New Zealand National Weeds Distribution Database: a feasibility study

Britta Basse[1, 2], Graeme Bourdôt[2], Jennifer Brown[1], Shona Lamoureaux[2]

[1]Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

[2]AgResearch Limited, Lincoln, Private Bag 4749 Christchurch, New Zealand.

28 October 2008

Report for Environment Southland

Envirolink medium advice 499-ESRC212

University of Canterbury Research Report UCDMS2008/8

CONTENTS

1. EXECUTIVE SUMMARY

Environment Southland (ES) aims to predict the environmental and economic risks posed by particular species of weeds that are already established in the region or that are currently absent but already naturalised elsewhere in New Zealand. A knowledge of national weed distributions would facilitate national and regional reporting on changes in weed distribution. This information on distribution changes can be used to monitor the success of weed management strategies and to assist in developing models for predicting weed risk. To this end ES contracted scientists at the University of Canterbury and AgResearch Ltd to establish the feasibility of developing a readily accessible, user-friendly, national database, to collate, hold and disseminate data about weed distributions throughout New Zealand.

The project was undertaken by first contacting 15 organisations in New Zealand that hold data on the locality of weeds (14 regional authorities and Department of Conservation (DoC) Hokitika) and asking how they collect, store and use weed spatial data, and what they consider the benefits of having a national weeds distribution database would be. The key results for this first part of the project are:

- Based on the recommendations listed by the biosecurity managers a national weed database should be kept simple, be accessible and reliable and show accurate current and historical distributions of weeds, enabling users to identify changes in the distribution over time.
- Without exception all organisations contacted saw benefits in having access to a national weeds distribution database.
- Weed occurrence data are currently collected by these organisations in a non-standardised way; sometimes locations of weeds within properties are recorded, and sometimes an entire property is 'marked' as having a weed present. Different co-ordinate systems are used.
- There is a wide range of data collection and storage systems employed by the different regional authorities and most are currently reviewing these systems. Many seem to be tending towards GIS data storage systems. The Integrated Research Information System (IRIS) project (Contact: Derek Postlewaight Environment Waikato) is an initiative currently underway that plans to have common database software across 7 subscribing councils.
- A minority of councils are very happy with their current collection and storage systems and have no desire to change them. Thus the collection of spatial data on a national scale would have to work in conjunction with individual requirements where possible.

The second part of the project investigated the resources required for implementing a national distribution database. The key results are:

- There are already several initiatives underway within New Zealand in relation to accessing spatial weeds data from different agencies and combining the data to create a national picture of weeds distributions. For example, The New Zealand Virtual Herbarium project (Contact: Aaron Wilton Landcare Research Ltd Lincoln) has a budget of $130K to collect and collate New Zealand herbarium weeds location data from a range of databases by December 2009.
- It may be possible to build on the technology and expertise of the above initiatives and others that are already underway. This would keep the costs and resources to a minimum. Further negotiations are required.
- It does not make sense to hold national weeds spatial data in one big database. In line with similar national and international initiatives, it would be more convenient for each weed collection agency to place their data on a web server

(some already do this). The data can then be electronically accessed by an authorised web server(s) when required. End users could then log onto a website and gain a national picture as depicted in Fig. 1.

- In order to gain a national picture of known weed presences we recommend as a minimum 4 compulsory fields: species, date, GPS co-ordinate of weed (either point or polygon) and map co-ordinate type. Most councils are currently collecting these fields but some standardisation is required.

- A workshop is required to bring key people and ideas together.

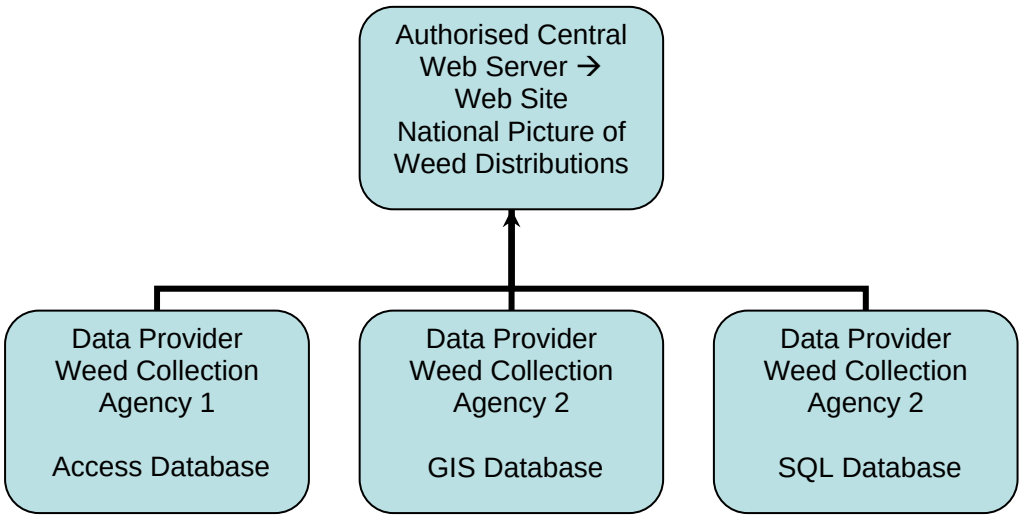

Figure 1. A schematic diagram of individual agencies providing weed spatial data to a authorised central web server.

