LANDCARE RESEARCH
MANAAKI WHENUA

National Environmental Monitoring Site Identification System

Envirolink Grant: 1729-HZLC137

National Environmental Monitoring Site Identification System

Alistair Ritchie and Juliana Osorio-Jaramillo

Landcare Research

Prepared for:

Horizons Regional Council

11-15 Victoria Avenue
Private Bag 11025, Manawatu Mail Centre
Palmerston North 4442
New Zealand

July 2017

Landcare Research, Gerald Street, PO Box 69040, Lincoln 7640, New Zealand,
Ph +64 3 321 9999, Fax +64 3 321 9998, www.landcareresearch.co.nz

Reviewed by:	Approved for release by:
David Medyckyj-Scott Technical Director - NLRC / RPAL / CL Landcare Research	Alison Collins Portfolio Leader – Characterising Land Resources Landcare Research

Landcare Research Contract Report:	LC2814

Disclaimer

This report has been prepared by Landcare Research for Horizons Regional Council. If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

Contents

Summary

Background

- There is a need for a nationally consistent convention for identifying environmental monitoring sites, regardless of their target domain or purpose.
- Unambiguous, globally unique identifiers are essential to ensure data for monitoring sites can be shared and analysed in multiple contexts and systems with confidence that the data for the same location and phenomena are being compared.

Objectives

- Conduct a workshop with a small but representative group of regional councils and Crown Research Institutes to establish the basic requirements for a national environmental monitoring identifier scheme.
- Compile a literature review of identifier conventions that have been previously used in New Zealand or are in use elsewhere.
- Propose a candidate convention for site identifiers and discuss any supporting systems and procedures that may be required.

Identifier Types and Conventions

- As digital systems are essential to the storage and publication of environmental monitoring data, the identifier convention shall be appropriate to both digital and written representation.
- Due to the distributed nature of the data, both of the location of the identified features and their stored records, identifiers need to last indefinitely and be globally unique. It may be useful for identifiers to carry some meaning, but not at the expense of persistence and uniqueness.
- Identifiers must be governed by clear policies for their creation and by an information system that ensures they are unique and can be used to recover data about a monitoring site.

Conclusions

- The workshop and literature review revealed a number of name or identifier conventions that were applicable to the problem. The World Meteorological Organisation's Integrated Global Observation System's station identifier convention can be used with only minor changes.
- A site can be defined as a Sampling Feature according to ISO19156:2011 (Observations and Measurements).

- The importance of governance cannot be understated and it is clear that any convention must be overseen by an authority representing the community of use. This will be greatly aided by information systems that manage the allocation of identifiers.
- It is advantageous to also provide information systems that support the discovery of the identified sites. Furthermore, the workshops showed the need to capture and provide important metadata about sites – these systems should therefore assist in the capture and maintenance of these data.

Recommendations

- The World Meteorological Organisation's Integrated Global Observation System's station identifier convention should be used as the basis for a national site identifier convention.
- A National Environmental Monitoring Standards committee should be established as a naming authority to oversee the allocation of identifiers.
- The naming authority should be supported by national monitoring site register – initially a database to help manage identifier values, but ultimately a digital register allowing discover of monitoring sites based on their identifiers and important metadata.

1 Introduction

New Zealand's environment is changing in response to many human and natural forces and to understand, manage, and adapt to these changes all parties involved must have access to data that describe the changing state of the environment. Increasingly, these data are being generated in a distributed monitoring network comprising sites that lie within different administrative areas and are operated by different parties, and are managed in a variety of information systems. Data about the same monitoring site, or monitored entity, may, however, be stored in more than one of these systems, meaning data must be aggregated from various sources to ensure a comprehensive set of data are created. This aggregation is compromised by ambiguous names and identifiers – how can an analyst be sure they are comparing data for the same thing, either in space or over time without an unambiguous linking identifier?

