Ministry for Primary Industries
Manatū Ahu Matua

Surveillance Evaluation Framework (SurF)

Main Document

17468 Deliverable: Draft Framework

MPI Technical Paper No: 2017/37

Prepared for Investigation and Diagnostic Centres
and Response Directorate
Ministry for Primary Industries
Pastoral House, 25 The Terrace
Wellington, New Zealand
www.mpi.govt.nz

by Dr Petra Muellner (Epi-interactive Ltd)
Prof Katharina D.C. Stärk (SAFOSO AG)
Jonathan Watts (MPI)

ISBN No: 978-1-77665-565-6 (online)
ISSN No: 2253-3923 (online)

March 2016

Disclaimer

While every effort has been made to ensure the information in this publication is accurate, the Ministry for Primary Industries does not accept any responsibility or liability for error of fact, omission, interpretation or opinion that may be present, nor for the consequences of any decisions based on this information.

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04 894 0300

This publication is also available on the Ministry for Primary Industries website at http://www.mpi.govt.nz/news-and-resources/publications/

Acknowledgements

The authors would like to thank Paul Stevens, Mark Bullians, Daan Vink, Anjali Pande, Paul Bingham, Tim Riding and Brendan Gould (all MPI) for their guidance and input.

Surveillance Evaluation Framework (SurF)

Prepared for

Investigation and Diagnostic Centres
and Response Directorate
Ministry for Primary Industries
as 17468 Deliverable: Draft Framework

by

Dr Petra Muellner (Epi-interactive Ltd)
Prof Katharina D.C. Stärk (SAFOSO AG)
Jonathan Watts (MPI)

March 2016

Contents

Tables

Page

Figures

Page

Glossary

Case definition[1]	A set of unified criteria for disease reporting to ensure uniformity and simplicity.
Design prevalence[2]	A standard hypothetical prevalence of a disease or risk organism against which to measure surveillance sensitivity. Can be specified at the herd (herd design prevalence), population or at the animal level.
Endemic	A disease or risk organism regularly found in a population or geographical area.
Epidemic[3]	An increase, often sudden, in the number of cases of a disease or the number of risk organisms above what is normally expected in a population or geographical area.
Exotic	Exotic disease or risk organisms are those not commonly present in New Zealand.
Evaluation framework	A document providing the conceptual basis for the planning and implementation of an evaluation project. It describes the activities, decisions and resources that are part of the evaluation process.
Evaluation project	The application of the evaluation framework to answer a distinct evaluation question.
Incursion	Detection of exotic land, freshwater and marine pests or exotic diseases in plants or animals.
Outbreak	Occurrence of cases of disease in excess of what would normally be expected in a population or geographical area. Carries the same definition as epidemic, but is often used for a more limited geographic area.
Performance indicator	A measure which is typically quantifiable and which can be used as a proxy for surveillance performance. A specific indicator should be linked to the assessment of a selected surveillance attribute (see below).
Sensitivity	Ability of a system or test to correctly identify positive events e.g. all animals or plants with a specific disease.
Specificity	Ability of a system or test to correctly identify negative events e.g. all animals or plants without a specific disease.
(Biosecurity) Surveillance	Biosecurity surveillance is the collection, collation, analysis, interpretation and timely dissemination of information on the presence, distribution or prevalence of risk organisms and the plants or animals that they affect[4]. At MPI a hierarchy of surveillance ranging from surveillance portfolio to surveillance system to surveillance programme to surveillance activity is being used.

[1] Adapted from http://wonder.cdc.gov/wonder/prevguid/m0025629/m0025629.asp
[2] Adapted from http://www.fp7-risksur.eu/terminology/glossary
[3] Adapted from http://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson1/section11.html
[4] http://www.biosecurity.govt.nz/files/pests/surv-mgmt/surv/biosecurity-surveillance-strategy-2020.pdf

Surveillance activity	A defined surveillance task that is a component of a surveillance programme or system.
Surveillance attribute	A characteristic used to evaluate an existing surveillance system or to conceptualise a proposed system.
Surveillance objective	The expected surveillance outcome, typically closely related to the policy decisions that will be informed by surveillance.
Surveillance portfolio	The high-level collection of a set of surveillance systems, programmes or activities that jointly serve a specific policy or strategic objective.
Surveillance system	A set of surveillance programmes and resources used to generate information relevant to its objective on the health or biosecurity status of a population.
Surveillance programme	Used in the context of the New Zealand biosecurity system to describe a specific surveillance activity or a group of related activities with a defined set of expected deliverables and outcomes.

1 Executive Summary

Surveillance of hazards to biosecurity is conducted to support and underpin New Zealand's biosecurity system. Surveillance evaluation is an integral part of the surveillance life cycle. It provides a means to identify and correct problems – as well as to sustain and enhance existing strengths of a surveillance system.

The Surveillance Evaluation Framework (SurF) presented here was developed to provide a consistent generic framework within which the Ministry for Primary Industries (MPI) biosecurity surveillance portfolio, and all its components, can be assessed. It is a novel cross-sectorial effort that aims to provide a common umbrella for surveillance evaluation in the animal, plant, environment and marine sectors. SurF builds upon and adapts previous national and international work conducted in the context of the evaluation of human and animal health surveillance; hence it builds upon existing good practice in surveillance evaluation.

SurF supports the conduct of the following four distinct components of an evaluation project: (i) Motivation for the evaluation, (ii) Scope of the evaluation, (iii) Evaluation design and implementation, and (iv) Reporting and communication of evaluation outputs. Further, SurF provides a visual output that allows for comparison of core performance between systems and within individual systems over time.

Case studies, prepared by MPI subject matter experts, are included in the framework to guide users in their assessment. The case studies were also used in the development of SurF in order to assure practical utility and to confirm usability of SurF in all included sectors.

Although SurF was developed for internal use by MPI, it could be applied to any surveillance system in New Zealand or elsewhere. It is anticipated that the structured approach and information provided by SurF will not only be of benefit to MPI but also to other New Zealand stakeholders.

2 Introduction

2.1 RATIONALE

The New Zealand Ministry for Primary Industries (MPI) invests significantly in a range of national biosecurity surveillance activities across the plant, animal, environment and aquatic sectors. These activities support and underpin New Zealand's biosecurity system, its ability to enable trade and to protect itself from biological risks through the early detection of pests and diseases, and the provision of evidence of pest or disease freedom. Given the importance of these activities to New Zealand, it is essential that the performance of MPI's surveillance systems and programmes can be assessed to provide assurances around the quality of delivery and the outputs of these programmes. It is also critical to ensure that they are responsive to change and continually evolve to meet changing biosecurity needs in an efficient manner. As concluded by Drewe and colleagues (2015), evaluation can be used to help both identify and correct problems – as well as to protect and enhance the strength of a surveillance system.

It is understood that evaluation forms an integral component of the surveillance life cycle managed by MPI (Figure 1). Surveillance activities, i.e. data collection, analysis and information dissemination, are repeated in each surveillance cycle. The duration of a cycle can vary depending on the policy context and the organism under consideration. Evaluation is typically conducted before a new cycle is begun, and is driven by organisational requirements such as annual reporting or financial accountability. Although evaluation can be conducted at the end of each cycle, it is expected that more extensive and formal evaluation would be conducted after several cycles were completed. Adjustments to the surveillance process can also be caused by triggers other than evaluation. For example, changes in the biosecurity situation, major adjustments following an outbreak or incursion, feedback from stakeholders or changes in policy and/or budget priorities can also affect the conduct of surveillance.

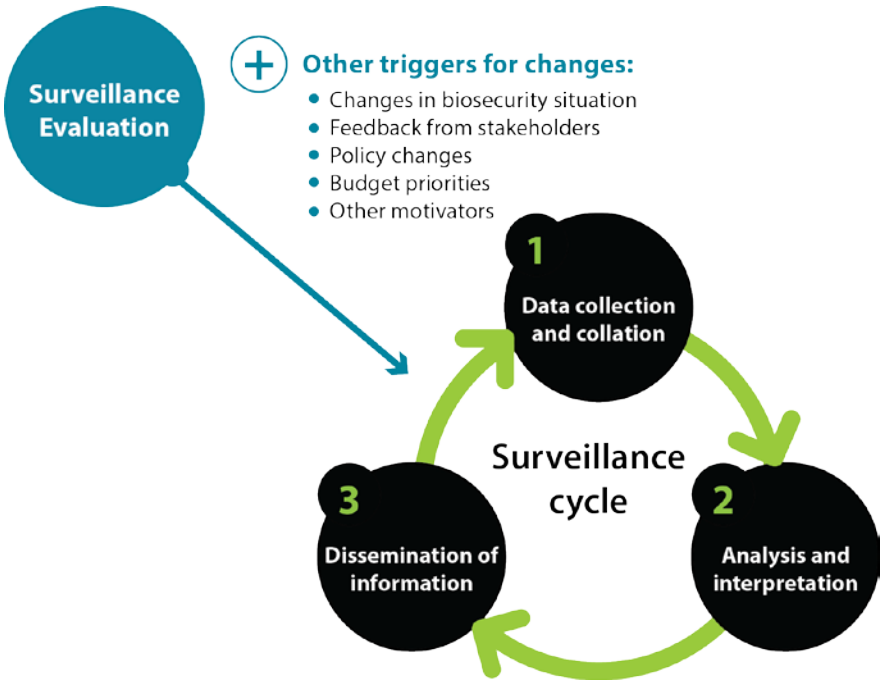

Figure 1: *Evaluation as part of the surveillance life cycle.*

2.2 THE MPI SURVEILLANCE EVALUATION FRAMEWORK

The Surveillance Evaluation Framework (SurF) was developed to provide MPI with a fit-for-purpose and efficient surveillance evaluation framework to meet defined needs across MPI's range of biosecurity activities. It builds upon previous recommendations made by a systematic scoping review (MPI, 2015), and delivers a structured and adaptable process that enables the systematic and comprehensive evaluation of MPI's surveillance activities, programmes, systems or portfolios.

SurF builds upon and adapts previous work conducted nationally and internationally in the context of the evaluation of human and animal health surveillance. This includes in particular the SERVAL Framework (Drewe *et al*. 2015), the recently published guidelines by the European Centre for Disease Control (ECDC, 2014), as well as the available information describing the preliminary version of the EVA tool (RiskSur Consortium 2013 & 2015), which at the time of writing is still under development. An effort was made to align this framework with the national standards proposed by the Aotearoa New Zealand Evaluation Association (ANZEA, 2015) where possible.