A recording system for known weed presences will only give an approximation of the true underlying distribution but will still be useful for modelling potential distributions and seeing how current distributions change over time. A better estimate of the true underlying distribution could only be achieved with standardised procedures for selecting sites and for collecting field data including collecting data on sites where weeds are absent.

A website which accesses data from individual databases (preferably where data are gathered under a standardised survey protocol) would provide information on national weed distribution. This would (1) enable risk analyses of particular weeds by comparing current and potential distributions, (2) facilitate national and regional reporting on changes in weed distribution, and (3) may be used to monitor the success of weed management strategies aimed at stopping spread.

The overall finding from this project is that the data and technologies to create a New Zealand national weeds distribution already exist. What is needed is standardisation in data collection across organisations and coordination and communication between organisations to realise this goal.

2. INTRODUCTION

Throughout New Zealand over 100 organisations (e.g., government institutions, regional authorities, private agencies and community groups) collect information about the location of weed infestations. Such spatial data, often in the form of map co-ordinates, are stored either on paper or computer and are often not accessed by the other agencies.

There is a need at the regional level to be able to predict the risks posed by weeds based on the species propensity to establish and spread. Predictions of weed potential distributions, spread and hence future distributions can be used to rank weeds in terms of management priority and to plan work programmes to proactively manage weeds in the most cost effective and timely way. This is particularly important for weeds which have not yet established in a region, are present in an adjacent region or have established at only a few locations. Knowledge of their current distribution and behaviour in other parts of New Zealand with similar physical and ecological characteristics can be used to identify the need for high priority control measures.

The distribution and spread of weeds is dependent on a wide range of factors, e.g., geography, climate, geology, species, ecosystem type, land use, etc. There are currently some models in existence which can be used to characterise sites in terms of these variables and their suitability for particular weed species (e.g., CLIMEX, LENZ, etc), thereby enabling prediction of the potential distributions of these species. However to provide most value to pest managers in determining the risk posed by a species (weed risk analysis), these model predictions need to be compared with records of the current and (if possible) historic geographical distribution of the particular weed species in New Zealand.

Most regional councils have some information on weed distribution however these records exist in a range of non-standardised formats and may only represent partial coverage of a region. In addition, national agencies such as the Department of Conservation have substantial but incomplete data on weed distribution. If all of this information were to be collated at a national level it would be possible not only to use it for weed risk analysis, but also to identify information gaps which need to be filled at regional and national levels to ensure robust weed risk analyses.

As stated in the Envirolink Medium Advice Grant proposal under which the project was funded (499-ESRC212), the aims of the project are:

> "To determine the costs and benefits of developing a readily accessible, user-friendly, probably web based, national database, to collate, hold and disseminate data about weed distributions throughout New Zealand."

This report is divided into two parts corresponding to the two steps described in the proposal required to achieve those aims:

1. **Benefits and stock-take survey:** Conduct a national survey of all regional councils and other agencies to establish what data are already in existence and what formats they are held in. The survey would also canvass regional councils and other potential users about the benefits of a national weeds database.
2. **Resource requirements:** Scope out the resource requirements for a work programme to collect and collate existing data and to create a mechanism to hold and disseminate it.

3. BENEFITS AND STOCK TAKE SURVEY

A national survey of the biosecurity managers at the regional authorities (Table 1) was conducted to establish what they perceived are the benefits of having a national weeds

distribution database, what data are already in existence and what format the data are stored in.

Table 1 Summary of regional authorities, biosecurity and IT contacts. All weed control operations on the West Coast are undertaken by DoC Hokitika and so Westland Regional Council were not surveyed.

