Australian Government
Department of Agriculture, Water and the Environment

cebra
Centre of Excellence for Biosecurity Risk Analysis

Ministry for Primary Industries
Manatū Ahu Matua

CEBRA Report Cover Page			
Title, ID, & Output #	*Identifying unexpected biosecurity risks, 1606B, Output #2*		
Project Type	*Standard*	**CEBRA Project Leader**	*Andrew Robinson*
Sponsoring Org.	*DAWE*	**Project Sponsor**	*Tina Hutchinson*
Project Leader/s	*Andrew Trainer*	**Collaborator/s**	*Simon Veitch (DAWR), Greg Hood (DAWR), Tony Arthur, (ABARES).*
Project Objectives	*The objectives of the project are to enable the department to:* *• better manage the compliance that matters most so that it can identify and deal with noncompliance and reward compliance* *• identify and manage unexpected biosecurity risks, as these often present a greater threat than biosecurity risks that are known and regularly managed* *• ensure that data collected by industry as part of Approved Arrangements are useful and utilised.*		
Outputs	A review of methods and tools that can be used for identifying unexpected biosecurity risk, with a case study (not documented here).		
CEBRA Workplan Budget	**Year 2016-17** *$100,000*		
Project Changes	*Nil*		
Research Outcomes	A critical review of methods and tools that can be used for identifying unexpected risk with reference to biosecurity risk management.		
Recommendations	That the department considers deployment of syndromic surveillance to augment data sources used for managing the biosecurity risk of pathways.		
Related Documents	*Nil*		
Report Complete	*Final report has undergone SAC and internal departmental reviews*		

Report Tracking:	SRC Reviews:	Date:
	DAWR/MPI Policy / Operational Reviews:	Date:
	BRSC Endorsement ( ) Yes ( ) No	Date:
	Report published to CEBRA Website ( ) Yes ( ) No	Date:

Identifying Unexpected Biosecurity Risks

Martina Hoffmann and Andrew Robinson, CEBRA

Executive summary

This report provides (i) a critical review of tools, and (ii) potential case study applications for identifying unexpected biosecurity risks. The report is part of CEBRA project 1606B: *Operational Imports Analysis on Compliance*, namely phase 1, milestone 2.

Project 1606B focuses on uses of data holdings, so this report reviews various tools that can be used for identifying unexpected biosecurity risks given data. However, data-free approaches, such as expert consultation, and those that increase but don't analyse data, such as increasing business knowledge, are briefly considered.

We conclude that syndromic surveillance, which focuses on monitoring the occurrence of symptoms of a pest or disease, has the potential to be a particularly useful tool for identifying unexpected biosecurity risks. However, by its nature, unexpected biosecurity risk is elusive, and recommendations about how best to detect and manage unexpected biosecurity risk will always be tentative.

CEBRA recommends that the Department (i) note the content of this report; and (ii) identify existing activities that act against unexpected biosecurity risk even if incidentally, such as the endpoint survey in international passengers and mail.

Table of Contents

1. Introduction

Weeds cost Australia around $1.5 billion a year in weed control activities and a further $2.5 billion a year in lost agricultural production[1]; vertebrate feral animals cost over $740 million in direct economic costs annually[2]. Prevention, and early detection of incursions, would mitigate the need for control measures.

Prevention and early detection of known biosecurity risks both require careful planning and investment of limited intervention resources. Prevention and detection of unknown and unexpected biosecurity risks is an even more complex problem. For example, it is much easier to justify investment in surveillance and mitigation of known risks than of unknown risks.

Unexpected biosecurity risks present a problem because they are difficult to monitor, and may be impossible to identify before an incursion. If a risk were expected, then a decision could be made to undertake biosecurity measures of some kind. Acting to detect unexpected risks presents a different challenge: to monitor activities in which risk is not expected, or to monitor already suspect activities in different ways.

This deliverable is a literature review into methods that might be useful in identifying unexpected biosecurity risks. This report covers a range of tools that can be broadly categorized as one of data-based surveillance, business knowledge and expert consultation. These tools are already being used to some degree by the Department to monitor known biosecurity risks, but there is some opportunity to expand their use around identifying unexpected biosecurity risk. This report is a deliverable within a broader project that focuses on the use of data resources, so it puts greatest weight on data-based surveillance.

1.1 Scope of the project

This report follows a literature review into methods for identifying unexpected biosecurity risk. The report:

1. explores the concept of 'unexpected biosecurity risk'
2. presents potential tools for identifying unexpected biosecurity risks.

1.2 Definition of unexpected biosecurity risk

For the purposes of this report, biosecurity risk is the risk associated with a biosecurity event or outcome. This encompasses both the likelihood of that event occurring and the consequence if that event were to occur.

A biosecurity risk is defined as *unexpected* if it relates to the presence of Biosecurity Risk Material (BRM) on a good, conveyance, or pathway upon which it is not reasonably expected or known to occur. This definition includes the presence of BRM on novel vectors (e.g., new shipping routes), and the presence of unexpected kinds of BRM on monitored vectors.

[1] http://www.environment.gov.au/biodiversity/invasive/weeds/weeds/why/impact.html
[2] http://www.invasiveanimals.com/research/phase1/goals/goal-12/12d6/

We next review several types of potential unexpected biosecurity risks, and then provide some examples of past incursions resulting from unexpected biosecurity risks.

Hitchhiker pests

Hitchhiker pests are organisms that have an opportunistic association with a commodity or item, not a biological association, or are associated with previous cargo.

Exposed goods/conveyances

Exposure risk refers to a person or thing that is defined as having been 'exposed' to another person or thing if the first-mentioned person or thing has been, or is likely to have been:

a) in physical contact with; or

b) in close proximity to; or

c) exposed to contamination, infestation or infection from; the other person or thing.

Note: this includes goods, a conveyance and premises. This concept is related to hitchhiker pests, but extends this idea to diseases.

Pest-free Status

Pest-free areas are zones that have been certified to be free of a particular pest or collection of pests. Below are some definitions from the International Plant Protection Convention (IPPC).

- Pest-free area (PFA): An area in which a specific pest does not occur as demonstrated by scientific evidence. Where appropriate, this condition is also officially maintained (ISPM 5, 2007).
- Pest status (in an area): Presence or absence, at the present time, of a pest in a specific area. Where appropriate, the pest status should also include its distribution, as officially determined using expert judgment based on current and historical pest records and other information (ISPM 5, 2007).

Importers may provide fraudulent documents that state that goods are from a zone that enjoys a pest free status, e.g., countries that hold pest free status for some but not all of its states. There may also be a change in the pest status that has not yet been detected. This would pose an unexpected biosecurity risk, without an element of fraud.

Mislabelling

Mislabelling of products, or other incorrect information relating to imports, is a potential source of unexpected biosecurity risk. Profiles or other risk-mitigation strategies may fail to capture BRM if the data that they rely on is incorrect.

Emergent diseases

These could be thought of as the unknown unknowns. These are animal, plant and human diseases that have not previously been identified or scientifically classified by the worldwide community. An example in the field of human disease is the coronavirus that caused SARS.

Black market or other illegal activities

This type of unexpected biosecurity risk is difficult to identify and catch at the border, or pre-border. An example is illegal pet keeping, in which specimens are often only caught in the event of an owner deciding that they no longer want to care for the animal, and releasing them into the wild.

Illegal pet keeping is not the same as illegal animal import and there are different biosecurity implications for each. Illegal import can be a big biosecurity risk. Illegal pet keeping may lead to a biosecurity risk in some cases, but the length of time the animal has been held can mitigate much of this risk. It won't be possible to know if import has occurred in many instances (as opposed to keeping progeny of a species imported many years ago) and it may be a state issue, or the risk will be one of 'environmental' biosecurity, and therefore an issue for a different jurisdiction.

1.3 Past incursions

Table 1 provides examples of incursions that were caused by unexpected biosecurity risks. The vector by which the species arrived in Australia, together with some details relating to economic cost, environmental impact, and current range is provided.

Table 1: Examples of incursions caused by unexpected biosecurity risks[3 4]

Species	Year	Vector	Details
Myrtle rust	Detected 2010	Unknown, possibly arrived from spores on imported plant material or attached to other imported goods/travellers.	Now established in Queensland, Victoria and NSW. Has had economic impact, with affected nurseries having to destroy contaminated stock. Potential risk to the timber industry.
Smooth newt	Detected 2011	Unknown, likely escaped/deliberately released from illegal pet keeping. Another possibility is it arrived as a hitchhiker with cargo or in a container.	Current incursion in Victoria, extent unknown. Potential ecological impacts to native wildlife.
Mexican feathergrass	1996	Imported as a nursery plant under	Eradicated. A highly invasive species with no grazing value to livestock.