SurF has been designed to provide sufficient flexibility to accommodate the needs of biosecurity surveillance evaluation in the different sectors defined by MPI (i.e. the animal, plant, environment and marine sectors). This includes surveillance for the wide range of endemic and exotic diseases and risk organisms that New Zealand's fauna and flora are potentially exposed to.

SurF consists of four components, each supporting a distinct phase in the evaluation (Figure 2). Each component describes the activities and decisions related to a phase within an evaluation project:

1. Motivation for the evaluation
2. Scope of the evaluation
3. Evaluation design and implementation
4. Reporting and communication of evaluation outputs

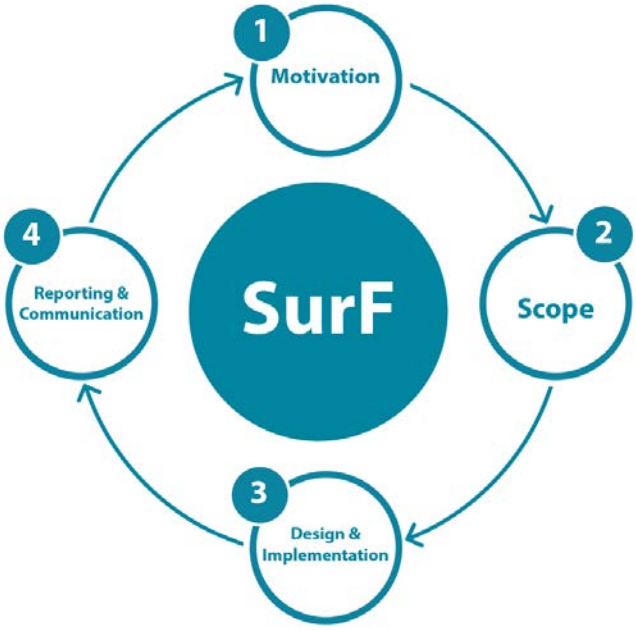

Figure 2: *The four steps of the Surveillance Evaluation Framework (SurF).*

The evaluation process and its four phases are described in detail in Chapter 2 of this document. All terminology of relevance to the framework is defined in Chapter 8. The use of consistent specified terminology facilitates communication and the implementation of any evaluation. The proposed terminology is based on current good practice of surveillance evaluation in an international context. However, it is noted that terminology is dynamic and can vary between sectors. It is therefore recommended that terminology is discussed and updated regularly as the framework is being used to assure a common understanding within MPI and among relevant stakeholders. Finally, for convenience, this document refers to surveillance systems throughout. It is, however, envisioned that the same framework and attributes could also be used for evaluation at the activity, programme or portfolio level.

3 Evaluation Process

The evaluation process as defined by SurF consists of four main phases, which are listed in Table 1. Each phase consists of a number of activities and produces specific outputs. These are described in the following sections of this document. The framework and the supporting guidance notes describe the aspects to be considered during the specific activity of the evaluation process. Depending on the situation and the system under evaluation it might not be possible to assess or describe all components in full detail; any abbreviations from the full protocol should always be documented to ensure consistency. Since SurF will be utilised to evaluate a wide diversity of surveillance systems, case studies from different sectors (i.e. marine, plant and animal; see Appendix 2 for details) have been prepared to guide users in their assessments in addition to the text and to illustrate the application of SurF in practice. Further, for convenience, SurF provides users with an Evaluation Template to guide capturing input for this section and also the following ones (Section 7).

Table 1: *Overview of the evaluation process described in SurF.*

IDENTIFICATION OF THE SYSTEM UNDER EVALUATION
I. MOTIVATION FOR THE EVALUATION
A. Evaluation trigger
B. Context
II. SCOPE OF THE EVALUATION
A. Evaluation objective
B. Evaluation question(s)
C. Time and resources
D. Evaluation intensity
E. Evaluation organisation and composition of evaluation team
F. Status of evaluation outputs
III. EVALUATION DESIGN AND IMPLEMENTATION
Design of the evaluation
A. Select attributes from master list
B. Choose methods to assess attributes
C. Make an inventory of available information sources about the system
D. Identify missing information
Implementation of the evaluation
E. Describe the surveillance system under evaluation
F. Describe the surveillance system's objective(s)
G. Describe the organisational structure

H. Identify and engage surveillance system users

I. Identify the target population and geographical coverage

J. Describe the design of the surveillance system

K. Describe the processes

L. Collect data and information

M. Assess the included attributes

IV. REPORTING AND COMMUNICATION OF EVALUATION OUTPUTS

A. State target audience

B. Report main findings

C. Summarise and synthesise results

D. Provide guidance for interpretation of results

E. Make recommendations

F. Facilitate plain reporting

3.1 IDENTIFICATION OF THE SYSTEM UNDER EVALUATION

As a first step before starting the evaluation process the surveillance system should be named.

3.2 I – MOTIVATION FOR THE EVALUATION

The first step of the framework focuses on the motivation for the evaluation. This consists of stating the evaluation trigger(s) and describing the surveillance context (e.g. what are the legal requirements, why is the organism under surveillance considered a problem, and what are the current national and international situations). Activities relevant to this phase of the evaluation are described below. Describing the evaluation trigger(s) is important as it clarifies the thinking around the most important factors driving the undertaking of the evaluation and may have a significant impact upon which attributes are chosen for evaluation

A. Evaluation trigger	Describe evaluation triggers. A trigger, or a series of triggers over time could lead to the decision to conduct a surveillance evaluation. For example evaluation can be planned or unanticipated. • Planned evaluation is required, for example, if the legal basis of surveillance activities requires evaluation to take place at regular intervals. This can also be due to quality assurance or other administrative processes implemented by the competent authority or its partners. • Unanticipated evaluation can be triggered by changes in the risk landscape, changes in the industry, or international triggers such as changes to World Organisation for Animal Health (OIE) or International Standards for Phytosanitary Measures (ISPM) rules or criteria of trading partners. Other factors include: the emergence of a disease or risk organism; changes in diagnostic techniques; concerns about the acceptability or representativeness of reporting; failure of the system to detect an outbreak or incursion; and inability of the system to properly quantify a problem. Declining resources or a requirement to link to other surveillance systems can also trigger an evaluation.

B. Context	Describe the context of the surveillance system to illustrate why surveillance is conducted and what drives its design. This may, where applicable, include:
	
	Why is surveillance required?
	
	• Characterisation of the risk organism(s).
	• Epidemiological profile of the diseases or risk organisms under surveillance.
	• A description of the population-at-risk and/or host range.
	
	Situational analysis (both national and international).
	
	• Current situation.
	  - Why is the organism or disease considered a problem?
	  - Briefly indicate the level of current knowledge.
	• Brief overview of historical situation.

3.3 II - SCOPE OF THE EVALUATION

The scope of the evaluation is informed by the motivation for the evaluation. In this phase, the objective of the evaluation should be specified and made explicit to ensure it is consistent with the motivation. A key task in this phase is to define and agree upon the evaluation question(s) and to agree upon an evaluation project plan. Depending on the motivation for the evaluation, the intensity should also be discussed. This is particularly relevant in the context of resources. Intensity and resources need to be aligned. If resources are limited, it may not be possible to conduct a full evaluation. The intensity will also depend directly on the motivation for the evaluation.

Organisational questions also need to be clarified in this phase of the evaluation. An evaluation can be conducted in-house or externally. This decision will depend on the regulatory context but also on budget considerations and on in-house capacity and evaluation competency. If the evaluation is to be tendered, this will impact on the time plan. In a scenario where the evaluation will be commissioned, most of the points listed below will be relevant and specified as part of the call for tenders.

A. Evaluation objective	As a first step towards narrowing down the evaluation question, one or more of the evaluation objectives suggested below could be selected: • To ascertain whether or not a surveillance system is meeting its current objectives or a proposed change in objective. If this is the objective of the evaluation then the objective of the surveillance should be stated. • To ascertain whether or not a foreign surveillance system is reliable enough to accept imports from that country, or if a domestic surveillance system is good enough to support export of animals, plants or their products. • To ascertain whether or not a surveillance system is providing value for money to the funder. • To determine how much benefit (monetary or otherwise) a surveillance system provides to its user groups. • To identify the strengths and deficiencies of a surveillance system. • To identify potential measures that could improve the performance, efficiency and productivity of a surveillance system.
B. Evaluation question(s)	Phrase the evaluation objective(s) into a specific question that can be answered by the evaluation (i.e. the main question to be answered by the evaluation). Where an evaluation is seeking to determine whether or not a surveillance system meets its objectives, the relevant surveillance objective should be clearly stated within the evaluation question. Specific expectations should be identified (e.g. surveillance should cost less). Examples of evaluation questions: • Is/are the surveillance activity(ies) or system(s) capable of meeting a technical objective or target? • How can specific surveillance attributes be improved? • What is the overall performance of the surveillance system? • What are the strengths and weaknesses of the surveillance system? • Is the surveillance system meeting its objective to [....]?
C. Time and resources	Specify the staff, funds and deadlines for the evaluation. The evaluation time plan may be impacted by other deadlines such as budget decisions, ministerial meetings etc. Identify the evaluation time frame, including the start date, delivery date and any interim deadlines with associated deliverables. Good practices as applied in general project management are applicable.

D. Evaluation intensity	Determine the expected evaluation intensity. The level of detail of the evaluation will depend on: • Motivation for the evaluation. • Evaluation objectives (see above). • Aspects that are intentionally excluded or considered out-of-scope (e.g. economics). • Available time and resources (see above).
E. Evaluation organisation and composition of evaluation team	Describe the organisation of the evaluation and composition of the evaluation team as well as their responsibilities. • Is this evaluation internal or external? Evaluation can be conducted in-house or contracted externally. The latter might be required when independence needs to be assured. • Describe the necessary knowledge and competencies required in the evaluation team. In particular, assessment of some of the technical attributes might require assistance, for example by an epidemiologist or other subject matter expert. Are individuals or organisations available to provide peer-review? • Specify the roles and responsibilities within the project, e.g. who will make decisions. Roles can include the following: - Leading and coordinating the evaluation (project manager, project secretary). - Providing input information (surveillance experts). - Clarification, interpretation and discussion of evaluation findings (stakeholders). - Dissemination of evaluation results. • Identify the people/organisations involved in and affected by the evaluation to identify communication needs.
F. Status of evaluation outputs	Specify the classification of evaluation data and results, e.g. some outputs might be confidential and access limited, while others should be accessible to a wide range of stakeholders. Specify the communication channels that will be used to disseminate the findings.