This is a long-standing problem with its roots in the pre-digital era, so conventions for names and identifiers have been established in the past. They do, however, vary as typically they are specific to an environmental domain or sampling activity. As consequence a convention for one domain, e.g. fresh-water hydrology, cannot be more broadly used elsewhere, e.g. in coastal marine environments. Furthermore, some domains have not established globally useful, or formally defined, conventions for their monitoring activities, making it very difficult to integrate data into larger monitoring datasets. This is the problem faced by agencies responsible for the monitoring and management of New Zealand's environment as they establish the shared systems that support New Zealand's Environmental Monitoring and Reporting.

This report has been commissioned to begin a process to establish a new nationally consistent convention for identifying monitoring sites, regardless of their target domain or purpose. It is the result of a workshop and interviews with staff from representative regional councils and Crown Research Institutes, and a subsequent review of previous and existing identifier conventions. It proposes the adoption of a convention based on the station identifier scheme for the World Meteorological Association's Integrated Global Observing System as it defines a pattern to create terse but enduring identifiers that can be readily used in documents and digital information systems.

No long-lived identifier convention is sound without well-defined, and enforced, policies for their use. We therefore recommend that the National Environmental Monitoring Standards steering group form a committee to oversee the governance, implementation, and operation of the site identification system. We also recommend the provision of a web-enabled national monitoring site register to support the creation of identifiers and the discovery of the identified sites. With a committee in place, and appropriate funding for the indefinite administration and maintenance of the convention and registry, a national convention for site identifiers is very possible with the associated significant benefits for New Zealand.

2 Assumptions

1. The identifiers discussed here identify physical entities, for example, a river monitoring site – not the stored data objects that describe them, for example, a PDF document or database record.

2. Because modern site data management systems and publication methods (e.g. lawa.org) are near-ubiquitously digital this document will consider the issue of identity primarily, but not exclusively, in the context of a distributed information system.

3. The identifiers discussed here are for a shared system – they are not expected to replace system or database identifiers used by data providers, these will continue to be defined and maintained independently according to the needs of the data provider and the technology they use.

3 Background

New Zealand's environment is changing in response to many human and natural forces, and to understand, manage, and adapt to these changes all parties involved must have access to datasets that describe the past, present, and future state of the environment. This Envirolink Medium Advice Grant was provided in the context of the environmental monitoring and sampling activities of New Zealand's regional councils and unitary authorities, stake-holding territorial authorities, and central government agencies. The three most relevant projects and systems are:

1. Environmental Monitoring and Reporting

   The Environmental Monitoring and Reporting (EMaR) Project is a collaboration between the Ministry for Environment (MfE) and regional councils. It involves exploring the standardisation of methods and sharing of data collection, management and exchange protocols to allow national scale interpretation of regional data. The end goal of the EMaR Project is to have environmental data collected by regional councils more widely available through Land, Air, Water Aotearoa.

2. National Environmental Monitoring Standards

   The National Environmental Monitoring Standards (NEMS) are a series of environmental monitoring standards, prepared by the NEMS steering group on authority from the Regional Chief Executive Officers (RCEOs) and MfE. The NEMS initiative is led and supported by the Environmental Data Special Interest Group (formerly known as LAEMG) to help ensure consistency in the application of work practices specific to environmental monitoring and data acquisition throughout New Zealand.

3. Land, Air, Water Aotearoa

   Land, Air, Water Aotearoa (LAWA) has been established by like-minded organisations with a view to helping local communities find the balance between using natural resources and maintaining their quality and availability. Initially, a collaboration between New Zealand's 16 regional councils and unitary authorities, LAWA is now a partnership between the councils, Cawthron Institute, the Ministry for the Environment, and Massey University and has been supported by the Tindall Foundation.

As these projects utilise a nationally distributed monitoring network comprising sites that lie within different administrative areas and are operated by different parties, the resulting data are managed in a variety of information systems. The use of the monitoring data generated at these sites, particularly by central government agencies, will likely cross all geographic, jurisdictional, operational, and system boundaries. The monitoring network is long-lived and stations have been operating, and are intended to operate, for time-periods that may exceed the lifetimes of projects, systems or agencies. It is, therefore, essential that all monitoring sites, environmental phenomena, monitoring methods, agencies, and other entities are clearly identified in such a way that supports unambiguous exchange and comparison of data in this national system. While the intent is to support exchange among government agencies, these identifiers will be of value when sharing data with other contractors and collaborators (for example, Crown Research Institutes, Universities, utilities providers and commercial contractors).

