Manaaki Whenua
Landcare Research

Collating soil quality and trace element State of Environment monitoring data

Envirolink Grant: 2050-HBRC254

Prepared for: Hawkes Bay Regional Council

July 2020

Collating soil quality and trace element State of Environment monitoring data

Contract Report: LC3803

Jo-Anne E Cavanagh, Brandon Whitehead, Shirley Vickers, Bryan Stevenson

Manaaki Whenua – Landcare Research

Reviewed by:	*Approved for release by:*
John Drewry	Chris Phillips
Research Scientist	Portfolio Leader – Managing Land & Water
Manaaki Whenua – Landcare Research	Manaaki Whenua – Landcare Research

Contents

Summary

Project and Client

- This project collated regional council site soil quality (including trace element) data for State of the Environment (SOE) monitoring up to 1 July 2018. This is an essential step towards improving the national consistency of soil quality and trace element monitoring and data management to support the aims of the Environmental Monitoring and Reporting initiative and the development of National Environmental Monitoring Standards.
- This project was undertaken for Hawke's Bay Regional Council, with funding from Envirolink (Advice Grant 2050-HBRC254).

Objectives

- Provide individual council datasets of site and soil quality monitoring (including trace elements) data from the start of the 500 soils monitoring programme in 1995 to 1 July 2018.
- Identify critical attributes useful to help streamline future data collation/addition to currently collated data.

Methods

- Data gaps in previously collated and rigorously cross-checked data (data from the original 500 soils programmes and previous collations of data for national reporting) that had been uploaded to National Soils Data Repository (NSDR) were identified through comparison of the data held with that indicated to be held by councils obtained from the sample inventory developed as part of a previous SOE monitoring review.
- A meeting was held with HBRC to refine the site information to be captured in the collated dataset, and the missing site and monitoring data was then requested from councils with a data template (excel workbook) provided to capture these data.
- All data received underwent a data check before uploading to the NSDR, and data exported for individual councils exported separately.
- To provide the final datasets, a text processing script was run on the exported data to develop both a council-specific and a national cross-council land use category from variably input data.

Results

- Data were collated for all 12 councils currently undertaking SOE monitoring to at least 1 July 2018. These datasets are largely complete but require independent cross-checking by councils to ensure robustness.
- Various observations were made during data collation, including the variable management of data across councils, challenges in obtaining site information, a varying range of analytical parameters captured by different councils and within the

dataset, and the more or less obvious conversion between volumetric data and gravimetric data; capturing original data is always preferred.

- Capture of site information was variable, with slope, landform, and precipitation most commonly missing.
- Land use remains a challenging parameter to capture, with varying categories used by councils, varying terminologies used within and by councils, and an overall lack of clarity about the basis for different classification.

Improvements to the dataset, and future data collation

- This data collation project has pulled together a great deal of historical data from different sources and the value of (and therefore the time required to do) a rigorous cross-check of the collated data to ensure their robustness for future use cannot be underestimated. These data have undergone scrutiny by MWLR, but a critical next step is an independent review by the individual councils – particularly those with a longer history of monitoring – to ensure that the collated data are robust.
- Given the gaps in site information in council records, and with some councils indicating that information has not been collected, it may be useful to assess the value of what the required site information offers in order to recommend whether missing site information should be retrospectively captured.
- The parameters captured in the current data collation should be evaluated to check whether all appropriate analytical parameters have been captured; water content at –5 kPa and –10 kPa are recommended for regular inclusion. Further consideration should be given to the extent to which council-specific additional parameters be captured in this dataset. Similarly, the extent to which one-off samplings and/or analysis (e.g. organic contaminants) undertaken at SOE sites should be captured in this dataset needs to be considered. This may be best determined through a workshop with councils.
- Other specific improvements that can be made to the collated data to help future data collation, are:
    - Ensure that at minimum, the month and year of sampling are captured alongside sampling results.
    - Where possible, direct upload of laboratory results would be preferred, or, if using spreadsheets, the laboratory identifier of the sample results should be included to aid any future cross-checking.
    - Consider how potential changes in site/sample naming conventions by a given council might be accommodated in data collations. Ideally, an appropriate naming convention is adopted, and then continues to be used, unchanged, to enable ready verification of data over time.
    - Consider the utility of NZLRI codes/classification for certain parameters, e.g. vegetation, slope, landform, and/or whether free-form entry might be more suitable for some parameters, e.g. vegetation.
- Land use remains problematic in terms of how information is captured, the [variable] land use classifications used, and the basis for delineating those categories. Going forward, it may be useful to consider explicitly what activities are considered to be most important for monitoring soil quality, and how these may vary for different land

use systems and regions, and use this information as the basis for grouping or delineating different current land uses. Care needs to also be taken with understanding when land *cover* vs land *use* is being used (noting that land use is often inferred from land cover).

- The Soils Description Handbook of Milne et al. (1995) is currently being updated and it would seem logical for this handbook and data captured for SOE monitoring to be consistent in terms of the parameters used. It is understood that the draft updated soil description handbook will be shared with councils shortly.
- Regardless of what format future data collation takes, a fundamental requirement to ensure the value of the SOE monitoring programme is realised is that the soil quality data (site and sample) is well maintained within individual councils. Once this is achieved, future data collations or reporting at national levels will be significantly easier, and more valuable for interpretation.

Next steps

- The most immediate next step required to provide a national, fully cross-checked dataset of SOE soil quality (including trace elements) is that individual councils cross-check the collated data to identify any missing data or data discrepancies, and where appropriate, provide any missing data. This will help ensure the robustness of the collated dataset.
- Further discussion with councils would be useful to inform some aspects for immediately progressing the national dataset development, e.g. confirming parameters captured, including more council-specific parameters and potentially parameters captured for site information. The latter may fall into a broader conversation for which the basis for land use classification, and what land use information is best captured to achieve the aims of SOE monitoring is a fundamental component for discussion. A workshop may be the most appropriate forum for discussion.

1 Introduction

Regional authorities and the Land Monitoring Forum (LMF) have been monitoring soil quality and trace elements since the Landcare Research '500 Soils' programme finished in 2000. A subsequent review by Hill et al. (2003) resulted in improvements and the publication of soil quality monitoring guidelines in 2009. However, data arising from these programmes have not been consistently captured between councils, and sometimes within councils. Thus, inconsistencies and gaps in monitoring and data management for various projects that have collated data for different purposes, e.g. Status of Cadmium in New Zealand (2014), Environment Aotearoa 2015 synthesis report, Land Domain 2018 report, have contributed to less than optimum national soil quality reporting, e.g. trace elements monitored by councils were not reported at all in the Environment Aotearoa 2015 synthesis report.

Collation of variably captured 'legacy' data is problematic and requires a much more 'manual' approach to bring together in a consistent manner, but it is an integral first step in developing a consistent dataset to facilitate reporting at a national level, such as is required under the Environmental Reporting Act (2015), as well as in identifying temporal trends. This project drew on previous data collations and undertook extensive data checking and obtained additional data to collate soil quality data (including trace element data) undertaken by regional council for State of the Environment monitoring programmes. The data captured extend from 1995, with the commencement of studies to develop a soil quality monitoring programme, i.e. pre-cursors to the 500 soils programme undertaken over 1998–2001, to 1 July 2018 (as a nominal end point). This current data collation is an essential step toward improving the national consistency of soil quality and trace element monitoring and data management to support aims of the Environmental Monitoring and Reporting (EMaR) initiative, knowledge transfer, and national Environmental Domain reporting. This project complements a previous review of SOE soil quality monitoring data (Envirolink grant number 1757-HBRC226, Cavanagh et al. 2017). This report captures the experiences in collating the legacy data and identifies critical attributes to be considered in the addition of future data to the collated dataset.

This project was undertaken with co-funding from the MBIE Endeavour funded programme Soil Health and Resilience: oneone ora tangata ora (C09X1613) and MBIE's Strategic Science Investment Fund.

2 Objectives

- Collate all regional council site and soil quality monitoring (including trace elements) data from 1995 to 1 July 2018 and provide individual council datasets of collated data for further data checking.
- Identify critical elements useful for future data collation.

3 Methods

Before the commencement of the Envirolink project, existing data held by MWLR (data from the original 500 soils programmes (Sparling et al. 1996, 1998, 1999, 2000, 2001; Sparling & Schipper 1997) and previous collations of data for national reporting had been uploaded to the National Soils Data Repository (NSDR). These data included a mix of results directly from laboratory data sheets (data which had been compiled for the earlier 'SINDI' project), or 'derived' data e.g. summary results from excel spreadsheets. These previously collated data have been rigorously cross-checked, additional site information added where available (e.g. sites in the 500 soils programme), and data gaps identified through comparison of data held with that indicated to be held by councils obtained from the sample inventory developed as part of the SOE monitoring review (Cavanagh et al. 2017).