3.4 III – EVALUATION DESIGN AND IMPLEMENTATION

The evaluation can make use of a range of approaches and methodologies to implement the tasks defined in the evaluation project plan (see Section 3.3. for details). Typically, it is a mix of quantitative and qualitative methods such as data analysis, calculation of indicators, interviews and conceptual analysis. Some tools are available to inform the decision on the best methodologies. The choice of methods will also depend on the type and extent of information, data and documents that are available for the specific evaluation subject. Ultimately, there is not a single right way of conducting the evaluation, but it should consist of a robust set of approaches that ensure the delivery of results that are adequate to answer the evaluation question(s) with the available resources. The design (3.4.1) and implementation (3.4.2) of the evaluation will be strongly driven by decisions taken in the previous phase. It should also follow relevant evaluation guidelines that are applied within the organisation, e.g. MPI. Attribute assessment is described in detail in Section 4 of this document, to introduce users to the overall process and structure of SurF. In the scenario where conduct of the evaluation is tendered, design and implementation are typically defined by the contractor. Design needs to be aligned with time and resource availability (see 3.3).

3.4.1 Design of the evaluation

Consult with experts in the relevant sectors(s) regarding the species affected or epidemiology of the risk organism or disease under surveillance, for assistance with selecting and assessing relevant attributes.

A. Select attributes from master list	Select attributes from master list. All core attributes should be included, unless they are excluded for a specific, documented reason. The choice of additional attributes lies with the assessor.
B. Choose methods to assess attributes	Decide which attribute is best assessed by which approach, once an overview of available information is established. This will also determine whether the attribute will be assessed with qualitative or quantitative approaches. The SurF Methods Catalogue (Appendix 1) lists a series of references describing methods of assessment for the different attributes used in SurF.
C. Make an inventory of available information sources	Prepare an inventory of available information sources. • Information sources can consist of documents such as legislation, guidelines, reports, meeting protocols or previous audits or evaluations. Surveillance data should also be included. If the latter are used, additional time and specific competencies are required in the evaluation team. • Identify relevant individuals to interview. All individuals involved and affected by a surveillance system are potentially relevant. It should be considered which perspective of the evaluation they could cover and to which attribute this would contribute.
D. Identify missing information	Based on the selected attributes and the available information sources, identify possible gaps and how the evaluation aims to address these gaps. The feasibility of information and data collection for the evaluation should also be considered in light of available resources.

3.4.2 Implementation of the evaluation

Complete Sections E–M step by step. Refer to Section 4 of this document for detailed guidance on Attribute assessment.

E. Describe the surveillance system under evaluation	Describe the surveillance system. The level of detail depends on the existing requirements and knowledge. It is suggested to keep this summary brief, typically not exceeding one page. Reference can be made to other existing and more detailed documents. This typically includes the following: • Name, legal status. • Evaluation objective and question as defined in Sections II.A and II.B. • Define the level at which the evaluation is being conducted: - Is the evaluation being conducted at the activity, programme, system or portfolio, level? ▪ If applicable, describe its components and how they relate to each other. - Where desired and resources are available the MPI Intervention Logic Model (ILM) could be used to describe the surveillance under evaluation. Further information on this approach can be provided by the MPI Assurance and Evaluation Group.
F. Describe the surveillance system's objective(s)	Describe the surveillance system's objective(s). • Are the surveillance objectives clearly defined and relevant to the actual situation of the disease or risk organism? • Choose one or several from the following list of six surveillance objectives: - Monitor the prevalence of a disease or risk organism: While usually aimed at endemic diseases or risk organisms, this is also applicable to new and re-emerging diseases and risk organisms and can form part of an assessment of the impact of control programmes on infection incidence. - Finding cases of a disease or risk organism: ▪ Detection of as many cases as possible of a known infection to facilitate control. The emphasis here is on finding those individuals, or locations, that are infected, or where the organism occurs, in order to intervene in some way, such as by culling, vaccination or delimitation surveillance. This will usually apply to an endemic disease or organism, i.e. an organism that is already present in the country. - Early detection of new or re-emerging disease or risk organism(s): Early detection could be defined as detection of infection before an outbreak or incursion becomes uncontrollable; this timeframe will vary by disease or risk organism(s) and should be estimated. If this objective is chosen, a statement should be included to define how early the system aims to detect infection. - Demonstrate freedom from a disease or risk organism: If this objective is chosen, a statement should be included to define the prevalence and associated confidence level, which are considered to indicate disease or risk organism freedom. - Identify changes in the population-at-risk or an organism range or host expansion: ▪ Here, risk factors, rather than a disease or risk organisms, are the target for surveillance. This might lead to identification of new population groups at risk or range/host expansion of an organism; targeted prevention measures could be considered. - Improve epidemiological understanding of a disease or risk organism: Generating knowledge about a disease or risk organism, for example academic research or hypothesis generation.

G. Describe the organisational structure	Describe who leads and manages the surveillance system being evaluated and briefly describe their roles. Identify whether there are suitable steering and scientific committees, where appropriate, and describe their roles and responsibilities. How are decisions being made?
H. Identify and engage surveillance system users	Identify and engage system users. • Identify the people involved in the surveillance system that is being evaluated: - Who pays for the surveillance? - Who provides the surveillance data? - Who analyses the surveillance data? - Who uses the resulting information? - Who benefits from any action resulting from the surveillance? - Who pays for risk organism or disease mitigation? - Who (if anyone) might lose out if a risk organism or disease is reported (e.g. it might be thought that famers' reputations could be tarnished if they declare disease in their herd or growers might lose millions in export markets if a new risk organism is discovered)? • Engage stakeholders and users: The engagement of stakeholders is essential and needs to be secured early in the process. A range of formats can be used to disseminate information on the conduct and objectives of an evaluation. For example, by using leaflets, email, or presentations at meetings. Identify how engagement will be secured. • Ensure to include surveillance system managers and implementing personnel.
I. Identify and describe target population and geographical coverage	Describe the target population, with reference to the population-at-risk, and the geographical coverage.
J. Describe the design of the surveillance system	Describe the design of the system. • Outline the surveillance design. • Describe the sampling frame. How is it decided? • Describe the general structure of the surveillance system including: - Origin of data (whether active, passive or enhanced passive). - Focus (whether disease or risk organism-specific, or general). - Survey design (e.g. case reports or continuous collection). - Sample size calculation and sampling strategy, including whether a risk-based strategy is used. - Where applicable describe calculation or statement of confidence and coverage/inference.
K. Describe the processes	• Processes Describe field operations/sampling and laboratory processes. Are quality control and assurance procedures (e.g. SOPs) followed and are audits/evaluations conducted? • Data Describe processes related to data collection, data management, data analysis and data dissemination. – Data collection. Assess use of appropriate data sources and collection methods and the existence of a case definition, where applicable, and data collection protocol. Consider each of the following: ▪ Who provides the data? ▪ Who collects the data? ▪ Where/when are data collected (space-time)? ▪ How are data collected? ▪ How are data recorded (e.g. on paper or electronically)? ▪ What types of data are being dealt with (e.g. active/passive, threat-specific/syndromic)?

	▪ Is there a data collection protocol? ▪ Are quality control and assurance procedures followed and are audits/evaluations conducted? ▪ How are staff trained to collect data? ▪ Is there a case definition? If so, please describe it. – Data management. Use and documentation of systems for processing information, including data processing protocols and data verification procedures. Consider each of the following: ▪ How are data managed? ▪ What data security measures are in place? ▪ How are data stored? ▪ How is data management documented? ▪ Are quality control and assurance procedures followed and are audits/evaluations conducted? ▪ Are there data processing protocols? ▪ Describe the data verification procedures. – Data analysis. Methods used for the analysis and interpretation of surveillance data. Consider each of the following: ▪ How are data analysed and interpreted? E.g. predictive models, risk factor analysis, prevalence estimation, summary measures. ▪ Are performance indicators used and if so, which ones and how are they calculated? E.g. numbers of reports received or samples collected per time unit, trend analysis or comparisons with results from other systems. – Data dissemination. Methods used for information exchange between people involved at all levels of the surveillance system. Consider each of the following: ▪ Which methods are used to exchange information between people involved in the surveillance system (providers, analysers and users of surveillance data)? These might include: case reporting cards, emails, letters, phone calls, interim reports of surveillance data, websites for disseminating information, and feedback given to the data providers. ▪ How frequently are data or reports disseminated? ▪ Do methods used (e.g. reports) adequately report the outputs from data collection, data management and data analysis? Is sufficient interpretation provided? ▪ To date, what actions (if any) have been taken as a result of the surveillance activity? These might include: details of mitigation measures imposed; decreased incidence of diseases or risk organisms; use of surveillance data for policy and programme decisions; and appropriateness of outbreak or incursion response. • Consider presenting the structure of the system in a flow-chart format.
L. Collection of data and information	Use the formats and sources identified previously, i.e. document review, interviews.
M. Assess the included attributes	Use the selected methods to provide quantitative or qualitative results (see Section 4 for details).

3.5 IV – REPORTING AND COMMUNICATION OF EVALUATION OUTPUTS

The output of the evaluation is typically captured in a written report that includes detailed descriptions of each of the sections listed above as well as results of attribute assessments. Information captured using the SurF Evaluation template (Section 7) will provide the basis for the report. In addition, the report provides the attribute evaluation results (see Section 4). All findings need to be discussed, interpreted and presented such that the reader is able to reach the conclusions related to the evaluation objective.

A. State target audience	Identify target audience(s) for evaluation outputs.
B. Report main findings	Reporting of main findings; will include descriptive parts as well as additional analyses. Results can be listed by attributes or by evaluation question(s).
C. Summarise and synthesise results	Summarise and synthesise results: • Describe the extent to which the system meets its objectives. • Identify the strengths and weaknesses of the surveillance system under evaluation. • Address the evaluation question(s).
D. Provide guidance for interpretation of results	Provide guidance for interpretation: • Assess system limitations, interfaces to other relevant activities such as interventions and control measures. • State information gaps, bias, uncertainties and assumptions. • Make recommendations to improve future evaluations.
E. Make recommendations (optional)	Depending on the evaluation objective, evidence-based suggestions for possible improvements to the surveillance system can be included (e.g. use of portable technology, risk-based requirement or sampling, review of sampling strategies including the sample size, pooling of samples, and integration of data from different sources. The value of surveillance might also be improved by changing the methods used to analyse or disseminate information).
F. Facilitate plain reporting	Provide plain English summary to support reporting of results to non-technical audience.