Well-defined and authoritative names for monitoring sites are essential, especially as the data captured are managed in digital information systems. Without them, it is difficult to query systems to find monitoring sites, and to establish environmental state and trends if observations cannot be confidently associated with a site. Also, the comparison of data held across systems or agencies is compromised by ambiguous names and identifiers. Horizons Regional council proposes to establish a working group to develop a convention as part of the National Environmental Monitoring Standards.

4 Requirements

4.1 Workshop

Landcare Research hosted a requirements gathering workshop attended by representatives of Horizons Regional Council, Auckland Council and NIWA at Palmerston North on 14 February 2017. The workshop aimed to:

- establish the scope of the literature review and final report
- define the requirements for any identifier convention
- identify existing conventions to be included in the review.

The conventions discussed at the workshop are reviewed in section 5.2. It was agreed that the scope of the identifier convention should be regional council environmental monitoring sites and the domain features being monitored. Initial work would focus on identifying sites, but any convention adopted should be readily extended to identify environmental phenomena (for example, rivers or animal species) and other artefacts of the monitoring activity (for example, property and method definitions).

When reviewing these requirements for the identifiers it was clear that expectations were two-fold, requiring the provision of:

1. a globally unique identifier that could be used in reports and digital information system
2. a set of basic meta-data allowing monitoring sites to be indexed and discovered. Broadly speaking, this should be based on what was being monitored, where and when.

Furthermore, it was clear that the identifiers were essentially a proxy for items that needed to be entered into a national monitoring site register. As a result, identifiers were to be unique in this context and the index meta-data represented the minimum data standards for entry in such a registry. Any proposed solution, therefore, has to consider such a registry, and associated aspects of its governance and implementation.

Participants agreed to the need for:

1. a defined convention for identifiers
2. a national register of monitoring sites and their features of interest managed in some form of authoritative on-line registry
3. a set of minimum metadata standards for registered entities (Table 1).

Table 1 A draft minimum set of meta-data attributes for a sampling site as defined during the workshop

Attribute	Description
identifier	A formal, persistent identifier for the site based on the convention adapted following this body of work.
location	A three-dimensional point locating the site. This is a structured value including the location's spatial reference system.
shape	A polygon or volume geometry describing the actual extent of sampling (for example, the cross-sectional profile of a combination weir stream gauging station). This is a structured value including the geometry's spatial reference system.
time	The time period (or periods, if sampling is periodically interrupted) over which measurements at the site were made.
label	A human-friendly label for a site, for example 'Manawatū River at Teacher's College'.
interested party	Any party with an interest in the station; this should include the role the party played ('maintenance', 'data consumer', 'land owner') and when.
purpose/activity	The purpose of the station ('stream gauge', 'vegetation plot', 'soil quality monitoring'), from a formally defined and governed vocabulary.
life cycle	The history of the station (when commissioned/decommissioned) and therefore current state (active/inactive).
description	An explanatory free text description of the site.
same as	A list of identifiers or names that have been assigned to the station by other parties or under earlier naming conventions.

4.2 Interviews

Following the workshop, participants were interviewed to understand the history of use of site identifiers in their organisation and their expectations for any new convention. Across the different institutions, various naming conventions have been used for an extended period of time. Experience has shown both the advantages of their use, but also the presence of issues to solve. Standardised naming conventions are seen to be important to help identify sites through space and time. They open the possibility to easily combine environmental observations of the same site.

Interviews were conducted using a questionnaire with a subsequent one-to-one phone call. The results are summarised below.

Types of naming conventions New Zealand institutions are using include:

- Internet URNs (Uniform Resource Names)
- A derivative of LUCAS system for terrestrial biodiversity plots
- Incremental numbers for marine sites and streams
- Alphanumeric conventions.

Requirements for a new site identifier convention:

- One location equals one identifier – regardless of multiple measurements and features there is still only one location that requires its own identifier
- It is considered as useful if the identifiers are meaningful, for example, based on location; however, it is accepted that this could be challenging as it is complex to define a convention for sites of different types and also because locations could change over time
- A standard method with a nationwide adoption.