A meeting was held with HBRC to refine the site information to be captured in the collated dataset, and the missing site and monitoring data was then requested from councils. A data template (excel workbook) was provided to councils to capture and collate site information. The parameters captured in the template are shown in Table 1. Data provided were cross-checked with laboratory data, where available, or other information sources, as required (and often required additional liaison with councils to confirm the data provided) and uploaded to the NSDR. All collated data were then exported in a spreadsheet format (similar to the data template) to provide collated site and sample information for individual councils.

Table 1. Summary of parameters captured in the data template and in data exports

Site properties	Sample details
Site name	Site name
other id	Sample id
Date of establishment	Date of establishment
Easting (NZTM)[1]	Easting (NZTM)
Northing (NZTM) [1]	Northing (NZTM)
Easting (NZMG) [1]	Easting (NZMG)
Northing (NZMG) [1]	Northing (NZMG)
Long WG84 (decimal degrees) [1]	Long WG84 (decimal degrees)
Lat WG84 (decimal degrees) [1]	Lat WG84 (decimal degrees)
Vegetation cover	Vegetation cover
Land use category	Land use category
Land use category 2 (if you use a second category, please specify purpose of different classification)	Land use category 2 (if you use a second category, please specify purpose of different classification)
Land use notes	Land use notes
Date of soil reclassification	Lab for soil quality analyses
NZSC soil classification	Chemical soil quality properties:, pH, total C, total N, Olsen P (mg/kg, mg/L), mineral N (KCl-extractable NO3-N and NH4-N), CEC, and bases – exchangeable K, Ca, Mg, Na, total saturated bases, volume weight
Method of classification	Biological soil quality properties – AMN, microbial biomass C, basal respiration
Soil series	Lab for soil physical analyses
Method for soil classification e.g. pit, SMAP	Physical soil properties: bulk density, particle density, total porosity, macroporosity (–5kpa), air-filled porosity (–10 kpa), water content, initial water content, readily available water, totally available water, aggregate stability
Slope	Lab for trace element analyses
Elevation	Trace elements Standard suite: total recoverable As, Cd, Cr, Cu, Ni, Pb, Zn, plus F and U Extended suite: Calcium-nitrate extractable Cd Total recoverable Al, Sb, Ba, Bi, B, Cs, Ca, Co, Fe, I, La, Li, Hg, Mg, Mn, Mo, P, K, Rb, Se, Si, Ag, Sr, Tl, Sn, V
Landform	Organochlorine pesticide residues (HBRC only)
Soil drainage class	
Parent material	
Annual precipitation	

[1]One original location was captured, with location converted to WG84 as primary projection in export

After the data were exported, further processing was undertaken to provide more consistent land use categorisation based on information captured as vegetation cover, land use and land use notes. Land use has been variably captured among, and within, councils in relation to: 1) the land use categories used within a council; 2) the apparently subjective application of the categories; and 3) the specific terminology used to refer to those categories, e.g. dairy, dairying, dairy farm, dairy pasture, etc. As such, a mapping approach was used to provide two normalized land use classification schemas:

- a set of council specific land use terminology and categories
- a set of national cross-council land use categories

The modelling exercise is semantic in nature, utilising the Simple Knowledge Organization Framework (SKOS) (Isaac & Summers 2009). A very basic relationship model was used to fit within the scope of the current project. This model identified linkages between variable terminologies and the specified categories, and a text-processing script was generated to utilize this mapping to populate the council-specific and national cross-council land use categories. The script uses a series of regular expression filters to find matching text values, either exact values or a closest match, using the Jaro-Winkler distance[1] method. Although this processing was not comprehensive to capture all variations in terminology, it did capture most.

4 Results

4.1 Collated datasets

Data were collated for all councils currently undertaking SOE monitoring:

- Hawke's Bay Regional Council
- Northland Regional Council
- Auckland Regional Council
- Bay of Plenty Regional Council
- Waikato Regional Council
- Horizons Regional Council
- Taranaki Regional Council
- Greater Wellington Regional Council
- Marlborough District Council
- Nelson City Council (2018 data only)
- Tasman District Council
- Environment Canterbury
- Environment Southland.

[1] https://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance

These data were exported on an individual council basis and require an independent cross-check by the councils to identify any discrepancies between the collated data and council records. We anticipate that some discrepancies will be identified, as during initial data checking we identified that for some early (500 soils) data and for some parameters (e.g. AMN, Olsen P, CEC) council-held data were volumetric, even though they had been identified as being gravimetric. We suspect this discrepancy has arisen because in the 500 soils reports, gravimetric laboratory data had been converted to volumetric units using bulk density (Sparling et al. 1995, 1997, 1998a,b; Sparling & Schipper 1997). Each council has been provided with a list of 'known' discrepancies or data gaps, to assist with their data cross-checking (see also Appendix 1).

4.2 Observations during data collation

Soil quality SOE monitoring data are, and have been, variably managed across councils and over time (often due to staff turn-over); this posed some challenges in receiving the requested missing data, in terms of both time required by councils to collate the requested missing data, and what data were able to be provided. On a positive note, most of the councils with less than optimal data management practices had recognised that improvement in the management of soil data and information was needed and were in the process of taking that forward; this project helped reinforce that need.

Some councils had changed site naming systems for their monitoring sites since the SOE review in 2017. Changing naming systems is arguably more problematic for historical data, for example, laboratory data sheets for sites originally sampled during the 500 Soils project are named with the original 500 Soils site names; however, later council resampling of these sites also used different naming conventions to identify the sites, which can provide challenges in ensuring sample data are assigned to the correct site. Changing naming systems also hampers the collation of new data to ensure new site names match the old site identities. For other councils, the site identifier was effectively a description of the site location, which can be cumbersome. A uniform naming system is strongly encouraged because it will make data verification and data provenance much more streamlined, i.e. ensures that existing site data can be easily correlated to the new sites to avoid more confusion; further information on naming conventions is provided in Ritchie et al. (2017).

A further observation of some council data was that information on land use at the time of sampling was not clearly linked to the results from that sampling. This has the potential to lead to erroneous classifications of land use in future data analyses.

Some more detailed observations are made below.

4.2.1 Site information

The site information requested from councils was based largely on the requirements specified in Hill and Sparling (2009) and is shown in Table 2. As was noted in the SOE monitoring review (Cavanagh et al. 2017), site information captured by councils is variable, with gaps resulting from information either not being collected or being lost over time. Information most consistently not captured was precipitation, closely followed by landform and parent material. The manner in which site information was captured also varied; some councils used NZLRI codes or categories to capture information on slope, vegetation, and parent material (NZLRI rock type), while others appear to be more free-form descriptions or specific slope angles.

Table 2. Site information captured in the current project, and by Hill and Sparling 2009

Current project	Hill and Sparling 2009
NZSC classification – as detailed as available	NZSC classification/soil series
soil series	
Method of determining NZSC soil classification	
Date of any soil reclassification	
Land use	Current land use
Land use notes	
Vegetation	Present vegetation
Slope	Slope
Elevation	Elevation
Landform (as per Milne 1995)	Landform (as per Milne 1995)
Annual precipitation	Annual precipitation
Parent material	Parent material
Soil drainage class (as per Milne 1995)	Soil drainage class (as per Milne 1995)
The nature and date of any extreme events such as flooding, landslips	The nature and date of any extreme events such as flooding, landslips

While site information is captured at the time of establishment, some parameters rarely seem to be used in interpreting the results of monitoring, e.g. slope, landform. However, these parameters can actually be useful in helping interpret data, for example, to delineate hill-country vs flatland drystock framing. Conversely, other parameters (precipitation) vary over time, thus the relevance of capturing some of this information is unclear. Given the gaps in site information in council records, and with some councils indicating that information has not been collected, it may be useful to assess the value of what the required site parameters provide in order to make recommendations about whether such information should be retrospectively captured where it is missing.

Accurate capturing of site and sample location is a critical component for robust SOE monitoring, and a key source of variability in results can be removed if there is confidence that the same location is being resampled. Confirming the location of original sites can be challenging as non-specific information about locations may be the only information available. Where more specific information is available it can be unclear whether that location was taken at the time of site establishment or at a later time, and therefore in a different location. A further factor confounding ready cross-checking of location information from different sources is the different projections for locations, e.g. NZTM, NZMG. Some location information was provided by councils in multiple projections but it was unclear which was the original projection, thus errors that may have arisen in converting projections may be introduced and lead to misleading information about where a sample was collected. For the current work, the WGS 84 projection, which is internationally relevant, is the primary reference used in the NSDR, and hence in the collated data; this was determined through conversion of the location from its original projection (typically NZTM or NZMG) during data upload.

