JNCC Report 737

Towards Indicators of Soil Health

Annex 2:

Complex Systems Review

Harris, M., Hoskins, H. and Way, L.

June 2023

For further information please contact:
Joint Nature Conservation Committee
Quay House
2 East Station Road
Fletton Quays
Peterborough
PE2 8YY
https://jncc.gov.uk/
Communications@jncc.gov.uk

This document should be cited as:
Harris, M., Hoskins, H. & Way, L. 2023. *Complex Systems Review. JNCC Report 737: Annex 2.* JNCC, Peterborough.
https://hub.jncc.gov.uk/assets/71cece04-eef3-4d34-b118-33ddad50912c#jncc-report-737-annex-2-complex-systems-review.pdf

Acknowledgments:
We would like to thank all JNCC and Environment Agency staff who were interviewed about their projects as part of the research for this report.

Evidence Quality Assurance:
This document is compliant with JNCC's Evidence Quality Assurance Policy
https://jncc.gov.uk/about-jncc/corporate-information/evidence-quality-assurance/

Summary

This document reports on a review of approaches used to produce indicators that capture complexity, to determine what could usefully be adapted and made use of during the development of an indicator of soil health, or other future projects. This took place through a series of interviews and a short literature review. Complexity covers both mathematical complexity (e.g. a network of many factors, interacting over a range of spatiotemporal scales) and conceptual complexity (e.g. subjectivities, factors that are difficult to define). The soil system is complex because it involves a large network of interacting biological, chemical, physical, and functional factors, as well as concepts that can be difficult to communicate to a non-specialist audience. Different actors also have different asks for soil and may therefore interpret a single 'health' indicator in different ways.

In terms of mathematical complexity, interviewed projects made use of a number of approaches to select both the inputs, interactions and outputs required, and the models or methods to be used to reach an assessment. Most projects took these kinds of decisions through a combination of literature reviews, expert input, project governance systems and iterative wider consultation. Factors considered important during these decision-making processes included data availability, whether the metric is widely used and accepted, alignment with other projects and stakeholder or policy priorities. A variety of modelling techniques can be applied, with BBNs particularly suiting data poor environments where integration of data sources and local knowledge is required. However, framework-based approaches can also be useful, especially in cases where a consistent assessment that cannot be based on a set methodology is required.

In terms of conceptual complexity, interviewed projects recommended communication as a key mitigating factor. This includes stakeholder engagement, clear and engaging presentation of outputs, and ensuring the clarity of definitions and scope. Within the presentation of final outputs, most projects chose to keep the complexity of the underlying system, data and modelling used hidden from the end user. This was often done by presenting a final figure that aggregated factors into a single metric, or a small set of metrics. All projects interviewed relied on an iterative process of building up knowledge as the project progressed; it was not expected that the complexity involved could be captured within the first attempt at putting together a model or assessment framework. This helped address issues around knowledge gaps and uncertainty. In terms of subjectivity, most projects either left interpretation to the user, provided flexibility in how outputs were presented, or created common standards of threshold values based on consensus among experts.

Interviewed projects varied greatly in terms of the development timeframes. However, none saw their work as complete and all saw potential for further development if funding allowed, again suggesting that an iterative development process is important.

Overall, projects involving complexity have many of the same requirements as simpler projects, only with more complex consequences if they go wrong. This report shows that there are many lessons that can be learned from previous projects (Figure 1). It is hoped that these can be made use of both in the development of an indicator of soil health, and future projects that also aim to produce indicators that capture aspects of other complex systems.

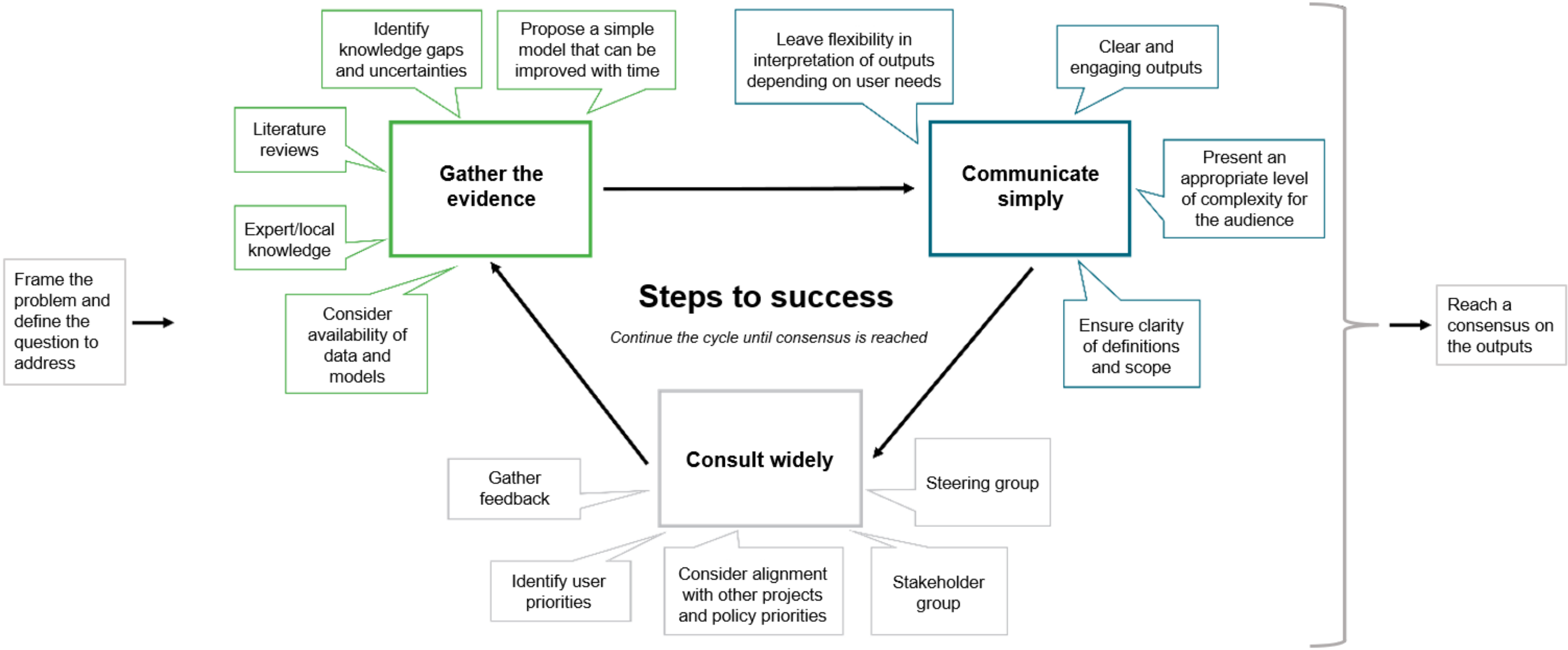

Figure 1. Steps to success. A graphical representation of lessons learnt from projects that were interviewed which are transferable to future projects.