4 Attributes

This section contains the list of attributes to be used in conjunction with SurF, the New Zealand Surveillance Evaluation Framework. The attributes, their definitions and recommended methods for assessment build on existing frameworks, in particular SERVAL and EVA, but also the review of Drewe *et al.* (2015) and the CDC and ECDC guidelines on surveillance evaluation and monitoring. SurF also includes some additional attributes, which were developed with the objectives and scope of SurF in mind. Furthermore, some previously proposed attributes were modified to give the framework sufficient flexibility to be used across the whole spectrum of New Zealand's biosecurity surveillance portfolio.

In SurF attributes are grouped into five 'Functional Attribute Groups' based on the logic presented in Figure 3:

A. Attributes assessing surveillance organisation and management;
B. Attributes assessing surveillance processes;
C. Attributes assessing the technical implementation of surveillance;
D. Attributes assessing surveillance outputs;
E. Attributes assessing the impact of surveillance.

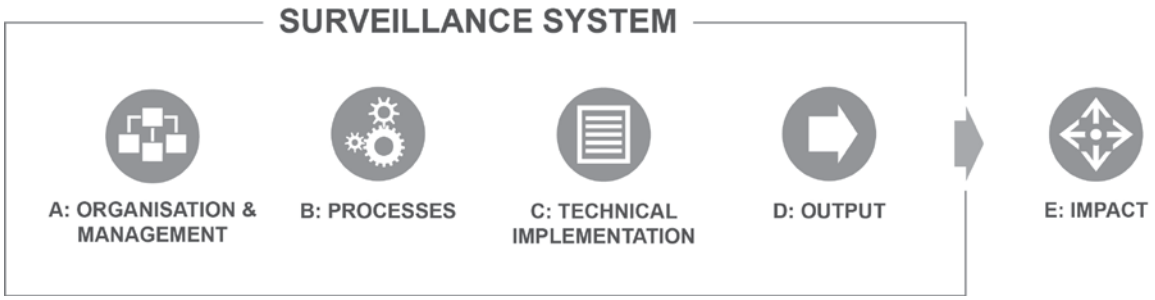

Figure 3: *Logic of Functional Attribute Groups (A–E) used in SurF.*

SurF includes a total of 29 different attributes (Table 2). Attributes are divided into core attributes (highlighted in **bold**; n=10) and accessory attributes (n=19). Each group includes at least one core attribute. The position of each attribute within a Functional Group is by no means a reflection of its importance, but reflects an alphabetic order. Core attributes assess essential aspects common to all surveillance systems, and it is recommended that they be included in all evaluations. If for any reason this has not been done, justification should be provided.

While it is recommended that core attributes are included in all assessments, the choice of accessory attributes is left to the evaluator and is not specified in SurF. The choice will ultimately be situation- and sector-specific and may be influenced by factors such as the evaluation question, the surveillance objective or the surveillance system's design.

Table 2: *List of attributes included in SurF (n=29). Core attributes are highlighted in bold.*

Functional Attribute Group	Attribute
A. Organisation & Management	1. Flexibility 2. **Organisation and management** 3. **Performance indicators and evaluation**
B. Processes	4. Data analysis 5. **Data and information collection** 6. Data management and storage 7. Field and laboratory services 8. Resource availability 9. **Technical competence and training**
C. Technical Implementation	10. Acceptability and engagement 11. Coverage 12. **Data completeness and correctness** 13. Interoperability 14. Multiple utility 15. RARR (Reliability, availability, repeatability, and robustness) 16. **Timeliness**
D. Outputs	17. Historical data 18. Negative predictive value 19. Positive predictive value 20. Precision 21. **Representativeness and bias** 22. **Sensitivity**[5] 23. **Specificity**[5]
E. Impact	24. Benefit 25. Decision support 26. Efficiency 27. External communication and dissemination 28. Internal communication 29. **Utility**

[5] At least one of the two attributes 'Sensitivity' and Specificity' is recommended to be included.

4.1 ATTRIBUTE ASSESSMENT

Traffic-light coding is used to provide a summary appraisal in SurF for each of the attributes, using the following standardised coding approach:

Traffic-light code	Description
●	Excellent or very good
●	Good, though room for improvement
●	In need of attention

Detailed guidance for the assessment of each SurF attribute is presented in the following sections. Sector-specific case studies are included in Appendix 2. These were specially compiled to provide guidance to SurF users by illustrating the use of the framework in practice.

4.2 RECOMMENDED METHODS AND REFERENCES

A wide range of qualitative and quantitative methods are available to assess individual attributes and the choice of method will always be situation-specific and dependent on the available data and information as well as the resource capacity of the assessor(s). Performance indicators, i.e. measures that are typically quantifiable and can be used as a proxy for surveillance performance, can be used for selected attributes. SurF provides references to recommended methods (details provided in Appendix 1 and case studies (Appendix 2)) to guide users in their assessment, but does not prescribe specific methods for attribute assessment. Not all methods will be applicable to specific surveillance objectives and/or contexts. For instance, negative predictive value is suitable for surveillance that aims to demonstrate freedom from disease, whereas sensitivity is relevant for surveillance that aims to detect diseases or risk organisms, or to find cases early (Drewe *et al.* 2015). Assessment of some attributes might require specialist support e.g. input from an epidemiologist to assess “Sensitivity”, “Specificity” or “Data Analysis”.

4.3 SURF ATTRIBUTES BY FUNCTIONAL GROUP

4.3.1 A: Attributes assessing surveillance organisation and management

Attributes in Group A are used to assess management and organisation of the surveillance system. They do not cover technical aspects of the system.

#	Name	Description	Guidance notes
1	Flexibility	Ability to adapt to changing information needs or operating conditions with little additional time, personnel or allocated funds.	Flexible systems can accommodate new events, changes in case definitions or technology, and variations in funding or reporting sources (CDC 2001). This attribute is determined more by the planning and management of the surveillance system than by the operation of the system. Simpler or more generic systems are likely to be more flexible. An evaluation of the flexibility of the system may be made by considering how the surveillance system has responded to changes in the past. Potential changes or events to consider include: • Changes in the information needs of the users of surveillance. • Changes in relevant national or international legislation or guidelines. • Changes in the demography of the target population. • Changes in the epidemiology of disease (including outbreaks), host range of an organism or the emergence of new disease or organism threats. • Changes or improvements to the methods of surveillance, including adoption of new technologies (e.g. development of new diagnostic methods). • Changes to behaviour or influences on behaviour of key actors and agents in the system (e.g. changes to reporting behaviour or the costs of diagnostic services). • An assessment of how likely it is that such changes may occur in the future and whether the surveillance system would be able to respond to these changes should also be made. Assessment of this attribute will be aided by consultation with key stakeholders of the system.
2	Organisation and management	How surveillance is organised and managed.	This attribute is based on an assessment of organisational structures of the surveillance system, including whether the objectives are relevant and clearly defined. Where applicable, describe existing formal steering and technical committees. Where committees exist their members should have appropriate expertise, clearly defined roles and responsibilities and should communicate regularly to oversee the function of the system. Members should also be representative of surveillance stakeholders. This also includes assessments of complexity and efficiency in meeting surveillance objectives.
3	Performance indicators and evaluation	Whether performance indicators are routinely used to monitor system performance and whether periodic external evaluation is used to assess the system outputs in relation to its objectives.	This attribute depends on whether performance indicators are routinely used to monitor system performance and whether periodic external evaluation is used to assess the system outputs in relation to its objectives. If indicators are used they should be named and described. Also any available results should be presented.

4.3.2 B: Attributes assessing surveillance processes

Group B attributes assess surveillance processes including the design of the surveillance system. They aim to provide a structured understanding of the methods and practices applied (e.g. during sampling or data analysis, as well as the technical competence and resources that support the surveillance system).

#	Name	Description	Guidance notes
4	Data analysis	Appropriate methods are used for analysis and interpretation of data.	The analysis is conducted with appropriate frequency and utilises suitable descriptive and analytical methods to produce valid results. Surveillance systems that perform well in this attribute will use analytical methods that are appropriate to the data and the information needs of users of the data whilst exploiting the data to its fullest extent. An evaluation of data analysis should include: • The identification of the analysis methods applied to surveillance data: – No analysis – Basic descriptive statistics – Examination of trends – More sophisticated statistical approaches (e.g. time series analyses, spatial analyses) • An assessment of whether the limitations of data have been understood and accounted for in statistical analyses? • An indication as to whether the body of data available is being fully exploited or could further use of data be made? It may help to review requirements for information made by users of the surveillance data in the past, to determine whether their needs were met by the methods applied.
5	Data and information collection	The use of appropriate data and information sources, sampling strategy and data collection methods.	A surveillance system that scores well on this attribute will have a clear and comprehensive case definition and risk organism description; make use of appropriate diagnostic tests; have a written protocol that describes collection of data (and samples); and the limitations of the collection methods will be clearly defined and understood. Questions to consider when assessing data and information collection include: • Is there (if applicable) a written case definition/organism description for this surveillance system that is clearly defined and complete, with specified inclusion and exclusion criteria? If so: – Does the case definition/organism description include relevant details of the case signalment, clinical and pathological signs and epidemiological information as appropriate? – Does the case definition include laboratory diagnosis? Alternatively, does the organism description include taxonomic ID? – Are the chosen diagnostic/taxonomic methods appropriate to the case definition/organism description, including in terms of samples being collected? – Are syndromes used and – if yes – defined in an appropriate way? – In those sectors where symptomatic surveillance is