Abbrev.	Council Name (Joe Bloggs' email format)	Biosecurity Contact	IT Contact
NRC	Northland Regional Council (bloggsj@nrc.govt)	Don McKenzie	Mike Podesta/Carol Cottam
ARC	Auckland Regional Council (joe.bloggs@arc.govt.nz)	Jack Craw/ Jonathan Boow	Phil Barlow
EW	Waikato Regional Council /Environment Waikato (joe.bloggs@ew.govt.nz)	Wendy Mead	Keith Hannon
EnvBoP	Bay of Plenty Regional Council /Environment Bay of Plenty (joeb@envbop.govt.nz)	John Mather	Jon Edney
GDC	Gisborne District Council (joe@gdc.govt.nz)	Trevor Freeman	
TRC	Taranaki Regional Council (joe.bloggs@trc.govt.nz)	Rob Phillips/ Bruce Pope	Catherine Law
HBRC	Hawke's Bay Regional Council (joe @hbrc.govt.nz)	Robin Packe/ Campbell Leckie/ Darin Underhill	Daryl Hall
MWRC	Manawatu-Wanganui Regional Council /Horizons Regional Council (joe.bloggs@horizons.govt.nz)	Don Clark	Des Armstrong
GW	Wellington Regional Council /Greater Wellington Regional Council (joe.bloggs@gw.govt.nz)	Richard Grimmet/ Pedro Jensen	Steve Moate
TDC	Tasman District Council (joe.bloggs@tdc.govt.nz)	Lindsay Vaughan	Jenny Eyles
MDC	Malborough District Coucil (joe.bloggs@marlborough.govt.nz)	Ben Minehan	Jamie Sigmund
WCRC	West Coast Regional Council (joe.bloggs@wcrc.govt.nz)	Simon Moran	
DoC HOKI	Doc Hokitika (jbloggs@doc.govt)	Jane Marshall	Clayson Howell (Wellington)
ECan	Canterbury Regional Council /Environment Canterbury (joe.bloggs@ecan.govt.nz)	Graham Sullivan	Ryan Elley
ORC	Otago Regional Council (joe.bloggs@orc.govt.nz)	Richard Lord	
ES	Southland Regional Council (Environment Southland) (joe.bloggs@es.govt.nz)	Richard Bowman	Randall Milne

A preliminary email was sent to each contact person and followed by a telephone call. Often, in addition, an IT person (Table 1) was contacted for more information about how data are stored electronically.

The questions asked in the survey were:

- Would a national weed distribution database be beneficial to the management of weeds in your region?
- How do you choose the sites where you collect data?
- When you go looking for weeds, what data do you collect in the field and how do you record the data in the field?
- How are these data stored back at the office?
- How is this information then used?
- Is this information available to other councils? Do they ever ask for it?

3.1 What are the benefits of a database on weed distribution?

All biosecurity contacts could see the benefits of having access to national weeds spatial data -along with some reservations about the practical implementation of such a scheme.

The positive comments received about a national database can be summarised as:

- It would be great for achieving consistency on how information is collected.
- It might help identify new total control weeds.
- It would give an idea of which weeds are heading our way.
- Such a database would help identify long-term weed distribution changes.
- It would assist in assessing weed risks and threats.
- It would help in identifying plants for Regional Pest Management Strategies because there is a need to see what is happening in other regions.
- It would help to identify errors in data.
- It would help highlight gaps in the data.
- It would identify risks in terms of any climate change impacts on weed distributions.

Regional authorities control non-overlapping tracts of land and therefore manage their own properties within the region. Indifferent comments in relation to this issue included:

- There is interest only from a curiosity point of view.
- It would be good from a national point of view but it would not make much difference at a local management level.

Some expressed reservations and were ‘nervous that it might turn into a monster’. Others said that money was an issue and they had no extra resources (dollars and/or personnel) to deal with such a scheme. A number of councils said that there were privacy issues with data going ‘public’ because weed-infestations could then be traced to a property. This matter was addressed at the Biosecurity Managers Group meeting held in Auckland on Sept 4th 2008; the conclusion reported to the authors of this report was that privacy issues may not be a problem because such information is already publicly available.

3.2 Which weeds?

Each regional authority has a Regional Pest Management Strategy (RPMS or PMS) with weeds classified according to their perceived management requirements, e.g., 'Surveillance' (S), 'Total Control' (TC), 'Boundary Control' (BC), 'Containment Control' (CC). A particular weed may have a different classification from one region to the next. In addition, classification schemes differ among regions. Location data typically are collected only for weeds in particular classes, e.g., some regional councils record data for Total Control weeds only.

3.3 How are the sites chosen?

We loosely define the word 'site' here as the location of a particular weed infestation. Site selection procedures differ between councils. In some cases, new sites are brought to the attention of biosecurity officers by a land owner or member of the public. In other cases, the biosecurity officer may have been visiting properties for many years and has good local knowledge of where weeds sites are or were historically. In yet other cases searches are made of the habitats where weeds are most likely to turn up e.g., certain weeds may be more likely to occur along riverbeds or in coastal regions. If a new weed infestation is found some councils conduct a delimiting survey in which a defined area around the infestation is searched for the species. Searching further afield continues until the limits of the infestation have been defined.