Contents

1 Introduction

1.1 Context

This document reports on a review of approaches used to produce indicators that capture complexity, to determine what could usefully be adapted and made use of during the development of an indicator of soil health. The task was time-limited and did not aim to be exhaustive, but rather to identify a number of solutions developed by and lessons learnt from a selection of accessible previous projects, and wider complex systems theory. It was planned in recognition of the complexities involved in the soil system (see Section 1.3). Research was undertaken through both time-limited literature searches and a series of focused interviews with contacts who had previously undertaken projects considered 'complex' (see Section 1.2). Lessons learnt from this task will also be useful for and applicable to any other future project with a need to deal with complexity.

The main report synthesises transferable approaches that could be considered during development of a soil health indicator (or other similar projects) that were identified during the review, whilst Appendix 2 provides case studies detailing further information about three key projects interviewed, and a list of all other projects that were consulted.

1.2 What is complexity?

Complexity is a concept that is difficult to define. Most definitions within the scientific literature focus on the mathematical aspects of complexity through 'complex systems theory.' However, complexity can also refer to conceptual complications. For the purposes of this project, both aspects will be considered.

Mathematically complex systems are generally considered to have most (but not necessarily all) of the following attributes (Filotas *et al*. 2014; Johnson 2009; Ramos-Martin 2003; Mitchell & Newman 2001):

- A network of many factors, interacting over a range of spatiotemporal scales.
- Feedback loops.
- Adaptability to external factors.
- A mixture of ordered and disordered behaviour.
- Emergent phenomena (surprising outcomes that come from the behaviour of the system as a whole, rather than directed in any way, for example the patterns formed by starling murmuration or schools of fish).
- Are influenced by a wider environment.

Conceptual complexity could cover confusion arising from:

- Definitions that are hard to understand and communicate.
- An incomplete or uncertain knowledge of the system under study.
- Subjectivities, for example in interpreting factors such as 'health' or 'condition', which may differ depending on the user and use case.
- Challenges to presenting outputs in an understandable way.

1.3 Complexity in the soil system

The soil system is complex both mathematically and conceptually. It involves a large network of interacting biological, chemical, physical, and functional factors (Straton 2006). These are influenced by wider systems, such as the ecological, geological, atmospheric, and anthropogenic systems, including by differing land uses and land capabilities. There is great variety in the different soil types (and conditions for the same soil type), and the micro-organisms that inhabit them. Trying to understand every possible factor involved and every interaction between each factor would not be possible. There is also great variety in measurement options. Communicating definitions, relationships and concepts involved can be challenging. Different actors have different asks for soil and may therefore interpret a single 'health' indicator in different ways if not well explained, or if not presented in a way that is flexible to the possible use cases. For example, a farmer might consider soils of 'good' condition or health to be those that lead to the highest possible crop yields, whilst a conservationist might consider 'good' condition to be that which supports the greatest range of biodiversity in an area.

The key focus for this report was to summarise practical lessons learnt from other projects that could be applied to developing indicators within the soil system, or other complex projects in future. For readers interested in a more academic analysis of the synergies between complex systems theory and soil science specifically, this has been discussed directly in previous publications (Faybishenko *et al*. 2016; Ritz 2008; Phillips 1998).

2 Solutions and mitigations to issues that complexity brings to a project

2.1 Mathematical complexity

2.1.1 Identifying inputs, outputs, and interactions

Identifying and selecting which inputs, outputs, and interactions to use within a model or indicator can be complex due to the sheer number of possibilities for inclusion. Most projects took these kinds of decisions through a combination of literature reviews, expert input, project governance systems and iterative wider consultation. Factors considered important during these decision-making processes included data availability, whether the metric is widely used and accepted, alignment with other projects and stakeholder or policy priorities.

Literature reviews ranged considerably in their rigour and resource-intensity. Some projects took a light touch approach in this area due to a known lack of literature within the projects' geographic context, only investigating the few known sources they were directed to by locals. Others completed a full systematic review with academic design. In one example, this led to just over 300 relevant sources from which to record evidence for subsequent decision making. Most projects, due to time and resource constraints, opted for an intermediate approach, with literature being reviewed in a time limited manner, aiming to gain the best understanding of the literature possible within a set time period.

Expert input took a number of forms. Many projects made use of local knowledge, especially in areas where other data sources were poor. This was typically done through workshops, or in one case a survey. In some cases, local experts were asked to also record their confidence in the answers they were giving, to better understand the likely accuracy, precision and uncertainty of information obtained. Accepted methods for undertaking expert elicitation (e.g. DELPHI) are further referenced in the JNCC EQA policy. Many projects also contacted specific subject matter experts of relevance, such as academics, specialised government agencies or relevant industry actors. For example, one project aiming to model coastal storm surges consulted the UK Hydrographic Office for expert advice (see project six in the list of projects interviewed below). Often, the project team would put together a first draft conceptual model (simply using logic and brainstorming), which they then asked for input from academic experts or local stakeholders with specific geographic expertise, to modify and inform interactions in the model. This provided a solid framework for ensuring input received was of most relevance to the project aims and scope, whilst ensuring accuracy of understanding by the project team. It could also be broken down into parts, so experts were only commenting on the part of the model of most relevance to them. For example, in one project modelling natural capital in the UK Overseas Territory of St Helena (see (see project nine in the list of projects interviewed below)), the island's water company were asked to review the water-related section of the conceptual framework, the forestry group were asked about the forestry section, and government representatives for agriculture and farming were asked about production related services. In some cases, experts were also asked to provide numerical inputs for certain interactions, such as the degree of slope at which it is no longer possible to get machinery in for forestry and the stocking rates for livestock. Many of the conceptual models developed within interviewed projects were based on a network approach, but a 'Theory of Change' style conceptual model was also highlighted as a common form to take.

Projects generally had a governance system that consisted of a project team (who undertook work for the project), a steering group (who were the key customer for the outputs) and in many cases a wider stakeholder group (who were thought likely to also make use of or be affected by the outputs, and/or who could provide expert insight and knowledge to the geographic area or subject matter). In general, working decisions were taken by the project team. However, for higher level decisions, key options (alongside their advantages and disadvantages) were presented to the steering group to consider, in order to ensure their involvement in and support of final outcomes. Stakeholders had more varied engagement across projects, feeding in priorities and information at an early stage (which helped to design many of the end products to suit their needs), and giving feedback on final outputs.

