are embedded throughout each of the **Understanding**, **Designing** and **Embedding** chapters.	
UNDERSTANDING		
	Key issues	**Questions to ask**
Complexity in Defra's policy domains	❖ Lack of knowledge of the issues and complexity challenges ❖ Potential for 'emergent' and unexpected outcomes ❖ New understanding of the system will come to light ❖ Need for regular review of the policy and its evaluation ❖ Need for rapid feedback to understand what is going on 'on the ground'	❑ Have you undertaken a mapping of the system, the policy and its delivery? ❑ What characteristics of a complex system do you recognise – in the policy or its context? ❑ How might these influence the way the policy is delivered or its outcomes? ❑ Have stages for review and feedback been built into policy design, implementation and evaluation plans?

Multiple & diverse stakeholders	❖ Different stakeholders may be able to contribute different information for the planning/design process ❖ Stakeholders may not have the same understanding of the system or agree on the best approach to the evaluation	❑ Have you identified the key stakeholder groups and communities affected by this policy and its evaluation? ❑ Have you actively involved stakeholders in the policy and evaluation design? ❑ To what extent is there agreement and lack of agreement about the policy itself, its outcomes or its evaluation?
DESIGNING		
	Key issues	**Questions to ask**
Complexity in Defra's policy domains	❖ Many evaluation designs can work well with complexity. However, there is no one-size fits-all design; the choice of evaluation design will depend on the complexity characteristics of the system, evaluation purpose and feasibility considerations ❖ The mix of approaches and methods selected may need to be adapted to changing circumstances ❖ Care must be taken in the choice of methods – methods that offer a high degree of certainty in straightforward contexts may mislead when applied to complex systems ❖ The evaluation design and plans need regular updating to address unexpected changes in policy and context ❖ Effective evaluation requires accurate, timely and relevant data – this is particularly important for complexity-aware evaluations ❖ Proportionality and affordability are key considerations when designing an evaluation, along with method and timescales	❑ Have you taken into account the complexity characteristics of the system in addition to the evaluation purpose and feasibility (available budget, skills and experience, timescales and data requirements) when selecting the overall evaluation approach? ❑ Are you clear about why your chosen approach is appropriate and what the limitations are? ❑ Has flexibility to review and change the evaluation design been built into the evaluation plan? ❑ Have you engaged stakeholders in the evaluation design? ❑ Is the proposed approach affordable? If not, what adjustments can be made, and what are the implications of these adjustments for the evaluation outputs?
Multiple & diverse stakeholders	❖ Different stakeholders may have different views about how the system should work, and how it is working ❖ Expectations of what the evaluation – and what different evaluation methods – can achieve need careful management ❖ Complexity-aware evaluations may need to adopt an iterative approach, while procurement rules may require external evaluation contractors to submit and deliver a fixed programme of work	❑ Have participative evaluation approaches and methods been considered? ❑ Have stakeholders committed to give the necessary time to the evaluation? ❑ Have stakeholders been primed to anticipate uncertainty in findings? ❑ Has flexibility been built into the commissioning of external evaluators, for example by using built-in decision points or contract options?

EMBEDDING		
	Key issues	**Questions to ask**
Complexity in Defra's policy domains	❖ When undertaking an evaluation within a complex system, results may be indicative rather than definitive ❖ Evaluation in a complex environment may only provide a snapshot; change may continue after an evaluation comes to an end ❖ Findings may not be transferrable due the specifics of a complex policy environment	❑ Have you talked about complexity with the potential audience(s) for the findings to manage expectations and identify the value evaluation can provide? ❑ Have difficulties in generating definitive and generalisable findings been discussed? ❑ Are there opportunities to feed findings back regularly to support implementation? ❑ Is the evaluation timed appropriately to take ongoing change into account?
Multiple & diverse stakeholders	❖ Given the complexity of Defra's policy domains, there is likely to be a diverse audience for findings who may want to use them in different ways ❖ Different stakeholders have different levels of satisfaction with provisional and indicative findings	❑ Have you considered multiple routes of dissemination? ❑ Does your plan include regular opportunities for discussing early findings?



3 UNDERSTANDING

This section explores the implications of complexity for **understanding**. This is the element of an evaluation concerned with understanding the intervention being evaluated and the system it operates within. In a complexity-appropriate evaluation, understanding is closely interwoven with the designing element (addressed in section 4). Both will continue to develop and be updated over the course of the evaluation as far as practicable, with understanding informing the evaluation design and vice versa.

3.1 COMPLEXITY IN DEFRA'S POLICY DOMAINS: ISSUES FOR UNDERSTANDING

- Knowledge about how a given complex system works may never be complete. There may be, for example, a substantial body of knowledge about certain aspects of the system, but a lack of information and understanding about other aspects, how different sub-components of systems interrelate, how to set the system boundaries, and what and who the system actually includes.
- Multiple interacting factors may be influencing outcomes; gathering data on all of these might be expensive and time consuming, or data may be absent or hard to find.
- These dynamics also underlie the potential for a system to transition across a threshold and experience a tipping point or regime shift to a new system state (e.g. lake eutrophication; coral bleaching; ocean acidification; pasture to scrub).
- There is potential for emergent[6] and unexpected outcomes. The level and type of change taking place might be unpredictable.

Example

Waste crime is a complex system whereby waste criminals adapt their behaviours and evolve new responses to overcome/circumvent enforcement measures. This makes evaluation challenging: evaluators are 'chasing a moving target' and it is difficult to know whether a policy is working effectively, and if so, how.

- Rapid changes or increased understanding of context may require regular review of the policy and its evaluation.
- Rapid information feedback will be needed to understand and respond to changes.

[6] See Emergence (page 45).

3.1.1 USEFUL QUESTIONS

- Has a theory of change been developed for the policy or programme in question?

Useful tools & resources

A theory of change describes how an intervention is understood to contribute to a chain of outputs and outcomes leading to its resulting impacts.

- Have you considered building a systems map of the policy and the context within which it is being implemented? Has it been used to inform or update the theory of change?

Useful tools & resources

Systems mapping can help identify possible impacts of external factors such as changes to other policies that can have a positive or negative impact on the effectiveness of a policy. Analysts in Defra's Future Farming team are pioneering systems mapping approaches as a tool for policy design, appraisal and evaluation.

- Do you identify any of the characteristics of a complex system as being present in the intervention itself, or the context in which it is being delivered?

Useful tools & resources

For a list of complexity characteristics, including examples to aid identification and key points, see Annex II: A Visual Guide to Understanding Complexity for Defra.

- In what way might any of these characteristics support the successful delivery of the policy and the achievements of its aims?
- Are there any characteristics which might make the outcome very unpredictable, or different in different settings?
- Are there any characteristics that might get in the way of successful delivery and achievement of outcomes?

Useful tools & resources

To better understand the system and to help anticipate unexpected system behaviour, speak to stakeholders and experts with experience of the system, both inside and outside of Defra. These might include individuals who are currently or were previously involved in the design, implementation or evaluation of policies, programmes or pilots in similar or relevant areas.

- Are there any external factors that might strongly affect the delivery and success of the policy? (e.g. change in other policy areas, introduction of new policies and programmes, changing environmental conditions)?
    - Have you considered what impacts these might have (positive, negative or neutral)?

Useful tools & resources

Ojeda-Martínez et al., (2009) employed the Driver-Pressure-State-Impact-Response (DPSIR) framework to investigate factors influencing and affecting the operation of marine protected areas.

- Have stages to review and update understanding of the system been built into policy design and implementation plans?
- Have you explored with the policy lead(s) the evidence base, methods and models used to inform the design of the policy?
- Has thinking moved on – are the original research methods still valid, and is there new work in this area that might be useful?
- Have stages to review and update understanding of the system been built into evaluation plans?
- During the evaluation new understanding may come to light and relevant changes in context may take place – have regular opportunities been built in to review and change the evaluation strategy?
- What implications does this have for the design of the evaluation and the resources required?

Useful tools & resources

For evaluation design considerations, go to Designing.

3.2 COMPLEXITY ARISING FROM MULTIPLE AND DIVERSE STAKEHOLDERS: ISSUES FOR UNDERSTANDING

- Different stakeholders may be able to contribute different kinds of information to aid with understanding how the system and/or the intervention work
- Stakeholders may have very different understandings of the system and its complexity
- Stakeholders may not agree on the purpose, approach and/or methods of the policy or its evaluation

3.2.1 USEFUL QUESTIONS

- Have you made a list of the key stakeholder groups and communities affected by this policy and its evaluation?
- Have you identified key areas of agreement and lack of agreement between different stakeholder groups (e.g. about the policy itself, its outcomes or its evaluation)?

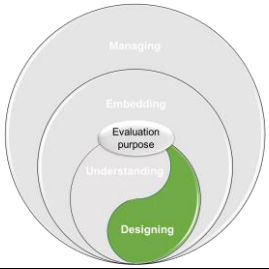
Useful tools & resources

Cultural consensus analysis can provide a useful tool for examining the extent to which stakeholder groups share similar mental models of the system, of resources, and the interactions and processes occurring between these components. As a form of analysis, it is compatible with systems mapping approaches such as group model building and can be used in conjunction with systems mapping to enhance the social-learning and knowledge sharing aspects (e.g. Mathevet et al., (2011)).

- Have you actively involved these stakeholders in the policy and evaluation design?

Advice from interviewees

In addition to their input in the evaluation design, include a broad range of expertise on procurement panels in order to reflect the considerations of a range of stakeholders and to ensure key priorities for the evaluation are covered.



4 DESIGNING

This section explores the implications of complexity for **designing**. Before reaching this point, the **Evaluation Purpose** should have already been explored, working through the considerations in Part 1.

Designing is the element of an evaluation concerned with choosing, implementing (including commissioning) and adapting the evaluation design. In a complexity-appropriate evaluation, designing is closely interwoven with the understanding element (addressed in section 3), and both will continue to develop and be updated throughout the evaluation. As new understanding about the intervention and system comes to light, more will be understood about how the intervention can best be evaluated, and vice versa.

4.1 COMPLEXITY IN DEFRA'S POLICY DOMAINS: ISSUES FOR DESIGNING

- There are a wealth of evaluation designs that work well with complexity, provided the evaluation is led and managed in a complexity-appropriate way[7].
- Most of these designs are 'method-neutral' – meaning a wide range of different data-gathering and analytical methods can be used. These can include surveys, case studies, focus groups, randomised control trials and qualitative comparative analysis. There is no required use of particular methods.
- There is no simple way to select the best design, and there is no gold-standard approach for complex evaluation. The choice will depend on the complexity of the intervention, characteristics of the system, evaluation purpose and the feasibility (including timescale and resource constraints) of the available designs and methods. If reviewing the purpose or methods of an existing evaluation, any proposed changes should be considered in light of the evaluation methods that have already been used.
- Tools to assist those who are designing evaluations in a complex context are discussed below. The design should seek to build on outputs and understanding of the intervention operation and context, developed at the 'Understanding' stage, such as system mapping.

[7] Complexity-appropriate evaluation is iterative and embedded throughout the policy cycle (see page 12).

Advice from interviewees

"It's about appropriate methods for different projects and different contexts. The strength [with randomised control trials], if they work, is that potentially you get clear quantitative findings about the effectiveness of your programme. Different forms of qualitative data capture the richness of the programme so that you're able to convey the strengths and weaknesses, what's worked and what hasn't in a more nuanced way." - Interviewee

- Often a hybrid design (a combination of designs) will be needed. Mixed-method approaches, combining qualitative and quantitative data, can act as a bridge to smooth the tension between attempts to simplify complexity into easily distilled measures and the need for a "full" holistic account of the system.