[3] https://invasives.org.au/projects/biosecurity/biosecurity-failures-australia-12-case-studies/
[4] Some of the incursions listed are environmental biosecurity issues.

		incorrect or out-dated names.	
Pigeon paramyxovirus	Detected 2011	Unknown, but likely through smuggling of infected pigeons.	Now endemic in Australia. Cost to racing pigeon industry, as well as potential to infest native bird species.
Jack Dempsey cichlid	Detected 2004	Probably illegally released from an aquarium, when the owner no longer wanted them.	An effort was made to eradicate, after which further incursions were detected. Some danger of competition with native fish.
Emerald furrow bee	Detected 2004	Unknown	Current incursion in NSW. Danger to biodiversity as it can assist in pollination of weed species.

Several of the incursions were possibly caused by illegal pet keeping (smooth newt), potential smuggling (pigeon paramyxovirus) or illegal release to the wild (Jack Dempsey cichlid). Myrtle rust, if it arrived by accompanying unexpected imported goods or travellers, could be considered a hitchhiker pest. Mexican feathergrass arrived multiple times, due to being a mislabelled product; in some cases, out-dated names were used, whereas in others, the product was labelled as a different (permitted) species.

2. Tools for identifying unexpected risks

The goal of using any diagnostic prediction tool should be to jointly maximise two factors (Brugere et al., 2017):

- **Sensitivity**: probability of an issue being reported given there is an issue (high sensitivity translates to a low number of false negatives), and
- **Specificity**: probability of not reporting, given there is no issue (high specificity means a low number of false positives)

Here, an 'issue' may be defined as unexpected biosecurity risk, or the presence of a pest or disease. In the case of biosecurity risks, timely detection of the issue is also critical. Detecting issues early can result in a better response outcome.

Sensitivity and specificity are at odds with one another, that is, increasing one commonly comes at the cost of reducing the other. The trade-off between them should be guided by risk aversion: how much do false negatives (low sensitivity) cost relative to false positives (low specificity)? Commonly, in biosecurity, false negatives are considered vastly more expensive than false positives, so high sensitivity is valued over high specificity.

Tools for identifying unexpected biosecurity risk should be considered with reference to maximising sensitivity and specificity. Although it might be possible to determine the sensitivity and specificity of surveillance and data processing by looking at past applications of these methods, it may be more difficult to assess the benefits of increasing business knowledge and expert consultation, as the effectiveness of these methods will likely be highly dependent on the nature of the business.

Next we discuss a range of tools that can be broadly categorized as one of business knowledge, data-based surveillance, and expert consultation.

2.1 Increasing business knowledge

Here, by business knowledge, we mean knowledge about the routes by which BRM might potentially arrive in, or travel around, Australia. This information is useful in targeting risks and in helping to identify the source of risk in the case of an incursion.

Ideally, all stakeholders would have perfect business knowledge, that is, complete knowledge of all routes for BRM into Australia. Business knowledge can be used to:

- trace back sources of problems;
- shorten response times when there is a detection (e.g., shutting down certain trade routes or importers); and
- target surveillance.

A recent example of a tool used to help to map travel routes for sheep and goats is the National Livestock Identification System (NLIS), electronic tags that were implemented at the start of 2017 to identify individual animals.

There are numerous ways in which business knowledge could be increased, for instance, implementing a comprehensive tracking system on all shipping containers would permit more fine-tuned risk-based intervention at the ports.

Increasing this kind of knowledge could assist in the detection of exposed goods/conveyances and hitchhiker pests.

2.2 Surveillance methods

While creating a complete map of international trade routes and national movements of exposed goods and populations would greatly assist biosecurity operations, there are some issues, such as black market activities, that increasing business knowledge will not help to uncover.

An alternative, albeit not infallible, system to identifying these types of risk, would be to monitor specific pests, diseases or symptoms using currently available or easily collectible data. In the case where a potential problem is flagged by the system, a response would be to pool resources into stopping spread or identifying the routes by which these might arrive in Australia.

Surveillance is an invaluable tool for detecting incursions in Australia and monitoring the disease and pest-status of countries overseas. In a sense, the Department already conducts surveillance for unexpected biosecurity risks, when it intervenes in apparently low-risk pathways. Examples include the endpoint survey in the international passengers and mail pathways, and the compliance verification exercise that operates in several cargo pathways.

There are several classes of surveillance methods that provide data that may be useful for identifying unexpected biosecurity risks. These include agent-specific, laboratory, event-specific, syndromic and sentinel surveillance.

When designing surveillance systems, several questions arise, including:

- How should surveillance be conducted?
- What data are available?
- How should the data be analysed?

Analysis of the data derived from surveillance should be performed with reference to a baseline or 'business as usual' state of the world. If surveillance is based on monitoring symptoms, there may be a fluctuating base rate of morbidity or mortality in certain species relating to the presence of known, established diseases that are being managed in the population, rather than a new disease. Also, a baseline may fluctuate due to external factors such as weather.

Active v passive surveillance

Surveillance can be active or passive. Passive surveillance can consist of regular, ongoing reporting by organisations of some statistic, for example, hospital admissions. Alternatively, it may consist of a campaign asking relevant parties to report if they observe something of interest, for instance, evidence of a specific pest or disease. Some examples of biosecurity-related passive surveillance are:

1. A poster campaign asking people to call a hotline if they see a specific weed, with a description and photograph of that weed
2. A poster campaign asking people to call a hotline if they see any weeds

These are both examples of passive surveillance, but the first case is agent-specific surveillance, while the data from the second is more general. Note that while there will likely be a greater amount of data collected in the second example, it is also more likely to produce false positives, that is, the surveillance system will have a lower specificity, as a tool for targeting any specific weed.

Active surveillance is defined as activities that are undertaking to actively determine the presence of a pest, disease or other quality, without necessarily relying on the participation of other parties. Some examples include going out into the field, or reviewing the records of other organisations in order to identify a specific pest, disease or condition.

Passive surveillance, informally, could be described as 'they come to you', whereas active surveillance is 'you go to them'.

When attempting to identify unexpected biosecurity risks, both passive and active surveillance systems are useful. A passive surveillance system such as the one described in the second example above is likely to capture a range of information, which is useful when we don't know what we're looking for. However, active surveillance for one pest could turn up another pest, so this is also a useful tool for identifying unexpected biosecurity risks.

Agent-specific surveillance

Often referred to as disease-specific surveillance, this is surveillance that targets a specific disease or pest. This includes reporting on the occurrence of specific diseases and pests from official sources and laboratories. In the case of unexpected biosecurity risks, an agent-specific surveillance system may not be the best option, given that, by definition, we want to target pests and diseases in unexpected places. This is not to say that agent-specific surveillance is not useful; in some cases, an unexpected pest might be uncovered during operations targeting a different pest. Also, using this method on novel pathways, or other pathways where the agent is not expected to be found, may result in identifying unexpected biosecurity risk.

Laboratory-based surveillance

Laboratory testing is a primary tool for disease identification in agent-specific surveillance, however, laboratory reports may also contain observations of symptoms, which could make this data useful for *syndromic surveillance* (see below).

Just like other surveillance systems, laboratory-produced data (e.g. test results) should always be considered in relation to sensitivity and specificity. A high rate of false positives or false negatives limits a test's usefulness as a tool for surveillance. Most laboratories have independent quality assurance in place (Ford et al., 2015), but lab-to-lab variability can affect the validity of the results (Jones et al., in preparation).

Syndromic surveillance

Syndromic surveillance (SS) uses a collection of tools that focus on monitoring the prevalence of syndromes (symptoms), rather than targeting a specific disease. In doing so, SS sacrifices specificity for sensitivity; the reasoning used to justify its application is that it is a much worse error to miss a pest than to act unnecessarily.