4.2.2 Sample information

As noted in the methods, the current data collation was drawn from a mix of laboratory results sheets (for standard soil quality chemical and physical properties) and previously collated data. While efforts were made to be comprehensive in the parameters captured in the recent data request, some parameters were over-looked. The most pertinent of these is volumetric water content (at –5 kPa and –10 kPa), which is a critical parameter (along with bulk density, total porosity, and particle density) for determining (the NZ Soils Bureau, McQueen (1988) defined parameters of macroporosity and air-filled porosity (or air capacity) at the respective tensions. Both these measures have been used to represent the 'macroporosity' indicator in the dataset and in reporting from the 500 soils data and by regional councils. The –5 kPa measure was used in the original 500 soils dataset and early regional council reporting, but since 2003, the air-filled porosity measured at –10 kPa (referred to as macroporosity in Hill and Sparling (2009)) has been the generally accepted tension for the macroporosity indicator. The distinction between these measures is further clarified in the draft National Environmental Monitoring Standards for Soil Quality and Trace Elements. The capture of volumetric water content at both –5 kPa and –10 kPa enables an additional check on reported macroporosity values on which tension was used and ideally as a means of calculating either parameter for the dataset throughout the entire sampling period. This is particularly important if the original laboratory reports are not available.

Beyond this, there was some variation between councils as to what parameters were routinely captured, e.g. Taranaki Regional Council routinely has calcium-nitrate extractable cadmium analysed; Horizons Regional Council has also undertaken visual soil assessments. More details on differences in the specific parameters captured by different councils are also captured in Appendix 2, and in Cavanagh et al. (2017).

As noted above, accurate information on sample location is critical in determining whether a source of variability in results could be due to differing locations being sampled. In data provided for collation, sample location was not often provided, or where it was, it is unclear whether the GPS location given is simply a cut and paste from standard site details, or the actual location taken at the time of sampling. Ideally, the location taken at the time of sampling is clearly recorded and used.

4.2.3 Other idiosyncrasies

Considerable effort was made to accurately capture the date of sampling (to the month and year), although for historical data this was difficult. In lieu of specific sampling date information, the date on laboratory data sheets was used and in some cases resulted in a change in the sampling year reported from council records – those sampling years appeared to be based on the sample ids used, which more accurately reflected the start of financial year in which the samples were collected. For more recent data, often only a year of sampling was provided, and in this case the date assigned was the start of the reported sampling year; at minimum the month of sampling would be useful to record. The format in which the date is captured needs to be considered, for example one council provided date as month in letters and year, which was problematic to incorporate into the dataset alongside conventional dates.

In the detailed data checking, it was observed that on occasion conversions between gravimetric and volumetric data were made, although it is sometimes not particularly evident where this has occurred. As mentioned above, it was identified that many councils hold volumetric data for some parameters (AMN, Olsen P, CEC) from the original 500 soils data, although it is captured in council records as gravimetric. In another case, Olsen P analysed as a volumetric measurement were captured in data collations as gravimetric, although the conversion factor was unclear – these data have been removed from the dataset. In the specific case of Olsen P, the conversion of volumetrically determined Olsen P to gravimetric data using bulk density appears to be relatively commonplace to enable comparison with laboratory data that have been determined gravimetrically. However, caution on this approach is urged. A more accurate conversion (in the context of comparing laboratory results expressed in different units) is through the use of volume-weight – the weight of a known volume of the dried-ground soil given that gravimetric analyses are undertaken using dried-ground soil.[2] This information is typically available for analyses expressed volumetrically.

4.3 Land use

As noted above, and previously (Cavanagh et al. 2017), land use is inconsistently classified by different councils, and sometimes within councils over time. Some of the difference is driven by regional differences in land use, e.g. mixed cropping (i.e. livestock grazing and cropping) systems are more prevalent in Canterbury. Variability over time is partly due to change in personnel, but may also arise from land use at the time of sampling, For example, a site on which grain crops are present at the time of sampling could be classified as an arable cropping site, or may be more appropriately classified as dry stock or dairy (or, ideally, a mixed cropping system) if the wheat crops are being grown as part of a mixed pasture cropping system. Similarly, a site on which kale is grown as a fodder crop could be classified as a market garden site, although it is probably more

[2] However, it should also be noted that there may be slight differences in the analytical method for gravimetric and volumetric analyses that may lead to differences in converted values compared with those determined using the specific method.

appropriately classified as dry stock or dairy. Regardless, the current variability in land use confounds reporting and interpretation of the monitoring data.

A mapping of land use terms in the data received with specific council categories and a nominal national category is shown in Appendix 2. This mapping is not comprehensive for all terms used by councils, but rather picks up the most common terms to illustrate a potential process that could be used to aid land use classifications, while capturing more detailed information about the land use. The nominal national categorisation (Table 3) builds on that provided in Cavanagh et al. (2017), but acknowledges that it can be difficult to distinguish between what might be considered arable cropping versus what might be considered intensive vegetable cropping (and may have significantly different inputs) at different sampling times, given the crops rotated can overlap. Hence, these two categories can be considered short-rotation cropping. While most council land use descriptions could be mapped to national categories, some could not, most notably ‘lifestyle block’ and ‘riparian’ classifications. The script run provided a means to automate the encoding of more consistent categories both within and across councils. Consistent, or comparable, categories provide the ability to develop an understanding and comparison of information on land use within the council as well as nationally (see Gehegan et al. 2011).

Table 3. Land use categorisations used for national categories of land use

Land use category	Comment on land use
Dairy	Includes organic dairy, irrigated and non-irrigated dairy. Needs to be clear whether this includes milking cows only. Needs to also consider whether a combined pastoral classification, perhaps based on input intensity, is more relevant.
Sheep/beef/deer	Includes sheep and beef, deer, goats, and is likely to be a mix of intensive and extensive systems. Ideally intensive and extensive systems could be identified with extensive (low input) systems including lifestyle blocks. Needs to be clear as to whether this includes dairy run-off (i.e. dry stock). Needs to also consider whether a combined pastoral classification, perhaps based on input intensity, is more relevant.
Pasture	Only if needed, and land use cannot be categorised as above. Also needs to consider whether a combined pastoral classification, perhaps based on input intensity, is more relevant.
Short-rotation cropping	Market gardens, vegetable crops, includes grain crops, hay, fodder crops
Perennial crop	Stonefruit, berry fruit, kiwifruit, grapes
Exotic Forestry	Plantation forestry, based on exotic species, primarily *Pinus radiata*
Indigenous vegetation	Indigenous forest, native scrub, reserves in non-urban areas, native tussock not used for grazing
Urban	Parks and reserves in urban areas (could overlap with some sites currently classified as indigenous vegetation)

A further complexity of cropping is considering where mixed cropping land use, i.e. where grazing pastoral rotations are included as part of crop-rotation, might fit. Cavanagh et al. (2017) suggested a potential ‘farm system’ classification for drystock and dairy systems to account for the difference in management practices between pastoral systems and livestock systems that include grain or other cropping, and potentially have greater land management activity, e.g. frequency of cultivation (Table 3). Further discussion is required on whether a minimum proportion of the farm or frequency of other crops needs to be specified to distinguish between pastoral and mixed cropping, or between different mixed cropping farms. The farm system concept had been suggested for wider use to address some potential land use classification issues, e.g. potential mis-identification of drystock or dairy sites as cropping sites as a fodder crop was growing at the sampling site at the time of sampling). However, to date, application of this concept appears to have rarely been used effectively and there is often confusion about the difference between land use and farm system. Thus, the farm system concept may be limited to application to livestock grazing systems, such as in Table 3.

Table 4 Potential land use categorisation for soil quality sites on dry stock and dairy farms[1]

Farm system	Description
Dry stock – pastoral	Pastoral farm system.
Dry stock – mixed cropping	Pasture and feed (grain, forage) crops (including dairy support) grown on a rotational basis. (Is there a need to consider frequency of crop rotation / pasture renewal or proportion of farm used for other crops to further delineate?)
Dairy – pastoral	Pastoral farm system
Dairy – mixed cropping	Pasture and feed crops grown on a rotational basis. (Is there a need to consider frequency of crop rotation / pasture renewal or proportion of farm used for other crops to further delineate?)