Most regional authorities do not use rigorous statistical surveying techniques. However, two exceptions are the probability-based surveys implemented by Environment Southland (ES) for a range of weed species and by Environment Canterbury (ECan) for nassella tussock. Both survey a different selection of sites each year (Bourdôt & Saville 2007). In the ES survey, sites are 2 x 50 m plots selected on the basis of ensuring good geographic coverage of Southland, with higher survey intensity in areas where weeds are known to occur. The process used to select sites in the 5-year plan is based on a generalised random tessellation stratified design (Stephens & Olsen 2004). The important distinction of the Environment Southland design is that each year reports can be produced for individual weed species, and for all species combined on weed distribution and abundance. Because the same survey design and protocol is being used each year in the ES survey, these annual reports are directly comparable among years and therefore can be used to monitor the success of weed management. This is also true for ECan's nassella tussock monitoring programme.

3.4 Data collection procedures in the field

A summary of 'on site' data collection procedures is given in Table 2. There are many data collection procedures including mental notes, notes jotted down in diaries, inspection sheets and other paper records, Personal Digital Assistants (PDA's) with GPS (Global Positioning System) facility, iPAQs (Pocket Personal Computers), laptops and tablets. Many regional authorities are in the process of updating these data collection procedures.

There is a trend towards GIS (Geographical Information System) software. A typical example is the Arc (Esri) software suite (ArcPad, ArcMap, ArcGIS) which allows regions of infestation to be hand drawn on a computer screen and then these boundaries are stored as the vertices of a polygon.

Table 2: Summary of data collection procedures in the field by regional authorities

Council	Location Data Collected?	Device Used?	Plans to change/upgrade?
NRC	Point data of infestations (might be the property entrance rather than the infestation site).	Paper records. Hand held GPS.	Yes.
ARC	Point data of infestation; the centre of the infestation is recorded if it is a patch.	PDAs with a GPS capability.	Yes, want to capture polygon data.
EW	Sometime GPS co-ordinate of where a weed is.	Paper records	Currently trying to standardise data collection procedures.
EnvBoP	Points or polygons of the infestation are plotted straight onto the tablet (often with the landowner there).	Tablets. These have GPS's and the database loaded	No, very happy with current system.
GDC	Point location of nodding thistle; Abundance recorded as high/med/low level of infestation.	Paper records; handheld GPS.	Yes, looking to go electronic.
TRC	Presence/absence; Point data marks the property (i.e., the whole property is marked).	Laptops; Hand held GPS used.	Yes, hoping to add an aerial map with polygon data.
HBRC	Ad hoc data collected.	No notes taken; GPS co-ordinates sometimes collected.	No.
MWRC	Map reference.	Paper records or mental notes.	Hoping to update to PDA's .
GW	Infestations matched up to land parcels; Size classes of infestations; Percentage coverage on large properties.	Paper records.	Yes, plans are to integrate PDA's with GPS.
TDC	Point data of approximate infestation is collected.	IPAQs loaded with GPS and ArcPad (Esri) software.	No.
MDC	Hard copy maps with infestation sketched on them (later digitised).	Paper records.	Hoping to update to an electronic system.
WCRC			
DoC HOKITIKA	GPS point data of infestation.	Paper records; Hand held GPS.	Yes, want to collect spatial polygons.
ECan	GPS location of infestation (sometimes just a point selected within the property from a topographical map).	Paper records.	Hoping to upgrade to a PDA web based system.
ORC	Hand drawn maps; GPS co-ordinates sometimes recorded for remote or large infestations; For urban infestations property address is recorded.	Paper notes GPS sometimes.	Yes, planning complete upgrade of data collection and storage.
ES	Location of infestation.	Handheld GPS device.	Yes.

Generally presence data only are recorded. Environment Southland record absences as part of their survey (see later). Taranaki Regional Council and Hawke's Bay Regional Council record absences for total control weeds. Other regional councils sometimes record that control operations have been completed at a certain date. The latter does not necessarily mean weeds are truly absent because weeds can return within a short time frame due to seed banks and root systems.

Qualitative measures of density such as low/medium/dense coverage are sometimes recorded. Sometimes actual densities such as '45 thistles', '0.001 ha' or percentage of property covered are recorded in a comments field but it was noted that these measures of density are prone to high variation between individuals. Environment Southland record

the number of 1 m^2 blocks inside 100 m^2 plots which have weeds (by individual species) present. This is then used as an index of density.

In many cases the weed location is marked by recording the GPS co-ordinate of the site. Either one co-ordinate or several are recorded depending on the size of the weed cover. For larger infestations, polygons are plotted on a map using Arc based software installed on a PDA, laptop or tablet. There were some discussions about whether there are any benefits to be gained from point data as compared to polygon data. In theory a national collection of spatial data should be able to encompass both types of data. The degree of accuracy of the site location is variable. Sometimes the whole property or the property letterbox is marked rather than the actual location of the weed.