Many of the projects viewed the selection of inputs, outputs, and interactions to include in the modelling as an iterative approach that could never be considered complete. Several of the projects interviewed were, or began with, a pilot study (typically lasting several months to a year). They therefore recognised the need for further consultation and improvements to be made in future (often over a period of several years) if the project continued. Some of the projects were more established and had already undergone such an iterative consultation process throughout many years, but still do not consider decisions to be final. One framework that aims to measure whether the marine environment is in 'Good Environmental Status' (see project eleven in the list of projects interviewed below) explained a process of outputs from initial work undergoing additional consultation through a series of testing rounds, bringing in wider international experts to review the work, engaging with a wider audience across policy and finally public consultation following publication. Each indicator therefore takes several years to develop, and additional indicators are being included with each assessment.

In terms of reasons behind the selection of which inputs, outputs, and interactions to include in modelling using these decision-making processes and governance systems, data availability was by far the most commonly mentioned factor. It was also generally the first factor thought of by interviewees and stated without hesitation. It is clear that without the relevant data to be able to determine values for inputs and their interactions with other factors within the model, it is not feasible to undertake the project. In some cases, this issue affected project scope. For example, one project modelling ecosystem services in Chilean vineyards (see project ten in the list of projects interviewed below) was unable to model the effects of management practices on fungal biodiversity due to data constraints. The resolution and frequency of update of the data was considered important as well – if these were not appropriate to the project's scope any data available could not be used anyway. Some projects had a requirement for using open data, which further affected data constraints. One project highlighted the importance of identifying and highlighting any gaps in the data, and what would be required to fill them in future (whilst making best use of currently available data in the meantime).

Alignment was considered another key factor by many projects. Metrics that were found to be widely used in the literature were typically considered more reliable and useful than unique or custom metrics. Several projects focused on using metrics that were based on international standards, as this implies reliability, rigour, and international comparability. For example, one of the indicators used within the Habitats Directive and UK Marine Strategy (see projects eleven and one in the list of projects interviewed below) was previously developed by the OSPAR Commission (which was an international collaboration between 15 governments). A number of projects were unable to answer the question of how they selected which inputs to use, as the approaches they used were based entirely on, or were adapted from, previously developed models, or set indicator methodologies (see Section 2.1.2). For example, one project modelling disaster resilience in the British Virgin Islands made use of the SciMap and SAGA models, which define the inputs required themselves (see project six in the list of projects interviewed below). One project highlighted the importance in their context of aligning the inputs selected with other related projects, to ensure the potential for comparability of results.

Stakeholder and policy priorities were considered particularly important when selecting the outputs to model. This input was generally received through the workshops and steering group sessions described above and was considered key to defining and scoping the task at an early stage. It also helped to provide a justification, rationale and use case for the work. In many cases, the selection of outputs helps narrow down the inputs required to those of relevance to the specific requirements of the project, rather than aiming to include all inputs to the whole system. In some cases, inputs were also selected based on stakeholder and policy priorities. For example, projects modelling the effects of management practices in the Brecon Beacons (see project two in the list of projects interviewed below) and in a viticultural region of Chile (see project ten in the list of projects interviewed below) were asked about the management practices that they already undertake or would be interested in trying, in order to model the effects of these on the ecosystem services of interest to each system.

2.1.2 Models and frameworks

Whilst a range of models and frameworks were used across the different projects, Bayesian Belief Networks (BBNs) were the most commonly used within the selection of projects interviewed. This may be due to an 'echo chamber' effect, whereby projects interviewed were already aware of each other, building on each other or involving researchers who worked on more than one of the projects within the selection. A BBN generally builds on a conceptual ecological model (a network of factors considered important to the system and their relationships to one another) but brings in numerical and probabilistic values in order to be able to start to estimate the effects of changes to inputs of the system and capture uncertainty. They can be set up using an R package. It is possible to add spatial data and run the model for each cell of a raster grid. It is also possible to use BBNs as an overall way to bring together other previously developed modelling approaches. For example, one project that modelled ecosystem services in Chilean vineyards (see project ten in the list of projects interviewed below) brought together the outputs from the InVEST Sediment Delivery Ratio model, a Species Distribution Modelling (SDM) based approach of estimating fire risk, and Soil and Water Assessment Tool (SWAT) modelling, into a BBN based framework to give an impression of the system overall.

BBN based approaches were seen to have a number of advantages over other modelling types. Firstly, they allow for the integration of different types of knowledge, including both quantitative data and stakeholder expertise, through the use of probabilistic factors. This was particularly important for projects with low data availability. Similarly, traditional models are harder to adapt for unique environments with unique ecological assemblages. BBNs allow for a more tailored approach and greater control over decisions around which inputs and outputs to include (see Section 2.1.1). The underlying conceptual model provides a good visual representation of what the BBN involves, whilst the more complicated mathematics is only introduced at a later stage. This is useful for stakeholder engagement and communication of outputs. However, it was also flagged that Bayesian approaches are often computationally intensive.

Other modelling techniques used within projects interviewed included multi-criteria analysis, SDM based approaches, and a variety of other previously developed models that were considered of relevance to the project's scope such as InVEST. It was highlighted that it is important to understand models that have been developed previously in the area, as making use of an existing model, or adapting an existing approach is significantly less time and resource intensive than creating something new. It also improves alignment and means that models have been tested prior to use. This is not possible in every case. For example, as described above unique environments are likely to require unique considerations within the modelling.

Previously developed models that may be of use in the project's context were largely identified through literature reviews (see Section 2.1.1), but sometimes also through expert input. As with the selection of inputs and outputs, availability of appropriate data was seen as the key factor in terms of whether a model could be applied within the context of a particular project. Therefore, the level of data a model required was also a key consideration. For example, one project was interested in modelling soil erosion (see project ten in the list of projects interviewed below). SWAT and InVEST were identified as options for this. However, SWAT needed more data inputs than InVEST, which in the geographic context of the project were not available. InVEST was also able to provide outputs with a higher data resolution. The project team therefore took the decision to use InVEST in this case. Approaches such as sensitivity analyses could also be useful to compare models where input data are available for both. Flexibility and the ability to adapt models based on additional information sources was another factor of importance to some. As with the selection of inputs and outputs, those that were most widely used within the literature were generally seen as more widely applicable than those that were less commonly cited. Other factors of importance were largely specific to each projects' context but were recorded through exclusion criteria and data extracted as part of the literature review process.