[1] It is anticipated that by considering farm system, sites previously identified as pasture would also be able to be assigned to dry stock or dairy and thus improve land use categorisation

Further, while Tables 2 and 3 presents dairy and dry stock (sheep/beef/deer) as separate categories, further consideration should be given to exactly what features of these systems are important to delineate. In other words, are these categories being primarily used as surrogates for intensity of inputs, or are there other reasons to justify the delineation, such as public interest in the effects of dairy? LMF (2009) distinguished between intensive and extensive pasture based on differences in LCDB grassland coverage classes, but this has not equated to 'on the ground' assessments of intensive and extensive pasture, as there is no agreed definition of what constitutes intensive or extensive systems for soil quality purposes. An additional consideration in the drystock/dairy vs intensive/extensive pastoral systems is where lifestyle blocks, which are becoming increasingly common, might most appropriately align. For livestock systems it may also be worth looking across to industry-classifications (e.g. https://beeflambnz.com/data-tools/farm-classes) to consider the relevant categories for SOE monitoring.

The preceding discussion serves to illustrate the complexity in consideration of land use and that there can be multiple legitimate ways to classify land use. Going forward it may be useful to consider more explicitly what activities, e.g. stocking rates, intensity of typical fertiliser application or cultivation, are considered to be most important for monitoring on soil quality, and how these may vary for different land use systems and regions, and to use this information as the basis for grouping or delineating different current land uses. An additional consideration is how easy it is to get required information.

5 Improving the dataset and recommendations for future data collation

The original intent for this project had been to generate a single, full, cross-checked dataset of collated monitoring data to 1 July 2018. However, a delayed start to the project and challenges with obtaining the identified missing data, including that some information had been lost or not collected in the first place, meant the revised output was a series of individual council datasets. While significant progress has been made in collating and integrating newer datasets with the original dataset, further progress is needed to ensure a nationally consistent dataset can be achieved. Such a dataset should avoid the need for additional data collations to address different needs; rather the required data should be able to be extracted from the nationally consistent data set – with the greatest utility arising if the dataset continues to be added to.

This collation has assembled much historical data from different sources, and the value of (and therefore the time required to do) a rigorous cross-check of the collated data to ensure its robustness for future use cannot be underestimated. These data have undergone scrutiny by MWLR, but a critical next step is an independent review by the individual councils – particularly those with a longer history of monitoring – to ensure that the collated data are robust.

The parameters captured in the current data collation should be evaluated to check whether all appropriate parameters have been captured; water content at –5 kPa and –10 kPa are recommended for regular inclusion. Further consideration needs to be given to the extent to which council-specific, additional parameters are captured in this dataset, e.g. visual soil assessment undertaken by Horizons Regional Council, calcium-nitrate extractable cadmium and zinc, undertaken by Taranaki Regional Council, organochlorine pesticide residues by Hawke's Bay Regional Council. Similarly, the extent to which one-off samplings and/or analysis (e.g. organic contaminants) undertaken at SOE sites should be captured in this dataset also needs to be considered. This may be best determined through a workshop with councils.

Other specific improvements that can be made to the collated data and should be considered for future data collation, are:

- Ensure that at a minimum, the month and year of sampling are captured alongside sampling results, which would add value to the interpretation of the results over time, given seasonal influences on some soil properties.
- Where possible, direct upload of laboratory results would be preferred, or, if using spreadsheets, the laboratory identifier of the sample results should be included to add any future cross-checking.
- Consider how potential future changes in site/sample naming conventions by a given council might be accommodated in data collations. Ideally, an appropriate naming convention would be adopted, and continue to be used, unchanged, to enable ready verification of data over time.

- Consider the utility of NZLRI codes/classification for certain parameters, e.g. vegetation, slope, landform, and/or whether free-form entry may be more suitable for some parameters, e.g. vegetation

Land use remains problematic in terms of how information is captured, the [variable] land use classifications used, and the basis for delineating those categories. Going forward, it may be useful to consider more explicitly what activities, e.g. stocking rates, intensity of typical fertiliser application or cultivation, are considered to be most important for monitoring on soil quality and how these may vary for different land use systems and regions, and to use this information as the basis for grouping or delineating different current land uses. In terms of land use information captured, to capture relevant information is to consistently capture vegetation cover alongside land use, and includes any specific comments on land use that have the potential to capture the most specific information but also allow re-classification if required, e.g. fodder crop in pastoral system. Care also needs to be taken with understanding when land *cover* vs land *use* is being used (noting the land use is often inferred from land cover).

While site information is captured at the time of establishment, some parameters rarely seem to be used to interpret the results of monitoring, e.g. slope and landform. Although these parameters can be useful for interpreting the nature of a site, e.g. hill-country vs flat drystock land use, ability to access this information can be limited. Some parameters (precipitation) vary over time, thus the relevance of capturing this information is unclear. Given the gaps in site information in council records, and with some councils indicating that information has not been collected, it may be useful to assess the value of what the required site information offers in order to provide recommendations on whether it should be retrospectively captured where it is missing.

The Soils Description Handbook of Milne et al. (1995) is currently being updated and includes details on information to adequately classify soils, e.g. slope, landform, parent material, as well as vegetation and land use. It would seem logical for this handbook and consequently data captured for SOE monitoring to be consistent in terms of the parameters used. It is understood that a draft of the updated handbook will shortly be shared with councils.

Regardless of what format future data collation takes, a fundamental requirement to ensure the value of the SOE monitoring programme is realised is that soil quality data (site and sample) is well maintained within individual councils. Once this is achieved, future data collations or reporting at national levels will be significantly easier, and more valuable for interpretation.

6 Next steps

The most immediate next step required to provide a national fully cross-checked dataset of SOE soil quality (including trace elements) is that individual councils cross-check the collated data to identify any missing data or data discrepancies, and, where appropriate, provide any missing data. This will help ensure the robustness of the collated dataset.

Further discussion with councils would be useful to inform some aspects for immediately progressing the national dataset development, e.g. confirming parameters captured, including more council-specific parameters, and potentially parameters captured for site information. The latter may fall into a broader conversation in which the basis for land use classification, and what land use information is best captured to achieve the aims of SOE monitoring, are fundamental components for discussion. A workshop may be the most appropriate forum for such discussion.

7 Acknowledgements

Thanks goes to Kiran Munir, and particularly to Suzanne Lambie (both MWLR) who assisted with nitty gritty checking of collated data.

This project was undertaken with co-funding from the MBIE Endeavour funded programme Soil Health and Resilience: oneone ora tangata ora (C09X1613) and MBIE's Strategic Science Investment Fund.

8 References

Cavanagh JE, Munir K, McNeill S, Stevenson B 2017. Review of soil quality and trace element State of the Environment monitoring programmes. Envirolink Advice Grant: 1757-HBRC226.

Isaac A, Summers E 2009. SKOS Simple Knowledge Organization System Primer. W3C Working Group Note 18 August. Available at: http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/.

Gehegan M, Smart W, Masoud-Ansari, Whitehead B 2011. Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies. doi.org/10.1145/2068976.2068977

DJ McQueen, DJ 1988. Glossary of Soil-Water Terms. NZ Soil Bureau Laboratory Report EP31. Department of Scientific and Industrial Research 1988.

Hill R, Sparling G. 2009. Soil quality monitoring, Land and Soil Monitoring: A guide for SoE and Regional Council Reporting; New Zealand. Land Monitoring Forum.

Milne JDG, Clayden B, Singleton PL, Wilson AD. 1995. Soil description handbook. Lincoln, New Zealand: Manaaki Whenua Press.

Sparling G, Schipper L, McLeod M, Basher L, Rijkse W 1996. Trialling soil quality indicators for the state of the environment monitoring. Supported from the Sustainable Management Fund, MfE Project Number 5001 – Research Report for 1995/1996. Landcare Research Contract Report LC9596/149.

Sparling G, Schipper L 1997. Trialling soil quality indicators for the state of the environment monitoring. annual research report for 1996/1997. Supported from the Sustainable Management Fund, MfE Project Number 5001. Landcare Research Contract Report LC9798/051.

Sparling G, Schipper L, McLeod M, Basher L, Rijkse W 1998. Trialling soil quality indicators for the state of the environment monitoring. Research Report for 1997/1998 Supported from the Sustainable Management Fund, MfE Project Number 5001. Landcare Research Contract Report LC9798/141.

Sparling G, Rijkse W, Wilde H, van der Weerden T, Beare M, Francis G 2000. Implementing soil quality indicators for land 500 Soils Project Report for 1998/1999. Supported from the Sustainable Management Fund, MfE Project Number 5089. Landcare Research Contract Report 9900/108.

Sparling G, Rijkse W, Wilde H, van der Weerden T, Beare M, Francis G 2000. Implementing soil quality indicators for land 500 Soils Project Report for 1999/2000. Supported from the Sustainable Management Fund, MfE Project Number 5089. Landcare Research Contract Report 0001/059.

Sparling G, Rijkse W, Wilde H, van der Weerden T, Beare M, Francis G 2001. Implementing soil quality indicators for land Research Report for 2000-2001 and Final Report for MfE Project Number 5089. Landcare Research Contract Report LC0102/015.