#	Name	Description	Guidance notes
			undertaken, are symptoms clearly defined for when samples need to be taken? • Is there a written sample and/or data collection protocol and are there appropriate assurance mechanisms to ensure the protocols are followed? • Have the sensitivity and specificity of the tests used been assessed (where relevant)? • Are there data collected that are not used in analysis, interpretation or surveillance management (redundancy)? • Are there information needs for which data are not currently collected and feasibly could be? • Are appropriate sampling strategies used, including the use of risk-based approaches and pooled sampling? This could include risk-based requirement calculations or risk-based sampling. The basis of the risks used in the design of the risk-based sampling strategy should be reviewed. Data collection methods should be clearly documented. It may help to review demands for information made by users of the surveillance data in the past, to determine whether their needs were met by the data available.
6	Data management and storage	How surveillance data is managed and stored.	Appropriate use and documentation of data management systems for processing information, including data processing protocols and effective use of data verification procedures, data storage and back-up protocols. Measures taken to assure authorised computer system access and to maintain confidentiality where needed. Is there a dedicated custodian? Data management is a broad area concerning the collation, storage and maintenance of data, including but not limited to matters of data quality, accessibility, usefulness and security. Assessing this attribute will require an intimate understanding of the data storage and management systems employed by the surveillance activity. More detailed references on assessing data management are provided in the Methods Catalogue of this document (Appendix 1). An assessment of this attribute should include: • Consideration of whether the database structure has been correctly designed: – Has each field of data been tightly defined to ensure correctness, conciseness and consistency across records? – Have primary keys, uniquely identifying each record, been assigned? – Has the database been normalised, to ensure data is stored in the most parsimonious, transparent and useable way? – Have validation constraints, preventing the input of invalid data, and internal cross-consistency checks been applied? – Is the data stored in a way that allows the required interrogation and analysis? • Consideration of whether documentation of the data is sufficient to facilitate interpretation and understanding of the data: – Is there a document providing a summary overview of the data and collection methods and

#	Name	Description	Guidance notes
			explaining any idiosyncrasies relevant to the analysis and interpretation of the data? – Is there a data dictionary that clearly defines each field? – Is there an entity relationship diagram that explains how the data relate? • Consideration of whether there are adequate and documented protocols for managing data quality and security: – Is the data management system covered by a data quality standard (e.g. ISO9000, Good Clinical Practice or Good Laboratory Practice)? – Are periodic data quality control checks implemented? – Are records management issues clearly defined, including policy on the retention of data? – Are data back-up and storage protocols in place?
7	Field and laboratory services	Field and laboratory activities are carried out using appropriate methods with quality assurance and timely and accurate production of results.	The tests used should have the required test sensitivity and specificity and be performed by accredited laboratories or personnel. Sampling should follow SOPs.
8	Resource availability	An assessment of the financial and human resources available and required for implementing the surveillance system.	The personnel have the required expertise and capability for conducting their tasks. There is sufficient laboratory capacity to allow turn-around of samples and reporting within acceptable (defined) time periods. Responsibilities for providing resources are clearly documented. Available resources match current requirements.
9	Technical competency and training	Technical skills of the personnel involved in the surveillance system, including access to relevant training.	The team providing technical management, guidance and day-to-day operation of the surveillance system should have adequate technical skills in relevant disciplines (such as epidemiology or ecology) to be able to perform the relevant analysis, interpretation and information dissemination. This includes the provision of adequate initial training and an on-going programme of training for those implementing the surveillance system, particularly those collecting the data.

4.3.3 C: Attributes assessing the technical implementation of surveillance

Attributes in Group C focus on technical aspects of surveillance and include characteristics such as timeliness, participation and coverage.

#	Name	Description	Guidance notes
10	Acceptability and engagement	Willingness of persons and organisations to participate in the surveillance system and the degree to which each of these users is engaging in the surveillance.	Acceptability/Participation examines the involvement or engagement of stakeholders in the planning, design and implementation of the surveillance activity. Poor engagement by some users might suggest a low level of motivation to become involved in surveillance activities, or a perceived lack of benefit. Technical, financial or knowledge issues could be other reasons for low levels of engagement. Reasons for low levels of engagement should be identified and described. The efficacy of any surveillance system that is greatly dependent on voluntary participation or human behaviour (e.g. passive surveillance activities) will be vulnerable to problems with engagement. This attribute could include an assessment of stakeholder awareness of the system and their understanding of it. One could also assess their beliefs about the benefits or adverse consequences of their participation in the system, including the provision of compensation for the consequence of disease/risk organism detection. Communication is known to be a key driver of engagement. This attribute includes an assessment of participation including identification of the factors likely to increase or decrease stakeholder participation and an assessment of the likely extent of impact of these factors on levels of participation. Qualitative or semi-quantitative social science approaches are likely to be of value in assessing participation. Consultation with all those involved in generating, analysing, reporting and using surveillance data will be valuable. Factors that may influence participation include: • What communication pathways exist internal to the surveillance system (e.g. between those collecting or providing data and those analysing and reporting the data)? Are these pathways formalised in any fashion? • Does information and feedback flow freely between those implementing surveillance and those using surveillance data? • How are each of the key stakeholders represented in the planning, design and implementation stages of the surveillance activity? • What are the incentives (e.g. compensation payments) or barriers (e.g. consequences of reporting) for participation?
11	Coverage	Proportion of the population of interest (target population) or proportion of areas of interest (e.g. specific habitats or high-risk sites) that is included in the surveillance activity.	The coverage of a surveillance system is related to representativeness, bias and sensitivity. Coverage can be particularly important in surveillance for the early detection of exotic or new (emerging) diseases or risk organisms. An assessment of coverage could include:

#	Name	Description	Guidance notes
			• Characterisation and qualitative comparison of the sampled and target populations. Alternatively, comparison of sampled areas or habitats versus areas or habitats of interest. • Where sufficient data on the target population/areas or habitats of interest exists, simple calculations of the proportional coverage can be made (e.g. 75% of the national herd and 45% of cattle holdings are sampled annually or 30% of the marine ports). • Where sufficient information on the background population, or high-risk areas or habitats respectively, is lacking, more sophisticated sampling designs might be employed (e.g. capture-recapture analysis or drop camera surveillance). • Considering whether the target population, or area of interest, has been adequately defined (i.e. whether the exclusion of certain animals or holdings or sites is merited). • The unit of interest – the level at which coverage is measured – is often the unit of interest of surveillance (e.g. animal, holding, high-risk site or specific marine habitat). If insufficient data exist, alternative perspectives might be desired. Coverage might then be assessed at other aggregate levels (e.g. geographical areas) or relevant intermediate steps in the surveillance pathway (e.g. the proportion of veterinary practices submitting diagnostic samples). • In certain contexts it may be worth establishing a timeframe of reference (e.g. annual coverage). The choice of timeframe should reflect the epidemiology of the disease or life history of a risk organism.
12	Data completeness and correctness	How complete and correct is the data obtained and recorded by surveillance.	Assessment of the proportion of data that was intended to be collected that actually was, and the proportion of data entries that are complete (i.e. include all variables) and correctly reflect the true value of the data collected. Includes assessment of data quality and documents if data validation is occurring. Completeness of surveillance data is the percentage of complete entries and should be considered at two levels: fields and records. Most commonly data completeness is measured as the proportion of records with complete and valid data in the data fields – where data fields are variables containing (where applicable) demographic, morphometric, taxonomic, clinical, pathologic or epidemiological information recorded for each sample. Key data fields (e.g. animal ID, holding of origin, test result etc.) should be identified and the proportion of completeness measured. Measurement of the proportion of records or observations that have been collated in the data system may also be considered. This will require comparison with an alternative source of data (e.g. the sample frame or paper records of sampling and test results). Poor data completeness may indicate problems in the following attributes: “data and information collection”, ”data management” or “internal communication” and “acceptability and engagement”.
13	Interoperability	Compatibility with and ability to	This is only relevant where such interfacing is a requirement to

#	Name	Description	Guidance notes
		integrate data from other sources and surveillance components.	assure utility. Most technical requirements for interoperability are nowadays standard characteristics of databases and information systems. Record keys are required to assure correct merging of records. These can for example be animal or sample IDs, holding IDs, postal codes.
14	Multiple utility	The ability of a surveillance system to capture information on several diseases, syndromes or risk organisms.	This is a measure of how generic a surveillance system is. For example, the collection of slaughterhouse records can provide information on the presence/absence of several diseases and risk organisms. Multiple utility in a system should be considered when examining the cost-effectiveness of a system. It may also be of benefit to assess the potential multiple utility which may provide recommendations on how to add value to the system currently implemented. An assessment of multiple utility could consider: • What additional information is or could be gathered during sample collection (e.g. on animal health or husbandry and demographics; or other endemic/native/cryptogenic/non-native organisms that are present or absent)? • What other types of samples are or could be collected at the time of sampling (e.g. environmental)? • What other diseases/risk organisms are or could be tested for with the samples collected? • How long are samples stored following testing and could they be used for other purposes (including other research purposes)? For a surveillance system to offer value to other diseases/risk organisms or information needs, the objectives and processes of the system should be aligned to other systems. So it may be expected that more generic systems are likely to have more potential for multiple utility. For example, a simple random survey of holdings or geographical locations, repeated annually and with good coverage and representativeness could be useful for various diseases/risk organisms; whereas a risk-based design aimed at a specific threat may be of limited value for other diseases/risk organisms with differing epidemiology, host-range, life history or habitat preferences.
15	RARR (Reliability, availability, repeatability, and robustness)	How reliable, available, repeatable and robust is the surveillance system.	Reliability means “does the system function without failure” and availability means “is the system operational when needed”. These attributes can be measured retrospectively by • Looking at the incidence of minor and major faults over a defined period of time or • Measuring the proportion of time that the system is fully functional. Assessment of these attributes will benefit from consultation with those involved in the generation, management and analysis of surveillance data. If performance indicators have been implemented in the surveillance process, historical data from these will give a good insight into the ongoing functioning of the system. Repeatability means “can the surveillance component performance be maintained consistently over time” or “how

#	Name	Description	Guidance notes
			consistently can the results be reproduced over time". Repeatability is often considered when validating diagnostic tests. A surveillance activity that performs well in this attribute produces data that are comparable across years and where changes to the data and data collection methods over time are clearly defined, understood and documented. In this context, one might consider changes to legislation; changes to diagnostic methods and sampling techniques, including improvements through adoption of new technology; changes to surveillance design; or influences on disease or risk organism reporting behaviour in passive surveillance activities. • How have these impacted on the comparability of surveillance data over the time period of interest? • Have these influences been identified and examined and can they be accommodated in interpretation of the surveillance data? Robustness means "the ability to obtain comparable results over time". It covers the ability of the surveillance system to produce acceptable outcomes over a range of assumptions about uncertainty by maximising the reliability of an adequate outcome.
16	Timeliness	The time between any two defined steps in a surveillance system.	The steps will vary according to the surveillance objectives so as to be epidemiologically or biologically meaningful. Commonly timeliness relates to the time interval between a relevant event/signal and its recording by the surveillance system. The timeliness of a surveillance system is especially important to surveillance for the early detection of emerging or exotic disease or organism threats – where the intention is to implement control measures as soon as possible. For example, for outbreak or incursion detection it might be important to consider the time delay from introduction to detection of the agent, or the time between when the agent should have realistically been first detected and the time when it actually was reported. On the other hand, for planning purposes, timeliness might be used to determine if a surveillance system detects and reports disease or risk organisms in time to initiate effective interventions before disease or risk organisms become widespread. Timeliness can be defined in various ways • This is usually defined as the time between any two defined steps in a surveillance system; the time points chosen are likely to vary depending on the purpose of the surveillance activity. • For planning purposes timeliness can also be defined as whether surveillance detects changes in time for risk mitigation measures to reduce the likelihood of further spread. The precise definition of timeliness chosen should be stated as part of the evaluation process.