Some of the projects interviewed did not use modelling at all, but rather a framework-based approach. This was not overly prescriptive in the methods used but provided a way in which different indicators and measurements could be brought together to give a common assessment. For example, in Common Standards Monitoring (an agreed approach to the assessment of condition on statutory sites designated through UK legislation and international agreements across the four countries of the UK - see project seven in the list of projects interviewed below), guidance on undertaking assessments is provided, but the exact methods for doing so are not. This allows flexibility in the system, whilst allowing for comparisons. Features (the species, habitats and geological and geomorphological characteristics for which sites are protected) are qualitatively ranked into categories of 'favourable,' 'unfavourable,' 'partially destroyed' and 'destroyed.' Guidance on assessing each feature through a variety of attributes (for example, bird populations as a feature require monitoring of population size, population density and habitat extent) has been developed through a programme of work involving JNCC and the four country nature conservation bodies (Natural England, Natural Resources Scotland, Scottish Natural

Heritage and the Department of Agriculture, Environment and Rural Affairs, Northern Ireland). In this way, it is possible for the vast complexity of monitoring environmental status to be reduced to categories that are easy to communicate and can be compared across features and systems, whilst also providing a rich variety of detail and intermediate outputs that more expert users can investigate further where appropriate. The avoidance of explicitly modelling the system means it is not possible to use the framework to make predictions or understand the likely effects of management, but it does allow for an assessment of current status in a way that reduces complexity for the end user considerably.

As with selecting inputs, outputs and interactions, the selection of models or frameworks to use was seen as a process requiring ongoing and iterative wider consultation.

Although not used within any of the projects interviewed, a number of other modelling approaches were mentioned in the complex systems theory literature that could be considered by future projects. These included agent-based modelling, discrete event simulation, evolutionary computation, lattices and networks, game theory, dynamical systems, (Newman 2011; Clancy *et al*. 2008; Mitchell & Newman 2001).

2.2 Conceptual complexity

2.2.1 Communication

Every one of the projects interviewed highlighted communication as a key solution to issues stemming from the subject area's complexity. This included stakeholder engagement, clear and engaging presentation of outputs, and ensuring the clarity of definitions and scope.

Stakeholder engagement was seen as a key part of the project process. It was recommended that this should begin at an early stage of the project and continue throughout. Stakeholders were seen as both essential sources of knowledge to feed into the project (see Section 2.1), but also key at reducing complexity by narrowing down local priorities and providing ideas for likely use cases of the outputs, which help the project team to focus on the areas of most interest to users, rather than get lost in the detail and complexity that the subject area could present if viewed without prioritisation. There was a perception that if stakeholders were involved in the decision-making processes (for example around which input factors to include and prioritise, or what they wanted output presentation to look like), they would also be much more likely to make use of the outputs due to a feeling of ownership over the process. It was seen as valuable to set aside resource to identify a target audience and build up a user base at the beginning of the project.

Most of the projects interviewed presented their outputs through some form of app, dashboard, or other interactive tool. Often this was R Shiny, although again this is likely due to the 'echo chamber' effect mentioned earlier. In some of these cases, users could turn on and off different inputs, and customise and run models within the dashboard without technical knowledge. For example, one project modelling fire risk in the Brecon Beacons (see project two in the list of projects interviewed below) under different management options allowed users to change the land cover type or grazing level in a location of their choice, to see how this would impact fire risk in the area. In another project modelling disaster resilience in the British Virgin Islands (see project six in the list of projects interviewed below), users could specify a storm category, storm path and rainfall data to see what the effects of such an event could be on flooding. In addition, users could change the underlying habitat data, and/or the friction values of those habitats (representing the condition of the habitats to providing flood mitigation), which allowed users to consider scenarios and assist in management and planning.

Within the presentation of final outputs, most projects chose to keep the complexity of the underlying system, data and modelling used hidden from the end user. This was often done by presenting a final number that aggregated factors into a single metric. For example, in the Brecon Beacons project (see project two in the list of projects interviewed below), a single aggregate metric of burn risk was presented, rather than any intermediate factors such as ignition risk. Similarly, the British Virgin Islands project (see project six in the list of projects interviewed below) created lots of intermediate outputs such as runoff risk, but only presented a combined flood risk value in the final map. In several projects, a single final output was a key policy request. Whilst in some other projects presenting a single final output was seen as infeasible, they generally adhered to the concept that simplifying final outputs to the greatest degree possible was desirable. For example, the Marine Strategy Framework Directive work (see project one in the list of projects interviewed below) considered that aggregating all indicators used would be the overall aim but is not possible. They therefore use a comparable framework to assess and report on all components of the marine system separately. This is seen as providing greater clarity than a single overarching aggregate. Similarly, a project identifying biodiversity indicators that could be used for nature-based solutions projects (see project five in the list of projects interviewed below) concluded that, due to the many and varied aspects of biodiversity, it was not possible to capture all relevant information in a single metric. A species-based metric and a habitat-based metric were therefore both included and presented separately.

One method many projects used to ensure that the final presentation remained simple, but the valuable depth and detail of information produced by the project could still be accessed by those who were interested, was to present intermediate outputs. In some cases, this formed a part of the application or dashboard, but was not the default setting. For example, in a project modelling marine natural capital, it was possible to click on each node presented to get further information about the linkages and data behind it. This information was also divided up into units such as ecological units, to ensure that the complexity was viewed in manageable units, rather than as the full network of nodes which would be intimidating to the user. In other cases, the more detailed information was presented in an accompanying technical report rather than through the main reporting or presentation mechanism. For example, in condition assessments for the Habitats Directive (see project eleven in the list of projects interviewed below), many intermediate outputs were reported on as part of the assessment process that needed to be submitted as evidence behind the overall estimates, but a matrix was used to determine the overall conclusion which was reported as the final headline estimate for each habitat. Similarly, there are many intermediate outputs that must be recorded throughout the process of monitoring a feature following the Common Standards Monitoring Guidance (see project seven in the list of projects interviewed below). Each feature has a list of attributes (for example, bird populations require monitoring of population size, population density and habitat extent). Each of these is associated with a target and a method of assessment. Outputs from each of these would be recorded at an intermediate stage, to inform the decision on the feature condition overall. One interviewee reflected that the aggregated headline results are most useful for reporting purposes, whilst the intermediate outputs are more likely to be used for understanding and informing specific policy responses.

Clarity of communication was also seen as a crucial aspect to the success of a complex project. Definitions should be as intuitive as possible, with explanations where there is any possibility for misinterpretation. Assumptions should be highlighted and made transparent to the user in order to ensure the information produced is not misinterpreted and misused. With many of the projects acting as decision support tools, there was often a need to communicate what this should mean in practice, to avoid the misinterpretation that the tool should replace decision making on the ground rather than provide additional information to support this.