Example

Malawska et al., (2014) highlight that agricultural and environmental policies often have unwanted and or unintended consequences as a result of simplistic assumptions. They call for integrated methods that bring together traditional agricultural and ecological models with system and human behavioural approaches, such as agent-based modelling.

- The mix of approaches and methods used may need to be adapted in response to changes as the evaluation progresses, such as changes in the system, intervention, or understanding thereof, or even in the evaluation purpose e.g. from learning (how do I make this work better?) to accountability (how well did it work?) and improving the knowledge base (how can I make similar policies work better in future?).
- When interacting with complex systems, change is a given, certainty is not possible. Methods that offer a high degree of certainty in straightforward contexts are liable to give a misleading sense of security when applied to complex systems.

"The perception of policy is often 'fire and forget'. There are ideas that polices cannot be modified once in place; that they need to be pre-formed in a perfect mould. We reinforce this with the idea that randomised control trials can provide a template for policy action when, in reality, they probably can only ever tell us about a rather narrow set of circumstances around policy implementation."

Ian Boyd, Defra Chief Scientific Advisor, 2018.[8]

- Using methods which do not engage with the dynamic and context-sensitive nature of complex systems may still be appropriate for certain narrowly-defined evaluation questions. In these cases, thinking about complexity may lead to a reasonable basis for an evaluation to use traditional experimental methods. In most cases however, the evaluation will require the considered use of complexity-appropriate methods and tools.
- In a complex, changing system, an evaluation may only provide a snapshot in time. Methods that can help policy colleagues look forward and backwards in time are likely to be particularly useful to users of evaluation results. In the face of very rapid change and uncertainty, evaluation approaches that are developmental and participative can support rapid feedback and build agents for change to support adaptive management.
- Effective evaluation requires accurate, timely and relevant data and this is particularly important for complexity-aware evaluations, where real-time data on project delivery activities, their impact, and any changes in the project's operating environment are vital inputs to the evaluation process.
- To establish effective monitoring and evaluation frameworks, evaluators and project managers need to work together to identify:
    - The key indicators required to monitor inputs, activities, outputs, outcomes and impacts; and
    - How, when and by whom data on change/s in key indicators are to be collected, verified, analysed, and reported.

  This will help to ensure project monitoring, performance management, and evaluation reinforce each other.
- Key considerations when designing an evaluation, along with method and timescales, are proportionality and affordability.
- A proportionate evaluation delivers findings that are of good quality and fit for purpose given the risks of getting the answer wrong.

[8] Boyd, I., (2018). Policy, evaluation and implementation, in *The Evaluator*, Autumn 2018, pp6-7, UK Evaluation Society.

- Currently within Defra, there is no prescriptive guidance on the percentage of programme spend that should be dedicated to evaluation; decisions are made on a case-by-case basis and take account of issues such as the innovative nature of an intervention, the scale of an intervention and the level of overall spend.
- Once the complexity characteristics of the evaluation are identified, the range of potential approaches should be reviewed, considering the best options for a specific evaluation, based on feasibility and affordability.
- Once a preferred evaluation approach has been identified, an estimate of the costs (including data collection, collation and analysis; project management; and reporting) should be produced.
- If the available resources are not adequate for the preferred design, then adjustments to the design and/or to the resources available need to be made.
- If additional resources cannot be secured, there are a number of ways to reduce costs, although each has potential implications for the evaluation outputs.

Examples of how to reduce costs

Ways to reduce evaluation costs while meeting evaluation objectives, and potential implications for the evaluation outputs:

- Work with stakeholders to prioritise the evaluation questions to be answered (bearing in mind that losing any evaluation questions could have implications for the evaluation to meet its purpose);
- Reducing sample sizes (this may result in reduced accuracy of estimates);
- Reducing the number of 'waves' of research (again, this may lead to reduced accuracy)
- Reducing the number of case studies to be undertaken (this may result in less depth of understanding of the system/intervention)
- Embedding data collection in the overall management of the intervention (this may result in poorer quality data); and
- Using alternative sources of data (this may have implications for data quality, accuracy or relevance).

See: Better Evaluation on implications of resources constraints: https://www.betterevaluation.org/en/node/5296

4.1.1 USEFUL QUESTIONS

- Have the complexity characteristics of the system been taken into account as far as possible when considering the evaluation design? For example:
  - Do you understand what may be influencing change? Can you detect if change is happening over the background 'noise' and what aspects of change are due to the policy and what are due to other influences?
  - Is the system and intervention still changing? How can you be sure that change will continue after the intervention ceases, or will continue to be sustained over time?
  - Can you identify levers to help push change through the system, or feedbacks that may inhibit or promote change?
  - What are the key indicators of inputs, activities, outputs, outcomes and impacts? And how, when and by whom are data to be collected, verified, analysed, and reported.

Useful tools & resources

For a list of complexity characteristics, including examples to aid identification, see Annex II: A visual guide to understanding complexity for Defra

- Has the evaluation purpose been considered, e.g. is the evaluation required for:
  - Listening and building trust – How can you ensure diverse voices are heard and build trust and legitimacy across stakeholders?
  - Learning – How is change happening? Why is change happening or not? How can you improve the implementation or impact of the policy? How can you feed learning back in a timely manner?
  - Accountability – Was the policy implemented as planned? Is it having the impact hoped for? Are there any unexpected benefits – positive or negative? Would change have happened anyway, in the absence of the policy?
  - Accountability – To what extent are quantitative measures needed or sufficient? Is additional information needed? To make sense of the results and increase their usefulness, do you need to ask how and why change is happening?
  - Building the knowledge base – How can you improve future similar policies? How can you help ensure that learning is transferable to other contexts?

- Is the evaluation purpose realistic and pragmatic in scope?

Example

"The evaluation of the Flood and coastal erosion resilience partnership funding (Defra project code FD2663) is an example of a robust evaluation of a potentially very diverse policy area which simplified by focusing on key outcomes – in this instance changes in the number of funded flood risk management projects. The evaluation did consider other factors, such as biodiversity and localised social effects, but as the intervention was at an early stage of deployment there was no outcome data for these topics and they were not a focus of the evaluation. The evaluation provided a rationale for this decision and was transparent about its focus."[9]

- Have you discussed with users their needs from the evaluation and how to meet these given the inherent uncertainties arising from complexity?
- How reliable do the findings need to be? What are the consequences of getting the answer wrong?
- Are the standards of rigour (and confidence in veracity of outcomes) being used appropriate to the evaluation purpose, resources and timescale?
- Has feasibility been taken into account?
    - Have the available budget, skills and experience and timescales been taken into account in the evaluation design?
    - Does the budget reasonably reflect the need – e.g. taking into account for example: the level of risk and innovation, the scale, value and profile of the policy, the availability of data?

Useful tools & resources

See guidance on resourcing evaluation in HM Treasury *Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation*

[9] Baker, J. (forthcoming) *Evaluating Environmental Interventions: Challenges, Options and Examples (EEICOE): Methodological inspiration for environmental evaluators*. Defra.

Useful tools & resources

Scoping studies can help to establish the feasibility of a methodological approach in relation to the purpose of an evaluation.

"We knew generally what were the research questions or what were the actions that we wanted out of this, and we knew potentially what might be some approaches, but to be able to develop them, flesh them out and to be able to ascertain what approaches weren't possible, that was the use of the scoping study." – Interviewee

- Have you explored whether there are any past Defra evaluations which might be relevant or useful? Are their approach and findings valid and meaningful in the context of your evaluation?

Advice from interviewees

"Evaluations of past programmes are a valuable source of evidence and analysis for evaluating current and future programmes." – Interviewee

- Have you discussed the possible range of evaluation approaches and their relative merits with others with evaluation expertise in your area?

Useful tools & resources

Stakeholders in other government departments can be a valuable source of information and expertise.

- Are you clear about why your chosen evaluation approach and methods are appropriate (given the evaluation purpose, resources and timescale)? Are you clear about what the limitations are and how they will affect the interpretation of results?

Useful tools & resources

There are a number of tools, methods or approaches that can work well with complexity, such as:

- Bayesian networks, which use quantitative data and / or human perception to identify probabilities for key variables and so model complex systems. These models can be improved over time as new evidence emerges. Bayesian networks require technical skills (and specialist software can be helpful) to elicit and calculate probabilities, but can be used in policy appraisal and evaluation. For example, the document linked above shows how it has been used in tracking Bovine TB.
- Agent based models model the behaviour and interactions of individuals, households, businesses or other 'agents'. These models then enable evaluations to handle feedbacks and detailed interactions between agents. The approach cannot predict the future of a complex adaptive system, but can be used to offer insights into the range of possible futures, e.g. in relation to adaptations in response to climate change.
- Qualitative Comparative Analysis enables systematic comparison across cases (usually between 10 and 50). It takes account of 'complex causation', where a combination of 'attributes' may produce a given outcome. It is particularly useful when evaluators have a small to medium number of cases of interventions which are similar but applied in different contexts. The Environment Agency used this approach to navigate the complexities in waste crime policy.[10]

Each of these approaches, has its own strengths and weaknesses with respect to complexity, as do other methods or tools.[11]

For further information on how to choose between methods, see The Magenta Book, plus Befani, B. *Choosing appropriate evaluation methods*. This resource provides a downloadable tool to identify appropriate evaluation methods. It uses a series of questions, to explore a method's ability to answer key evaluation questions (such as "What was the additional/net change caused by the intervention?"); and its ability to meet additional needs (for example the need to generalise evaluation findings outside the case/sample used for the analysis). It also explores the conditions that need to be met in order for a method to be applied in practice (for example "To what extent is it possible to control who does and doesn't receive the intervention?"). The tool provides a summary report on the most appropriate method/s given the information provided.

[10] https://www.cecan.ac.uk/case-studies/environment-agency-enforcement-on-waste-crime
[11] For an overview of these strengths and weaknesses, see HM Treasury Magenta Book 2020 Supplementary Guide*: Handling Complexity in Policy Evaluation*

For further information on relevant evaluation and research methods and good practice, see:

- CECAN Evaluation Policy and Practice Note Series (EPPNs)
- CECAN syllabus: Qualitative Comparative Analysis; Systems Mapping; and Agent Based Modelling

- Is there flexibility to change the evaluation approach to respond to changing conditions? Have you considered the opportunities for flexible evaluation designs within the current parameters of commissioning rules? Explore the range of options through discussion with procurement colleagues.
- Have you identified a preferred approach based on the options available?
- Are the estimated costs of the preferred approach affordable?
- What adjustments to the design can be made to ensure the evaluation is affordable?
- What are the implications of these adjustments (for example on data quality or accuracy), and have these been discussed with stakeholders?

4.2 COMPLEXITY ARISING FROM MULTIPLE AND DIVERSE STAKEHOLDERS: ISSUES FOR DESIGNING

- Stakeholders may have different mental models and views regarding what the system is, how the system should work and how (and if) it is working.

Example

Cisneros-Montemayor et al., (2018) identify the difficulties that marine system environments present for assessing progress towards environmental sustainability, due to the multiple and often unclear objectives of different stakeholders and the inherent viability of marine ecosystems and the problems of directly observing those systems.

- Different stakeholders from different research traditions may have very different views on the best approach to take to the evaluation.
- Expectations of what can be achieved in an evaluation needs careful management.
- For a better understanding of complex systems, you need to involve the stakeholders actively in the evaluation. This can lead to tensions between the separation required to demonstrate objectivity and the immersion needed to develop understanding.