4.3.4 D: Attributes assessing surveillance outputs

Group D attributes assess the outputs of surveillance, to gain an understanding of their limitations and qualities.

#	Name	Description	Guidance notes
17	Historical data	Quality and accessibility of archived data.	Maintaining historical data is more important to surveillance activities designed to provide evidence for freedom from disease or risk organisms or for monitoring trends in prevalence of endemic disease or risk organisms. Historical data can also be valuable for research and trend analysis. This attribute is related to "data management and storage" and "RARR". Questions to consider include: • How many years of data are stored? • How complete and reliable are the data? • Are the data stored in a way that allows the required interrogation and analysis? • Is there a summary overview of the data and collection methods explaining key idiosyncrasies of the data and changes to the data or collection methods over time? • What use is currently made of historical surveillance data?
18	Negative predictive value	The probability that no disease/risk organism is present given that none is detected by the system.	The negative predictive value expresses the chances of missing the presence of a disease or risk organism. It can therefore be considered a reflection of the risk of false-negative surveillance outcomes. False negative results can be very costly, for example in export testing. Most surveillance systems therefore aim to maximise the negative predictive value (ideal value is 1). This attribute is mainly influenced by the test characteristics (sensitivity and specificity) as well as the prevalence of disease. Alternatively, where applicable, negative predictive value can be influenced by the methods of surveillance and the density and geographic spread of the risk organism.
19	Positive predictive value	Probability that a disease/risk organism is present given that it is detected by the system.	The positive predictive value expresses the probability that a disease/risk organism is present, given that that it has been detected by the surveillance system. It can therefore be considered a reflection of the risk of false-positive surveillance outcomes. Such false positive results can be costly. However in the situation of a very severe outcome (e.g. highly contagious diseases such as FMD or the presence of a voracious pest such as *Carcinus maenus*), they are often acceptable as long as the control measures and trade disruptions can be managed. The ideal value of the positive predictive value is 1. This attribute is mainly influenced by the test characteristics (sensitivity and specificity) as well as the prevalence of disease. Alternatively, where applicable, positive predictive value can be influenced by the methods of surveillance and the density and geographic spread of the risk organism.
20	Precision	How certain a numerical estimate obtained from the study population is or –	A precise estimate has a narrow confidence interval. Precision is influenced by sample size, the chosen confidence level and

#	Name	Description	Guidance notes
		alternatively – how large the uncertainty of an estimate is.	data completeness and correctness. Precision in surveillance activities designed to monitor prevalence or density is a measure of the degree of certainty around the point estimate of prevalence, incidence or density (i.e. the confidence interval or standard error). NB: A related concept in surveillance designed to provide evidence for freedom from disease or risk organism is the measure of confidence in disease or risk organism freedom derived from the sensitivity of the surveillance system. The precision of point estimates in epidemiological studies is dependent upon disease or risk organism prevalence, sample size and the approach to sample selection (i.e. the design effect). In ecological studies the precision of the estimates is dependent on the area covered (sample size) and the search methodology (e.g. search efficiency). The precision of a surveillance activity will determine how sensitive the surveillance system is to changes in prevalence or density. The desired level of precision will be defined by the epidemiology of the disease or risk organism, the surveillance objectives, the risk that the disease or risk organism poses and the availability of resources.
21	Representativeness and bias	Extent to which the frequency of features of the population of interest are correctly reflected in the surveillance data that are collected.	A surveillance system that is representative accurately describes the distribution of disease/risk organisms in the population or area of interest. Bias describes the extent to which a prevalence or population density estimate produced by the surveillance system deviates from the true prevalence or population density value. Bias is reduced as representativeness is increased. The representativeness of a surveillance system is related to coverage and bias; it is a comparison of the sample and target populations or specific areas of interest with regard to a number of key features or risk factors. Features taken into account when considering representativeness could include, for example, production type, species, geographic location, habitat preferences and environmental parameters. Bias can be divided into two main types: information and selection bias. Information bias results from systematic differences in the way information is collected, for example on the presence or absence of a disease or a risk. Selection bias occurs when there is a systematic difference between the individuals/samples/transects included and those that are not. Some potential sources of bias to consider include: • Sensitivity/specificity of the methods applied. • Under-reporting in passive surveillance activities. • The sample source population: For example in animal surveillance, sampling at abattoirs may lead to an under-estimate of the prevalence of many diseases as these animals are from a healthy (and younger) sub-population, whereas sampling fallen stock may lead to an over-estimate of burden. Similarly, in marine surveillance sampling a receiving location may lead to an underestimate of risk organism incursions, whereas sampling vectors or pathways will provide an over

#	Name	Description	Guidance notes
			estimate of invasion success. • Selection bias may also be introduced in terms of geography, habitat, population size, organism traits, species, age, sex or purpose. As such, the first step will be to identify key characteristics of the target population or areas of interest upon which to measure representativeness. These characteristics might be risk factors for the disease or organism threat – knowledge of the associations between these characteristics, selection in the sample population/environment and disease/organism will inform the understanding of bias. Examples of relevant features include: • Species • Population or group size (e.g. herd size) • System type or focus (e.g. production type) • Age, sex or purpose • Geographic location • Habitat • Presence/absence or frequency of vectors The second consideration in assessing representativeness is whether there are sufficient and accurate data on the identified features in both the target and sample populations. Bias may lead to erroneous conclusions about the burden or distribution of disease in the population or an organism in an environment. For some surveillance activities – such as risk-based surveillance aimed at detecting cases to facilitate control – surveillance may be intentionally biased toward sub-groups of the population at higher risk of disease or geographic locations and habitat strata where the occurrence of risk organisms is more likely. So the context and objective of surveillance will determine whether bias is acceptable or not. Bias in the surveillance output can be examined and quantified by several methods (see Methods Catalogue: Appendix 1). If bias is deemed to be significant and unacceptable and cannot be satisfactorily corrected for during analysis and interpretation of the data, one might consider reviewing the design and implementation of the surveillance activity.
22	Sensitivity[6]	Proportion of true events correctly classified as such.	Evaluation of the sensitivity of a surveillance system is especially important for surveillance activities designed to detect outbreaks and incursions. Sensitivity of a surveillance system should be considered according to the objective of the surveillance activity: • Surveillance sensitivity (case detection) refers to the proportion of individual units (e.g. animals, plants) that have the condition of interest that the surveillance system is designed to detect that are correctly identified as such, • Surveillance sensitivity (outbreak or incursion detection) refers to the probability that the surveillance system will detect a significant event. This requires a clear definition of what constitutes a

[6] At least one of the two attributes 'Sensitivity' and Specificity' should be included.

#	Name	Description	Guidance notes
			significant event (e.g. an outbreak, incursion, or host expansion). • Surveillance sensitivity (presence) refers to the probability that an event will be detected if present at a certain level (prevalence) in the population. Sensitivity is the most commonly assessed attribute of surveillance systems. Combined with timeliness, it is of particular importance to surveillance for early detection of outbreaks or incursions. Along with representativeness it is frequently scrutinised when evaluating surveillance activities intended to provide evidence for disease or risk organism freedom. When monitoring the prevalence of endemic diseases or risk organisms, poor sensitivity will contribute to bias in the surveillance outputs. Some considerations when assessing the sensitivity of surveillance include: • The probability of selection into the surveillance system must be defined and quantified. This may be a simple random sample of animals from a single homogenous population or a complex pathway of epidemiologic and behavioural factors describing the observation, reporting and subsequent investigation of a notifiable disease or risk organism (i.e. passive surveillance). • The probability of diagnosis (i.e. the sensitivity of the diagnostic protocol, including that of laboratory tests). • The choice of design prevalence or assumed density (i.e. the expected prevalence of disease or density of a risk organism that the system is designed to detect) is a key assumption. • In ecology, when the main aim is detection of a single pest organism, search efficiency (ability to detect the organism if it is there) could be taken into account.

#	Name	Description	Guidance notes
23	Specificity[7]	Proportion of true non-events correctly classified as such.	Evaluation of the specificity of a surveillance system is especially important for surveillance activities designed to detect outbreaks or incursions and cases because it is related to the misdirection of resources: i.e. expenditure on a disease or risk organism investigation and mitigation measures that are needlessly applied. The specificity of many surveillance activities will be very high or complete (100%), because of the consequences of confirming disease; this is especially true for surveillance for exotic diseases carrying implications for trade. Specificity can be considered at several levels, depending upon the epidemiology of the disease or organism and the objectives and design of the system: • The specificity of pre-diagnostic indicators of disease (e.g. clinical signs). • The specificity of screening and confirmatory diagnostic tests applied. • The rate of false-positive signals raised by detection algorithms applied to surveillance data. • The proportion of reports of suspect cases of a disease or risk organism that are subsequently negated (NB: this metric actually concerns the Positive Predictive Value of a system; a related concept which has been assessed in some evaluations). Assessment of specificity should include the false alarm rate, i.e. the proportion of wrongly suspected outbreaks or incursions. False alarm rate is the inverse of the specificity (i.e. the proportion of true non-events correctly classified as such) but is by some more easily understood than specificity.

[7] At least one of the two attributes 'Sensitivity' and Specificity' should be included.