Sparling G, Schipper L, McLeod M, Basher L, Rijkse W 1996. Trialing soil quality indicators for the state of the environment monitoring. Supported from the Sustainable Management Fund, MfE Project Number 5001 – Research Report for 1995/1996. Landcare Research Contract Report LC9596/149.

Ritchie A, Osorio-Jaramillo J 2017. National environmental monitoring site identification system. Envirolink Grant: 1729-HZLC137 Available at: https://envirolink.govt.nz/assets/Envirolink/1729-HZLC137-National-environmental-monitoring-site-identification2.pdf

Appendix 1 – Known data discrepancies or gaps

Table 5. Details of known data discrepancies or gaps, that can be corrected in a next stage of dataset development

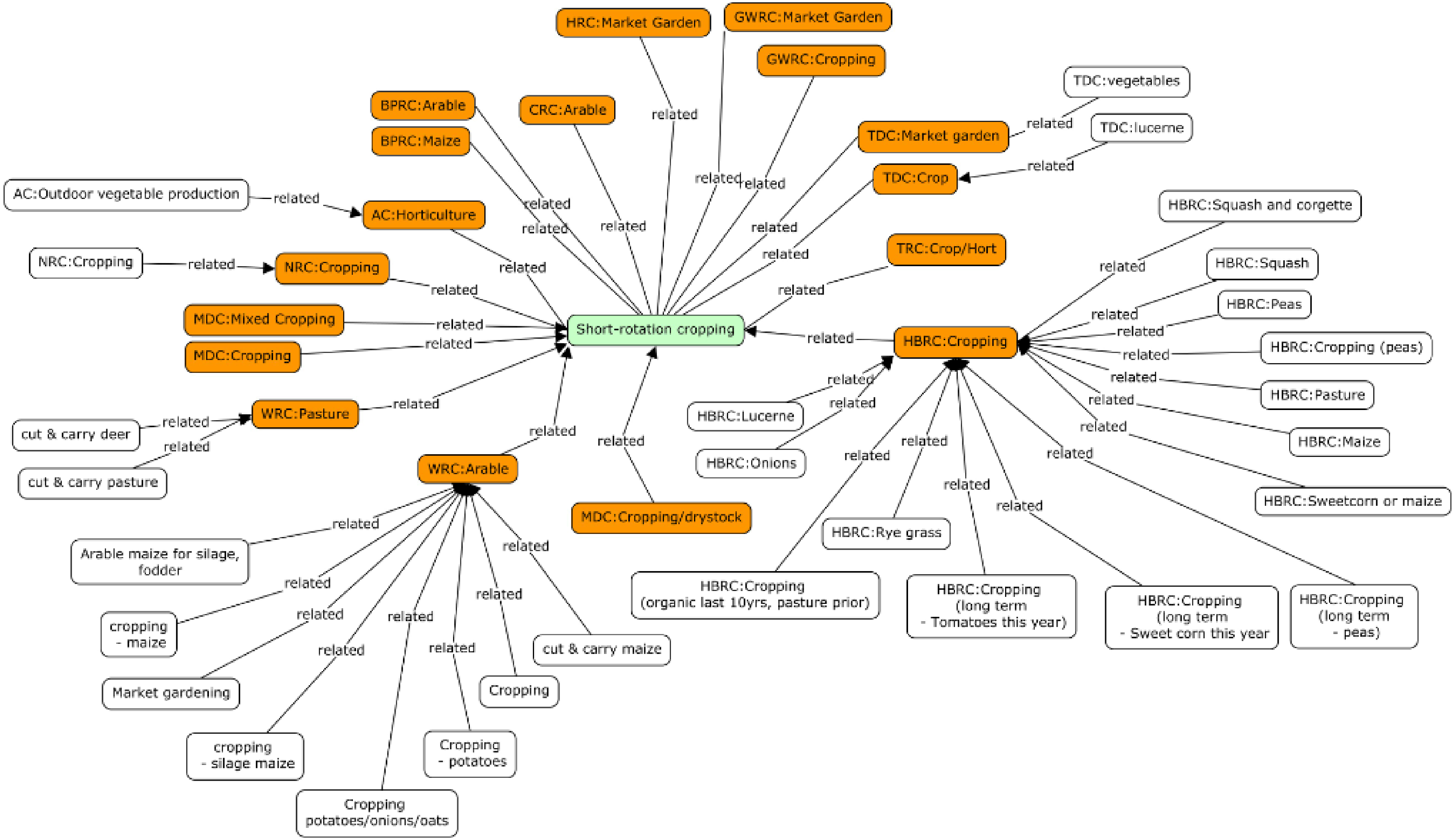
Council	Data collation status notes	Additional comments
Hawke's Bay	No site information or data for Landsites 47 and 48 Recent sampling dates based on start of reported sampling year?	
Northland	Confirmation of the actual sampling dates for 2010/2011 samplings for specified results (differing information in NRC report, and laboratory data sheets). Recent sampling dates based on start of reported sampling year	
Auckland	Some early 500 soils data held by council is volumetric - converted from original gravimetric lab results using bulk density (affects AMN results, potentially Olsen P) Missing data for urban parks sampled in 2012 – to be uploaded New site identifiers not currently used as main site identifier in collated data Recent sampling dates based on start of reported sampling year	
Bay of Plenty	Some early 500 soils data held by council is volumetric - converted from original gravimetric lab results using bulk density (affects AMN results, potentially Olsen P) Some site information from 500 soils documents currently not included (sites sampled in 1998, and 1999) Misalignment of data for selected soil physical measurements (bulk density, particle density, total porosity, macroporosity –5 kpa, air-filled capacity (–10 kpa) for sites sampled over 2006 to 2012 during upload, hence these data show in the wrong columns). Multiple site 'pages' for a small number of sites; these show as multiple lines in the export, and will be merged. Discrepancies between nominal sample collection data and date in from lab for some sites, e.g. for site 23/1 for 2005/2006 the sample schedule indicates samples were collected in March 2005, but the sample in date to lab is 12 May 2006 (there were sample provided to lab in March 2005, but not 23/1). Need some independent verification of when samples were most likely collected (within a few months of collection is sufficient – to understand whether scheduled sample time or date into lab is more accurate.	Date format (month as 3 letters, and year in separate columns) provides challenges in provided a comparable date with other councils
Waikato	Some early 500 soils data held by council is volumetric - converted from original gravimetric lab results using bulk density (affects AMN results, CEC and potentially Olsen P) handling of data for sites 99/130 to be improved (these sites were merged over time)	
Horizons	No known data gaps or discrepancies	
Taranaki	Some site information from 500 soils documents currently not included (sites sampled in 1998) Missing site information for sites in 2012 Recent sampling dates based on start of reported sampling year?	

Council	Data collation status notes	Additional comments
Greater Wellington	No known gaps or discrepancies	
Marlborough	Recent sampling dates based on start of reported sampling year	Land use not clearly linked to sample
Nelson	No known data gaps or discrepancies	Some data (Olsen P, TC, TN) is gravimetric converted from volumetric using bulk density
Tasman	No known data gaps or discrepancies Recent sampling dates based on start of reported sampling year?	
Environment Canterbury	Stone content is not currently captured	Only data from sites established under 500 soils regime (and up to 2006) are included in this data collation
Environment Southland	No known data gaps or discrepancies Recent sampling dates based on start of reported sampling year?	Site names provided are long (descriptors of site), hence some sample ids were used as site ids.