Clarity of communication around project scope in particular was also seen as essential. Understanding exactly what the customer wants (e.g. creating a list of questions that they hope the work will answer) is key to defining this and designing a successful project. Ensuring the project team and all stakeholders are aware of project scope from an early stage and that this is well defined is key to preventing scope creep, which can often be an issue when dealing with complexity. Communicating when projects are at an early stage or have produced an interim product was also seen as important, with most projects asking for feedback from users at this point.

One project highlighted an additional layer of communication that was necessary if using contractors. It is important that their methods can be replicated at a later date, so transparency is important. Recording the decision making is just as important as recording the mechanics of what was done to ensure thought processes are not lost.

2.2.2 Dealing with incomplete knowledge and uncertainty

All projects interviewed relied on an iterative process of building up knowledge as the project progressed; it was not expected that the complexity involved could be captured within the first attempt at putting together a model or assessment framework. Most began with an initial conceptual model of how the system worked. Additions to this were made through literature reviews and expert input (see Section 2.1). Often this was then simplified again to best fit with stakeholder priorities, data available and for simplicity of presentation. Many interviewees seemed confused when asked how they dealt with incomplete knowledge of the system, as they considered it impossible to know everything about a system and the relationships within it. Identifying and being transparent to clearly communicate evidence gaps was seen as important but trying to ensure a complete knowledge of the system was not seen as pragmatic. Adding knowledge at a later date following incomplete knowledge in the early stages of a project was not seen as an issue, but rather a necessity. Issues often emerge when the model is first run which need to be resolved through further research, so undertaking this step early on in the project was advised.

Offshore marine monitoring for the Habitats Directive reporting was an example of a project that faced particularly significant issues around incomplete knowledge and a lack of data availability (see project eleven in the list of projects interviewed below). This was mitigated by the fact that the assessment is open to whatever data, monitoring and indicators are available at the time, so it is possible to add to as more knowledge becomes available between reporting periods. The reporting also provided an option to include expert opinion where data were not available. They also made use of indicators rather than direct monitoring data. For example, in many of these offshore areas, direct surveys of species and habitat condition are not available. However, Vessel Monitoring System satellite data is available on the location of fishing boats, speed and gear type being used when trawling. Parts of the fishing gear in contact with the seabed is calculated on the width of the fishing gear multiplied by the average vessel speed and time fished. This gives a swept-area value per grid square. This was used as a proxy for pressure and was overlaid onto habitat/biotope distribution maps and habitat sensitivity maps with defined rules for predicting damage levels. Work is being done to incorporate pressures from other human activities into this indicator.

Far fewer projects were found to have implemented procedures to understand and communicate confidence and uncertainty in results. However, all projects agreed that this was important and would need to form part of the project's future work. Some projects recorded confidence when capturing data sources, whether through a literature review or expert input. For example, in a project modelling marine natural capital (see project twelve in the list of projects interviewed below), they aimed to include all identified factors within the modelling, but mark any with high levels of uncertainty as of low confidence, which led to differing levels of confidence reported against final outputs. Another project which mapped fire risk based on differing land use change and grazing intensities in the Brecon Beacons (see project two in the list of projects interviewed below) asked stakeholders to provide not only their perception of how easily different types of habitats ignited and burnt under different conditions, but also their level of confidence in the answer. This scoring was incorporated into the model, through averaging all scores for each habitat type and weighting burn risk assigned to each habitat type based on the survey results. In some cases, such as in the Habitat Directive assessments (see project eleven in the list of projects interviewed below), expert knowledge was used to determine whether to record something as 'unknown' or to input an assessment with low certainty.

2.2.3 Managing subjectivity

Most projects involving subjectivity left interpretation to the user. For example, one project that aimed to use Earth Observation to detect changes in habitat condition reported on normalised burn ratio (among other results - see project eight in the list of projects interviewed below). This could be seen as a graduated scale indicative of 'bad' condition, but this was not commented on within the results.

In many of the applications developed, there was flexibility for users to control the inputs and outputs they saw depending on their wishes, which allowed them flexibility to make subjective decisions themselves about what to report on. For example, a farmer making use of the data may wish to see different outputs (e.g. productivity) to a conservationist (e.g. biodiversity), and could do so with the click of a button.

Producing interpreted outputs that did suggest positive or negative implications, or reasoning behind results, were seen as work for the later stages of a project. They were not seen as necessary, but rather an improvement that could be added after there was high confidence in other aspects of the project. In some cases, this stage was not reached for many years. Even with these kinds of outputs, it would not be possible to categorically claim that something is definitely 'good' or 'bad' – communication would be a key factor once again to show that this is just a possible interpretation, and validation through other work such as field surveys would be required. One interviewee also flagged that 'good' and 'bad' condition would change depending on which species you were interested in – some species actually like degraded or disturbed habitats. Another project aiming to select biodiversity indicators for nature-based solution projects (see project five in the list of projects interviewed below), highlighted that some indicators, such as species abundance, could be both 'good' or 'bad' for biodiversity overall depending on the species being measured (e.g. an indicator species vs a non-native invasive).

The Habitats Directive, Marine Strategy Framework Directive (now the UK Marine Strategy in the UK) and Common Standards Monitoring (see projects eleven, one and seven in the list of projects interviewed below) directly aimed to report on whether an area was in 'good' condition, 'good environmental status', and 'favourable' condition respectively. This naturally includes a degree of subjectivity that cannot be fully eliminated. Generally, this was managed by wide consultation among experts, both within the UK and internationally when developing the assessment system and being very explicit in audit notes when reporting. Where possible, the frameworks aligned with international standards to mitigate risks. Determining thresholds (the point at which a habitat switches from being 'good' condition to 'bad' condition) was seen as a particularly subjective point within the methods developed. Transparency and communication were once again seen as essential for managing subjectivity. If the method is described clearly, everyone should be able to follow it consistently. The aim of the approach is to standardise what is considered 'good' or 'favourable' in order to reduce the subjectivity and provide a fair comparison across all sites assessed. If the limitations are described clearly, then any criticism or disagreement is already dealt with. Providing separate guidance for each feature or habitat was also seen as important for reducing potential for any issues caused by favourable/unfavourable condition varying by context.

2.2.4 Mitigating system traps

One project analysed within the literature review aspect of the research behind this report (the National Food Strategy) raised an interesting consideration that was not identified by any of the projects interviewed. This was that understanding how to influence a complex system is very different to understanding how it works. If developing an indicator of condition or health, it can be assumed that the rationale behind this is to lead to management scenarios or behavioural changes that can improve this. However, when these take place within complex systems, they often experience one of a number of 'system traps.' These are "archetypal ways in which systems can go wrong" (Dimbleby 2021).