- Complexity-aware evaluations may need to adopt flexible or iterative approaches. However, procurement rules may require external evaluation contractors to submit and deliver a fixed programme of work. This can present challenges for complexity-aware evaluations. The following suggestions may assist in maximising the opportunities for flexible evaluation design within common parameters of commissioning rules:
    - Speak to procurement colleagues at the earliest opportunity when planning an evaluation to identify the degree of flexibility that can be built into the process and what options are available.
    - Within tenders, recognise that circumstances may change. Tenderers could be asked to demonstrate how the evaluation approach can accommodate and adapt to changing circumstances. The evaluation purpose should be regularly reviewed.
    - Consider building a feasibility or scoping period into the evaluation that can be used to define the following stages.
    - Consider a staged approach to commissioning, with built-in decision points and options to reconsider the approach and/or contractor. Contractors can be asked to prepare the brief for the next stage of work as part of their remit. Where the evaluation is likely to be managed through a mix of internal and external resources, a break-point type approach could also be used; giving responsibility to a study Steering Group as to whether to proceed at certain stages with external contractors or conduct the work internally.
    - There may be a need to expand or revise the scope of future phases of an evaluation. This could also be achieved through commissioning additional studies, if helpful.
    - Working with contractors on a 'co-creation' basis can sometimes be the best vehicle for delivering an evaluation that is sufficiently flexible and adaptive to cope with complexity. This can be built into contracts by stating that a co-creation work style is expected and/or by using a time and resources (rather than fixed price) contract style. Call off contracts can enable evaluation expertise and advice from policy experts and external consultants to be brought in at short notice.

Advice from interviewees

"Another thing that I think is quite crucial to have within any evaluation contract is some provision for either extension or a cool-off; there will be cases that you're not able to think of everything; having that ability to expand or extend certain areas is quite important." - Interviewee

4.2.1 USEFUL QUESTIONS

- Have you planned to involve evaluation users and other stakeholders to obtain their input into the evaluation and its design, and to share findings?

Useful tools & resources

Participatory and emancipatory methods that involve stakeholders actively in the evaluation, and that encourage the participation of quieter voices will be particularly useful. Participatory evaluation can facilitate spaces for diverse groups of stakeholders to come together to learn and collaborate with each other and share experiences by improving their ability to interact and appreciate different actor perspectives (e.g. Daw et al., (2015)).

Example

Mathevet et al., (2011) in the case of the Camargue Biosphere Reserve in the Rhône river delta, France, attempt to capture the cognitive mental models of how people represent their interaction with water-related systems, in order to determine a shared common understanding of the system for the purposes of coordinating management goals.

Useful tools & resources

Participatory scenario modelling is a tool for integrating ecological simulations with participatory approaches, and enabling stakeholders to navigate complex trade-offs, promote discussion and identify long-term management strategies.

- Have evaluation stakeholders committed to give the necessary time to the evaluation?
- Has the acceptability of the method been taken into account? In particular evaluation stakeholders' and users' preconceptions about:
    - The rigour or quality standard of different specific methods – which will be quite different in complex environments (for example randomised control trials are often impractical or even impossible)
    - The achievable level of certainty in the results, and
    - The appropriate level of objectivity i.e. separation between the evaluator and those being evaluated?
- Have the standards of certainty and rigour required for this evaluation, given the resources and purpose, been clarified and agreed with stakeholders and policy

colleagues? Is it appreciated that it will be impossible to resolve all uncertainty, even where sophisticated evaluation approaches are used?

- Have you identified the degree of flexibility enabled by the procurement process?
- Does the tender ask bidders to demonstrate how their proposed approach can be adapted to changing circumstances?
- Have you considered a staged approach to commissioning, with built-in decision points and options to reconsider the approach and/or contractor?

5 EMBEDDING

This section explores the implications of complexity for **embedding**. This is the element of an evaluation concerned with feeding back understanding and learning to evaluation users and participants, and embedding these into relevant processes both inside and outside of the evaluation (dissemination and use).

5.1 COMPLEXITY IN DEFRA'S POLICY DOMAINS: ISSUES FOR EMBEDDING LEARNING

- It can be hard to communicate complexity.

Example

Rogers et al., (2013) identify the importance of embedding and translating the key issues of social-ecological complexity between researchers and stakeholders for decision-making using a participatory process.

- Evaluation findings, particularly in relation to complex policy delivery, may be 'provisional and indicative' rather than definitive. Change may continue after the evaluation comes to an end – and so, more than in other contexts, evaluation reports could simply provide a snapshot in time.
- Rapid evaluation feedback to meet the needs of complex policies may require different reporting standards to a major evaluation report.
- The particular sensitivity of complex systems to their context means that it can be hard to generate results that are transferrable from one context to another. If transferability is required, consider the primary and target contexts and think about how to generate results that are as transferrable as possible. It is impossible to capture in a report all of the nuances of the analysis and synthesis carried out, so report authors and users must also be particularly disciplined in the way they present and use such findings.
- In complex policy domains, it is particularly difficult to generalise evaluation results to other contexts and it is not always possible to have the original evaluator around to interpret the results for a new context. While reading across evaluations can be informative, it is likely that different contexts will require bespoke evaluation.

5.1.1 USEFUL QUESTIONS

- Has there been discussion with stakeholders about the complexity in the policy areas and the evaluation challenges related to this?

Useful tools & resources

For a set of relevant complexity characteristics to frame and inform discussions with stakeholders, see Annex II: A visual guide to understanding complexity for Defra.

- Have difficulties in generating definitive and generalisable findings been discussed?
- Is the evaluation timed appropriately to take ongoing change into account?
- Is the evaluation timed appropriately to feed results usefully into decision-making?
- Does the evaluation suggest that more or different data should be collected to monitor the policy effectively?
- Have processes through which the evaluation findings can feed back to policy makers on a regular basis been considered?

Useful tools & resources

If a systems mapping has been conducted or is planned, evaluation findings can be embedded by using new understanding to revise and update the systems map. A system's map regularly updated in this way becomes a living tool to support both evaluation and effective policy design; it has the potential to provide value to a given policy area over time across multiple policies and evaluations.

Useful tools & resources

Process evaluations can be a valuable way in which lessons can be fed back quickly to the policy process. See Public Health England (2018) Guidance on process evaluation for more information.

Advice from interviewees

Evaluation design and governance arrangements that bring together evaluators and policy analysts can help to embed learning.

For example, in one case where an evaluation method needed to be co-designed with the intervention at the outset:

"Because they require some changes in programme design quite often, that opens up quite an interesting conversation for evaluation and evidence people to become more involved in the design of policies." – Interviewee

- Are the appropriate governance processes in place to allow for rapid responses to early or emerging evidence and findings, or changes to the policy-making environment? This is a key concern for complexity-appropriate evaluation, and in particular for managing commissioned evaluations.

Advice from interviewees

"There's real value in having more flexible, agile evaluation in these circumstances so that we can get information quickly to policy makers and also the ability to change the focus of research or target different areas depending on the results that earlier research gives." – Interviewee

5.2 COMPLEXITY ARISING FROM MULTIPLE AND DIVERSE STAKEHOLDERS: ISSUES FOR EMBEDDING LEARNING

- It may be hard to communicate provisional and indicative findings from complex evaluations to stakeholders
- Given the complexity of Defra's policy domains, there may be a very diverse audience for evaluation findings who may want to use them in different ways

5.2.1 USEFUL QUESTIONS

- Have you considered multiple routes of dissemination?
- Have stakeholders been primed to anticipate uncertainty in findings?

Useful tools & resources

To help explain where and why uncertainty might arise from complex systems, consider sharing Annex II: A Visual guide to understanding complexity for Defra.

Useful tools & resources

Approaches such as participatory systems mapping and group model building can provide a participative tool to build a common understanding of system complexity across different stakeholder communities and enable a structured means of communicating the learning around these issues (e.g. Vugteveen et al., (2015)).

- Have there been opportunities to discuss early findings with key stakeholders?

Advice from interviewees

A three-year evaluation of nature improvement areas produced annual progress reports that stakeholders found valuable. Much effort was expended by stakeholders in the nature improvement areas in collecting data on impacts of the policy at local level. This was valuable, because it helped to secure their buy-in to the policy.

- How will stakeholders be involved in decision-making processes? Have these processes been designed and/or discussed with stakeholders? How will their involvement be managed?

Useful tools & resources

Participatory co-engineering can be used to look at the interactional behaviours between stakeholders and aid in collective decision-making processes (Daniell et al., (2010))

Useful tools & resources

Knowledge co-production approaches afford a means of identifying stakeholder configurations (i.e. specific roles they play in the system), as well as opportunities to bring different actors together in a shared space to work collaboratively, exchange ideas and experiences and jointly plan future management strategies, increasing the likelihood of more consensual and integrated decision-making (e.g. Reyers et al., (2015)).

- Has the purpose of the evaluation, and the purpose of dissemination, been considered when deciding the format and content of the evaluation report?
- How will the findings be interpreted? By whom? What expertise is needed for this? What role does the evaluation team need to take in communicating the findings back to other stakeholders?

ANNEXES: TOOLS AND FURTHER RESOURCES

ANNEX I: KEY ISSUES AND QUESTIONS

The table below summarises the 'questions to ask' and 'key issues' presented throughout the Complexity Evaluation Framework. It can be used as a checklist of complexity-aware considerations. It is also available in this format in Chapter 2 (page 15) in the document. Further detail on each of the questions and key issues can be found in Chapters 3 to 5.

	MANAGING
	This framework is written predominantly for those commissioning or managing evaluations. Therefore, the **Managing** element of an evaluation is the overarching perspective from which the issues in this framework are explored. As such it has no chapter of its own; considerations for managing complexity-appropriate evaluations are embedded throughout each of the **Understanding**, **Designing** and **Embedding** chapters.

	UNDERSTANDING	
	Key issues	**Questions to ask**
Complexity in Defra's policy domains	❖ Lack of knowledge of the issues and complexity challenges ❖ Potential for 'emergent' and unexpected outcomes ❖ New understanding of the system will come to light ❖ Need for regular review of the policy and its evaluation ❖ Need for rapid feedback to understand what is going on 'on the ground'	❑ Have you undertaken a mapping of the system, the policy and its delivery? ❑ What characteristics of a complex system do you recognise – in the policy or its context? ❑ How might these influence the way the policy is delivered or its outcomes? ❑ Have stages for review and feedback been built into policy design, implementation and evaluation plans?
Multiple & diverse stakeholders	❖ Different stakeholders may be able to contribute different information for the planning/design process ❖ Stakeholders may not have the same understanding of the system or agree on the best approach to the evaluation.	❑ Have you identified the key stakeholder groups and communities affected by this policy and its evaluation? ❑ Have you actively involved stakeholders in the policy and evaluation design? ❑ To what extent is there agreement and lack of agreement about the policy itself, its outcomes or its evaluation?

	DESIGNING	
	Key issues	**Questions to ask**
Complexity in Defra's policy domains	❖ Many evaluation designs can work well with complexity. However, there is no one-size fits-all design; the choice of evaluation design will depend on the complexity characteristics of the system, evaluation purpose and feasibility considerations	❑ Have you taken into account the complexity characteristics of the system in addition to the evaluation purpose and feasibility (available budget, skills and experience, timescales and data requirements) when selecting the overall evaluation approach? ❑ Are you clear about why your chosen approach is appropriate and what the limitations are?