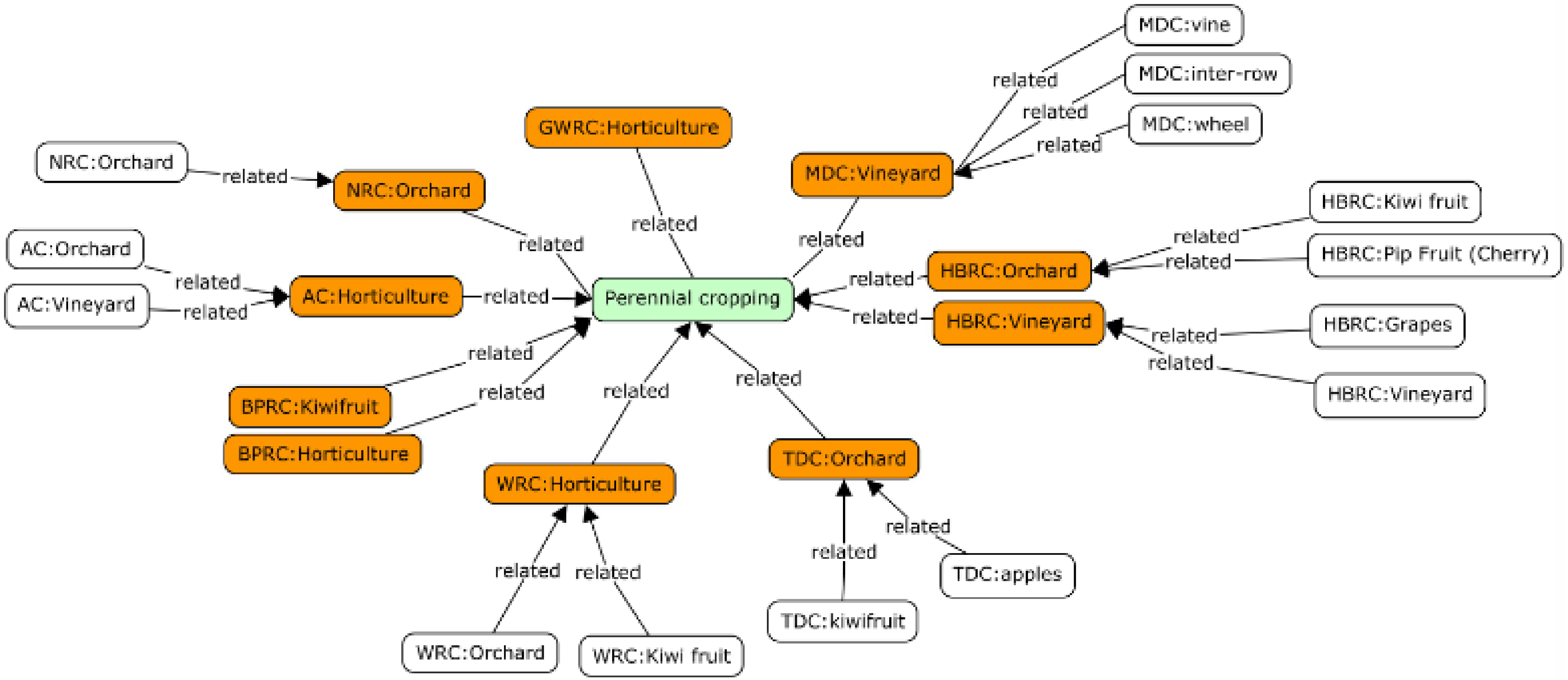
*Date gaps here refer to known collected data and EXCLUDES missing information where it is known not to be available, e.g. certain site characteristics

Appendix 2 – Land use classification mapping

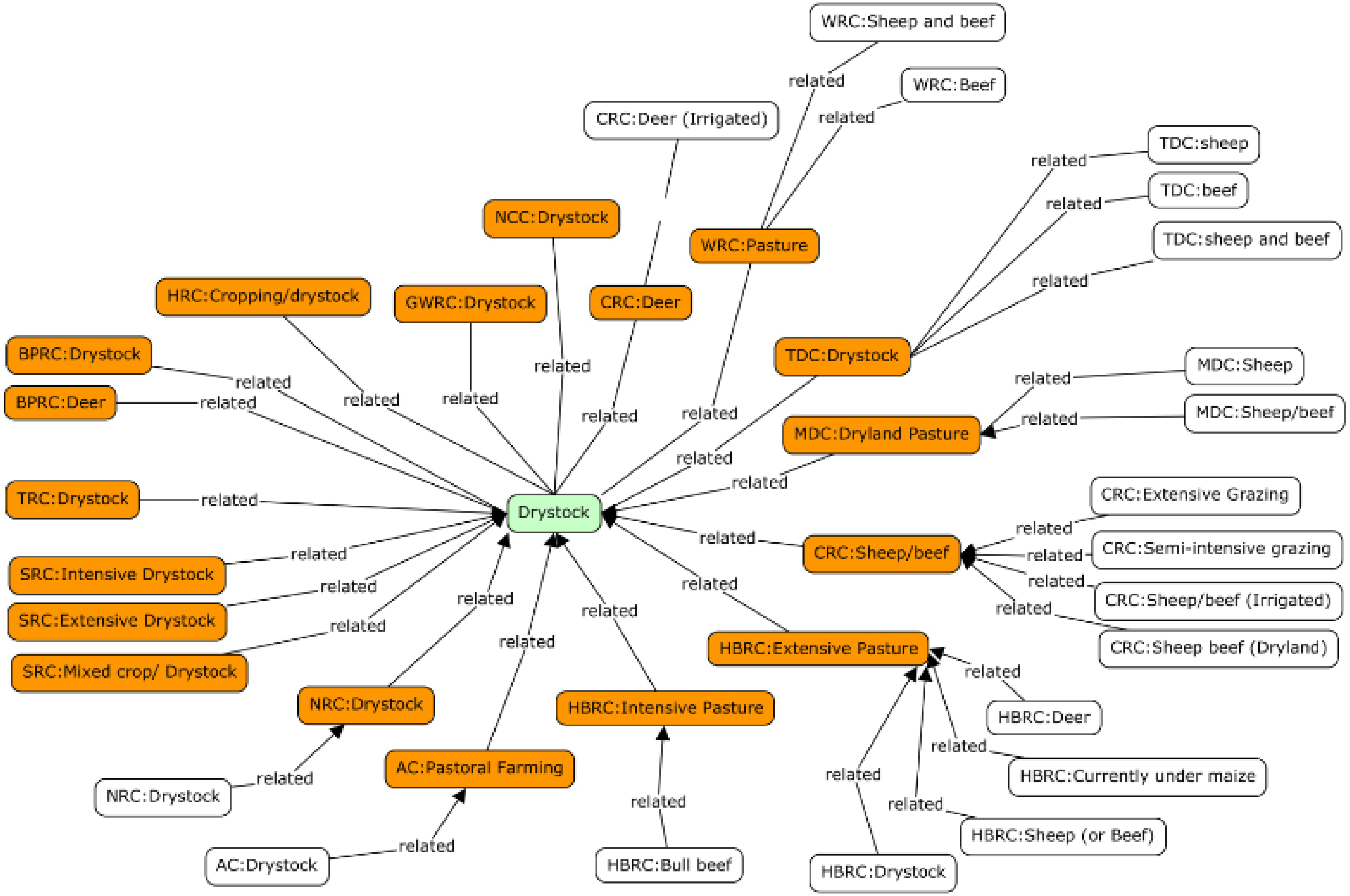
The following figures provide an illustration of the relationship between land use information provided in council data (land use classification and any associated notes on land use) used to identify council specific classifications (orange) or the nominal cross-council classification (green). Note: this land use mapping is not comprehensive for all terms used, but rather captures most terms used.