One well known example of a 'system trap' is the tragedy of the commons, whereby a common resource is used up at a rapid rate due to the incentive to use it before it runs out, when in reality sustainable use could ensure the resource regenerates as quickly as it is used. Whilst unlikely to be directly included within a model or framework focusing on assessing soil condition or health, it is an important extension to consider in terms of interpretation and next steps. Which of the inputs assessed are finite? What are their regeneration rates? How can people be incentivised to use them responsibly?

Other 'system traps' explored within the National Food Strategy (Dimbleby 2021) that could lead to similar considerations include:

- 'Seeking the wrong goal': an issue based on unclear definitions or interpretations of outputs (e.g. 'health' vs 'quality' vs 'productivity', etc.). Similarly, where multiple goals exist (e.g. food production vs diverse habitat (re)creation in the context of soils), balancing the trade-offs and needing to take into account different factors depending on the goal can be challenging.
- 'Policy resistance': feedback loops constantly balance the system so efforts to create change are counteracted.
- 'Drift to low performance': when a decrease in condition is taking place so slowly that the baseline gets underestimated. Knowing the response in the system is important (e.g. the critical threshold at which soils no longer function or deliver ecosystem services).
- 'Escalation': when a system's goal is relative rather than absolute (e.g. a farmer wanting to get greater productivity from the soil each year).
- 'Shifting the burden to the intervenor': solutions treating a 'symptom' rather than a 'cause' (e.g. over reliance on fertilisers masking the problem when soils become depleted).
- 'Rule beating': avoiding following the rules created to enforce change.

3 Development timeframes

The shortest project interviewed was the work modelling ecosystem services in Chilean vineyards (see project ten in the list of projects interviewed below). This only took three months in total to develop. However, this was only a pilot project, leading to many caveats and potential future work directions if funded further. The longest project interviewed was the Habitats Directive reporting (see project eleven in the list of projects interviewed below). The first round of assessment for this was 1994 to 2000, and the approach is still open and continuously developing today. Although the UK are no longer likely to be involved as the reporting went to the EU, reporting to the Bern Convention will continue, which has a very similar (but not identical) reporting format. Most projects interviewed were several years into development. None saw their work as complete (just as fit for intended purpose); all saw potential for further development if funding allowed. Most were tied into annual funding cycles, which can be restrictive in terms of planning for larger scale projects and brings uncertainty to predicting timeframes for future work.

4 Conclusions and recommendations

Overall, projects involving complexity have many of the same requirements as simpler projects, only with more complex consequences if they go wrong. An effective governance system and early stakeholder engagement will support any decisions made – which in the case of a complex system include how to simplify it and select the factors of most relevance and use to those involved. Good evidence underpinning the project is also key – which in the case of a complex project may involve combining scientific literature with expert or local knowledge to account for data deficiencies within the project's context. Clear and engaging communication ensures project outputs are understood – which in the case of a complex project avoids the higher risk of misinterpretation, mitigates potential conflict around subjectivity, and allows for an appropriate degree of complexity to be shown for the audience to avoid confusion whilst allowing more experienced users to dig into the detail. A variety of modelling techniques can be applied, with BBNs particularly suiting data poor environments where integration of data sources and local knowledge is required. However, framework-based approaches can also be useful, especially in cases where a consistent assessment that cannot be based on a set methodology is required. Key recommendations are outlined in Table 16.

Table 16. Key recommendations.

Theme	Recommendation
Identifying inputs, outputs, and interactions	• Perform literature reviews. Rigour and resource required will depend on project scope. • Make use of expert input and local knowledge where appropriate, for example in data poor contexts. • Set up a governance system with well-defined roles to support decision making. For example, this may consist of a project team, a steering group, and a wider stakeholder group. • View the developmental process as iterative, with continual consultation and reassessment. • Consider the feasibility of including each input, output and interaction based on data availability and evidence gaps. • Consider alignment with policy and stakeholder priorities, and other relevant projects.

Theme	Recommendation
Models and frameworks	• Make use of an existing models or adapt an existing approach, where possible. • Consider using Bayesian Belief Networks, particularly if your project is associated with low data availability or requires integration of data and expert knowledge. • Consider using framework-based approaches, particularly if both flexibility and comparability are required. • The selection of models or frameworks to use should be seen as a process requiring ongoing and iterative wider consultation.
Communication	• Ensure early and continued stakeholder engagement. • Present outputs in a clear and engaging way, such as through an app, dashboard, or other interactive tool. • Present an appropriate level of complexity for the audience. This may require the production of intermediate outputs that experts can investigate further, and a selection of high-level aggregated outputs for non-technical audiences. • Ensure clarity of definitions and scope.
Dealing with incomplete knowledge and uncertainty	• Rely on an iterative process of building up knowledge as the project progresses; do not expect that the complexity involved can be captured within the first attempt. • Allow for flexibility in data requirements to obtain the most complete knowledge possible. • Record and communicate confidence and uncertainty where possible.
Managing subjectivity	• Where consensus has not be reached about a subjective threshold, leave interpretation of data to the user, and simply report the numerical values. • Provide flexibility for users to control the inputs and outputs they see depending on their interests. • Where subjective assessments are made, communicating the wider context around these assessments and how the conclusion was reached is key. • Where subjective assessments are made, ensure they are the result of wide stakeholder consultation and consensus building, and be explicit in audit notes when reporting.
Mitigating system traps	• Consider wider ways in which the system may be influenced, as well as just how the system currently works.

References

Clancy, T.R., Effken, J.A. & Pesut, D. (2008). Applications of complex systems theory in nursing education, research, and practice. Nursing Outlook, Special Issue Informatics: *Science and Practice* 56, 248-256.e3. Available at: https://doi.org/10.1016/j.outlook.2008.06.010 [Accessed 05.10.2021]

Dimbleby, H. (2021). National Food Strategy. The Plan. Independent Review. Available at: https://www.nationalfoodstrategy.org/ [Accessed 05.10.2021]

Faybishenko, B., Hubbard, S., Brodie, E., Nico, P., Molz, F., Hunt, A. & Pachepsky, Y. 2016. Preface to the Special Issue of Vadose Zone Journal on Soil as Complex Systems. *Vadose Zone Journal* 15. Available at: https://doi.org/10.2136/vzj2016.01.0005 [Accessed 05.10.2021]

Filotas, E., Parrott, L., Burton, P.J., Chazdon, R.L., Coates, K.D., Coll, L., Haeussler, S., Martin, K., Nocentini, S., Puettmann, K.J., Putz, F.E., Simard, S.W. & Messier, C. (2014). Viewing forests through the lens of complex systems science. *Ecosphere* 5, art1. Available at: https://doi.org/10.1890/ES13-00182.1 [Accessed 05.10.2021]

Johnson, N. (2009). Simply Complexity: A Clear Guide to Complexity Theory. Oneworld Publications.