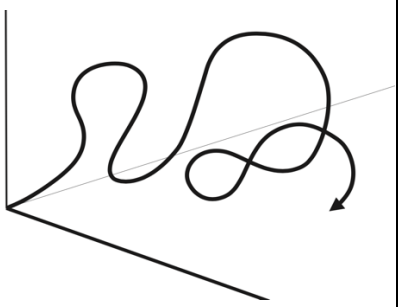
	❖ The mix of approaches and methods selected may need to be adapted to changing circumstances ❖ Care must be taken in the choice of methods – methods that offer a high degree of certainty in straightforward contexts may mislead when applied to complex systems ❖ The evaluation design and plans need regular updating to address unexpected changes in policy and context ❖ Effective evaluation requires accurate, timely and relevant data – this is particularly important for complexity-aware evaluations. ❖ Proportionality and affordability are key considerations when designing an evaluation, along with method and timescales.	❑ Has flexibility to review and change the evaluation design been built into the evaluation plan? ❑ Have you engaged stakeholders in the evaluation design? ❑ Is the proposed approach affordable? If not, what adjustments can be made, and what are the implications of these adjustments for the evaluation outputs?
Multiple & diverse stakeholders	❖ Different stakeholders may have different views about how the system should work, and how it is working ❖ Expectations of what the evaluation – and what different evaluation methods – can achieve need careful management ❖ Complexity-aware evaluations may need to adopt an iterative approach, while procurement rules may require external evaluation contractors to submit and deliver a fixed programme of work.	❑ Have participative evaluation approaches and methods been considered? ❑ Have stakeholders committed to give the necessary time to the evaluation? ❑ Have stakeholders been primed to anticipate uncertainty in findings? ❑ Has flexibility been built into the commissioning of external evaluators, for example by using built-in decision points or contract options?
EMBEDDING		
	Key issues	**Questions to ask**
Complexity in Defra's policy domains	❖ When undertaking an evaluation within a complex system, results may be indicative rather than definitive ❖ Evaluation in a complex environment may only provide a snapshot; change may continue after an evaluation comes to an end ❖ Findings may not be transferrable due the specifics of a complex policy environment.	❑ Have you talked about complexity with the potential audience(s) for the findings to manage expectations and identify the value evaluation can provide? ❑ Have difficulties in generating definitive and generalisable findings been discussed? ❑ Are there opportunities to feed findings back regularly to support implementation? ❑ Is the evaluation timed appropriately to take ongoing change into account?

Multiple & diverse stakeholders	❖ Given the complexity of Defra's policy domains, there is likely to be a diverse audience for findings who may want to use them in different ways ❖ Different stakeholders have different levels of satisfaction with provisional and indicative findings	❑ Have you considered multiple routes of dissemination? ❑ Does your plan include regular opportunities for discussing early findings?

ANNEX II: A VISUAL GUIDE TO UNDERSTANDING COMPLEXITY FOR DEFRA

Adapted from CECAN's *The Visual Representation of Complexity*[12]

Below are some of the characteristics and behaviours that complex systems in Defra might exhibit. These complexity characteristics can be seen in both complex ecological and social systems; indeed, a system of interest to Defra may comprise a combination of the two.



NON-LINEARITY

A system is non-linear when the effect of inputs on outputs are not proportional. Outputs may change exponentially, or even change direction (e.g. after increasing for some time, they may begin decreasing), despite small or consistent changes in inputs.

Examples

- Increasing payment rates for land management does not translate into a corresponding increase in their uptake. Land managers do not behave as the rational agents of traditional economic theory; there are other factors at play.
- A new product may be slow to take-off, but after a certain point, sales accelerate, before slowing again as the market is saturated.

Key points

- In social settings, few things are actually linear.
- Non-linearity can mean that the relationships between things can be just as powerful in determining outcomes as the structure of interactions.
- In non-linear systems when we double or halve an input, the output will not be double or half its original value and may be completely different.

[12] Boehnert, J. et al., (2018)

FEEDBACK

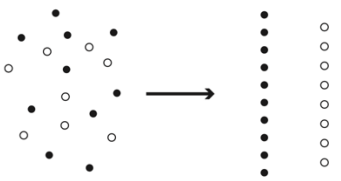
When a result or output of a process influences the input either directly or indirectly. Feedback can accelerate or suppress change.

Examples

- Positive feedback leads to accelerating change. For example, as the climate changes, permafrost melts and releases more greenhouse gases, contributing further to climate change.
- Negative feedback creates stability in systems. For example a thermostat, or the human body shivering/sweating, maintains a constant temperature.
- Feedbacks operating between resources, actors and governance. For example, environmental monitoring data such as the changing status of different fish populations can be used to inform policymaking leading to specific management interventions that can modulate fishing activities and behaviour (e.g. the use of particular types of fishing gear; fisheries closures), which in turn reduces off-take lowering fishing pressure and restoring declining fish stocks (Martone et al., 2017; Fujitani et al., 2018). Or, the influence of economic policy drivers on agro-ecosystems such as the move towards bioenergy crops and the subsequent (positive and negative) impacts this transition can have on land-use change and biodiversity (e.g. farmland bird species) (Malawska and Topping, 2017).

Key points

- Feedback loops can lead to runaway effects or can create inertia through dampening of effects - two extremes.
- Positive feedbacks are reinforcing and accelerate change.
- Negative feedback suppress change and are stabilising/regulating.
- Feedback processes can be slow and fast



SELF-ORGANISATION

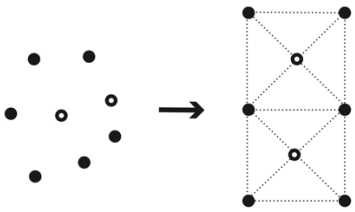
Regularities or higher-level patterns can arise from the local interaction of autonomous lower-level components.

Examples

- Shoals of fish, flocking of birds
- Multiple individuals locally clearing non-crop species leading to large-scale habitat fragmentation.
- Sheep paths - these informal paths across land have no architect; they are formed by erosion caused by the footfall of individuals over time. Patterns of paths emerge as each individual chooses their own route.

Key points

- Simple and autonomous behaviour can create order at larger scales.
- This higher-level order requires only local (or lower-level) interactions.
- Order arises spontaneously without top down control and hence can often remain in place even if part of the system is disrupted.



EMERGENCE

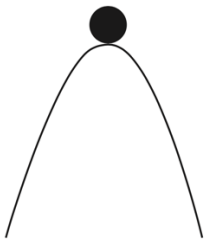
New, unexpected higher-level properties can arise from the interaction of components. These properties are said to be emergent if they cannot easily be described, explained, or predicted from the properties of the lower level components.

Examples

- E.g. Community resilience – a community's capacity to function in and respond to shocks and extreme events – is an example of emergence; it is shaped by and arises from interactions between human and environmental components (Faulkner et al., 2018).

Key points

- Completely new and unexpected properties or things can arise simply from the interaction of lower level entities. These new properties can be difficult and sometimes impossible to predict.
- Emergence and self-organisation are closely related concepts. Self-organisation can cause emergent phenomena, but emergent phenomena do not have to be self-organised.



TIPPING POINTS

The point beyond which system outcomes change dramatically. A threshold is the point beyond which system behaviour suddenly changes.

Examples

- A species' population reducing in numbers such to the extent that it cannot re-establish itself in the wild.
- Building new relationships between industry partners can lead to sudden step-changes in collaboration or knowledge sharing, developing grounds for innovation.

Key points

- Sudden change takes place unexpectedly.
- Knowledge of tipping points can be used to affect change in a system. We can aim to get a system past a tipping point (as also described in the 'domains of stability' definition).
- A system may be pushed towards and past a tipping point by positive feedback of some kind.

DOMAINS OF STABILITY

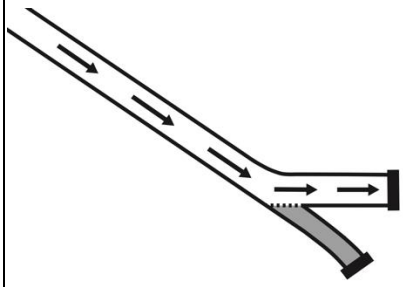
Complex systems may have multiple stable states which can change as the context evolves. Systems gravitate towards such states, remaining there unless significantly perturbed. If change in a system passes a threshold, it may slide rapidly into another stable state, making change very difficult to reverse.

Examples

- Land management improvements in a specific environment may not lead to increases in bird populations, because birds require multiple habitats (e.g. over-wintering, nesting and chickling habitat).

Key points

- Knowledge of domains of stability can be used to effect change in a system. If we can push a system into a different, more desirable, stable state with a policy intervention then we have changed the system in a robust way.
- We do not need to put in continuous effort to keep the system in the new state.
- We may try to use policy to change the positions of domains of stability.
- What is possible in a system is often discontinuous and sticky. Not everything is stable and change can be hard to reverse.



PATH DEPENDENCY

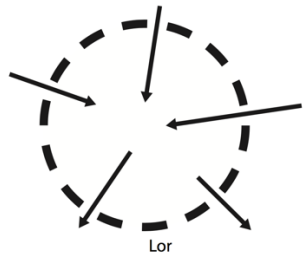
Current and future states, actions, or decisions depend on the sequence of states, actions, or decisions that preceded them – namely their (typically temporal) path.

Examples

- The organisation chosen to lead a new policy initiative influences which other organisations also become involved.
- Species which colonise a habitat first have "founder effects", determining ultimate community composition.
- When introducing a policy that leads to behaviour change, we might see that as the first actors start to change their behaviour, others follow, and the behaviour becomes ‘the new normal’. The actions/choices of the first adopters can determine the behaviours of those who follow.

Key points

- ‘History matters’; it may be difficult or impossible to revert to a previous path once certain changes have been enacted.
- When appraising different policy options, consider what path-dependencies these might lead to.



OPEN SYSTEM

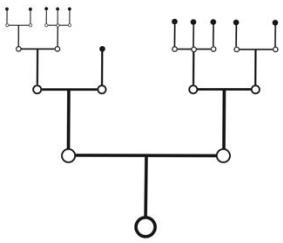
An open system is a system that has external interactions. These can take the form of information, energy, or material transfers into or out of the system boundary. In the social sciences an open system is a process that exchanges material, energy, people, capital and information with its environment.

Examples

- A food production company changes in response to changes in food fashions or the cost and availability of ingredients.

Key points

- Open systems are impossible to bound.
- Open systems mean that we must be alert to outside influences.



DISTRIBUTED CONTROL

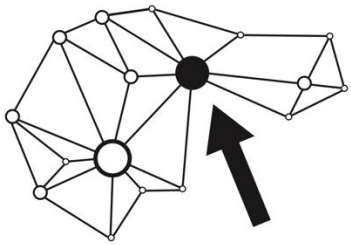
Control of a system is distributed amongst many actors. No one actor has total control. Each actor may only have access to local information.

Examples

- An intervention's success may be determined by enforcement officers 'on the ground', rather than the central agency.
- Central groups and their distributed branches may conduct work in contradictory ways.

Key points

- True top-down control is not possible in complex systems. Decisions and reactions happen locally and the interactions of all these lower-level decisions can give us system-level properties such as stability, resilience, adaptation or whole system emergent regulation.
- The best we can do is to "steer" the system.



LEVERS AND HUBS

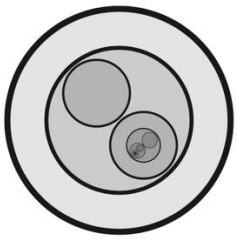
There may be components of a system that have a disproportionate influence because of the structure of their connections. How these behave can help to mobilise change, but their behaviour may also make a system vulnerable to disruption.

Examples

- If a keystone species becomes extinct there may be cascading extinctions among other species.
- Across the food system, the operations of key manufacturers and retailers in the supply chain can have a disproportionate impact on producers (e.g. farmers) in terms of the quality, quantity, type and wholesale prices of food as well as consumer behaviour.
- Statutory instruments, markets, regulations and protocols are examples of policy levers that can be used to produce significant social and environmental outcomes.

Key points

- Identifying hubs and levers can help identify best places to intervene in complex systems.
- Structure matters; knowing the structure of interactions in a system is crucial to understanding how it will behave, change or fail.