Widely applied in epidemiology, SS involves the collection, analysis and interpretation of health-related data, generally to identify and model emerging diseases. By focusing on symptoms, SS can quickly identify the presence of an abnormal trend that may warrant further investigation, linking up potential cases without having to wait for definitive individual diagnoses. In this way, SS can avoid the lag in data collection that is associated with clinical diagnosis or laboratory confirmation. A 2013 survey of 500 local health departments (local/state entities) in the U.S. found that 62% employed some form of electronic SS (Chughtai et al., 2016).

Syndromic surveillance is also used in veterinary science, where it focuses on the signs of disease that can be detected by the human senses. In human medicine, it has been extended to also include the physical or emotional symptoms reported by the subject (Brugere et al., 2017).

The Department has trialled an experimental SS system in monitoring the biosecurity risk of imported ornamental finfish (ACERA project 1206A, CEBRA projects 1305A, 1405A, 1505A). Samples are taken randomly from bags of apparently healthy fish and subjected to histopathological analysis, which includes recording of symptoms regardless of whether they can be linked to diseases. Trigger warnings are set in the database to alert the pathway manager upon a given number of successive positive observations of symptoms.

Although SS is generally understood to focus on symptoms, there is no universally agreed upon definition. Some interpretations, such as those used by Abat et al. (2016), include monitoring purchasing data for non-prescription medication, carcass condemnations and submissions to laboratories as falling under SS, whereas Brugere et al. (2017) consider these types of data an extension and omit them.

Timeliness may be considered a factor: the local health department data analysed by Chughtai et al. (2016) came from a SS system that updated data at least once every 24 hours.

An advantage of SS over disease or pest-specific surveillance is its potential to survey a larger part of a population for the same effort as an agent-specific system. It can also identify emergent pests and diseases (the unknown unknowns), that is, those that have not previously been classified. Several diseases that have evolved recently were discovered using SS. An example is Severe Acute Respiratory Syndrome (SARS); the Center for Disease Control (CDC) issued a global alert for a severe form of pneumonia, based on symptoms, twelve days before laboratories suggested that a new coronavirus may be the cause of SARS[5]. This is a syndromic discovery as the alert was raised because of a change in the prevalence of respiratory-related symptoms.

Syndromic surveillance is increasingly being recognised as the most cost-effective approach to the detection of new and emerging diseases. However, the sensitivity of SS will vary in different contexts. Abat et al. (2016) noted that SS

[5] https://www.cdc.gov/about/history/sars/timeline.htm

data are non-specific but highly sensitive. Syndromic surveillance will naturally fail in cases where an issue exists but does not present the syndromes being targeted, for instance, infection can occur in the absence of clinical signs (Brugere et al., 2017). Several papers argue that the goal of SS should be to 'enhance, rather than replace, traditional approaches to epidemic detection' (Brugere et al., 2017, Abat et al., 2016).

Sentinel surveillance

Sentinel surveillance generally involves using specific people, organisations (e.g., hospitals), or biota (e.g., chicken flocks) to act as sentinels, that is, as sources of information relating to some specific pest, disease or symptom that is being monitored. Because sentinel surveillance can target either specific pests and diseases, or symptoms, it can be either syndromic or agent-specific surveillance.

There are many people who could act as sentinels for unexpected biosecurity risks, including anyone working within proximity of an airport or ship port such as stevedores, vessel masters, police and Australian Border Force personnel. Also, doctors, veterinarians or rangers working in an area in which incursions could establish may be good candidates to identify risks post-border. A specifically chosen selection of ships (journeys) or flights that are representative of all traffic on regular routes, might act as sentinels.

Data processing and analysis

The following is a more detailed description of some tools that are useful for data processing and analysis.

Syndromic categories through text analysis

Human-generated data, for instance, Facebook posts, are not always in an easy-to-digest form, and may need some processing before they can be searched for clinical signs or symptoms in a human health context. In a biosecurity context, they could be searched for signs indicating the presence of a pest or disease, such as unexpected fatalities of animals or plants.

There are any number of symptoms that might be focused on for SS. For human-related data, chief complaints can be grouped into SS categories manually, or using algorithms (Abat et al., 2016). Ali et al. (2016) match chief complaints to these categories using fuzzy logic and machine-learning, together with weather data.

In another example, Torii et al. (2016) consider the use of online consumer product reviews as a source of health-related data. About 1.3 million Amazon.com reviews on Grocery and Gourmet Food products were scoured for health-related information, using a natural language processing system called nQuiry.

The nQuiry system processes language using the following steps:

- tokenization
- sentence chunking
- part-of-speech tagging
- syntactic parsing
- phrase extraction

- concept candidate selection
- concept searching

Torii et al. (2016) applied this process to a subset of data and the concepts that were created from this process were then manually reviewed and labeled as relevant or irrelevant as sources of medical-related information. They found that there were many false extractions, so they did some manual filtering to deal with these. Rules based on this manual filtering were fed into a machine-learning algorithm to further divide information, based on phrases.

The relevance of phrases was tested using logistic regression. Naïve Bayes and several other techniques were also examined, and were found to generally agree with the decisions made using logistic regression.

This type of procedure could be used analogously to process other free text field data such as information from Twitter and Facebook, with a biosecurity risk focus. The Department has done some work in this area, mining Twitter to look at the spread of information relating to specific biosecurity-related incidents[6].

When looking for information relevant to biosecurity operations, search methods could target terms relating to human, plant or animal health. For instance, discolouration on plant foliage is a symptom of any number of plant diseases.

Comparison with baseline

The key to identifying unexpected biosecurity risks is to look at changes in the system with regard to a baseline. If there is suddenly a series of reports indicating, say, a 'higher than usual' morbidity (fever, in this example) in cattle, it may indicate the presence of a foot and mouth outbreak. This approach raises several questions, amongst them: what is 'higher than usual', and how should the baseline be calculated? Control charts provide a framework for resolving these questions.

Control charts

A control chart is a simple statistical tool that can be used to determine whether the variation in a measurable quantity is within the normal behaviour of the system, or is unusual and warrants special attention. Control charts are designed to be used in real time with continuously updated data. They are a straight-forward implementation that could be incorporated into existing systems without substantial changes to operating conditions or software. A rigorous discussion of control charts is provided in Fox (2009) and we include a brief description here.

The simplest control charts assume when a process is *in control*, its observations vary around a central mean or baseline according to a random process with constant underlying variance.

In some control chart applications, the mean and standard deviation of the underlying distribution are estimated. The upper and lower 'control limits' are

[6] The department's Principal Data Scientist, Greg Hood, has developed a method of mining Twitter.

then set at plus and minus 3 standard deviations from the mean. The system is assumed to be in control unless (for instance):

- The measurable quantity falls outside of the control limits (i.e., above the upper limit or below the lower limit)
- The measurable quantity sits above or below the mean for several time steps in a row
- The measurable quantity increases or decreases for several time steps in a row

An example is provided in Figure 1.

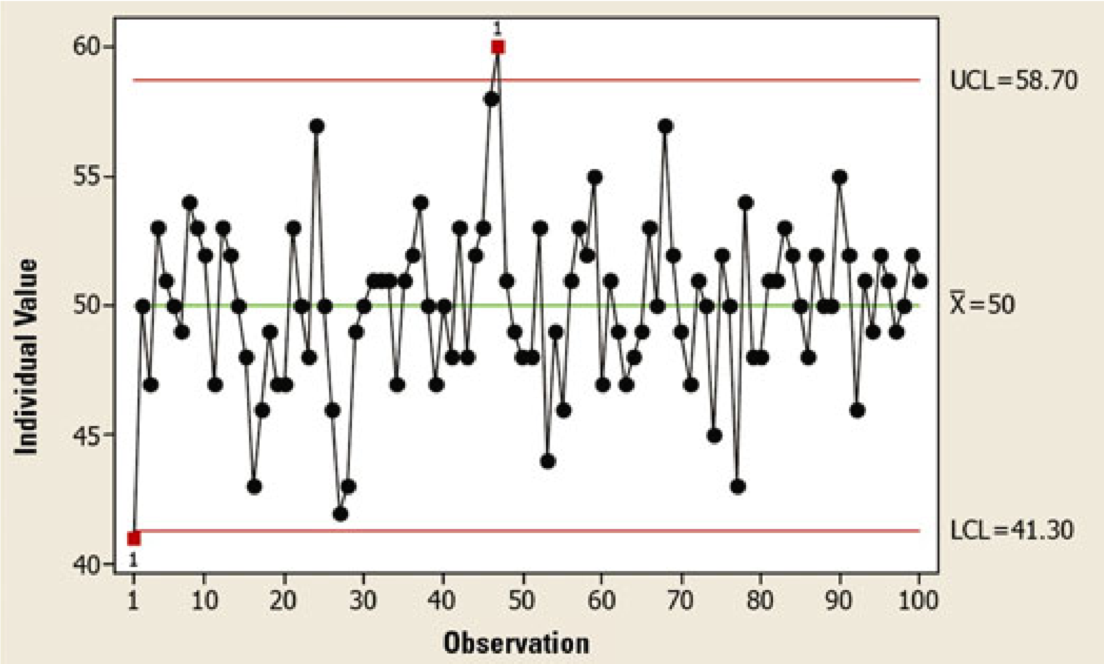


Figure 1: A representative control chart showing the mean (green line) and upper and lower confidence levels cut-off points (UCL and LCL) at three standard deviation (red lines)[7]

To provide a reliable estimate of the parameters for the underlying distributions requires substantial data. Splitting the data across smaller time periods would create more data points (e.g. monthly numbers of reports, etc.) but in many instances, the presence of seasonal effects can compromise this approach, and would require extra modelling steps.