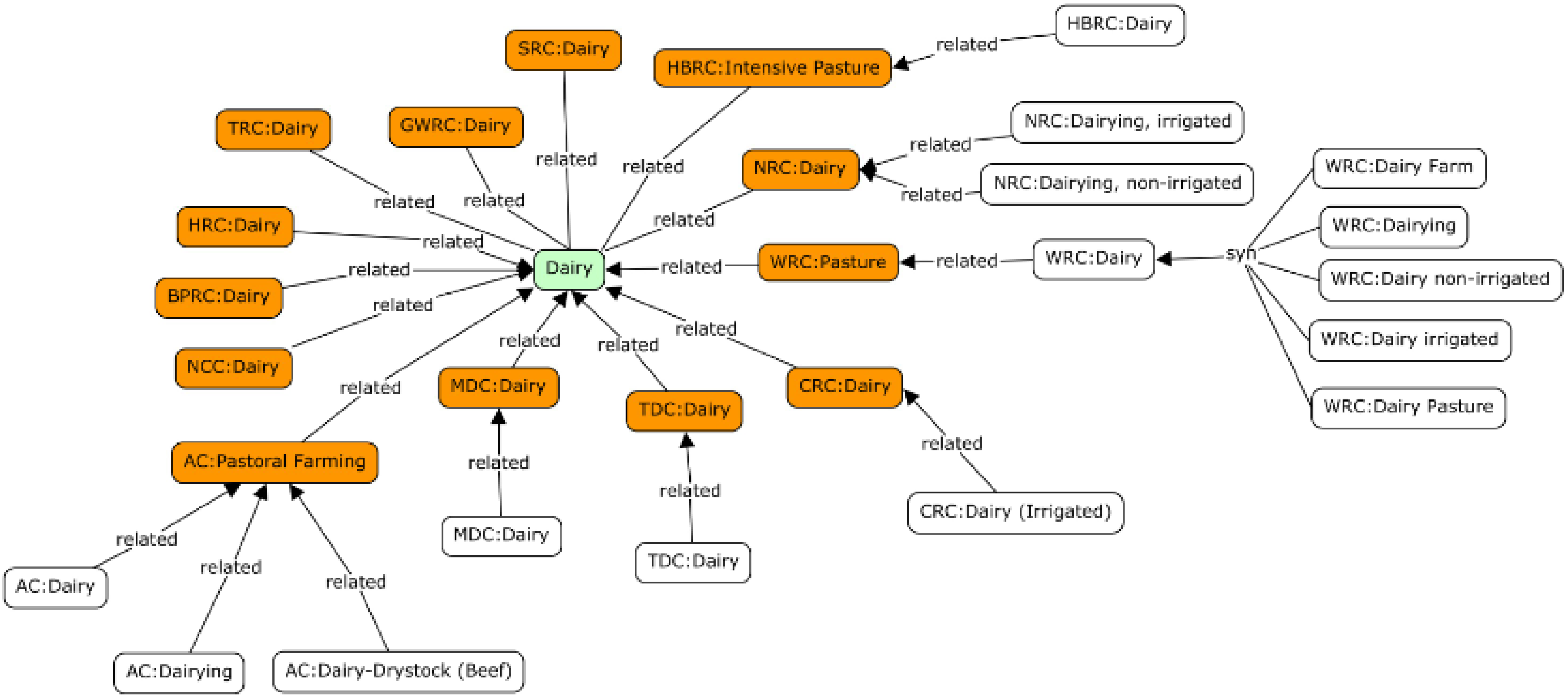


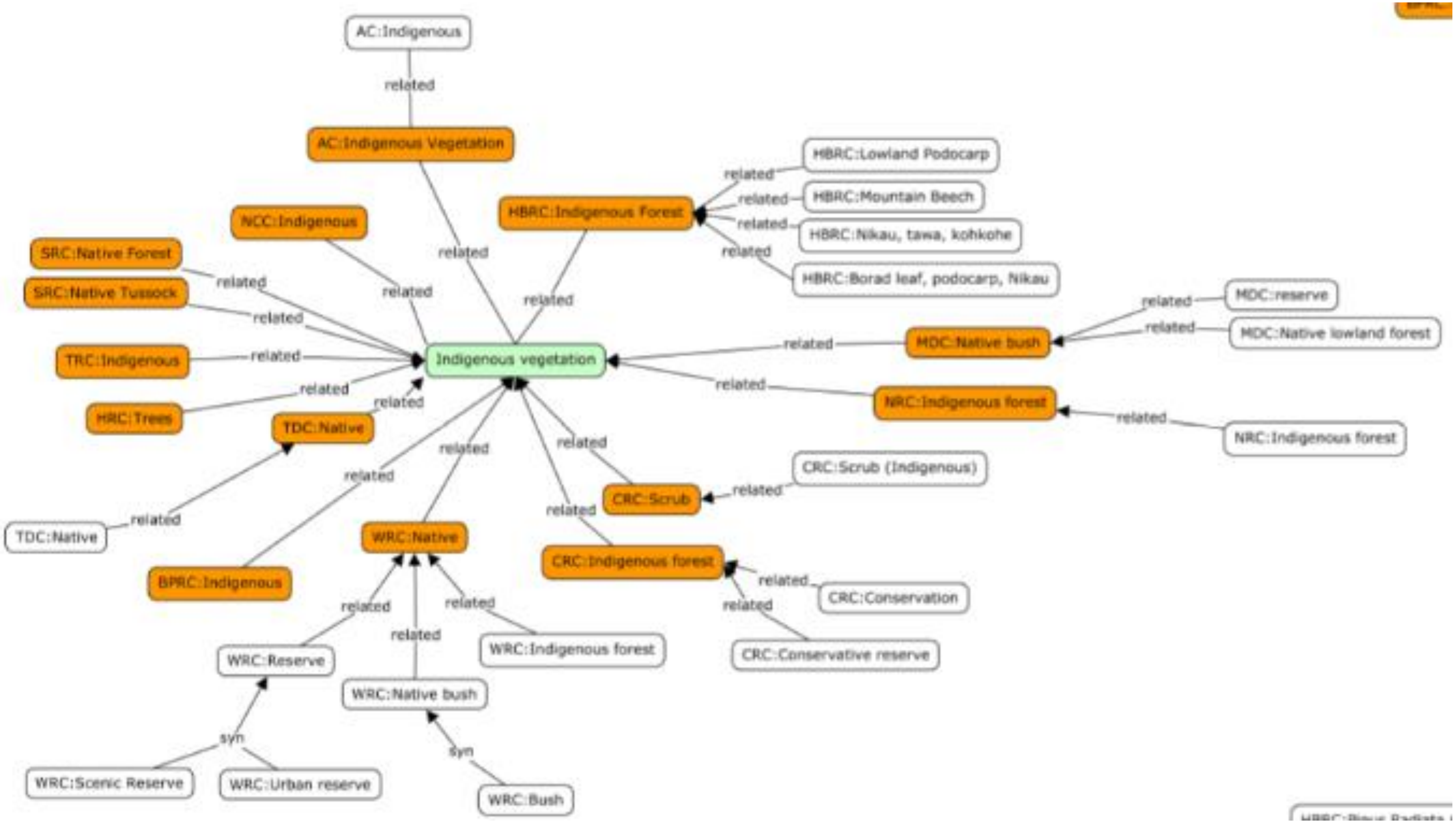
Short-rotation cropping
HRC:Market Garden
GWRC:Market Garden
GWRC:Cropping
TDC:vegetables
TDC:lucerne
TDC:Market garden
TDC:Crop
BPRC:Arable
BPRC:Maize
CRC:Arable
AC:Outdoor vegetable production
AC:Horticulture
NRC:Cropping
NRC:Cropping
MDC:Mixed Cropping
MDC:Cropping
TRC:Crop/Hort
HBRC:Squash and corgette
HBRC:Squash
HBRC:Peas
HBRC:Cropping
HBRC:Cropping (peas)
HBRC:Pasture
HBRC:Maize
HBRC:Sweetcorn or maize
HBRC:Lucerne
HBRC:Onions
HBRC:Rye grass
HBRC:Cropping (organic last 10yrs, pasture prior)
HBRC:Cropping (long term - Tomatoes this year)
HBRC:Cropping (long term - Sweet corn this year
HBRC:Cropping (long term - peas)
WRC:Pasture
cut & carry deer
cut & carry pasture
WRC:Arable
MDC:Cropping/drystock
Arable maize for silage, fodder
cropping - maize
Market gardening
cropping - silage maize
Cropping potatoes/onions/oats
Cropping - potatoes
Cropping
cut & carry maize
related



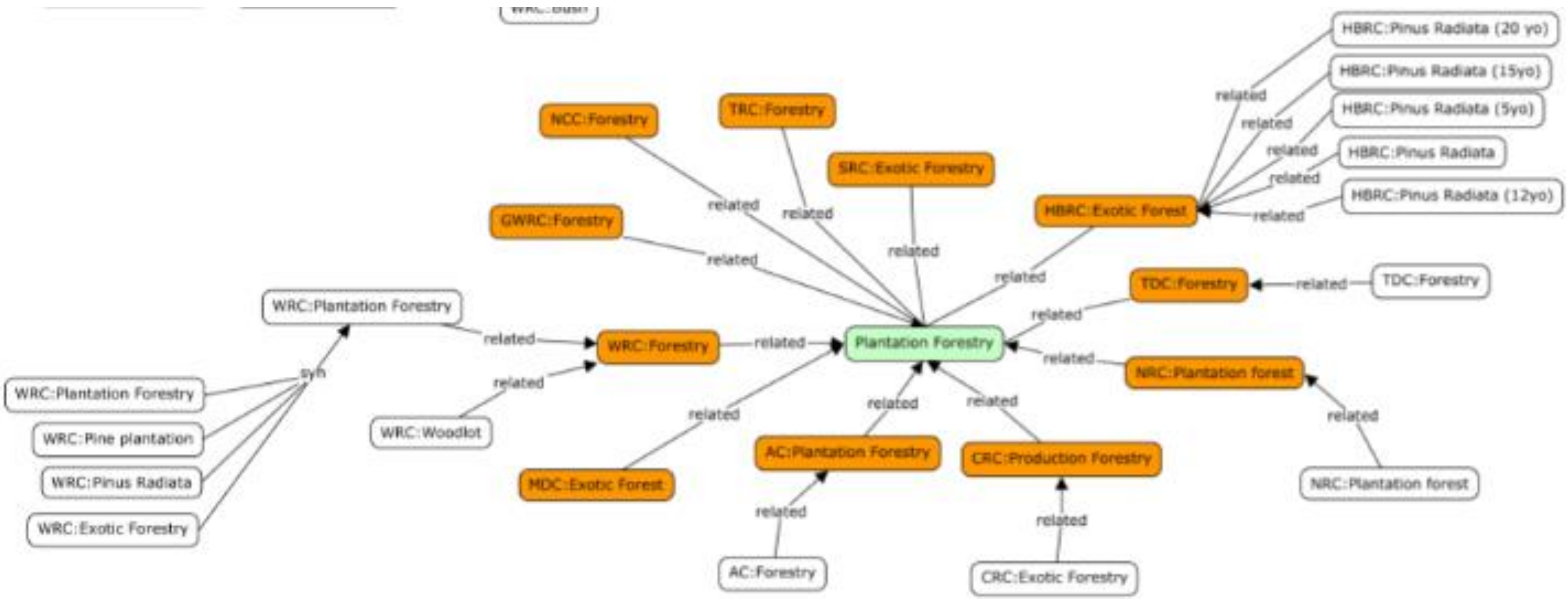
Perennial cropping
GWRC:Horticulture
NRC:Orchard
NRC:Orchard
AC:Orchard
AC:Vineyard
AC:Horticulture
BPRC:Kiwifruit
BPRC:Horticulture
WRC:Horticulture
WRC:Orchard
WRC:Kiwi fruit
TDC:Orchard
TDC:kiwifruit
TDC:apples
HBRC:Orchard
HBRC:Kiwi fruit
HBRC:Pip Fruit (Cherry)
HBRC:Vineyard
HBRC:Grapes
HBRC:Vineyard
MDC:Vineyard
MDC:vine
MDC:inter-row
MDC:wheel
related