NESTED SYSTEMS

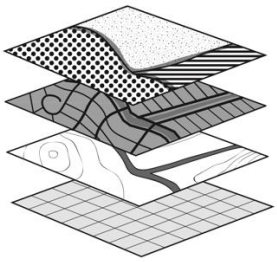
Complex systems are often nested hierarchies of complex systems (so-called 'systems of systems').

Examples

- An ecosystem is made up of organisms, made up of cells, made up of organelles which were once free-living bacteria, made up of complex metabolic processes intertwined with genetic systems (each nested level is a complex system).

Key points

- When studying a particular system, it is useful to be aware of the larger system of which it is part, or the smaller systems operating within it.
- Mechanisms of change (as in realist evaluation) may be taking place at a higher or lower level to the one where an intervention is taking place.



MULTIPLE SCALES AND LEVELS

Actors and interactions in complex systems can operate across scales and levels. For this reason, systems must be studied and understood from multiple perspectives simultaneously.

Examples

- Land managers may operate at a local scale on their own property, which is embedded in a catchment in a particular climatic zone subject to global change and within the context of local communities, local and national and trans-national governance and global economic trends.
- Interventions involving public sector organisations could involve national, regional and local levels as well as individual actors. Organisations such as the National Health Service are complex multi-level agencies in their own right.

Key points

- Usually more than one domain is required to fully understand a problem.
- We need to think broadly about systems at multiple scales and fields as properties or dynamics of one scale often feed up or down to affect other domains.

UNKNOWNS

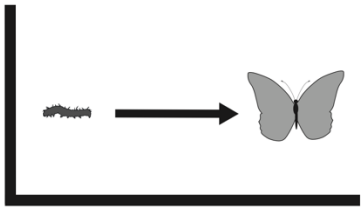
Because of a complex system's nonlinear causal structure and the number of interactions between its components as well as with the system's wider context, there are likely to be many factors which influence (or have the potential to influence) a system of which we are not aware. The inevitable existence of such unknowns mean we often see unexpected indirect effects of our interventions.

Examples

- A powerful social grouping operating in a policy area not anticipated by a policy maker.
- An undiscovered plant in a rainforest with numerous potential health applications.

Key points

- Expect the unexpected.
- Be prepared to learn as the system unfolds it will become apparent that it might influence or be influenced by completely unexpected things.
- A new technology might enable a fundamental change, leading to widespread social effects.



CHANGE OVER TIME

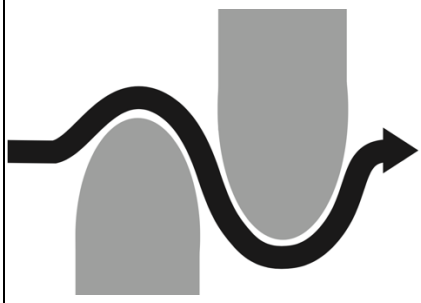
Complex systems inevitably develop and change their behaviour over time. This is due to their interconnectedness and adaption, but also the fact that these systems are usually out of equilibrium and are continuously changing.

Examples

- Ecosystems undergo succession over time: e.g. from annual plants, to scrub, to woodland.
- Social norms evolve over time.

Key points

- We cannot automatically assume that complex systems have reached a stable state.
- Do not rely on the system being the same in the future.



ADAPTATION

Components or actors within the system are capable of learning or evolving, changing how the system behaves in response to interventions as they are applied. So, for example, in social systems people may communicate, interpret and behave strategically to anticipate future situations. In biological systems, species will evolve in response to change.

Examples

- When bacteria evolve to become resistant to antibiotics.
- When an individual or organisation finds a way to circumvent a new tax or regulation, for example developing new chemical compositions that circumvent safety regulations

Key points

- The rules of the game change as you play it.
- We have to be prepared to adapt our interventions in response to how the system reacts to previous input.
- We should be aware of the pressures to adapt that we are putting in place in systems.
- We also need to be prepared for individuals - and systems - to adapt in response to an intervention in ways we didn't anticipate.

UNPREDICTABILITY

A complex system is fundamentally unpredictable. The number and interaction of inputs, causes, mechanisms and feedbacks mean it is impossible to accurately forecast system behaviour with precision. Random noise can have a large effect. Complex systems are fundamentally unknowable at any point in time - i.e. it is impossible to gather, store and use all the information about the state of a complex systems.

Examples

- In the economy and other systems, it is impossible to know the intentions and interactions of all actors.

Key points

- We can't forecast the future, instead we must explore uncertainty with rigour.
- Predictive models will always be limited in complex systems, however they can be used to explore and compare potential scenarios, and system behaviours.
- Precise prediction is impossible in the long term.

ANNEX III: TOOLS AND RESOURCES

The table below collates the tools and resources mentioned in this framework together in one place. Where particular methods are mentioned, these are examples and not the only methods you should consider. For more on selecting methods, see the chapter entitled "Designs, approaches, methods and tools" in the HM Treasury Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation (details below). Also see *Choosing Appropriate Evaluation Methods Tool* (Befani, 2016), details below.

	Resources to consult	**Why**
UNDERSTANDING	ANNEX I: A VISUAL GUIDE TO UNDERSTANDING COMPLEXITY FOR DEFRA	An illustrated list of characteristics and behaviours that complex systems in Defra might exhibit, with key points and examples. Use it to help identify where and how complexity might be affecting your system.
	Theory of change From: betterevaluation.org	A theory of change describes how an intervention is understood to contribute to a chain of outputs and outcomes leading to its resulting impacts.
	Participatory systems mapping From: cecan.ac.uk	A systems map is a graphical representation of the components in a system and the causal relationships between them. Participatory systems mapping is a participative approach for building and analysing systems maps; it can help to build a common understanding of system complexity across different stakeholder communities and enable a structured means of communicating the learning around these issues. Participatory systems mapping is part of a wider family of approaches which includes causal loop analysis, systems dynamics, conceptual mapping and group model building.
	Stakeholders and experts with experience of the system, both inside and outside of Defra	Speak to these individuals to better understand the system and to help anticipate unexpected system behaviour. E.g. those who are currently or were previously involved in the design, implementation or evaluation of policies, programmes or pilots in similar or relevant areas.
	Cultural consensus analysis Cultural Consensus Theory: Applications and Frequently Asked	Cultural consensus analysis can help to examine the extent to which stakeholder groups share similar mental models of the system, resources, and the interactions and processes between these components.

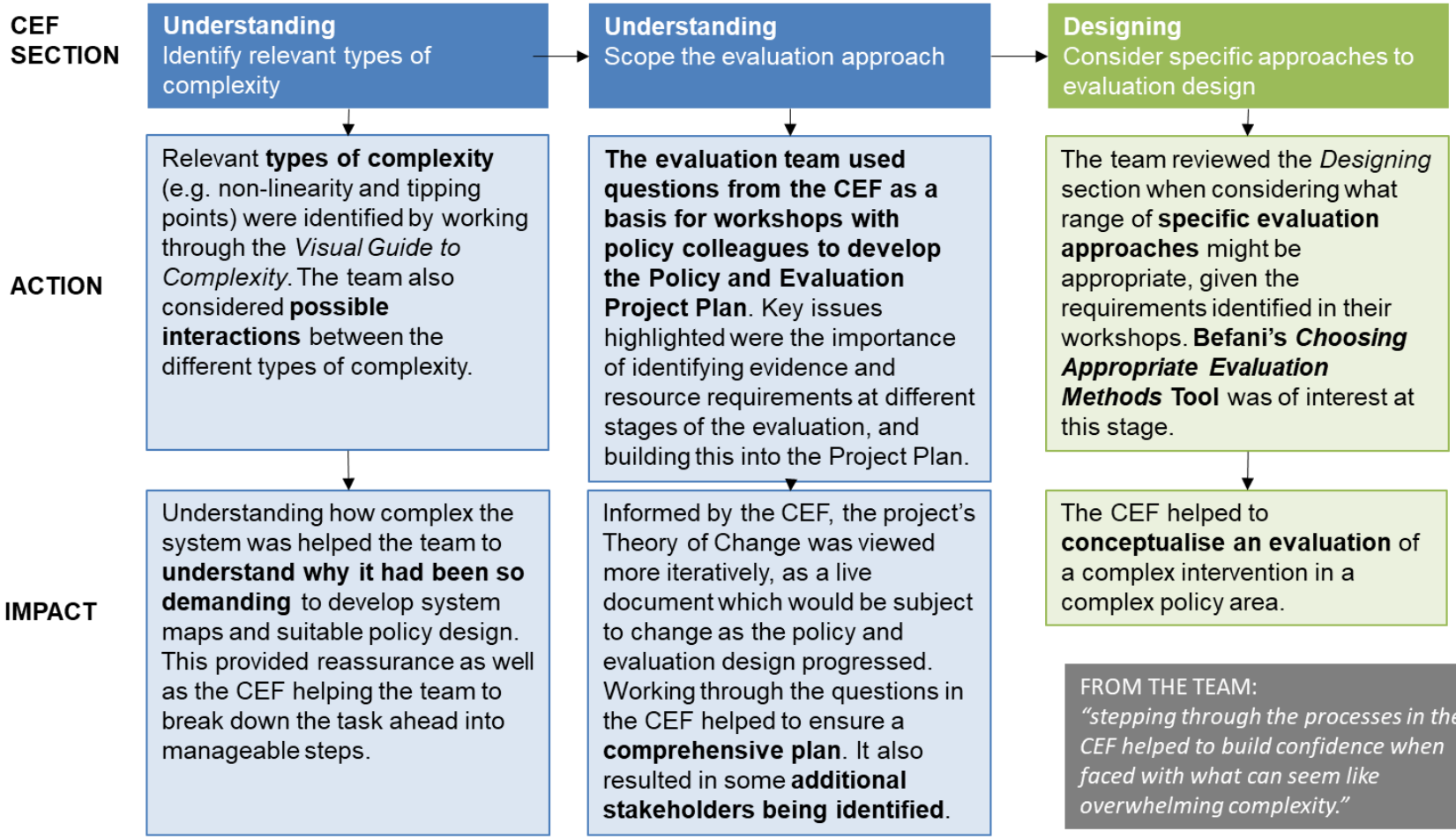
	Questions by Susan C. Weller	
DESIGNING	Chapter “Designs, approaches, methods and tools” in *Magenta Book 2020* Supplementary Guide on complexity *: Handling Complexity in Policy Evaluation*	There is no simple way to select the best evaluation design for complexity. The choice will depend on the complexity characteristics of the system, evaluation purpose and the feasibility of the available designs and methods. This chapter of the Magenta Book Complexity Supplementary Guide on complexity provides further guidance on selecting designs.
	Appendix 1 “Overview of available approaches and methods” in *Magenta Book 2020 Supplementary Guidance: Handling Complexity in Policy Evaluation*	There are a number of tools, methods or approaches that can work well with complexity. Each has its own strengths and weaknesses with respect to complexity; these are outlined here.
	Befani, B. Choosing appropriate evaluation methods tool From: bond.org.uk	The Choosing Appropriate Evaluation Methods tool is an accessible aid to help you understand evaluation methods and choose the right ones for your purposes.
	Stakeholders in other teams and departments with evaluation expertise	Stakeholders in other teams and government departments can be a valuable source of information and expertise with respect to the possible range of evaluation approaches and their relative merits.
	Agent-Based Models From: cecan.ac.uk	Agent-Based Models can handle feedbacks and detailed interactions.
	Bayesian networks From: cecan.ac.uk	Bayesian networks can combine both quantitative data and human perception.
	Qualitative Comparative Analysis (QCA) From: cecan.ac.uk	QCA has been used by the Environment Agency to navigate the complexities involved in waste crime policy interventions.
	CECAN syllabus From: cecan.ac.uk	The CECAN syllabus highlights a range of essential reading for those interested in the evaluation of complex policy and programmes. It includes sections on Qualitative Comparative Analysis, Systems Mapping and Agent-Based Modelling.