In practice, control charts can avoid the need to estimate a mean and standard deviation by using any arbitrary cut-off. For instance, infection preventionists typically perform outbreak detection on hospital data using simple rules such as 'three or more new cases of a single pathogen within two weeks, in a single ward' (Lin and Trick, 2016).

While initially the control chart limits and rules can be more or less arbitrary, based on experience with similar systems elsewhere or using statistical conventions such as three standard deviations. However, as data accumulate, so will experiences in identifying false positive and false negative decisions.

[7] http://www.dummies.com/careers/project-management/six-sigma/how-to-use-control-charts-for-six-sigma/

A disadvantage of control charts arises due to assuming stationarity, that is, that the statistic has a constant mean over time; and appropriate scope, that is, that the statistic is representative of the full population. Such approaches do not allow for natural (that is, non-problematic) variation in the prevalence of the subject being measured, and can miss related issues if the monitored data does not represent the full population. For instance, a study monitoring human disease cases in a single ward could miss outbreaks spanning multiple hospital units (Lin and Trick, 2016).

Cumulative sum

The cumulative sum (CUSUM) algorithm is a form of control chart (Page, 1954) that has been applied in various surveillance schemes (Wang et al., 2010). It keeps track of the accumulated absolute deviation between expected and observed values, and flags an aberration in the data when this sum goes above some threshold.

The accumulated deviation S_t is defined as:

$$S_t = \max\left(0, S_{t-1} + \left(\left(X_t - (\mu_0 + k\sigma_{xt})\right)/\sigma_{xt}\right)\right)$$

where X_t are the observed values and μ_0 and σ_{xt} are the mean and standard deviation during the baseline period (Wang et al., 2010).

Variants of the CUSUM method, namely C1-mild, C2-medium and C3-ultra, are standard routines that have been used by the Centre for Disease Control (CDC). They differ as follows (Jackson et al., 2007):

- C1: baseline period is the last seven periods, that is (t-7, ..., t-1)
- C2: baseline period is the last seven periods, starting from two periods ago (t-9, ..., t-3)
- C3: this uses the C2 algorithm, but with a test statistic of $S_t + S_{t-1} + S_{t-2}$

The period used by Jackson et al., 2007 is one day.

Exponential weighted moving average

The exponential weighted moving average (EWMA) method builds on the constant average control chart described above by allowing for a changing mean and variance over time. Assuming that the individual observed values are $X_i \sim N(\mu, \sigma^2)$ during some period (say, daily), the average weighted periodic counts over some specified area at time t>0 are modelled as:

$$Z_t = \lambda \bar{X}_t + (1-\lambda) Z_{t-1}$$
$$with\ Z_0 = X_0$$
$$UCL = \mu + k\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\left(\frac{\lambda}{2-\lambda}\right)\left(1-(1-\lambda^{2t})\right)}$$

where n is the number of observed values that make up the mean and k is the 'control limit coefficient' (Wang et al., 2010). Generally, k is set between 0 and 3 (Linnet, 2006).

Generalised Linear Models

Generalised linear models (GLMs) are a well-known statistical method used for prediction. These can be employed for aberration detection by fitting a model to the data with time as a predictor.
Jackson et al., (2007) discuss fitting a Poisson-errors GLM with terms for day of the week, monthly, linear time trends, and holidays, based on three years of data. The test statistic in this case is the probability from a Poisson distribution of observing at least X_t cases, given a mean of $E(X_t)$. If the probability is lower than some threshold, for instance, 5%, this would flag an aberration.

Scan statistics

Scan statistics (Naus, 1964) are widely used for early disease outbreak detection in surveillance systems, by identifying geographical disease clusters (Ali et al., 2016). Rather than evaluating/identifying disease clusters after the fact, it works prospectively in real time by looking at daily or weekly data feeds. It compares historical counts of data, looking for clusters across space-time, adjusting for purely temporal or spatial variation in counts (Costa and Kulldorff, 2014). Taking into consideration spatial elements helps to avoid missing connections between different locations, for instance.

The traditional prospective space-time permutation scan statistic was developed by Kulldorff et al. (2005). The scanning window used in this method is cylindrical in shape, which ignores potential interactive factors for disease spread, such as information about roads and landscapes (Costa and Kulldorff, 2014). Costa and Kulldorff, (2014) propose two alternative, irregularly shaped space-time permutation scan statistics using dynamic cluster geometry.

Surveillance software can perform automated cluster detection (Lin and Trick, 2016). Scan statistics have been widely used for studying early disease detection, through the use of 'SaTScan' software developed by Kulldorff (1997). Ali et al. (2016), after processing the data to identify syndrome categories from chief complaints, used SaTScan statistics to identify disease outbreaks and clusters using a number of case studies such as dengue fever.

Huang et al. (2010) identified clusters of pathogens using a space-time permutation scan statistic (WHONET-SaTScan) approach. Over 5 years, such an approach identified 59 clusters (including all those previously identified by the hospital's infection control program), 95% of which were deemed by the hospital epidemiologists to merit consideration or warrant active investigation/ intervention.

Method comparison

Wang et al. (2010) compared the exponential weighted moving average (EWMA), C1-MILD (C1), C2-MEDIUM (C2), C3-ULTRA (C3) and the traditional space-time permutation scan statistic model with a cylinder-shaped scanning window. Their conclusion was that the space-time permutation scan statistic had a specificity of 99.9% and a detection time of less than half a day, while the specificity of the EWMA was 95.2% and the C3 method had the lowest specificity at 73.7%. The exponential weighted moving average exhibited the shortest detection time (0.1 day), while the modified C1, C2 and C3 methods exhibited a detection time of close to one day. Also, the performance of the algorithms was correlated with parameter values chosen for the models, and that this may affect performance.

Jackson et al. (2007) compared three control-chart-based statistics (CUSUM C1, C2 and C3), two exponential weighted moving average methods, and a generalized linear model. The conclusion was that all of the algorithms tested had poor sensitivity, particularly for outbreaks that did not begin with a surge of cases. The generalized linear model had the highest sensitivity, detecting 54% of the simulated epidemics when run at alert rate of 0.01. Of those tested, the model with the lowest sensitivity was found to be the CUSUM C3, with 44.5%.

Costa and Kulldorff (2014) developed two new scan statistics (mentioned above) with improved computational performance. Simulation work aimed at evaluating the methods further is currently underway.

Surveillance systems around the world

There is a multitude of different surveillance systems for human, animal and plant diseases and pests in place around the world. A few of these surveillance systems are described briefly here. They use many tools for collecting, extracting, processing and analysing data, as described above. These include:

- Text processing
- Natural Language Processing (NLP)
    - nQuiry
    - AlchemyAPI
- Syndromic category classification
- Machine-learning
- Aberration detection methods
    - Control charts
    - Cumulative sum (CUSUM)
    - Exponential weighted moving average (EWMA)
    - Generalised linear model (GLM)
    - Spatial scan statistics
- Data visualisation

IBIS

IBIS (international biosecurity intelligence system), developed by CEBRA, is an intelligence network for plant and animal (aquatic and terrestrial) biosecurity surveillance. IBIS is an example of an SS system that captures data useful for identifying unexpected biosecurity risks.

IBIS has two parts: automated SS for information gathering and a crowdsourcing aspect for data classification and quality control.