Mitchel, M. & Newman, M. (2001). Complex systems theory and evolution. Available at: https://melaniemitchell.me/PapersContent/EncyclopediaOfEvolution2002.pdf [Accessed 05.10.2021]

Neitsch, S.L., Arnold, J.G., Kiniry, J.R. & Williams, J.R. 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute Technical Report No.406. September 2011. Available at: https://swat.tamu.edu/ [Accessed 05.10.2021]

Newman, M.E.J. (2011). Complex Systems: A Survey. *American Journal of Physics* 79, 800–810. Available at: https://doi.org/10.1119/1.3590372 [Accessed 05.10.2021]

Phillips, J.D. (1998). On the relation between complex systems and the factorial model of soil formation (with Discussion). *Geoderma* 86, 1–42. Available at: https://doi.org/10.1016/S0016-7061(98)00054-8 [Accessed 05.10.2021]

Ramos-Martin, J. (2003). Empiricism in ecological economics: a perspective from complex systems theory. *Ecological Economics* 46, 387–398. Available at: https://doi.org/10.1016/S0921-8009(03)00191-5 [Accessed 05.10.2021]

Ritz, K. (2008). 'Soil as a paradigm of a complex system' in Ramsden, J. and Kervalishvili, P. Complexity and security. IOS Press.

Straton, A. (2006). A complex systems approach to the value of ecological resources. Ecological Economics, Think Tank 2003 56, 402–411. Available at: https://doi.org/10.1016/j.ecolecon.2005.09.017 [Accessed 05.10.2021]

Appendix 1: List of projects interviewed

1. Defra (2019). Marine Strategy Part One: UK updated assessment and Good Environmental Status. https://www.gov.uk/government/publications/marine-strategy-part-one-uk-updated-assessment-and-good-environmental-status.

2. Ecosystems Analysis Team, JNCC. Fire Risk Modelling in the Brecon Beacons (not published).

3. Ecosystems Analysis Team, JNCC. Species Distribution Modelling work (not published).

4. Ecosystems Analysis Team, JNCC. Developing Indicators for the Tree Health Resilience Strategy Environmental Goals (not published).

5. Harris, M., Woodcock, P., Wright, E., Britton, A., Matear, L., Smith, A., Vina-Herbon, C., and Cheffings, C. Biodiversity Indicators Review – International Climate Finance Evidence Project. JNCC, Peterborough.

6. Jones, A., Trippier, B., Sym, E., Jones, G. & Wright, E. 2020. An assessment of the value of natural capital in the protective service against coastal and inland flooding in the UK Overseas Territory of the British Virgin Islands: BVI Flood Resilience Modelling Tool – Technical Report. JNCC Report No. 649, JNCC, Peterborough, ISSN 0963-8091.

7. JNCC and the CNCBs. Common Standards Monitoring. https://jncc.gov.uk/our-work/common-standards-monitoring/.

8. Lightfoot, P., French, G., Hassall, I., Hecker, L., Jones, A, Trippier, B. & Robinson, P. 2021. Using Sentinel data to monitor change in habitats and historic landscape features. Technical report for the Caroline Herschel Framework Partnership Agreement for Copernicus User Uptake (Work Package Six). JNCC Report No. 687. JNCC, Peterborough, ISSN 0963-8091.

9. Smith, N. 2019. St Helena Natural Capital Assessment: Final Report. Natural Capital in the UK's Overseas Territories Report Series (South Atlantic Region). Contracted report to JNCC.

10. Trippier, B., Jones, A., Jones, G., Barbosa, O., Colson, D., Duran, A.P., Godoy, K., Harris, M., Harrison, S., Tierney, M., Smith, M. & Wright, E. 2019. A Natural Capital Approach to Landscape Planning: a Pilot Project in Colchagua Valley – Technical Report, Chile. JNCC Report No. 635, JNCC, Peterborough, ISSN 0963-8091.

11. Marine Ecosystems Team, JNCC. Habitats Directive assessment (not published).

12. Marine Ecosystems Team, JNCC. Marine Natural Capital Modelling (not published).

Appendix 2: Case Study Projects

Marine Strategy Framework Directive (MSFD)

The MSFD is a framework that aims to address the complexities of monitoring the marine ecosystem as a whole. It covers all biodiversity and food webs, and all drivers of biodiversity change, within the marine environment, from the sea floor to the water column. It is split between the biodiversity elements (species, habitats) and the pressure elements (physical abrasion, eutrophication, contaminants, hydrodynamic impacts, noise, litter). The ultimate aim of the framework from a policy perspective is to achieve good environmental status (GES). The framework includes a series of high-level targets (e.g. no habitat loss of a specific habitat type, reduced pressure on cetaceans from bycatch, etc), each of which has a threshold beyond which it is considered as of GES.

The complexity in this project arises from the number of indicators considered within the system and the knowledge gaps associated with the system. Marine systems also have extra complexity, as interactions are in multiple directions throughout the water column, so are more three dimensional than most terrestrial interactions (although this is similar in some respects to soils).

Approaches and solutions taken by MSFD that could be transferable to other complex projects include:

- Defining and justifying what you want to measure. For example, all work for MSFD relates to legislation and policy targets, so these are the main drivers for this step.
- An iterative developmental approach, making use of a network of both subject area experts and potential users of the output (e.g. policy decision makers):
    - Indicators are first developed by a panel of experts.
    - There is then a series of testing rounds which help to shape the indicator, identify the metrics that will be used, assess whether it will be successful and determine how it will operate. This testing will also consider the scale that the indicator should be.
    - Input is provided by wider international experts.
    - Stakeholder consultation identifies any potential consequences, whether it will require changes to current data streams and how realistic it is to make the indicator operational.
    - The indicator is published for public consultation.
- A framework-based approach:
    - Each indicator tries to capture a single element of the system as it is not possible to measure all elements at once.
    - There are 11 high level descriptors, each of which can be broken down into multiple different indicators. For example, biological diversity is a high-level descriptor, which can be broken down into many different habitats and species-based indicators such as bird population abundance and bird breeding success.
- Numerical outputs with tailored and defined thresholds above which a measurement is considered to be of GES.
- A pragmatic approach:
    - Accepting that every indicator has knowledge gaps and being clear that the selected approach is the best that can be done with current data and resources.