WRC:Sheep and beef
WRC:Beef
CRC:Deer (Irrigated)
TDC:sheep
TDC:beef
TDC:sheep and beef
NCC:Drystock
WRC:Pasture
HRC:Cropping/drystock
GWRC:Drystock
CRC:Deer
BPRC:Drystock
BPRC:Deer
TDC:Drystock
MDC:Sheep
MDC:Sheep/beef
MDC:Dryland Pasture
TRC:Drystock
Drystock
CRC:Extensive Grazing
CRC:Semi-intensive grazing
CRC:Sheep/beef
CRC:Sheep/beef (Irrigated)
CRC:Sheep beef (Dryland)
SRC:Intensive Drystock
SRC:Extensive Drystock
SRC:Mixed crop/ Drystock
HBRC:Extensive Pasture
NRC:Drystock
HBRC:Intensive Pasture
HBRC:Deer
AC:Pastoral Farming
HBRC:Currently under maize
NRC:Drystock
HBRC:Sheep (or Beef)
AC:Drystock
HBRC:Bull beef
HBRC:Drystock
related

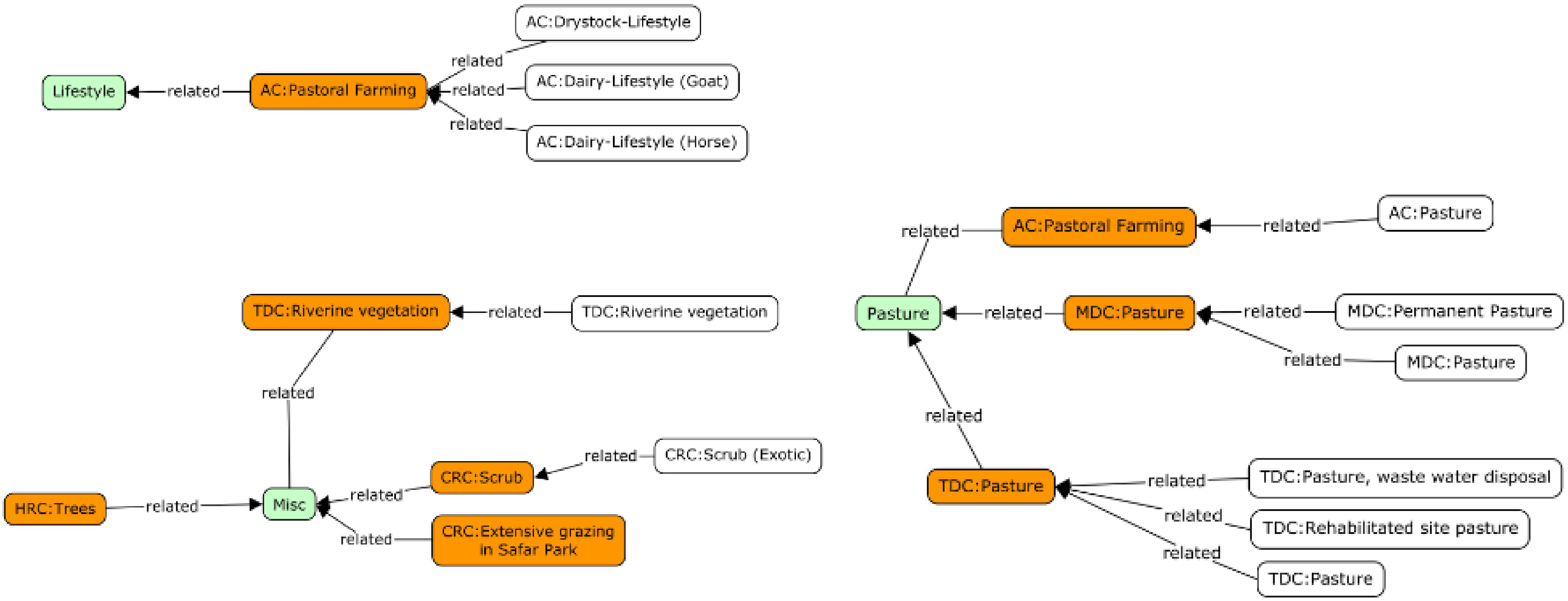

SRC:Dairy
HBRC:Intensive Pasture
HBRC:Dairy
TRC:Dairy
GWRC:Dairy
NRC:Dairying, irrigated
NRC:Dairy
NRC:Dairying, non-irrigated
WRC:Dairy Farm
HRC:Dairy
WRC:Dairying
Dairy
WRC:Pasture
WRC:Dairy
syn
BPRC:Dairy
WRC:Dairy non-irrigated
NCC:Dairy
WRC:Dairy irrigated
MDC:Dairy
CRC:Dairy
TDC:Dairy
WRC:Dairy Pasture
AC:Pastoral Farming
CRC:Dairy (Irrigated)
MDC:Dairy
TDC:Dairy
AC:Dairy
AC:Dairying
AC:Dairy-Drystock (Beef)
related



AC:Indigenous
AC:Indigenous Vegetation
HBRC:Indigenous Forest
HBRC:Lowland Podocarp
HBRC:Mountain Beech
HBRC:Nikau, tawa, kohkohe
HBRC:Borad leaf, podocarp, Nikau
NCC:Indigenous
SRC:Native Forest
SRC:Native Tussock
TRC:Indigenous
HRC:Trees
TDC:Native
TDC:Native
BPRC:Indigenous
Indigenous vegetation
MDC:reserve
MDC:Native bush
MDC:Native lowland forest
NRC:Indigenous forest
NRC:Indigenous forest
CRC:Scrub (Indigenous)
CRC:Scrub
CRC:Indigenous forest
CRC:Conservation
CRC:Conservative reserve
WRC:Native
WRC:Reserve
WRC:Native bush
WRC:Indigenous forest
WRC:Scenic Reserve
WRC:Urban reserve
WRC:Bush
related
syn



Plantation Forestry
NCC:Forestry
TRC:Forestry
SRC:Exotic Forestry
GWRC:Forestry
HBRC:Exotic Forest
HBRC:Pinus Radiata (20 yo)
HBRC:Pinus Radiata (15yo)
HBRC:Pinus Radiata (5yo)
HBRC:Pinus Radiata
HBRC:Pinus Radiata (12yo)
TDC:Forestry
TDC:Forestry
WRC:Plantation Forestry
WRC:Forestry
NRC:Plantation forest
NRC:Plantation forest
WRC:Plantation Forestry
WRC:Pine plantation
WRC:Pinus Radiata
WRC:Exotic Forestry
WRC:Woodlot
MDC:Exotic Forest
AC:Plantation Forestry
CRC:Production Forestry
AC:Forestry
CRC:Exotic Forestry
related
syn


Lifestyle
related
AC:Pastoral Farming
related
AC:Drystock-Lifestyle
related
AC:Dairy-Lifestyle (Goat)
related
AC:Dairy-Lifestyle (Horse)
TDC:Riverine vegetation
related
TDC:Riverine vegetation
related
HRC:Trees
related
Misc
related
CRC:Scrub
related
CRC:Scrub (Exotic)
related
CRC:Extensive grazing in Safar Park
related
AC:Pastoral Farming
related
AC:Pasture
Pasture
related
MDC:Pasture
related
MDC:Permanent Pasture
related
MDC:Pasture
related
TDC:Pasture
related
TDC:Pasture, waste water disposal
related
TDC:Rehabilitated site pasture
related
TDC:Pasture