Information collection

IBIS employs a software robot that scours the internet every day looking for the following things:

- Plant and animal disease reports
- Industry reports
- Articles from relevant journals

- Other articles and comments that may be relevant to biosecurity

Resources such as search engines, RSS feeds and Twitter are scanned. Once an article has been found, AlchemyAPI (a NLP service) extracts the following elements:

- Title
- Text
- Language
- Author
- Location

Quality control

The crowdsourcing aspect of IBIS provides an important quality control function. Members of the IBIS community perform two main tasks:

- Evaluate whether an article collected by the robot is relevant to the site or not
- Edit the article and ensure that it has been categorised and tagged correctly, that the publish date and places are correct, and so on

Each article in IBIS has a status (Trash, Raw, Keep, Promoted and Alert) indicating its relevance and important. Figure 2 shows the status that articles can have within the IBIS system and how these are related.

IBIS community

The IBIS network and database is growing daily with members devoted to collecting and organising information used for tracking and forecasting diseases and following emerging disease trends. Users can interact by submitting articles, commenting on collected articles, and engaging in research and/or analysis activities. The functions of different users are spelled out more clearly below[8]:

- Member: receives a (personalised) daily digest, can search the database, add comments to articles, submit articles to the community
- Article Evaluator (AE): in addition to the member's rights, the AE can also evaluate the articles therefore changing their status, edit their content, promote them or alert the community, assign articles to issues
- Search Editor (SE): can create search queries, can add and edit: search sources, pests, diseases, hosts, qualifiers and blocked sites.

[8] https://ibisbiosecurity.org/welcome/

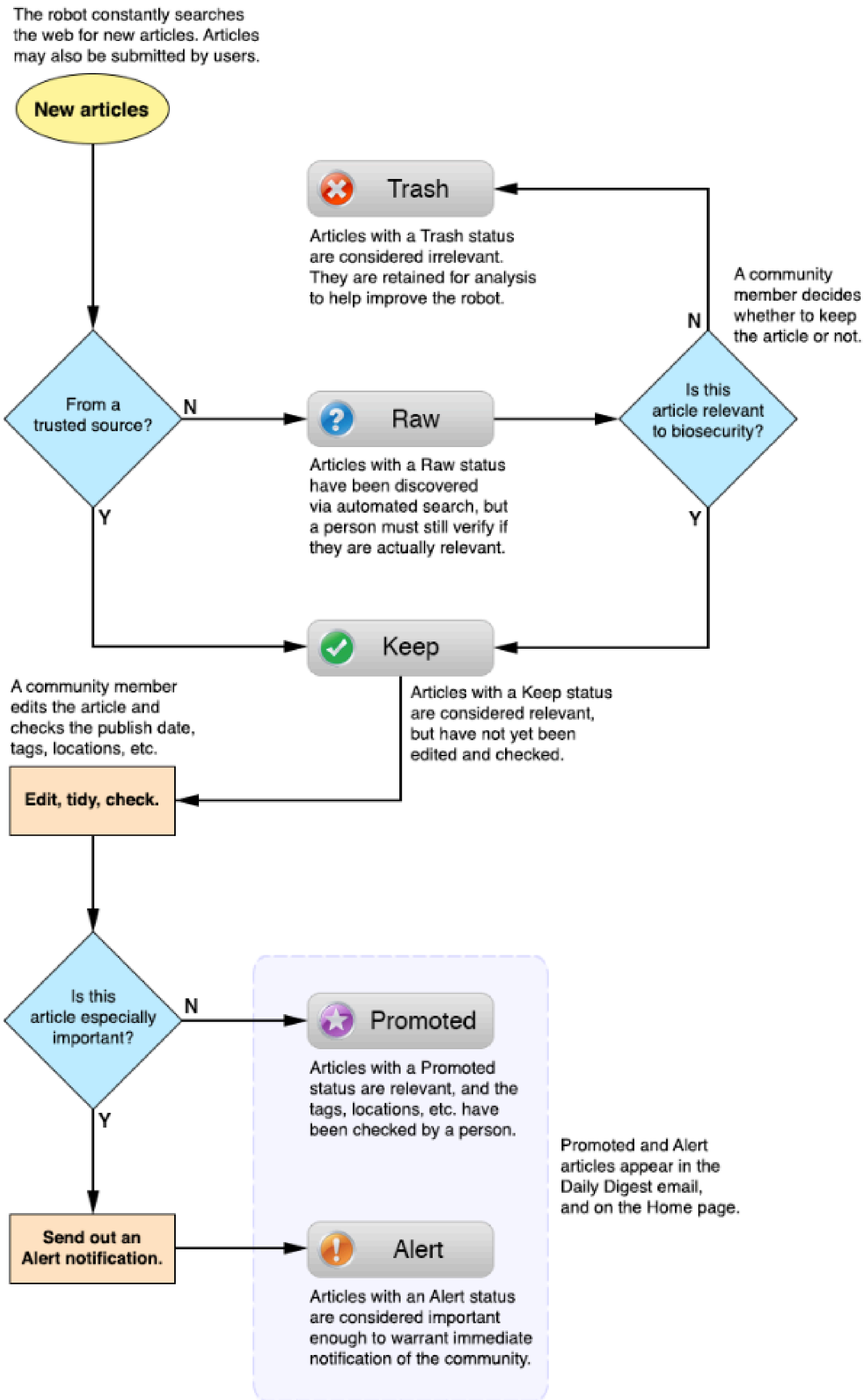


Figure 2: IBIS flowchart

BioSense

BioSense is a secure, cloud-based platform maintained by the US CDC as part of the National Syndromic Surveillance Program (NSSP) to collect electronic syndromic (human) health data from across the US, sharing data to allow stakeholders to get a better picture of health. It encompasses:

- ESSENCE (Electronic Surveillance System for the Early Notification of Community-based Epidemics)
- Adminer: allows basic queries of MS SQL data stored on the BioSense Platform
- Access & Management Center (AMC): for assigning rights, access control, and data-sharing privileges to ESSENCE (and eventually other platform tools)
- R Studio Professional: a statistical tool that lets users analyze data beyond what's available through ESSENCE

By using common resources readily available on the BioSense Platform (e.g., networks, servers, software, tools, and storage), users have limited need for additional IT support. Users of the BioSense Platform benefit through efficiencies gained, cost reductions, and information-sharing capabilities.

Also, the BioSense Platform offers local and state users free secure data storage space, an easy-to-use data display dashboard, and, most importantly, a shared environment where they can collaborate and exchange knowledge of SS.

ESSENCE

ESSENCE was created and developed through the US Department of Defense (DoD), and was originally designed for early detection of bioterrorism attacks following September 11, 2001.

Later, it was adapted by the CDC and has since been adopted by the DoD military healthcare system as well and in numerous health departments in the US and in other countries. ESSENCE was the most widely used system for surveillance reported in a survey of 500 US local health departments, with 27% of the 109 local health department respondents that used SS listing it as the system they use (Chughtai et al., 2016).

Data

ESSENCE incorporates traditional reportable disease surveillance, sentinel surveillance and SS using multiple data such as[9]:

- Patient location (residency)
- Facility (hospital) location
- Chief complaints
- Diagnosis discharge information
- Weather (temperature, rainfall, etc.)
- Pharmaceutical sales

Chief complaints are grouped using syndrome categories. There are multiple

[9] https://www.cdc.gov/nssp/biosense/docs/biosense-platform-quick-start-guide-for-essence.pdf

versions of ESSENCE (Lombardo, 2004), and the syndrome categories monitored over time have varied slightly.

Ali et al. (2016) claims that BioSense provides monitoring support for eleven syndromic, and seventy-eight sub-syndromic categories (but doesn't describe these). A 2014 source[10] lists twelve syndromic categories:

- Botulism-like
- Exposure
- Fever
- Gastro-intestinal Illness
- Hemorrhagic Illness
- Influenza-like Illness
- Injury
- Neurological
- Rash
- Reportable Diseases
- Respiratory
- Shock/coma

Alert algorithm

A 30-day baseline of data is used to calculate alerts. The user guide for ESSENCE notes that 'Poisson/Regression/Exponentially Weighted Moving Average (EWMA) switch algorithm is the default temporal alerting algorithm used for ESSENCE'[11]. Other methods for aberration detection that are available include C1, C2 and C3 CUSUM methods.