	Past evaluations	Evaluations of past programmes are a valuable source of evidence and analysis for evaluating current and future programmes.
	Participatory evaluation From: betterevaluation.org	Participatory evaluation can facilitate spaces for diverse groups of stakeholders to come together to learn and collaborate with each other and share experiences by improving their ability to interact and appreciate different actor perspectives.
	Participatory scenario modelling	Participatory scenario modelling is tool for integrating ecological simulations with participatory approaches, and enabling stakeholders to navigate complex trade-offs, promote discussion and identify long-term management strategies.
EMBEDDING	ANNEX I: A VISUAL GUIDE TO UNDERSTANDING COMPLEXITY FOR DEFRA	An illustrated list of characteristics and behaviours that complex systems in Defra might exhibit, with key points and examples. Use it to frame and inform discussions with stakeholders.
	Process evaluation & Guidance on process evaluation From: gov.uk	Process evaluation can be a valuable way in which lessons can be fed back quickly to the policy process.
	Participatory systems mapping From: cecan.ac.uk	Participatory systems mapping can help to build a common understanding of system complexity across different stakeholder communities and enable a structured means of communicating the learning around these issues. It is part of a wider family of approaches which includes causal loop analysis, systems dynamics, conceptual mapping and group model building.
	Participatory co-engineering From: ecologyandsociety.org	Participatory co-engineering can be used to look at the interactional behaviours between stakeholders and aid in collective decision-making processes.
	Knowledge co-production approaches	Knowledge co-production approaches identify stakeholder configurations (i.e. the specific roles stakeholders play in the system), and opportunities to bring different actors together in a shared space to work collaboratively, exchange ideas and experiences and jointly plan future management strategies, increasing the likelihood of more consensual and integrated decision-making.

ANNEX IV: USER CASE STUDIES

CASE STUDY 1: USING THE CEF TO CONSIDER MULTIPLE LAYERS OF COMPLEXITY

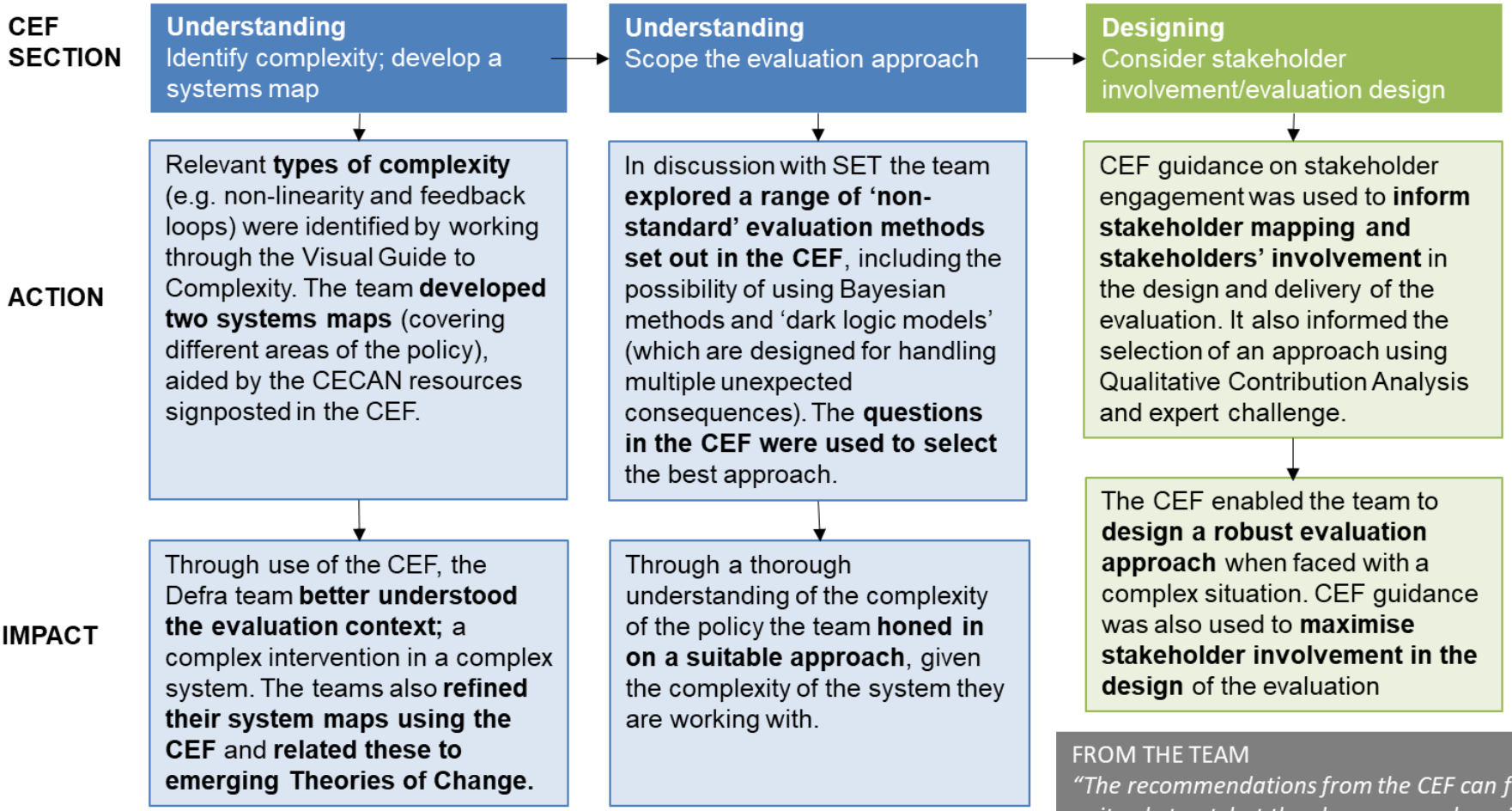
Evaluation Stage: Early stages of an evaluation design, including scoping and conceptualisation
Type of Evaluation: A complex intervention in a complex policy area
Use of CEF: The team held a session with colleagues from the Strategic Evaluation Team (SET), where they used the *Understanding* section of the CEF to identify complexity characteristics. They then also used the CEF to design workshops.



FROM THE TEAM:
"stepping through the processes in the CEF helped to build confidence when faced with what can seem like overwhelming complexity."

CASE STUDY 2: USING THE CEF TO CONSIDER OTHER EVALUATION METHODS

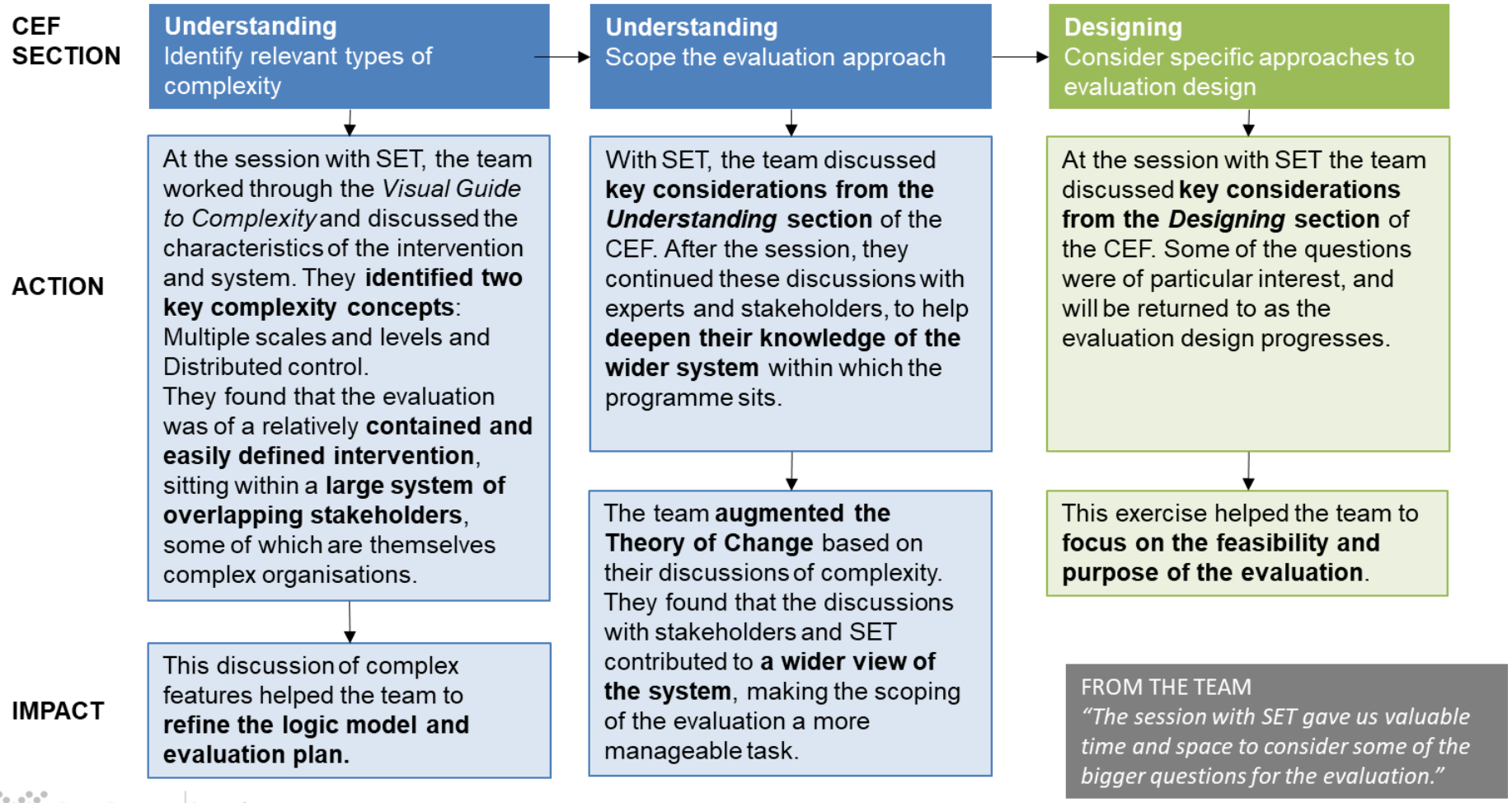
Evaluation Stage: Early stages of an evaluation design, including scoping, conceptualisation and choosing methods
Type of Evaluation: A complex intervention in a complex system
Use of CEF: Following an initial discussion with the Strategic Evaluation Team (SET), the team used the CEF to inform systems and stakeholder mapping; it also highlighted a range of evaluation methods that could be used widening the team's understanding of potential approaches.



FROM THE TEAM
"The recommendations from the CEF can feel quite abstract, but they become much easier to interpret when applied to a real-life situation."