Different co-ordinate systems are used between councils with the main ones being NZMG (New Zealand Map Grid) and NZTM (New Zealand Transverse Mercator).

3.5 In office data storage

A summary of the 'in office' storage facilities is given in Table 3. Data are stored in a range of ways including mental notes, diary jottings, paper records, Excel spreadsheets and electronic databases. There are a range of electronic databases: Access, Oracle, Sequel (SQL) databases and Spatial Database Engines (SDE's). Most councils are in the process of updating their electronic data storage systems and those that don't have an electronic database are hoping to get one. There is an initiative underway called the Integrated Research Information System (IRIS) project (Contact: Derek Postlewaight Environment Waikato) where councils will have a common software suite – including database software. Seven councils have subscribed to this project.

Table 3. In office data storage

Council Abbrev.	Database Name	Database Type	User Interface	Main Info Stored	Weeds	Updating
NRC	WORKSMART	SQL server database (user configured table).	Property based.	Species/date/rough GPS location of infestation (sometimes just a point on the property)/ area inspected; infested area; number of plants (if re-growth after eradication); plant density (qualitative);owner and property details; etc...	Those where control and effort are required.	Yes, currently under review.
	Lantana	Excel Spreadsheet.	Excel Spreadsheet	Columns: Region; location; northing; easting; officer; area; density (low, medium, dense); date cleared; etc...	Lantana.	
ARC	BIOMAP	Esri (SDE Spatial Database Engine).	User interfaces are Esri products. Property based information.	species/date/GPS point data of centre of infestation on the property/time/officer/presence data (no absence data).	Focus is on TC weeds.	Yes, small changes likely. Want to capture polygon data.
EW	BIS(Biosecurity Information System).	Oracle.	Property based user interface with mapping capability.	Weed name/inspection date/loosely where the weed is/sometimes the GPS co-ordinate/maybe estimate of the density (all these depend on the officer).	Focus is on Containme nt weeds.	Yes, parts of it under review.
EnvBoP	Land Management Database: Pest Plant Module.	SQL Server 2005.	MS ACCESS with GIS map objects interfaced.	points and polygons of infestations/property details/information on visits/work required/comments/time frames for control.	Weeds prioritised according to the RPMS.	Yes, want to upgrade the GIS component and integrate Document Management

Council Abbrev.	Database Name	Database Type	User Interface	Main Info Stored	Weeds	Updating
						- Long term replace ACCESS database with .NET applications.
GDC	No electronic database -paper records kept in filing cabinets.			Nodding thistle GPS co-ordinate of infestation/ keep inspection sheet with infestation details for other weeds.	Weeds based on biosecurity officers experience.	Yes, want an electronic database with GIS facility.
TRC	TUMBLEWEED	Microsoft SQL.		Presence data recorded by whole property being tagged/ presence-absence data recorded for TC weeds.	Species listed in the RPMS.	Yes, planning to record aerial photographs into the database so that areas of infestations can be recorded as polygons.
HBRC	CLOVER	SQL 2000	MS ACCESS with GIS Arc 9.2 interfaced.	Date/officer/plant species/presence/absence/density (low, med, high)/GPS guestimate of infestation/comments about abundance.	TC weeds (about 17 of them).	No.
MWRC	INCIDENTS	SEQUEL	GIS software for mapping; crystal reporting system.	Species/visit date/ map reference/ha of infestation/ qualitative density/management methods.	Weeds based on the experience of the biosecurity officers.	Yes, currently talking to GW about their upgrades. Want to have PDAs with ArcGIS.
GW	PLANTS	Oracle	ACCESS	Fields of database are based on inspection sheets e.g.: size class of infestation/percentage cover of infestation.	Species listed in the RPMS.	Yes, to a Origen Ozone product.
TDC	EXPLORE PESTS	Esri SDE Spatial Database Engine.	Internal browser application (Esri product).	Pest code/date/ GPS point(s) of approximate infestation/property valuation number.	Mainly monitoring weeds but biosecurity officers are always on the lookout for other weeds.	Yes, want to check in and check out into a personal geo-database type structure.
MDC	BIOSECURITY DATABASE	ACCESS	Property based user interface.	Weeds information from property inspections.	TC, CC and S weeds as described in RPMS/ other ad hoc weeds are also recorded if seen.	
		ArcGIS maps.	emap application.	Digitised version of hand drawn map with infestation marked as a polygon or point.		Hoping to update to an Esri product 'deco' this year.
DoC HOKI	BIOWEB	SEQUEL server.		Point data of the weed location for surveillance weeds. These infestations are usually small so no need to record area.	Surveillance Weeds.	Yes .
ECan	SURVEILLENCE	SEQUEL Server (tables).	Web based forms. ArcGIS.	Individual plants (GPS location) or infestations (GPS point + description).	Total Control weeds.	Yes, want to change to a web based spatial system that utilises PDA's. Currently
	BIOSECURITY	SEQUEL Server (tables).	Web based forms. Property Based.	Prime record is valuation number. Also recorded: weed type/area/GPS location of infestation (sometimes just a	Control weeds (e.g. broom, gorse,	