Data visualization

There are various tools for data visualisation within ESSENCE, such as time series graphs of daily data counts over time including alert and warning signals in different colours (see Figure 3) and shaded maps indicating warning and alert areas.

[10] https://public.health.oregon.gov/DiseasesConditions/CommunicableDisease/PreparednessSurveillanceEpidemiology/essence/Documents/userguide.pdf

[11] https://www.cdc.gov/nssp/biosense/docs/biosense-platform-quick-start-guide-for-essence.pdf

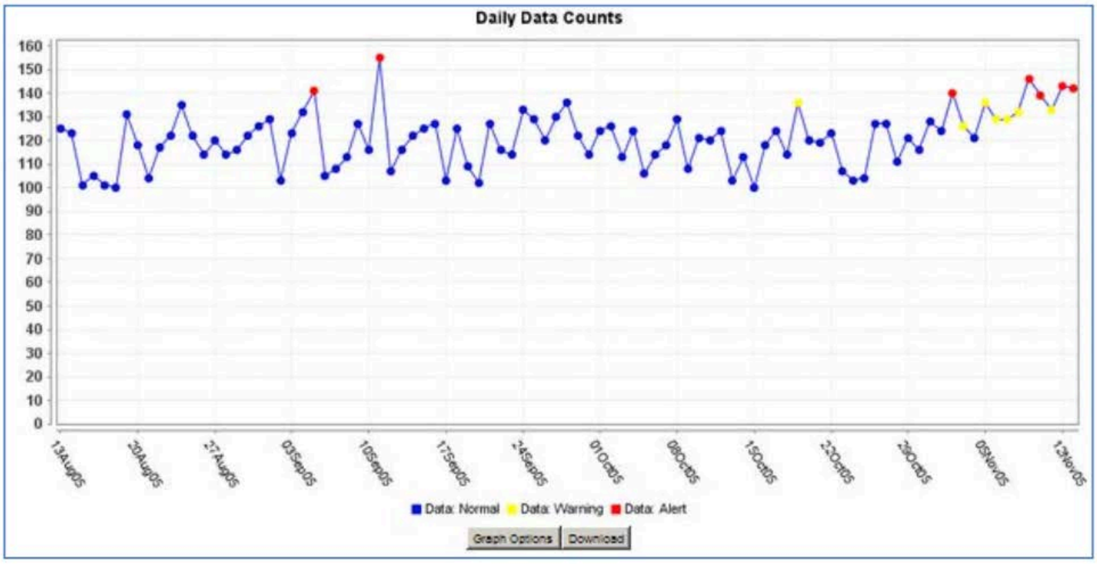


Figure 3: Example graph of output from the ESSENCE system, monitoring one syndrome over time (source: https://public.health.oregon.gov/DiseasesConditions/CommunicableDisease/PreparednessSurveillanceEpidemiology/essence/Documents/userguide.pdf)

Other surveillance systems for human disease

While ESSENCE was reported as the most commonly used system for monitoring disease in the US in the study conducted by Chughtai et al. (2016), other systems for monitoring human disease are also available.

RODS

RODS (real-time outbreak and disease surveillance) uses Naive-Bayes algorithm to classify chief complaints data into seven syndromic categories (Ali et al., 2016). RODS was implemented by 12% of 109 survey respondents who used and had access to SS (Chughtai et al., 2016).

EARS

The early aberration reporting system (EARS) provides surveillance for forty-two syndromes and implements C1, C2 and C3 CUSUM methods (Ali et al., 2016). EARS was implemented by 7% of 109 survey respondents who used and had access to SS (Chughtai et al., 2016).

NNDSS

The National Notifiable Diseases Surveillance System (NNDSS) co-ordinates collection of notifiable human disease data by state and territory health authorities in Australia.

The NNDSS provides fortnightly reports on notifiable diseases in Australia and internationally[12]. The reports provide statistics for the reporting period, quarter and year to date. The reports compare these figures to the same data from the previous year, rolling means and recording by how much figures exceed rolling means (if they exceed the rolling mean by more than two standard deviations). This is a control chart type comparison.

[12] http://www.health.gov.au/cdnareport

Visualisation tool: infectious disease

Traditional and syndromic surveillance of infectious diseases and pathogens (Abat et al., 2016) considers 252 disease-specific and 10 SS systems that use laboratory data, and are described in available PubMed articles for monitoring infectious disease from 2009 to June 2014. These are depicted in Figure 4.

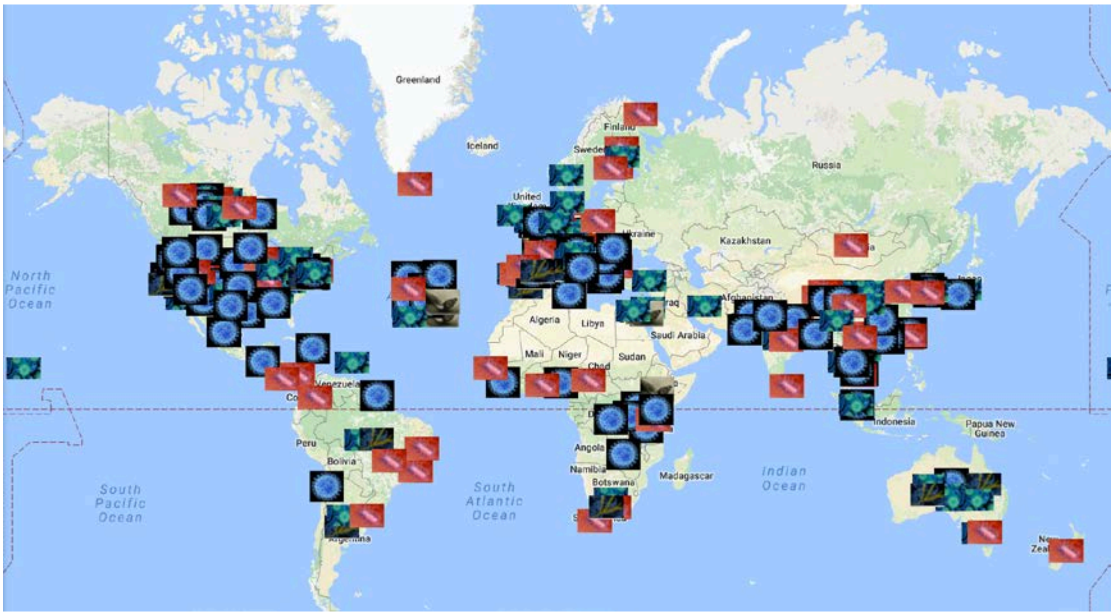


Figure 4: Infectious disease surveillance systems (source). The virus image represents surveillance systems focusing on viruses, the bacterium image represents surveillance systems focusing on bacteria, the fungus image represents surveillance systems focusing on fungi, and the polymicrobial image represents surveillance systems monitoring various different pathogens.

2.3 Expert Consultation

An alternative to surveillance is expert consultation. As this report focuses on tools that use data, we cover expert consultation only briefly. There are several methods that may be useful to assist in the identification of unexpected biosecurity risks. Two methods that are commonly deployed in risk management include:

- Horizon scanning
- Prospective hindsight & pre-mortem

Horizon scanning

Horizon scanning is a systematic examination of information to identify potential risks. Using expert consultation to conduct horizon scanning to identify unexpected biosecurity risks would involve a group of experts or stakeholders sitting down to discuss what information is available and what these risks might be, given the current system.

Prospective hindsight and pre-mortem

The idea of the 'pre-mortem' was popularised by Klein (2007)[13]. Instead of stakeholders discussing where the gaps may be in the current system, and what

[13] https://hbr.org/2007/09/performing-a-project-premortem

incursions this might lead to, the idea is to start with the incursions, or worst case scenario, and work backwards to see how this could have occurred. The theory is that, by imagining that the incursion has already occurred, the stakeholders will feel more comfortable in revealing potential gaps in the system.

2.4 Conclusion: Literature review

There is a variety of tools for identifying unexpected biosecurity risks, including increasing business knowledge, surveillance methods, and expert consultation.

Increasing business knowledge can assist in identifying routes for BRM into Australia. Several types of surveillance methods are potentially useful for identifying unexpected biosecurity risk, including active, passive, agent-specific, syndromic and sentinel surveillance. Data derived from these surveillance methods can be monitored using a variety of control chart and scan statistic methods to check for deviations from business as usual, which can indicate the presence of unexpected biosecurity risks. Expert consultation, particularly prospective hindsight, could be a potentially useful tool for identifying potential incursions and tracing these back to unexpected biosecurity risks.