Problems with the current naming conventions include:

- In some institutions they depend on the type of observation, this means that each type of observation has its own standard to define a site's name. This approach creates synonyms for the same site that cannot be easily managed without an information system
- Some conventions depend on the sampled domain, and the sampling strategy and intensity
- In location-based conventions 'Z' coordinates for meters and sensors are not taken into account
- The IDs for all the naming conventions are human-generated; there are no systems to help with the governance.

Issues for consideration:

- If a new convention is implemented, how do systems trace the history of the previous names and link them to the new names?
- Not only sites should have a naming convention. There should be naming conventions for samples, measurements, parameters and other artefacts of monitoring systems.
- How are sub-sites or sub-divisions of a site named?
- How does a naming convention scale to support national systems?

Useful, but non-essential, features of any new convention:

- Meaningful identifiers for sites
- Capacity to infer a hierarchy of sites from the identifier
- An information system to support governance of the naming convention.

Overall, it was seen to be important that the naming convention to be implemented should have the capacity to be used with any type of observation or measurement. It was also important to define which metadata should accompany the identifier when it is included in a register of monitoring sites. Finally, a proper governance regime is essential to ensure the identifier convention is used properly (each site is identified once, and each identifier is unique) and to define how the historic systems can be linked into any new system.

5 Review of identifier conventions

5.1 Criteria for review

5.1.1 What is being identified

The ISO19101 General Feature Model [1] defines the concept of the feature and these are useful when discussing what is to be identified. Features can be organised into feature types ('classes of features having common characteristics') that describe physical things (e.g. rivers, bridges) or abstract notions (e.g. terms in a vocabulary). Feature types have properties that can provide useful tokens in an identifier (based on aspects such as their type or defining characteristics).

For a monitoring system, the ISO19156 Observations and Measurements [2] specification defines a useful set of feature types by which to organise the things under consideration.

- *Domain Features* define the features being monitored – the 'features of interest' – such as a river reach, aquifer or soil body.
- *Sampling Features* define the location and the extent of sampling activities (for example a river monitoring site or borehole) aiming to provide representative measurements of properties of the domain features. Sampling features can be organised into sampling feature complexes, where a parent site may be a collection of monitoring platforms; perhaps be sampled and sub-sampled for off-site analysis; or may be otherwise subdivided according to a domain-specific sampling strategy. How these subdivisions are described can vary widely and is deliberately deferred to the domain.
- *Processes* define the methods used to measure values of the properties of interest - for example, an instrument mounted on a platform located by the sampling feature.
- *Observations* are the results, with an optional estimate of uncertainty, of these measurements as generated by a process at a particular time.

The definition of domain feature types and the ways of measuring their properties requires input from experts in the domain. This document will focus on the core, reusable components of a monitoring system, the sampling features – specifically a monitoring site as a type of sampling feature. A monitoring site may take a variety of forms as appropriate to the investigation. They may be a physical platform on which one or more sensors are mounted, a fixed position that is a base from which observations are made (e.g. a bird count), or may be co-located with the domain feature of interest in that the extent of the site may be that of the feature (e.g. a wetland site covers an entire wetland system). As a result, in ISO19156 the spatial representation of a sampling feature is therefore a shape that may be a point, line, polygon or solid according to the true extent of sampling (a simple location parameter can be defined to provide a convenient point for display on maps).

5.1.2 Types of identifiers

NEMS defines an identifier as a 'unique label for a site, which may be numeric or alphanumeric'. Identifiers may be very simple, intended only to be unique in a limited context, or more complex, aiming for cross-system uniqueness and possibly the carriage of some meaningful information within the label. Identifiers can be typed according to the extent of the uniqueness and how they are structured – the result can be due to a combination of these aspects.

Uniqueness

Document or Table Identifier: an identifier that is guaranteed to be unique within the document (for example a heading or database primary key) but likely to be repeated in other contexts. These are of no practical value in a distributed system. Example: an auto-incrementing integer key in a database table.