Council Abbrev.	Database Name	Database Type	User Interface	Main Info Stored	Weeds	Updating
				point selected within the property from a topographical map).	ragwort, nodding thistle).	talking to Eagle technology.
ORC		Excel Spreadsheet, one per species.	Map Info software.	Address/compliance/date/GPS co-ordinates of infestation sometimes.	Weeds on PMS.	Yes, currently looking at an entire IT review for data collection and storage.
ES	SURVEILLANCE PEST PLANTS	Access		Date/weeds name/ northing/easting/area infested/measure of abundance/comments	Those listed on the PMS. Some data on potential pests.	Currently negotiations underway with IRIS.
	PEST PLANT COMPLIANCE	SQL (created by Datacom).	Web based form.	Property Inspection information.		
		ArcGIS.		Survey weeds.	Murihiku region surveillance weeds.	

3.6 In office data usage

Data are used by biosecurity officers in a number of different ways; relocating weed-infested sites when returning to a property, keeping track of property inspections, compliance and enforcement, report writing, mapping infestations, and RPMS reviews.

Interestingly in some cases the database information was not used directly by biosecurity officers. Often their local knowledge is such that there is no need to use the database. This may mean that there is not much incentive to actually record their data in the database. Others reported that the existing database in the council wasn't easy to use and they kept their own database information in Excel spreadsheets. In one case a council had updated its database to include a mapping component which had made the whole data entry procedure much more difficult. In some extreme cases there was a notable difference in the descriptions of the usefulness of the council's database when talking to IT people (who thought it was great) and biosecurity officers (who were less enthusiastic).

These points re-iterate the importance of a national database being accessible and practical.

3.7 Data sharing

Usually the data are available to other councils but they rarely ask for it. Most information is passed to other councils at BMG meetings. Sometimes data are given to a CRI, other research institutions or DoC. When data on an infestation are linked to a property, these data are sometimes reluctantly shared with DoC.

4. RESOURCE REQUIREMENTS

For the second part of the feasibility study it was necessary to scope out the resource requirements for a programme to collect, collate and disseminate the data at a national level. This was done by identifying and investigating initiatives already in existence.

4.1 Initiatives already underway

During the first part of the study it became apparent that there are a number of applicable national and international weed distribution and data collection initiatives already underway. Information from the following projects (listed below) was obtained via websites and by questioning people from corresponding organisations about their views on collecting national weeds distribution data.

1. **New Zealand Virtual Herbarium (NZVH):** (Contact Jerry Cooper, Aaron Wilton, Landcare Research Ltd Lincoln). This project is the New Zealand version of the **Australian Virtual Herbarium** (see http://www.rbg.vic.gov.au/avh/). Plant data including GPS location data (longitude, latitude) will be collated from approximately 10 New Zealand Herbaria and made available via a New Zealand Virtual Herbarium website. This involves electronically pulling information as required from a variety of databases e.g., Access and SQL databases (note that there are no GIS databases). This project makes use of the **New Zealand Organism Register (NZOR)** for ensuring plants are assigned standardised names. **Tapir** software (http://www.tdwg.org/activities/tapir/) will be used to translate in house database fieldnames to standardised fieldnames as described by the **Darwin Core** (http://wiki.tdwg.org/DarwinCore). The standardised data is then made available to the outside world via each herbarium's in house web server. Eventually all standardised data collected on the New Zealand Virtual Herbarium website will be submitted to the **Global Biodiversity Information Facility (GBIF)** website (NZ Contact: Jerry Cooper Landcare Research Lincoln, http://data.gbif.org/countries/NZ).

2. **One-Land:** (Contact Jim McLeod Environment Waikato (EW)). A data commons project to access mainly GIS data from local and regional councils and other authorities. Data is not restricted to weeds but any type of GIS data collected. This would be an automated process whereby the authority would place its GIS database information on a web server which would be accessible by a commons server either regularly or upon request.

3. **DoC BioWeb:** (Contact Custodian Clayson Howell DoC) Clayson has recently lead a project to collect and collate data from the DoC database, various herbaria and Regional Council databases to create a grid overlay of presences for certain plants within New Zealand.