In the hunt for unexpected biosecurity risks, syndromic surveillance across wide range of data seems to be the best match, as such risks come from an unexpected source. However, a significant challenge lies in the resourcing of surveillance for unexpected risks, in the light of the recognized need for surveillance for expected risks.

Bibliography

1. Abat, Cédric, et al. "Traditional and syndromic surveillance of infectious diseases and pathogens." *International Journal of Infectious Diseases* 48 (2016): 22-28.

2. Ali, M. A., et al. "ID-Viewer: a visual analytics architecture for infectious diseases surveillance and response management in Pakistan." *Public health* 134 (2016): 72-85.

3. Brugere, Cecile, Dennis Mark Onuigbo, and Kenton Ll Morgan. "People matter in animal disease surveillance: Challenges and opportunities for the aquaculture sector." *Aquaculture* 467 (2017): 158-169.

4. Butler, Declan. "When Google got flu wrong." *Nature* 494.7436 (2013): 155.

5. Carlier, M., et al. "Oyster farmers' perception of spat mortality outbreaks: more a firm than a farm issue." *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine, Madrid, Spain* (2013): 226-237.

6. Chughtai, Sarah, et al. "Assessment of Local Health Department Utility of Syndromic Surveillance: Results of the 2015 Biosurveillance Needs Assessment Survey." *Journal of Public Health Management and Practice* 22.Suppl 6 (2016): S69.

7. Costa, Marcelo A., and Martin Kulldorff. "Maximum linkage space-time permutation scan statistics for disease outbreak detection." *International journal of health geographics* 13.1 (2014): 20.

8. Elliot, A. J., et al. "Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system." *Epidemiology and infection* 143.16 (2015): 3416-3422.

9. Ford, Laura, et al. "Approaches to the surveillance of foodborne disease: a review of the evidence." *Foodborne pathogens and disease* 12.12 (2015): 927-936.

10. Fox, David R. "Statistical Methods for Biosecurity Monitoring and Surveillance." *Australian Centre of Excellence for Risk Analysis* (2009)

11. Halliday, Jo, et al. "Bringing together emerging and endemic zoonoses surveillance: shared challenges and a common solution." *Phil. Trans. R. Soc. B* 367.1604 (2012): 2872-2880.

12. Huang, Susan S., et al. "Automated detection of infectious disease outbreaks in hospitals: a retrospective cohort study." *PLoS Med* 7.2 (2010): e1000238.

13. Jackson, Michael L., et al. "A simulation study comparing aberration detection algorithms for syndromic surveillance." *BMC Medical Informatics and Decision Making* 7.1 (2007): 6.

14. Kulldorff, Martin. "A spatial scan statistic." *Communications in Statistics-Theory and methods* 26.6 (1997): 1481-1496.

15. Kulldorff, Martin, et al. "A space–time permutation scan statistic for disease outbreak detection." *PLoS Med* 2.3 (2005): e59.

16. Lawson, Becki, Silviu O. Petrovan, and Andrew A. Cunningham. "Citizen science and wildlife disease surveillance." *EcoHealth* 12.4 (2015): 693-702.

17. Lin, Michael Y., and William E. Trick. "Informatics in Infection Control." *Infectious disease clinics of North America* 30.3 (2016): 759-770.

18. Linnet, Kristian. "The exponentially weighted moving average (EWMA) rule compared with traditionally used quality control rules." *Clinical Chemical Laboratory Medicine* 44.4 (2006): 396-399.

19. Lombardo, Joseph S., H. Burkom, and J. Pavlin. "ESSENCE II and the framework for evaluating syndromic surveillance systems." *Morbidity and Mortality Weekly Report* (2004): 159-165.

20. Morgan, Kenton L., et al. "Accuracy of herdsmen reporting versus serologic testing for estimating foot-and-mouth disease prevalence." *Emerging infectious diseases* 20.12 (2014): 2048.

21. Naus, Joseph Irwin "Clustering of random points in line and plane." *Harvard University* (1964)

22. Page, E. S. "Continuous inspection schemes." *Biometrika* 41.1/2 (1954): 100-115.

23. Robertson, Colin, et al. "Spatial data issues in geographical zoonoses research." *The Canadian Geographer/Le Géographe canadien* 60.3 (2016): 300-319.

24. Torii, Manabu, et al. "Mining health-related issues in consumer product reviews by using scalable text analytics." *Biomedical informatics insights* 8.Suppl 1 (2016): 1.

25. Welvaert, Marijke and Caley, Peter (2016) “Citizen surveillance for environmental monitoring: combining the efforts of citizen science and crowdsourcing in a quantitative data framework.” *SpringerPlus*

26. Wang, Xiaoli, et al. "Comparing early outbreak detection algorithms based on their optimized parameter values." *Journal of biomedical informatics* 43.1 (2010): 97-103.

27. World Health Organization. *International Health Regulations (2005)*. World Health Organization, 2008.

28. https://www.daf.qld.gov.au/__data/assets/pdf_file/0003/56658/IPA-Asian-Toad-Risk-Assessment.pdf

Appendix: Potential sources of data

Numerous sources of data may be useful in assisting in the identification of unexpected biosecurity risks.

Official records

There are a number of records kept by government and businesses that may be useful in identifying outbreaks of pests and diseases. The advantages of such data are that they are generally routinely recorded in a standardised format and hence may be easier to process. They include:

- DAWR data
    - notifiable diseases
    - reportable diseases
    - quarantine facility information
    - ABARES farm surveys data
    - NLIS (National livestock identification tag data)
- Australia Post data
- Customs and immigration data
- APVMA data
- Shipping-related information
- Medical data
    - notifiable diseases
    - hospital admissions
    - chief complaints
- Absenteeism (related to human health)
- Laboratories
- Fumigation services by pest removal businesses
- Purchase data for pharmaceuticals and medicines (animal and human) and herbicides and other plant-related products
- Carcass condemnation data
- Botanical garden and park records
- Zoological vet networks

Notifiable disease in animals in Australia

In Australia, notifiable diseases[14] in animals must be reported immediately to agricultural authorities. Reports can be made by contacting a veterinarian, state/territory department of primary industries or agriculture, or phoning the *Emergency Animal Disease Watch Hotline*. The requirement to report is contained in individual states and territories legislation and the list of notifiable diseases changes slightly between states and territories. Similar arrangements are in place for reportable diseases in aquatic animals[15].

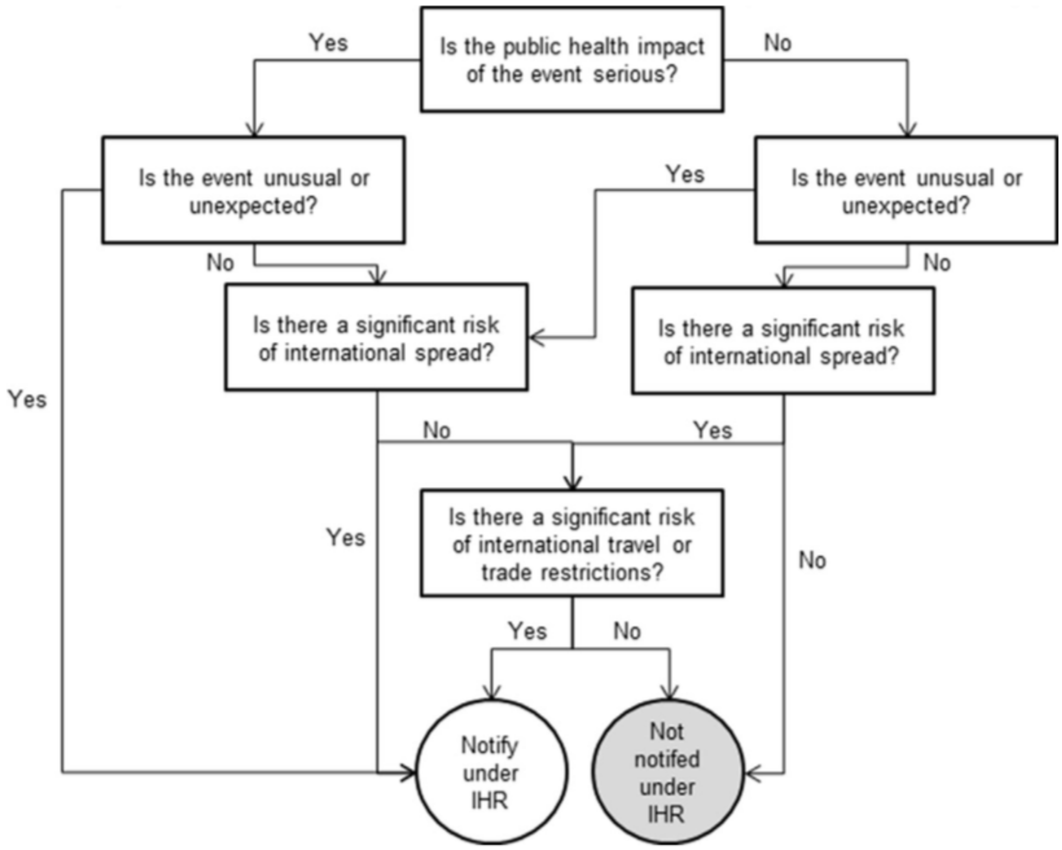
[14] http://www.agriculture.gov.au/pests-diseases-weeds/animal/notifiable#national-list-of-notifiable-diseases-of-terrestrial-animals-at-november-2015
[15] http://www.agriculture.gov.au/pests-diseases-weeds/animal/reportable-aquatic