- Transparency and communication:
  - The method is described as a series of steps so it is clear how the data will be analysed, what type of outcome the indicator hopes to achieve and what the key caveats and limitations are. If everyone applies the same method in the same way, they should all get the same results. This reduces the subjectivity inherent in the threshold-based approach to the greatest extent possible.

The framework itself was enacted in approximately 2010, with the first assessment taking place in 2012, and a second in 2016. Further assessments are planned every 6 years. The indicators are being developed alongside this in a gradual process, with each indicator requiring several years of testing – starting with local datasets and building it up to be operational.

Common Standards Monitoring

Common Standards Monitoring was developed to provide an agreed approach to the assessment of condition of statutory sites designated through UK legislation and international agreements across the four countries of the UK. The output aims to give an impression of overall site condition. The framework provides guidance around the assessment of 16 key features (8 habitat features and 8 species features), each of which is based on gathering data against several defined attributes. 'Features' are the species (e.g. seals, breeding birds), habitats (e.g. woodlands, heathlands) and geological / geomorphological characteristics (e.g. fossils, landforms) for which sites are protected. Each feature is qualitatively ranked into categories based on the flexible methods within the guidance. Based on the latest statement, these categories are favourable (when its condition objectives are being met), unfavourable (when its condition objectives are not being met), partially destroyed (sections or areas of the feature have no hope of being reinstated) and destroyed (the entire interest feature has been affected to such an extent that there is no hope of recovery). Optionally, trend qualifiers (e.g. recovering, declining, no change, recovered) can also be applied to each feature.

The complexity of this project arises from the need for both flexibility and consistency in terms of monitoring methods across different administrations, the need to reduce subjectivity in terms of what is considered ‘favourable’ and the large number of possible input variables.

Approaches and solutions taken by Common Standards Monitoring that could be transferable to other complex projects include:

- The use of an inter-agency working group:
  - To produce the statements and carry out the body of work that produced the guidance for each feature.
- Engagement with a wide range of experts:
  - This was considered a key reason the framework has not been superseded.
  - Where there was disagreement between experts, bilateral meetings were seen as a crucial mechanism to fully understand and take into account all views and reasoning.
- The use of intermediate outputs to ensure complexity is captured and available to those for whom it is of interest, but not presented as the headline.
  - Each feature has a list of attributes (for example, bird populations require monitoring of population size, population density and habitat extent).
  - Each of these is associated with a target and a method of assessment.

    - Outputs from each of these would be recorded at an intermediate stage, to inform the decision on the feature condition overall.
- The framework-based approach:
    - Guidance is provided that allows for comparisons to be made whilst remaining flexible to the monitoring methods used.
    - Whilst the flexibility has caused some problems around how comparable and reliable measurement methods involving new technologies are, it is considered to have kept the framework 'steady' over the years despite a changing data collection landscape.
- Standardising aspects that could be considered subjective by providing detailed definitions, to allow for fair comparisons.

Separate guidance is provided for each feature, reducing potential for any issues caused by favourable/unfavourable condition varying by context.

The framework was established in 1998. The first reporting period for the whole site series took place in 2005. Guidance documents were completed by 2005, but some have been updated since. In particular, a new statement was released in 2019, replacing the previous statement. This allows for inclusion of novel technology such as EO and eDNA in monitoring systems.

Water Framework Directive (WFD)

The WFD was published by the EU in 2000. It aimed to ensure integrated river basin management for Europe. The Environment Agency chair the UK Technical Advisory Group and led the process of developing a framework to implement this within the UK. It was based on the concept of measuring against what the ecological community in a given area "should" be (the deviation from a reference system).

The advisory group is formed of a number of task teams (e.g. alien species group, chemistry task team, freshwater task team). Each of these has developed their own implementation mechanisms and tools, and interpreted Directive wording for their context. Within this, reference systems were developed for each water body type (for example, coastal, estuarine) and categorised further within each type (e.g. coastal macrotidal, exposed coast, etc). For some categories, such as benthic habitats, an additional scale was required, based on salinity and sediment size (e.g. benthic fine mud), as there is so much variation and heterogeneity within the larger category. Each reference system aimed to describe the expected ecological community for that context. The output has a numeric value of between zero and one. Each tool is written up as a practitioner's guide, so anyone with data could use it, although some tools are more complex than others. For this report, a member of the Marine Benthic Team was interviewed.

Approaches and solutions taken by WFD that could be transferable to other complex projects include:

- Each indicator is made up of subunits which capture the detail and complexity but are reported as a single index of quality to allow for comparisons to be made and results to be easily interpreted.
    - These are different parts that are combined together to give the zero to one value.
    - For example, in the benthic marine case, these include the taxa number, the AZTI Marine Biotic Index (a measure of sensitivity to disturbance) and Simpson's evenness.

    - Each sample is scored in this way separately, and statistics are used to combine samples to give a score for the water body overall.
- An iterative developmental approach:
    - In order to define the boundaries between the five classes of condition (high, good, moderate, bad and poor ecological status), they went through a cyclic cycle of getting expert judgement from benthic ecologists, readjusting and reconsulting.
    - Comparisons were made across the EU member states for each water category and physical characteristic. Where little difference was found in terms of implementations developed, this suggested a consensus.
    - They have tried to improve the approach each reporting cycle. For example, the reference models have recently been updated and targeted monitoring is helping to fill in evidence gaps.
- Definitions:
    - Normative definitions were given in the Directive itself. These specified what each condition class should represent and outlined a methodology (e.g. stating which parts of the ecological community assessments had to look at).
- Based on a legal framework:
    - The Directive specified which parts of the ecological community were legally required to be included, driving forward action.
- Comparison to a reference system:
    - Reference systems for each class were based on collecting vast quantities of historic monitoring data and bringing this together with supporting parameters to identify a top-level community expected to be present.
    - In theory, reference systems could be based on a site with no or little anthropogenic impact, but often this is impractical, so relying on historical data is necessary.
    - The type of sampling used was also built into the reference system, as you would get very different results between, for example, a grab sample and a core sample.
- Communication:
    - It is important to be clear on what the indicator can and cannot do.
    - Lots of people interpret the result as an exact value but differing levels of data and monitoring lead to differing levels of uncertainty.
- A pragmatic approach:
    - The interviewee advised focusing on what can be done and recognising that it is not possible to do everything in an indicator.
    - It is also important to accept the limitations of using an indicator. When you create indicators, you are generalising by definition, whereas in reality every context it is applied in will have its own nuances. If you wanted to take measures based on results from the indicator, you would need to follow up with an investigation taking into account the context of the location.
- Using tools at the scale they are designed for:
    - A tool designed for use at a water body scale will not work well if applied to a quadrat sized sample.

The approach took around 20 years to develop in total. Work began in 2001.