4. **Plant Conservation Network:** (Contact Jon Sawyer DoC www.nzpcn.org.nz) This website has a portal for members of the public to record locations of pest plants. Future plans include plotting distributions of pest plants on a national scale.

5. **Weeds & Invasive Plants (WIP) website:** (http://www.agis.agric.za/wip) A South African project which uses data from **SAPIA (South African Plant Invaders Atlas)** database. It shows a grid overlay of known weed presences and uses a grid size of .25 of a degree (approx 25 km).

6. **Ministry of Agriculture and Forestry Biosecurity New Zealand (MAF-BNZ):** (Contact: Andrew Harrison, John Sanson). Have collected GIS data of water weeds

(didymo) invasions throughout New Zealand to obtain a national picture of infestations.

7. **Global Invasive Species Database (GISD):** (Contact Mick Clout Auckland University, http://www.issg.org/database/welcome/). The website includes portals for New Zealand pest plant distributions.

8. **New Zealand Biodiversity Recording Network (NZBRN):** (Contact Colin Meurk Landcare Research Ltd. Lincoln, http://www.nzbrn.org.nz/). This project allows members of the public to log in and upload co-ordinate data of plant location and other plant data.

9. **Dataversity Website:** (http://dataversity.org.nz/): This website hosts a private discussion group "for people who manage data relating to terrestrial and freshwater biodiversity in local government".

4.2 Expected Costs

A number of the above projects are funded either entirely or partially by the **Terrestrial and Freshwater Biodiversity Information Fund (TFBIS)** (Contact: Alan White DoC).

A logical approach would be to build a national weeds distribution as an extension of one of the above projects rather than starting from scratch. Most of the people spoken to were open to such ideas but further discussions and communication are required. A workshop or conference would be most useful since a national approach would greatly benefit with input from some of the people involved in the above projects.

Holding all national data in one big database was not advised by Aaron Wilton and Jerry Cooper (the New Zealand Virtual Herbarium project), Jim McLeod (One-Land project) and Clayson Howell (DoC). Instead each data provider should retain their own individual databases but make data available to those authorised to access it by placing data on a web server. A number of regional councils are already in the process of doing this to allow database information to be accessed while biosecurity officers are inspecting properties. A central authorised server could then pick up the data and display it (on a webpage) upon request to give a national distribution picture (Fig. 1). It would be necessary to have the data provided in accordance with some standard data format protocol. These format protocols would have to be agreed upon by the data providers. International format standards already exist as mentioned in section 4.1.1 above.

As an estimate of the costs that might be involved, the New Zealand Virtual Herbarium project has budgeted $130,000 to create a web page interface (similar to the Australian Virtual Herbarium) by June 2009 that can access the herbarium records of four New Zealand herbaria (Auckland Museum, Te Papa, Landcare Research and Rotorua). Once this has been set up the cost of adding each remaining individual herbaria would presumably be a small fraction of this although it is too early to say exactly what it would be.

The One-Land Project of Jim McLeod at Environment Waikato has estimated costs of $614,000 over a 17 month period for a running pilot programme which accesses a vast variety of GIS data from a number of government agencies.

With negotiations it may be possible for a website accessing national weeds distribution data to ride on the existing efforts of other national initiatives thereby reducing costs. The main costs for councils will be in deciding what fields of data to collect (see next section) and ensuring that the data are collected with consistency across New Zealand. In

addition, each council would need to ensure that the data are entered into an electronic database that can be accessed externally.

It may be that in house database field names differ among data providers. For example one council might have ‘plant species’ as the field name while another might have ‘plant name’. It is possible to run database records through a software program which can standardise the field names. An example is Tapir software (http://www.tdwg.org/activities/tapir/) as used in the New Zealand Virtual Herbarium project which converts in house database field names to standard fieldnames as prescribed in the Darwin Core (http://wiki.tdwg.org/DarwinCore). Such software is free, easy to install and use according to Jerry Cooper (New Zealand Virtual Herbarium Project).

4.3 Compulsory data to be collected

A *minimum* of four compulsory fields would be sufficient to create a presence map. These four fields should all relate to the presence of a weed and are:

1. Species
2. Date
3. GPS co-ordinate(s) of weed location(either a point(s) or polygon(s))
4. Map Co-ordinate Type.

Most of these fields are already collected by the regional authorities in various formats, but it would be ideal if they could be stored in a standardised way. For example the name of a species should be selected from a published (and agreed) species database (for example NZOR), date of observation should include day, month and year, and GPS coordinates should be given to mark the centre of its infestation. If the infestation is large then a number of points should be plotted or a polygon of the infestation boundary can be plotted. Weed collections agencies need to agree on the exact standards.