4.3.5 E: Attributes assessing the impact of surveillance

Attributes in Group E focus on the assessment of surveillance impact, considering the benefits provided by the system and what data and information are communicated to stakeholders.

#	Name	Description	Guidance notes
24	Benefit	Direct and indirect advantages provided by the information generated by the surveillance system.	The benefit of surveillance quantifies the monetary and non-monetary positive direct and indirect consequences produced by the surveillance system and assesses whether users are satisfied that their requirements have been met. This includes financial savings, better use of resources and any losses avoided due to the existence of the system and the information it provides. These avoided losses may include the avoidance of: • Primary industry production losses • Human mortality and morbidity • Economic losses • Decrease in consumer confidence • Threatened livelihoods • Harmed ecosystems • Utility loss • Loss of sociocultural values Often, the benefit of surveillance estimated as losses avoided can only be realised by implementing an intervention. Hence, it is necessary to also assess the effect of the intervention and look at surveillance, intervention and loss avoidance as a three-variable relationship. Further benefits of surveillance include maintained or increased trade, improved ability to react in case of an outbreak of a disease or incursion of a risk organism, maintaining a structured network of professionals able to react appropriately against a (future) threat, maintaining a critical level of infrastructure for disease/risk organism control, increased understanding about a disease or risk organism, and improved ability to react in case of an outbreak of a disease or incursion of a risk organism. The benefits of a surveillance activity should be listed and, where possible, quantified. An evaluation of the benefits of a surveillance activity may include: • A characterisation of all the potential benefits of the surveillance activity • A description of benefits as perceived by the relevant stakeholders • Where possible, quantify market benefits in financial terms • Where possible, quantify non-monetary benefits by alternative methods • Consider how the benefits are distributed among stakeholders, including: producers, consumers, the livestock industry or society Points to consider whilst assessing the benefits of surveillance include: • Surveillance and disease or risk organism control are

#	Name	Description	Guidance notes
			often integrated: That is to say, surveillance provides information that informs control and so many benefits of surveillance are often realised by control measures. As with costs, it is important to understand the benefits of surveillance in the broader context of disease or risk organism mitigation. Benefits of surveillance may be considered as disease or risk organism losses and mitigation costs avoided by detection of a disease or risk organism. So it may be useful to begin by listing all the losses and costs resulting from a disease or risk organism, and disease or risk organism mitigation measures. It may be difficult in some instances to distinguish between the direct benefits of surveillance and those arising from mitigation. • The benefits of surveillance for early detection of disease or risk organism outbreaks can be quantified as the losses and costs avoided through earlier detection and control. • The primary benefit of surveillance providing evidence of disease or risk organism freedom is access to international markets. The economic value of international trade can be attributed as a benefit to surveillance. Officially recognised disease or risk organism-free status often permits the disease or risk organism-free country/region to maintain border security measures against introduction of the disease (e.g. restriction on trade and movement of risk goods or requirement for pre-export testing) – thus mitigation of risk of incursion is also a benefit of surveillance for freedom from disease or risk organisms. • Surveillance for case-detection and monitoring prevalence of endemic disease or presence of risk organisms provides information for the improved control and management of disease and risk organisms; including prioritisation of diseases and risk organisms and allocation of resources. • Improved public health is an obvious advantage to surveillance for zoonoses. Increased consumer confidence is another – although consumer confidence may also be of significance to other high-profile, non-zoonotic diseases. • Consider potential indirect or secondary benefits of surveillance; e.g. externalities or spill over of benefit to other sectors or industries. It may be helpful to consider potential benefits both upstream (e.g. animal feed producers) and downstream (e.g. value-added producers) of the production system. Examining the value chain will aid in this.
25	Decision support	The direct link between the information created by surveillance and decision-making.	Includes an assessment of the availability of the information created by surveillance to relevant decision-makers. For example, describes how surveillance infrastructure is used to provide decision-support during outbreaks or incursions or is used for priority setting. Includes assessment of reporting of surveillance outputs to decision-makers.

#	Name	Description	Guidance notes
26	Efficiency	Link between the resources implemented and the results obtained. An efficient system will accomplish a job with minimum expenditure of time, human effort and cost.	Conducting surveillance incurs costs, for example, salaries, consumables, travel. These costs can be compared against the outputs from surveillance such as reports, disease or organism detections and notifications, or other signals. A surveillance system can be considered efficient if there is an optimal balance between economic investments and its output, the latter achieving the desired quality attributes (e.g. precision, timeliness). Risk-based surveillance can – where appropriate in terms of the surveillance objective – provide efficiency gains in surveillance systems.
27	External communication and dissemination	An assessment of the data and information provided to relevant stakeholders outside of the surveillance system.	An assessment of the data and information provided to those outside the surveillance system including the timeliness and types of output produced. The efforts made to disseminate these outputs including the use of web-based systems should also be assessed. Communication concerns the dissemination of information and provision of feedback into the system. Communication in a surveillance system is often related to various other factors such as participation, timeliness, stakeholder interest and system impact. Relevant to the assessment of both "internal communication" and "external communication and dissemination" an assessment of communication should consider: • A list of the outputs that are generated from the surveillance data; who are these intended for and do they meet all information needs of the target audience? • An assessment of who has access to the surveillance outputs; are all stakeholders represented? • An assessment of whether the surveillance outputs are produced sufficiently frequently. Do they contain up-to-date data of sufficient quality? Are the data presented with sufficient discussion of its meaning, limitations and biases from an epidemiological perspective? • A list of other feedback provided to those contributing to the surveillance system e.g. data quality checks. Qualitative or semi-quantitative social science approaches (e.g. stakeholder interviews, focus group discussions) are likely to be of value in assessing this attribute. Consultation with the key stakeholder groups of the surveillance system will be useful, including: • Providers of surveillance data (e.g. farmers, veterinarians, laboratory staff, taxonomists, field workers etc.). • Those analysing and interpreting the surveillance data (i.e. generating information and knowledge from the data and disseminating it). • Users of surveillance data, including the direct customer (funder) but also other beneficiaries of the information as appropriate (e.g. government, industry or academia).
28	Internal communication	An assessment of the data and information provided to relevant stakeholders inside the surveillance system.	An assessment of the methods used and ease of information exchange between all those involved in providing, managing, analysing and disseminating information for the surveillance system. The methods used to provide feedback to data providers and to increase their awareness about hazards and surveillance activities should also be assessed.

#	Name	Description	Guidance notes
			Relevant to the assessment of both "internal communication" and "external communication and dissemination" an assessment of communication should consider: • A list of the outputs that are generated from the surveillance data; who are these intended for, do they meet all information needs and are they are at the required level for the target audience? • An assessment of who has access to the surveillance outputs; are all stakeholders represented? • An assessment of whether the surveillance outputs are produced sufficiently frequently. Do they contain up-to-date data of sufficient quality? Are the data presented with sufficient discussion of its meaning, limitations and biases from an epidemiological perspective? • A list of other feedback provided to those contributing to the surveillance system e.g. data quality checks. Qualitative or semi-quantitative social science approaches (e.g. stakeholder interviews, focus group discussions) are likely to be of value in assessing this attribute. Consultation with the key stakeholder groups of the surveillance system will be useful, including: • Providers of surveillance data (e.g. farmers, veterinarians, laboratory staff or taxonomists). • Those analysing and interpreting the surveillance data (i.e. generating information and knowledge from the data and disseminating it). • Users of surveillance data, including the direct customer (funder) but also other beneficiaries of the information as appropriate (e.g. government, industry or academia).
29	Utility	Describes how useful, profitable, or beneficial surveillance is in relation to its objectives and describes the changes that have been made based on the outputs provided by the surveillance system.	This attribute consists of an integrated appraisal of the actions taken as a result of the information provided by the surveillance system, e.g. changes in protocols or behaviour and changes in mitigation measures and particularly changes in disease or risk organism occurrence. Even not taking action can be considered a valid conclusion based on surveillance information provided. The attribute is mostly assessed in a descriptive (qualitative) way. However, more comprehensive assessments are possible, including the simulation and economic assessment of outbreaks and incursions that may have been avoided thanks to surveillance-based interventions. The attribute describes the extent to which surveillance objectives are achieved and includes an assessment of stakeholder uptake and acceptance. Stakeholder input is relevant to this attribute.

5 Integrating Surveillance System Performance at a Glance

Attribute assessment by SurF is supported by a visual output (Figure 4). At the individual evaluation level this allows quick assessment of a system's strengths and weaknesses and, in addition to the case report form, standardises the reporting of SurF results across different evaluations.

An additional element of SurF is the framework's ability to support the assessment of the performance of MPI's surveillance systems and programmes to provide assurances around the quality of delivery and the outputs of MPI surveillance programmes. This may include business intelligence reporting requirements such as the number of MPI surveillance systems that have elements in need of attention, or the percentage of systems with the majority of attributes rated as good or excellent. However, this functionality should be applied with caution as it assumes that all attributes have the same weight. This is almost certainly not the case. Furthermore, previous results could be used to benchmark performance over time. We recommend using this feature mainly for providing a quick overview. Users should make sure to also consider all attributes in detail to make sure that minor deficiencies in highly critical attributes are not missed.