CASE STUDY 3: A STRAIGHTFORWARD INTERVENTION WITHIN A COMPLEX SYSTEM OF OVERLAPPING STAKEHOLDERS

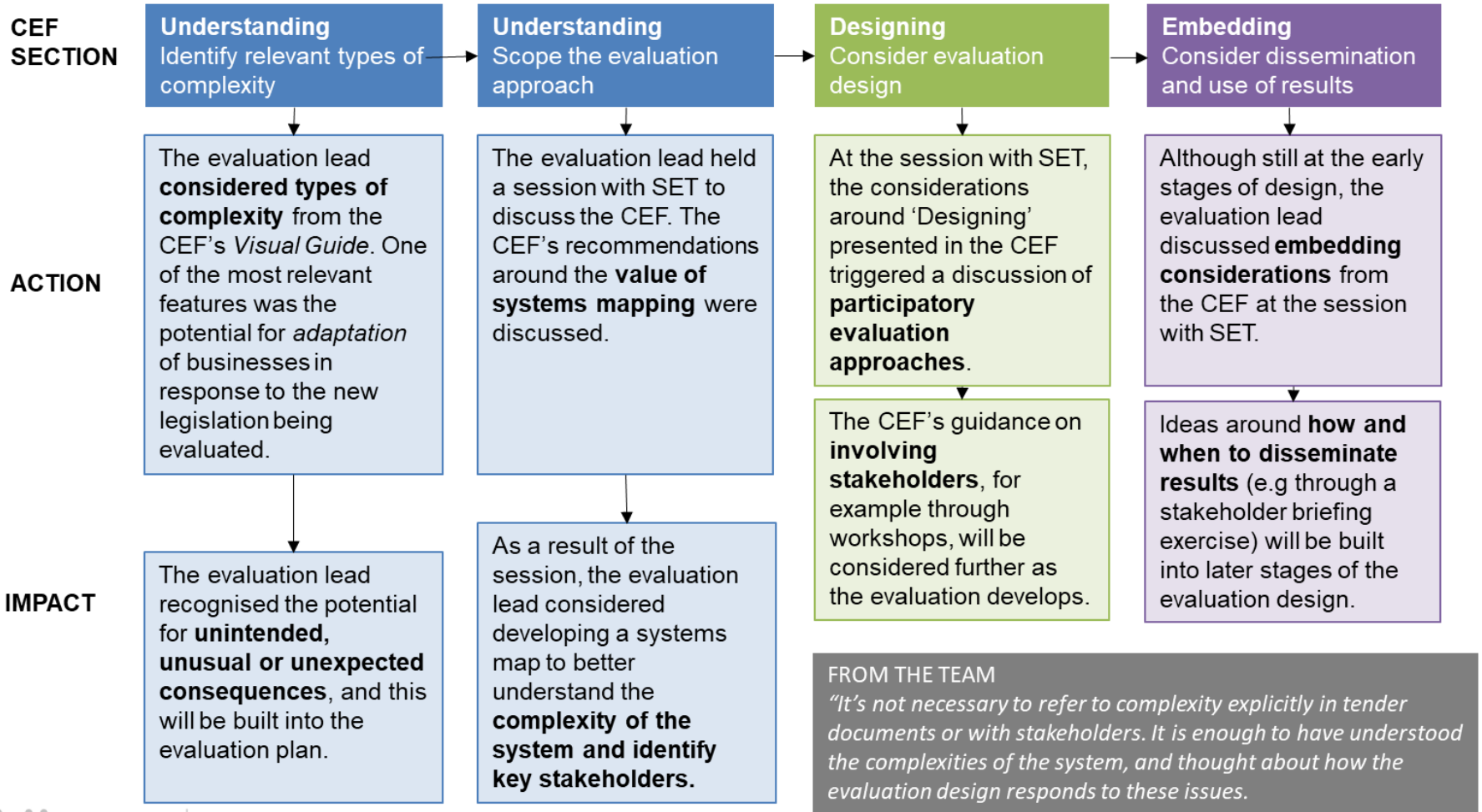
Evaluation Stage: Beginning of the evaluation, when scope and evaluation purpose were still being defined
Type of Evaluation: A relatively simple intervention operating in a complex system
Use of CEF: The team held a session with the Strategic Evaluation Team (SET) to discuss a complexity-aware approach to the evaluation. The CEF was used to help the team consider the 'big questions' for the evaluation around scope and purpose.



FROM THE TEAM
"The session with SET gave us valuable time and space to consider some of the bigger questions for the evaluation."

CASE STUDY 4: AN ADAPTIVE SYSTEM

Evaluation Stage: Early stages of an evaluation design. The evaluation lead was scoping the evaluation, but also considering options for methods and possible strategies for embedding.
Type of Evaluation: An evaluation of an adaptive system
Use of the CEF: A session working through the CEF with the Strategic Evaluation Team (SET) was held to discuss the implications of complexity

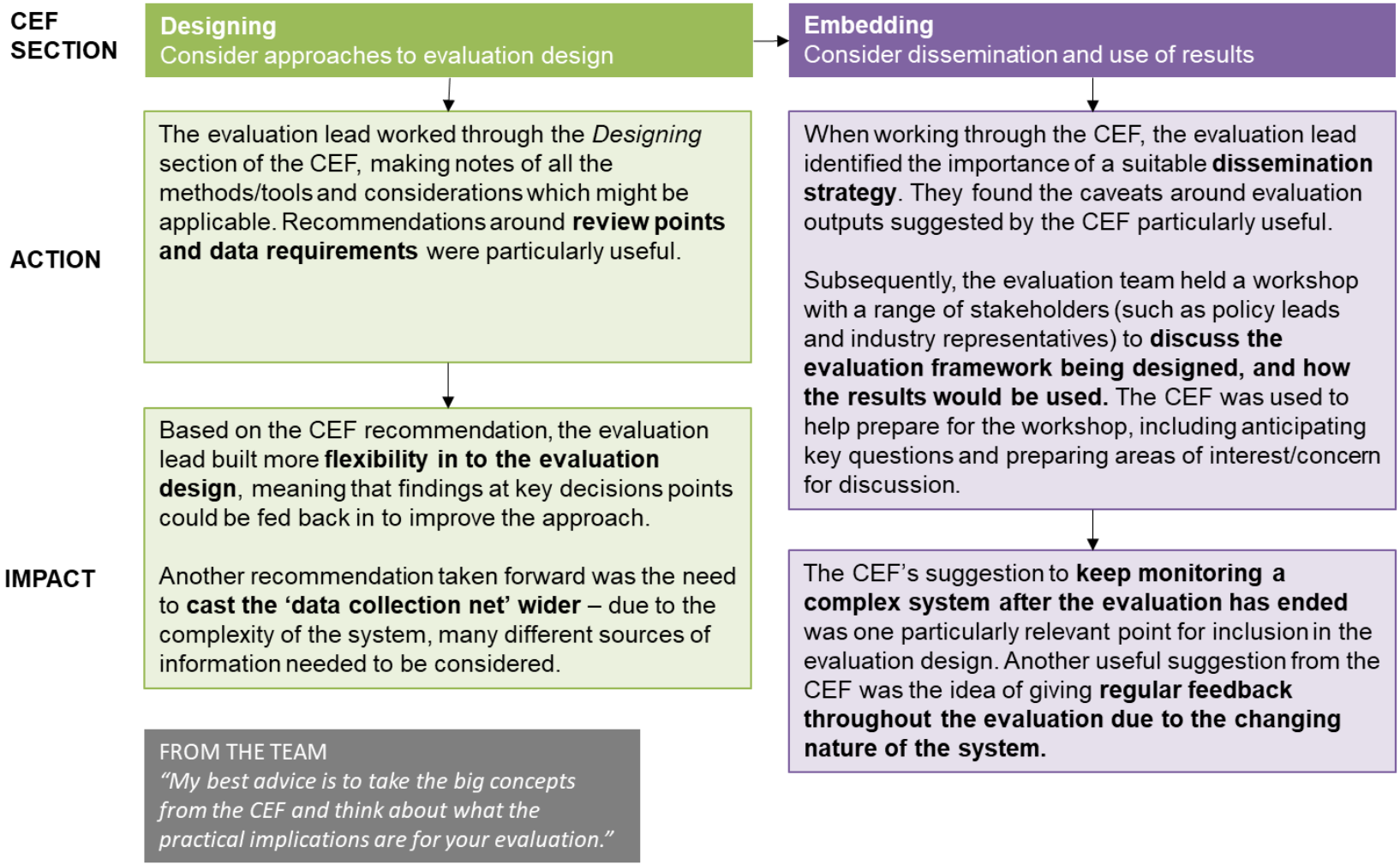


FROM THE TEAM

"It's not necessary to refer to complexity explicitly in tender documents or with stakeholders. It is enough to have understood the complexities of the system, and thought about how the evaluation design responds to these issues.

CASE STUDY 5: FOCUS ON EMBEDDING COMPLEXITY

Evaluation Stage: A Defra evaluation lead designing a monitoring and evaluation framework for a set of new policy interventions.
Type of Evaluation: A range of different interventions within a complex system.
Use of the CEF: Independent use of the CEF to consider the relevance of complexity to an existing evaluation plan.



ANNEX V: HOW THIS FRAMEWORK WAS DEVELOPED

This framework has been commissioned by Defra and delivered by CECAN Ltd. It builds on an existing body of work by CECAN[13] and others in the fields of complexity science and policy evaluation. This framework has been tailored to fit Defra's specific needs and context through five meetings and workshops[14], a Rapid Evidence Review and a series of interviews with Defra Group evidence leads, policy leads and commissioners of evaluation which took place between January and April 2019.

INTERVIEWS

Ten interviews were held with target users of the framework and other key stakeholders from across Defra and two arm's-length bodies, representing different roles and aspects of Defra evaluation activities, including commissioning, undertaking and using the results of evaluation. Collectively, interviewees covered a broad range of Defra's policy domains including: natural environment; environmental land management; farming; marine and fisheries; resources and waste; environmental regulation and enforcement; and official development assistance (ODA) (overseas aid). Interview topics addressed a range of issues relevant to the design of this framework, including:

- The types of complexity that arise in Defra's policy domains and the implications for Defra's work in these areas.
- The role evaluation plays in helping to address complexity in policy development and implementation and how this might be enhanced in Defra.
- The evaluation culture in Defra and any barriers and enablers to the uptake of complexity-appropriate evaluation.
- Useful current resources for evaluating and navigating complexity.

RAPID EVIDENCE REVIEW

A Rapid Evidence Review explored literature from a range of academic and non-academic sources to inform this study. Searches sought to identify complex system properties, behaviours and challenges within journals whose breadth of output cut across Defra policy areas. Searches returned approximately 3,500 articles, which were screened for eligibility based on a range of factors including: language (English); geography (predominantly the UK and EU, although other regions were still included if the focus of the article was

[13] The Centre for the Evaluation of Complexity Across the Nexus www.cecan.ac.uk
[14] Comprising: 2 CECAN Ltd workshops; 2 meetings with Defra's Strategic Policy Evaluation and Social Research (SPESR); and a further workshop with SPESR, potential users of the Complexity Evaluation Framework and other key stakeholders from Defra and its agencies.

relevant); focus (the complexity of a given system and its implications both conceptually, methodologically, and for governance, management and policy); and publication date (since 2009). Analysis of the resulting 172 articles comprised a thematic assessment of the abstracts, which produced a set of high-level emergent thematic areas drawn from the commonalities of issues identified in the articles. This evidence-base has been used to shape the overall way the CEF has been framed, and has shed light on the implications of complexity for the theory and practice of policy evaluation, the nature and challenges of policy in Defra, and the range and scope of issues within the Defra policy remit. In addition to the materials included in the review, other sources have been drawn upon to inform the development of this framework, including CECAN's Annex to the HM Treasury *Magenta Book* on complexity and evaluation, CECAN's own body of research in this area and previous scoping reviews[15].

SECOND PUBLICATION OF THE CEF

The first published version of the CEF was evaluated by an independent evaluator, Steer Economic Development (Steer-ED), from October 2019 to April 2020. Steer-ED's evaluation made recommendations to help maximise the value of the CEF, in response to interviews and observations conducted by Steer-ED with users of the first published version of the CEF. A separate published report provides details of Steer-ED's evaluation methodology, findings and recommendations.

This second version of the CEF incorporates changes agreed between Steer-ED and Defra, based on the evaluation findings. These include revisions, minor deletions and the addition of new content (for example user case studies, new introductory text, and additional guidance for evaluation commissioners).