System Identifier: an identifier that is guaranteed to be unique across a single system or network (for example a database or registry). It requires a central registration process. Example: DSIR National Soils Database 'SB' numbers (e.g. 'SB09877').

Universal (or Global) Identifier: an identifier that has such a low probability of being duplicated and causing a conflict that it is essentially unique and therefore does not require a central registration process. Example: the 128-bit Universally Unique Identifier (UUID) (e.g. 'bddcbb3a-20d6-442f-a574-0284159075b8').

Structure

Opaque: these identifiers focus on uniqueness and do not provide any obvious information about the feature being identified. Sometimes, meaning is obliquely conveyed due to the convention used to create the identifier, but this is not intentional, and should not be considered a reliable source of information about the feature.

Natural: natural identifiers are constructed by aggregating property values to create a unique value, for example, a combined project code, location, and date, or a group of integer keys. This makes for robust system identifiers (e.g. database primary keys) and if simple enough can be memorable, but they will change as data change. In distributed systems that cannot ensure a cascading update to all other values that use the ID as a reference, a failure to update the references will break the links the define the network.

Semantic Identifiers: a variation on natural identifiers in that the aim to be 'human readable' through facets that carry meaning. Appropriate construction can allow for the creation of good universal IDs, ensure they are memorable, and make them easy to work with (for example identifiers can be guessed).

Duration

Transient: transient, or local identifiers, are not designed or governed in a way that guarantees longevity, or uniqueness beyond the scope of their use. There may be changes at any time, and the systems that use them are small enough that updating every instance of an identifier is possible.

Persistent: the concept of persistent identifiers has been around for a long time, but has come to prominence with the advent of the internet. In a distributed system, nodes (e.g. a web page) may be referred to by any other node (web page) in the system without being aware of the reference, it will therefore unable to update references if the value changes. Persistent identifiers address this by never changing – aiming to endure indefinitely – and therefore protecting against broken references or comparisons.

As the monitoring system is distributed and long-lasting the following identifier types will be most useful:

- System or Universal
- Natural or Semantic
- Persistent

5.1.3 Minimum requirements for identifiers

RFC1737 [3], the Internet Engineering Taskforce's minimum set of requirements for Uniform Resource Names (URNs) provides a specification for identifiers that can be applied to a monitoring site network and its environmental features of interest. Both the internet and the monitoring network are a distributed system comprising entities that may be managed or referred to by any number of agencies. Also, a primary use of the identifiers will be in digital information systems. The following requirements are taken wholesale from RFC1737 with only minor changes for clarity (for example references to 'URNs' have been replaced with 'identifiers').

Requirements for functional capabilities of identifiers:

- Global scope: An identifier is a name with global scope. It has the same meaning everywhere.
- Global uniqueness: The same identifier will never be assigned to two different things.
- Persistence: The lifetime of an identifier shall be permanent. That is, the identifier will be globally unique forever, and may well be used as a reference to things well beyond the lifetime of the thing it identifies or of any naming authority involved in the assignment of its name.
- Scalability: identifiers can be assigned to any entity that might conceivably be accessible, for hundreds of years.
- Legacy support: The identifier scheme must permit the support of existing legacy naming systems, insofar as they satisfy the other requirements described here. For

example, LSID identifiers, IGSN numbers, and WMO site identifiers seem to satisfy the functional requirements and allow an embedding that satisfies the syntactic requirements described here.

- Extensibility: Any scheme for identifiers must permit future extensions to the scheme.
- Independence: It is solely the responsibility of an identifier-issuing authority to determine the conditions under which it will issue an identifier.
- Resolution: An identifier will not impede resolution (translation into a URL or query string). To be more specific, for identifiers that have corresponding Uniform Resource Identifiers (URIs), there must be some feasible mechanism to translate an identifier to a URI.