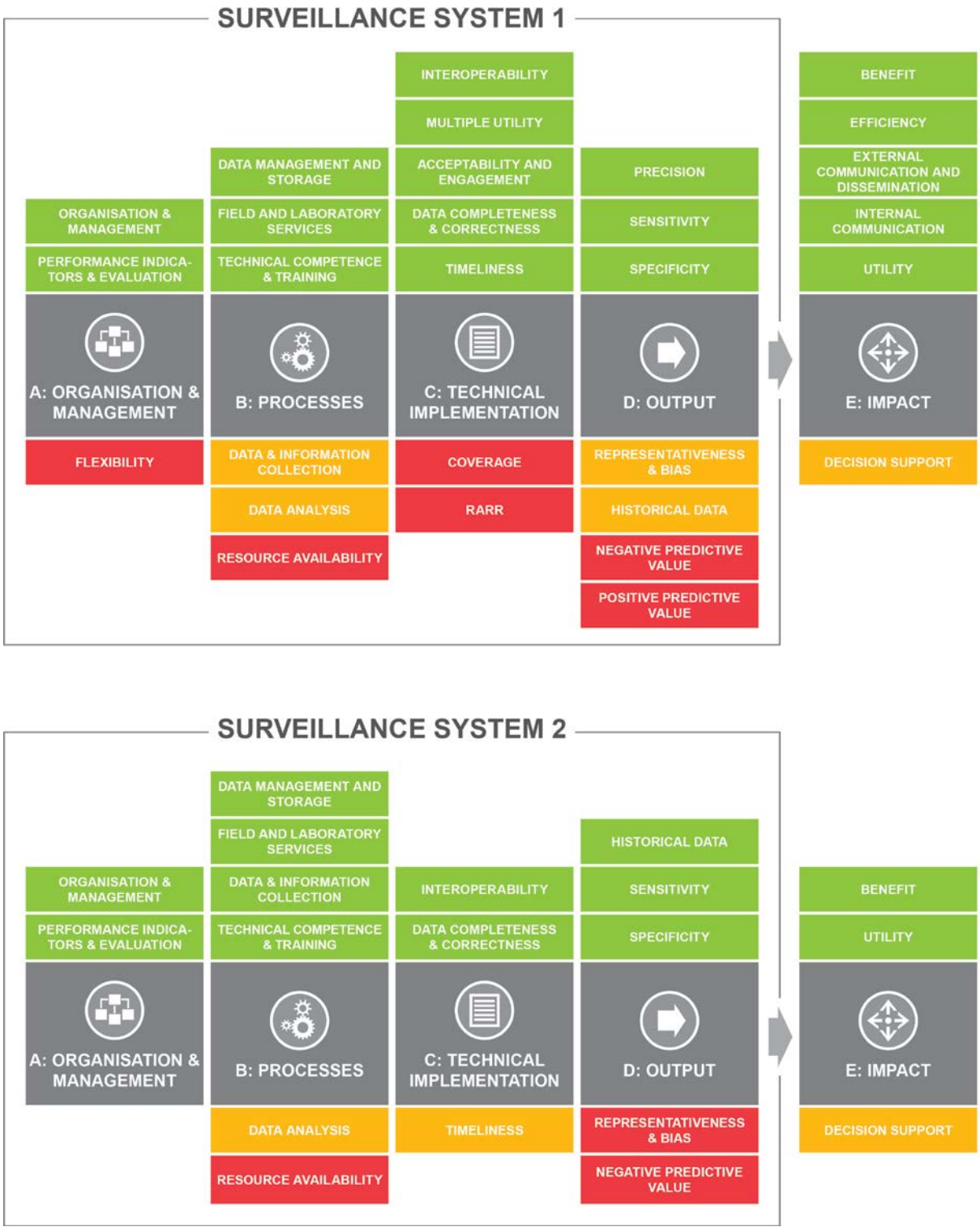

Figure 4: *Visual outputs of performance assessment of attributes using the SurF framework. The format allows comparison between different evaluations or systems (described here as 'System 1' and 'System 2'). Attributes assessed positively are always placed at the top of the process box, while those in potential need of attention are placed below.*

6 Important Considerations for Users

The MPI evaluation framework was designed to ensure consistency in the evaluation of different biosecurity surveillance systems by providing a robust process that is not sector- or context-specific. This should also make results of evaluations comparable. Although SurF was developed for internal use by MPI, it could be applied to any surveillance system, including, for example, surveillance conducted by regional or city councils in New Zealand. MPI will consider external (independent) evaluation or consultation, in addition to SurF, where required. SurF draws from existing surveillance frameworks and in many parts adopts what has been developed elsewhere. Its greatest novelty lies in the extension to surveillance in the context of plant, environment and marine biosecurity surveillance and combining this with biosecurity animal health surveillance under a common umbrella.

The aim was to develop a generic framework that is not overly specific and allows sufficient flexibility to be used across the wide range of MPI surveillance systems and to compare and assess system performance. While the standardised assessment of core attributes provides consistency between the assessments of different systems, the choice of accessory attributes allows users to tailor the evaluation to the individual context. SurF is a very generic framework, which allows a large amount of flexibility around attribute selection and as such differs from recently published animal surveillance frameworks, which emphasise alignment of attributes with specific surveillance objectives e.g. freedom from disease (Drewe *et al*. 2015; EVA 2013, 2015). Further, a substantial number of attributes are included to accommodate the diversity and unique context of MPI's surveillance systems.

Ideally the structured approach and information provided by SurF will not only be of benefit to MPI but also to other New Zealand stakeholders. Outputs of surveillance evaluation can be a useful tool to communicate operating principles of surveillance and the value they provide can be a source of assurance and credibility (Drewe *et al*. 2015). In this sense, outputs provided by SurF could be used to inform both national and international stakeholders. Outputs can also be of interest in a quality assurance context.

Case studies were prepared by MPI subject matter experts between September and December 2015, using data and information that was already available (see Appendix 2 for details). The objective of the case studies was to provide a proof of concept approach, which shows that the framework appears robust, complete and user-friendly across the different biosecurity sectors it is targeting. The evaluations provide a comprehensive appraisal of the selected systems using available data and include the following MPI programmes:

- National Apiculture Surveillance Programme (NASP)
- Marine High Risk Site Surveillance Programme (MHRSS)
- Forestry High Risk Site Surveillance (HRSS)

Importantly, these case studies were developed with the goal of testing SurF and providing applied guidance to future SurF users. As such they provide non-peer-reviewed example evaluations to illustrate the framework at use, rather than finalised assessments.

7 SurF Evaluation Template

Evaluation section	*Details*
Identification of Surveillance System	
Name	
I. Motivation for the evaluation	
A. Evaluation trigger(s)	
B. Context	
II. Scope of the evaluation	
A. Evaluation objective	
B. Evaluation questions	
C. Time and resources	
D. Evaluation intensity	

Evaluation section	*Details*
E. Evaluation organisation and composition of the evaluation team	
F. Status of evaluation outputs	
III. Evaluation Design & Implementation	
Design of the evaluation	
A. Select attributes from master list	
B. Choose methods to assess attributes	
C. Make an inventory of available information sources about the systems	
D. Identify missing information	
Implementation of the evaluation	
E. Describe the surveillance system under evaluation	
F. Describe the surveillance system's objective(s)	

Evaluation section		*Details*	
G. Describe the organisational structure			
H. Identify and engage surveillance system users			
I. Identify the target population and geographical coverage			
J. Describe the design of the surveillance system			
K. Describe the processes			
L. Collection of data and information			
M. Assess the included attributes[8]		Traffic-light code[9]	Details
A. Organisation & Management	1. Flexibility	● ● ●	
	2. Organisation and management	● ● ●	
	3. Performance indicators and evaluation	● ● ●	

[8] Note: Core attributes (**in bold**) assess essential aspects common to all surveillance systems, and should be included in all evaluations. If for any reason this has not been done, justification should be provided.

[9] Rate each attribute according to the traffic light code, where ● = Excellent or very good ● = Good, though room for improvement ● = In need of attention. Either circle the appropriate colour OR delete those that do not apply.

Evaluation section		*Details*	
B. Processes	4. Data analysis		
	5. Data and information collection		
	6. Data management and storage		
	7. Field and laboratory services		
	8. Resource availability		
	9. Technical competence and training		
C. Technical Implementation	10. Acceptability and engagement		
	11. Coverage		
	12. Data completeness and correctness		
	13. Interoperability		
	14. Multiple utility		

Evaluation section		*Details*	
C. Technical Implementation cont.	15. RARR (Reliability, availability, repeatability, and robustness)	● ● ●	
	16. Timeliness	● ● ●	
D. Outputs	17. Historical data	● ● ●	
	18. Negative predictive value	● ● ●	
	19. Positive predictive value	● ● ●	
	20. Precision	● ● ●	
	21. Representativeness and bias	● ● ●	
	22. Sensitivity[10]	● ● ●	
	23. Specificity[10]	● ● ●	
E. Impact	24. Benefit	● ● ●	

[10] At least one of the two attributes 'Sensitivity' and Specificity' is recommended to be included.

Evaluation section		*Details*	
E. Impact cont.	25. Decision support	● ● ●	
	26. Efficiency	● ● ●	
	27. External communication and dissemination	● ● ●	
	28. Internal communication	● ● ●	
	29. Utility	● ● ●	
IV. Reporting & Communication of Evaluation Outputs			
A. Target audience			
B. Report of main findings			
C. Summarise and synthesise results			
D. Provide guidance for interpretation of results			
E. Make recommendations			
F. Plain reporting			

8 References

Aotearoa New Zealand Evaluation Association (ANZEA). Evaluation standards for Aotearoa New Zealand (2015); available at: http://www.anzea.org.nz/wp-content/uploads/2014/08/ANZEA-Superu-Evaluation-standards-final-020415.pdf; last accessed 28 June 2015.

Center for Disease Control (CDC). Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group (2001). *Morbidity and Mortality Weekly Report* (MMWR) 50:1–35.

Drewe JA, Hoinville LJ, Cook AJC, Floyd T, Gunn G, Stärk KDC. SERVAL: A New Framework for the Evaluation of Animal Health Surveillance (2015). *Transboundary and Emerging Diseases* 62:33–45.

European Centre for Disease Prevention and Control (ECDC). Data quality monitoring and surveillance system evaluation - A handbook of methods and applications (2014). Available at: http//www.ecdc.europa.eu/en/publications/Publications/Data-quality-monitoring-surveillance-system-evaluation-Sept-2014.pdf; last accessed 16 February 2015.

Ministry for Primary Industries (MPI), New Zealand. Surveillance Performance Evaluation Framework (2015). Deliverable 1: Scoping Review.

The RISKSUR Project Consortium. The EVA tool: an integrated approach for evaluation of animal health surveillance systems (2013). RISK SUR Task 14 Development of an integer evaluation matrix. Available at: http//www.fp7-risksur.eu/sites/fp7-risksur.eu/files/documents/Deliverables/RISKSUR%20(310806)_D14.pdf; last accessed 10 February 2015.

The RISKSUR Project Consortium. The EVA tool: an integrated approach for evaluation of animal health surveillance systems (2015). RISKSUR Task 5.18 Evaluation framework and methods for economic assessment. Available at: http://www.fp7-risksur.eu/sites/default/files/documents/Deliverables/RISKSUR%20(310806)_D1.4.pdf; last accessed 11 February 2016.

9 Appendices

9.1 APPENDIX 1: SURF METHODS CATALOGUE (PROVIDED IN SEPARATE FILE)

9.2 APPENDIX 2: SURF CASE STUDIES (PROVIDED IN SEPARATE FILE)