[15] Watson, B. et al., (2014)

ANNEX VI: REFERENCES

Baker, J., (2015). *Evaluating Environmental Interventions: Challenges, Options and Examples (EEICOE): Methodological inspiration for environmental evaluators.* Defra, London (forthcoming)

Befani, B. (2016), *Choosing appropriate evaluation methods* [tool], London: Bond. Online at: https://www.bond.org.uk/resources/evaluation-methods-tool [last accessed 21/5/19]

BetterEvaluation, (2014), *Scope of Work* [web page]. Online at: https://www.betterevaluation.org/en/evaluation-option/scope_of_work [last accessed 21/5/19]

BetterEvaluation, (no date), *Participatory Evaluation* [web page]. Online at: https://www.betterevaluation.org/en/plan/approach/participatory_evaluation [last accessed 21/5/19]

Bicket, M., Christie, I., Gilbert, N., Hills, D., Penn, A., Wilkinson, H., (2020). *Magenta Book 2020 Supplementary Guide: Handling Complexity in Policy Evaluation.* HM Treasury. Online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879437/Magenta_Book_supplementary_guide._Handling_Complexity_in_policy_evaluation.pdf [last accessed 7/7/2020]

Boehnert, J., Penn, A., Barbrook-Johnson, P., Bicket, M., Hills, D., (2018). *The Visual Representation of Complexity – Definitions, Examples and Learning Points* [poster], available at https://www.cecan.ac.uk/sites/default/files/2018-07/JB%20online%20pdf%20The%20Visual%20Communication%20of%20Complexity%20-%20May2018%20-%20EcoLabs.pdf [last accessed 21/5/19].

Boyd, I., (2015), *Evaluating Complex Policy Interventions in Defra* [online presentation slides]. Online at: https://esrc.ukri.org/files/funding/funding-opportunities/centre-for-evaluating-complexity/cec-defra-presentation/ [last accessed 21/5/19]

Boyd, I., (2018). Policy, evaluation and implementation, in *The Evaluator*, Autumn 2018, pp6-7, UK Evaluation Society.

Byrne, D., (2016). *Qualitative Comparative Analysis: a pragmatic method for evaluating intervention. CECAN Evaluation and Policy Practice Note (EPPN) No. 1 for policy analysts and evaluators.* CECAN, Surrey. DOI: 10.15126/00850609 [last accessed 21/5/19]

CECAN, (2017). *Evaluation of Complex Policy and Programs: A CECAN module for future policy analysts and evaluators*. Version 1.0. Online at: www.cecan.ac.uk/resources [last accessed 21/5/19]

CECAN (2019) *Participatory Systems Mapping: a practical guide*. Online at: https://www.cecan.ac.uk/sites/default/files/2019-03/PSM%20Workshop%20method.pdf [last accessed 16/8/19]

CECAN, (various dates). *CECAN Policy and Practice Note Series (CECAN EPPNs)*. Online at: www.cecan.ac.uk/resources [last accessed 21/5/19]

Centre for Disease Control and Prevention, (2018). *Evaluation Briefs - Developing Process Evaluation Questions* [web page]. Online at: https://www.cdc.gov/healthyyouth/evaluation/pdf/brief4.pdf [last accessed: 21/5/19]

Cinner et al., (2012). Comanagement of coral reef social-ecological systems. *Proceedings of the National Academy of Sciences Apr 2012*, 109 (14) pp5219-5222; DOI: 10.1073/pnas.1121215109. Online at: https://www.pnas.org/content/109/14/5219 [last accessed 21/5/19]

Cisneros-Montemayor, A.M., Singh, G.G., Cheung, W.W.L., (2018). A fuzzy logic expert system for evaluating policy progress towards sustainability goals. *Ambio*, Springer Netherlands, Volume 47, Issue 5, pp595–607. https://doi.org/10.1007/s13280-017-0998-3 [last accessed 21/5/19]

Daniell, K. A. et al., (2010). Co-engineering participatory water management processes: theory and insights from Australian and Bulgarian interventions. *Ecology and Society* 15(4): 11. Online at: http://www.ecologyandsociety.org/vol15/iss4/art11/ [last accessed 21/5/19]

Davidson, J. L., van Putten, I. E., Leith, P., Nursey-Bray, M., Madin, E. M., Holbrook, N. J., (2013). Toward operationalizing resilience concepts in Australian marine sectors coping with climate change. *Ecology and Society* 18(3): 4. http://dx.doi.org/10.5751/ES-05607-180304 [last accessed 21/5/19]

Daw, T. M. et al., (2015). Evaluating taboo trade-offs in ecosystems services and human well-being. *PNAS Proceedings of the National Academy of Science of the United States of America.* 112 (22) pp6949-6954. Online at: https://doi.org/10.1073/pnas.1414900112 [last accessed: 21/5/19]

Defra, (2018). *A Green Future: Our 25 Year Plan to Improve the Environment - Annex 1: Supplementary evidence report*. London. Online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673492/25-year-environment-plan-annex1.pdf [last accessed 21/5/19]

Defra, (no date). *Policy Design Framework*. [internal document]

Defra and the Environment Agency, (2018). *Our waste, our resources: a strategy for England*. London, p143. Online at https://www.gov.uk/government/publications/resources-and-waste-strategy-for-england [last accessed: 21/5/19]

Djenontin, I.N.S., Meadow, A.M., (2018). The art of co-production of knowledge in environmental sciences and management: lessons from international practice.

Environmental Management. 61(6): pp885-903. Online at: https://doi.org/10.1007/s00267-018-1028-3 [last accessed 21/5/19]

Faulkner, L., Brown, K., Quinn, T., (2018). Analyzing community resilience as an emergent property of dynamic social-ecological systems. *Ecology and Society* 23(1):24. Online at: https://doi.org/10.5751/ES-09784-230124 [last accessed: 21/5/19]

Fujitani, M. L., Fenichel, E. P., Torre, J., Gerber, L., (2018). Synthesizing ecological and human use information to understand and manage coastal change. *Ocean and Coastal Management*, Elsevier. https://doi.org/10.1016/j.ocecoaman.2017.10.001 [last accessed 21/5/19]

Henriques, C., Garnett, K., Weatherhead, E.K., Lickorish, F.A., Forrow, D., Delgado, J., (2015). The future water environment — Using scenarios to explore the significant water management challenges in England and Wales to 2050. *Science of The Total Environment*, Elsevier Volumes 512–513, pp381-396. https://doi.org/10.1016/j.scitotenv.2014.12.047 [last accessed 21/5/19]

HM Treasury, (2020). *The Magenta Book - Guidance for evaluation*. London. Online at https://www.gov.uk/government/publications/the-magenta-book [last accessed 17/04/20]

Holdschlag, A., Ratter, B.M.W., (2013). Multiscale system dynamics of humans and nature in The Bahamas: perturbation, knowledge, panarchy and resilience. *Sustain Sci* 8(3): pp407-421. https://doi.org/10.1007/s11625-013-0216-6 [last accessed 21/5/19]

Ojeda-Martínez et al., (2009). A conceptual framework for the integral management of marine protected areas. *Ocean and Coastal Management*, Elsevier Volume 52, Issue 2, pp89-101. https://doi.org/10.1016/j.ocecoaman.2008.10.004 [last accessed 21/5/19]

Malawska, A. et al., (2014). Why do we need to integrate farmer decision making and wildlife models for policy evaluation?. *Land Use Policy*, Elsevier Volume 38, pp732-740. Online at: https://www.sciencedirect.com/science/article/abs/pii/S0264837713002202 [last accessed 21/5/19]

Malawska, A., Topping, C. J., (2017). Applying a biocomplexity approach to modelling farmer decision-making and land use impacts on wildlife. *J Appl Ecol*. 2018, 55, pp1445– 1455. Online at: https://doi.org/10.1111/1365-2664.13024 [last accessed 21/5/19]

Martone, R. G., Bodini, A., Micheli, F., (2017). Identifying potential consequences of natural perturbations and management decisions on a coastal fishery social-ecological system using qualitative loop analysis. *Ecology and Society* 22(1): 34. Online at: https://doi.org/10.5751/ES-08825-220134 [last accessed: 21/5/19]

Mathevet, R., Etienne, M., Lynam, T., Calvet, C., (2011). Water management in the Camargue Biosphere Reserve: insights from comparative mental models analysis. *Ecology and Society* 16(1): 43. Online at: http://www.ecologyandsociety.org/vol16/iss1/art43/ [last accessed 21/5/19]

Mulazzani, L., Trevisi, R., Manrique, R., Malorgio, G., (2016). Blue Growth and the relationship between ecosystem services and human activities: The Salento artisanal fisheries case study. *Ocean and Coastal Management*, Elsevier, Volume 134, pp120-128. https://doi.org/10.1016/j.ocecoaman.2016.09.019 [last accessed 21/5/19]

Parrott, L., Meyer, W. S., (2012), Future landscapes: managing within complexity. *Frontiers in Ecology and the Environment*, 10: pp382-389. doi:10.1890/110082 [last accessed 21/5/19]

Perry, R. I., Ommer, R. E., Barange, M., Werner, F., (2010). The challenge of adapting marine social–ecological systems to the additional stress of climate change. *Current Opinion in Environmental Sustainability*, Elsevier Volume 2, Issues 5–6, pp356-363 Online at: https://www.sciencedirect.com/science/article/pii/S1877343510001107 [last accessed 21/5/19]

Public Health England, (2018). *Guidance – Process Evaluation* [web page]. Online at: https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/process-evaluation [last accessed 21/5/19]

Reyers, B., Nel, J. L., O'Farrell, P. J., Sitas, N., Nel, D. C., (2015). Navigating complexity through knowledge coproduction: Mainstreaming ecosystem services into disaster risk reduction. *PNAS Proceedings of the National Academy of Science of the United States of America.* 112 (24) pp7362-7368. Online at: https://doi.org/10.1073/pnas.1414374112 [last accessed 21/5/19]

Rogers, K. H. et al, (2013). Fostering complexity thinking in action research for change in social–ecological systems. *Ecology and Society* 18(2): 31. http://dx.doi.org/10.5751/ES-05330-180231 [last accessed 21/5/19]

Rogers, P., (2014), *Theory of Change*, UNICEF. Online at: https://www.betterevaluation.org/sites/default/files/Theory_of_Change_ENG.pdf [last accessed 14/8/19].

Vennix, J. A. M., (1999). Group model-building: tackling messy problems. *System Dynamics Review*. Volume 15 Number 4, pp379-401. Online at: http://www.iwrms.uni-jena.de/fileadmin/Geoinformatik/projekte/brahmatwinn/Workshops/FEEM/Vennix_1999_Group_model_building.pdf [last accessed 21/5/19]

Vugteveen, P., Rouwette, E., Stouten, H., van Katwijk, M. M., Hanssen, L., (2015). Developing social-ecological system indicators using group model building. *Ocean & Coastal Management*. Elsevier. Volume 109, pp29-39. Online at:

https://www.sciencedirect.com/science/article/pii/S0964569115000587 [last accessed 21/5/19]

Watson, B., Watson, T., Elliott, B., Vanner, R., Shaw, B., Morris, S., (2014). *What range of methods is available to evaluate the implementation and impact of complex environmental policies on complex systems? A report to the Department for Environment, Food and Rural Affairs.* Policy Studies Institute. Defra, London. Online at: http://www.psi.org.uk/site/project_detail/REA_evaluating_impact_and_implementation_of_complex_environmental_policies [last accessed 21/5/19]

Weller, S. C., (2007). Cultural Consensus Theory: Applications and Frequently Asked Questions. *Field Methods*. Sage. Volume: 19 issue: 4, pp339-368. Online at: https://doi.org/10.1177/1525822X07303502 [last accessed: 21/5/19]