Notifiable disease in humans in Australia

Medical data can potentially be useful for identifying outbreaks of pests and diseases that affect humans, including food-borne pathogens such as hepatitis A and zoonotic diseases such as rabies. A complete list of notifiable diseases in Australia is maintained by the Department of Health[16]. Notifications are made to state or territory health authorities, and include a unique record reference number, state or territory identifier, disease code, date of onset, date of notification to the relevant health authority, sex, age, Indigenous status and postcode of residence[17]. This is co-ordinated by the Commonwealth's National Notifiable Diseases Surveillance System (NNDSS). Computerised, de-identified records are then supplied (daily) to the Department of Health.

Notifiable diseases internationally

The World Health Organization (WHO) requires disease reporting to perform its global surveillance and advisory role. In order to do this, it has developed the *International Health Regulations 2005*, which identifies a number of specific human diseases, as well as a limited set of criteria to assist in deciding whether an event is notifiable to WHO. Any 'event that may constitute a public health emergency of international concern' is notifiable (WHO, 2008). This is displayed in Figure 5.



[16] http://www.health.gov.au/casedefinitions
[17] http://www.health.gov.au/internet/main/publishing.nsf/Content/cda-surveil-nndss-nndssintro.htm

Figure 5: Decision tool for reporting detected events relating to public health emergencies of international concern, adapted from the WHO *IHR 2005* (source: Roberston, C., 2016)

Specific animal diseases are monitored globally by the OIE (World Organisation for Animal Health). Mandatory reporting is required of member states against 117 terrestrial and aquatic animal diseases[18].

Several countries including Australia, Brazil, Canada, UK and the US have their own regulation relating to notification[19].

Laboratories

As well as reporting related to many human and animal-related diseases, laboratories in Australia are also required to notify if they detect any of the following micro-organisms[20]:

- Campylobacter spp.
- Cryptosporidium spp.
- Salmonella spp.
- verotoxin-producing Escherichia coli (VTEC)
- Vibrio spp.
- Giardia cysts
- Listeria monocytogenes
- Cyclospora spp.
- hepatitis A
- norovirus

APVMA

The Australian Pesticides and Veterinary Medicines Authority (APVMA) maintain the adverse experience reporting program (AERP). This contains information on lifestock, pets and plants that have had an adverse experience following the application of some pesticide or veterinary medicine. The purpose of this dataset is to monitor reactions to individual products. This dataset could be useful for biosecurity purposes as there is a possibility that the veterinary medicine or pesticide has been mistakenly linked to adverse symptoms that are actually caused by a pest or disease.

Telephone

There are several helplines that may be useful sources of data for surveillance. One disadvantage of using telephone data is the uneven distribution of phones, particularly in developing countries (Brugere et al., 2017). Some areas may not be coverable under a phone-based scheme.

Web-based

These are user-driven records and include:

- Web searches
- Social media

[18] http://www.oie.int/animal-health-in-the-world/
[19] https://en.wikipedia.org/wiki/Notifiable_disease#cite_note-2
[20] https://www2.health.vic.gov.au/public-health/infectious-diseases/notification-procedures

    - twitter
    - facebook
- Online reviews
- Identification tools
    - Self diagnosis tools (for human-related disease)
    - Weed identification tools
    - Animal identification tools

One advantage of web-based data is that they are produced in real-time. Also, the unofficial nature of these data may encourage reporting. Elliot et al. (2015) examined medical data collected by the National Health Service (NHS) in the United Kingdom, and found that prevalence predictions made using the telephone diagnosis helpline lagged behind those made using the online self-diagnosis tool.

A disadvantage of using web-based information, particularly free-text information available from social media services like Facebook and Twitter is that it may be of poor quality due to low-level literacy skills of some users (Brugere et al., 2017). This can imply that a lot of pre-analysis processing is required. Also, as with telephone-based information, internet-based data can suffer from coverage issues associated with limited internet availability. Also, Butler (2013) found that predictions based on web-related data are more likely to be affected by changes in people's search behaviour.

People

People who come into contact with a pest or disease are potential sources of information, and can also act as sentinels in a sentinel surveillance system. These include:

- Government staff such as DAWR, customs & immigration and Australia Post workers
- Importers & brokers
- Point of entry staff such as port and airport workers
- Farmers
- Rangers
- Veterinarians (for animal-related diseases and pests)
- Hospital staff (for human-related diseases and pests)
- Citizens that come into contact with pests and diseases post-border

Such people may record information through official records, online or telephone services.

Farmers

Brugere et al. (2017) argue that farmers are well-placed (perhaps even more than veterinarians) to recognise changes signalling disease in their animals, given that they come into most contact with them. Morgan et al. (2014) found that in the case that the clinical signs of disease are pathognomonic (that is, specific to a particular disease or pest), the effectiveness of farmer diagnosis can be as high as that of laboratory testing.

There are many factors that impact the success of surveillance systems involving farmers. Farmers who report disease should be kept anonymous (to avoid being ostracised). Also, farmers should be informed as to the use of the data, and also compensated for any slaughter of their stock. Programs in which farmers are compensated and well-informed have higher sensitivity (Brugere et al., 2017).

Carlier et al. (2013) studied French oyster farmers and found that unclear trigger notification guidelines such as 'increased mortality' also hampered reporting, and also that often farming practices aimed at preventing disease emergence were overridden by individual profit-maximisation.

Citizen science

In *Citizen Science and Wildlife Disease Surveillance* (2015), Lawson et al. discuss the usefulness of citizen science teams to act as a surveillance team. Citizen science schemes have been employed to target the detection of emergent plant diseases, but the majority are aimed at diseases of vertebrates. A common source of data in this field is road kill carcass collection. Some of the problems with this may arise in lack of funding and appropriate training.

Welvaert and Caley (2016) differentiate between intentional and unintentional reporting, and opportunistic and controlled detection when comparing citizen science to crowd sourced data such as information from facebook. They argue that making useful inferences depends on where data lie within this framework.

Environmental

This consists of:

- Environment
- Weather
- Data on water pollution
- Data on air pollution

Environment

Environment is pertinent to investigating unexpected biosecurity risks. Consider, for instance, the arrival of a pest species that has a low heat tolerance arriving in one of Australia's northern ports.

Weather

ID-Viewer: a visual analytics architecture for infectious diseases surveillance and response management in Pakistan (Ali et al., 2016) is concerned with creating visual tools that assist in identifying outbreaks of infectious diseases in Pakistan by looking at chief complaints (i.e. SS data). It was found that certain combinations of symptoms signalled particular diseases only after taking into consideration seasonal weather conditions.

Similarly to environment, weather can affect biosecurity risks. For instance, seasonal pests such as Asian Gypsy Moth are only likely to establish under certain conditions. A potential issue arises, however, when the supply chain, as complicated as it is, is not fully understood, and this pest ends up making its way into an environment in which it can establish.

Water and air pollution

A cryptosporidium parasite outbreak in Milwaukee in the US was traced back to using infected ice using water pollution testing (Abat et al., 2016).