We stress this is the minimum required to create a presence distribution map. In addition to these four fields it would be useful to have a unique identifier associated with each record and some form of record metadata which could include who entered the record, estimates of density, general comments on the quality of the data etc.

The gaps in such a map would be ‘unknowns’, however a more accurate description of absences could be gained by using a survey design with some statistical rigour (see the discussion and conclusion section). Options should therefore also be created for completing fields for common descriptions on the field methods used at the site to determine weed presence or absence or abundance, numerical measures of density or ground cover, and for common descriptions on the survey design used to select the site.

5. DISCUSSION AND CONCLUSION

Without exception all the biosecurity managers were positive (with some reservations) about having some form of national weed distribution information. A range of data collection and storage systems is used by the regional councils. These include mental notes and paper records, laptops taken into the field with the database installed on them, and polygons of infestations marked on a map. By placing a small number of database fields collected in a standardised way and making them available to the outside world a centralised server could pick up data on request from individual weed agencies and then display a national picture of weed presences.

5.1 Recommendations

The following recommendations would allow regional councils and other weed collection agencies to become data providers, and enable a national distribution data base for weed occurrences to be created;

1. A minimum of four data fields (collected with some standardisation agreed upon) for weed presence should be stored in the database namely: Species; Date; GPS location co-ordinate(s); Map Co-ordinate Type.
2. Location co-ordinates stored in the database should mark the infestation site as closely as possible.
3. If infestation is large then mark it with either a few points that roughly cover the infestation or enclose it in a polygon if possible.
4. Plan software upgrades with a national data collection picture in mind.
5. Run database records through software so that fieldnames are standardised. Put this standardised data on a web server.
6. A centralised server picks up data on request and displays it on a website.

These recommendations could be implemented via a three phase plan:

Phase 1. Run a workshop: Agree on compulsory fields for data collection and a standardised data format protocol. Decide who will create the website for end users and what the user interface will be.

Phase 2. Weed agencies standardise data field names and make this standardised data available on the web servers

Phase 3. A centralised server collects weed presence data from weed agencies. A web interface then shows current national weed presence data and historical data. This will enable gaps in the data to be identified and addressed.

5.2 The future – true absences and densities

We have provided recommendations that we believe would enable a national picture of known occurrences of weeds to be created. This picture would (1) enable the risk posed by these weeds to be quantified and (2) show how the distributions of these areas of known weed occurrences are changing over time. It would not however provide information about densities or absences.

Ideally information on weed distribution would mean developing databases that contain both presence and absence data for weeds. In addition, if this data were collected using a rigorous survey design, then measures of weed infestation such as the proportion of the total area in the region infested, average weed density, proportion of the total area infested over a certain density, could be readily estimated (along with estimates of uncertainty). If addition, if such surveys were repeated at regular time intervals, then these estimates of weed distribution could be tracked through time. This would allow reporting on changes in weed distribution, and potentially enable the success of weed management strategies aimed at stopping spread. Information on weed distribution (both presence and absence information) could assist in developing information models to predict risk areas and species for targeting weed management and control.

If data were to be collected in a nationally co-ordinated system then it would need to be standardised at the collection point by developing a standard field protocol (e.g., as has been developed by the National Possum Control Agencies for possum monitoring (http://www.npca.org.nz/). Such a protocol would need to be developed with input from all councils so it was practical. It would include standard definitions of sites, timing of

surveys, intensity of site searches, and so on, along with standardised reporting forms. A standard field protocol would go some way to reduce differences among councils, and among individual field-staff to ensure consistency in data quality.

Such a national weed survey protocol would need commitment from all councils and considerable investment in time. The information gathered would greatly enhance our knowledge of national weed distributions.

6. ACKNOWLEDGEMENTS

We thank the Foundation for Research Science and Technology, for an Environlink Medium Advice Grant 499-ESRC212. We also thank the people from the various regional and district councils, Auckland Museum, DoC, MAF-BNZ and Landcare Research Ltd who participated via telephone or in person, in this project. All were extremely helpful and friendly and this was greatly appreciated.

7. REFERENCES

Bourdôt G, Saville D 2007. Monitoring nassella tussock (*Nassella trichotoma*) under Environment Canterbury's Regional Pest Management Strategy - Year 10 (2006 - 2007). AgResearch. Pp. 27

Carver J. 2008. Environment Data 2.0. Building the digital platform for a sustainable future. A draft environmental sector data management vision and action plan. A discussion document prepared for MoRST (Ministry of Research Science and Technology).

Cooper J, Carver J. 2005. A Survey of Regional Council Databases. Report prepared for Terrestrial and Freshwater Biodiversity Information System Fund.

Stephens, D. L., Olsen, A. R. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*. 99: 262-277.