Requirements for identifier encoding – in addition to requirements on the functional elements of identifiers, there are requirements for how they are encoded in a string:

- Single encoding: The encoding for presentation for people in clear text, electronic mail and the like is the same as the encoding in other transmissions.
- Simple comparison: A comparison algorithm for identifiers is simple, local, and deterministic. That is, there is a single algorithm for comparing two identifiers that does not require contacting any external agent.
- Human transcription: For identifiers to be easily transcribed by humans without error, they should be short, use a minimum of special characters, and be case insensitive. (There is no strong requirement that it be easy for a human to generate or interpret an identifier; explicit human-accessible semantics of identifiers is beneficial but not essential.)
- Transport friendliness: An identifier can be transported unmodified in the common Internet protocols, such as HTTP, TCP, SMTP, FTP, etc., as well as printed paper.
- Machine consumption: An identifier can be parsed by a computer.
- Text recognition: The encoding of an identifier should enhance the ability to find and parse identifiers in free text.

5.1.4 Authority and governance

System or global identifiers are assigned in a particular context and must be managed to ensure stability and accuracy. There must be a known governing body, or authority, that assigns identifiers. This authority will define the policies that govern the creation of an individual identify and ensure that an identifier is not misused, primarily by ensuring they are not duplicated within the system. They will also be responsible for matters of use: defining any restrictions on who may use the system; what – if any – licenses users are subject to; and, ideally, a system for managing and registering identifiers.

5.2 Review of identifier and naming conventions

The following conventions were identified by participants at the requirements gathering workshop and during background searches for information. The list is by no means exhaustive – identifiers are found in any organised system – and these were selected because of known relevance, and because they might offer possible patterns for implementation, or represent a novel approach. Some, such as the conventions defined by the Soil Conservation and Rivers Control Council, were considered but omitted from the review as they predated the widespread use of digital information systems. The conventions are described and evaluated according to criteria developed from section 5.1:

- The *origin* refers to the body or bodies that defined the convention
- The *domain* covers what discipline or type of features the convention was applied to
- The *type* of identifier refers to those defined in section 5.1.2
- If possible, a breakdown of the *structure* of the identifier is provided, along with an *example*
- The *users* are the agencies or community that used the convention. This criterion considers the period of use or, if that is unknown, whether it is in use or defunct
- *Digital* describes whether or not the convention is explicitly defined to be used in digital systems
- *Governance* describes who is responsible for the convention and assigning identifiers and if there is a documented creation policy
- *Scalability* addresses how many features can be identified and whether there are restrictions on the type of features that can be identified
- *Restrictions* describe any caveats on use, for example, licenses or availability to a limited community of users
- *Suitability* is based on the value of the convention to a national environmental monitoring network according to the preceding criteria.

5.2.1 National systems

Australian National Environmental Information (NEII) Monitoring Sites Register

Origin	Bureau of Meteorology (Australia) in collaboration in members of the NEII Reference Group.	
Domain	All environmental domains	
Type	Not applicable – at present the register does not mandate a formal site identifier convention.	
Description	*'The National Environmental Monitoring Sites Register (NEMSR) provides a consolidated view of Australia's environmental monitoring sites. It brings together a diverse range of networks across environmental domains, including seismic monitoring stations, ocean radars, long-term weather observation sites, flux stations, and ground cover reference sites. Other nationally-important networks are coming soon.* *'Users can select one or more networks and display them through the data viewer. Clicking on a site shows its location, site metadata and provides a link to the data itself. NEMSR can also be accessed programmatically through [service] endpoints [...].' [4]*	
Users	Government; Research Institutes; Universities	In use
Digital	Not applicable	Not applicable
Governance	NEII Reference Group	Not applicable
Scalability	Not applicable	
Restrictions	Australian datasets	
Suitability	None. Does not define an identifier convention; is intended for Australian data. The concept of a centrally governed register to manage and publish information is applicable and could be applied in New Zealand.	

USGS National Water Information System

Origin	United States Geological Survey	
Domain	Hydrology	
Type	System; Semantic (but intended to be opaque); Persistent	
Description	*'The USGS creates a unique station (site) number to distinguish a site from other locations where hydrological data is collected. [...]* ***Surface water sites*** *Stations are assigned numbers according to a downstream order system. In assigning station numbers in downstream order, no distinction is made between partial-record stations and other stations; therefore, the station number for a partial-record station indicates downstream-order position in a list made up of both types of stations. Gaps are left in the series of numbers to allow for new stations that may be established; hence, the numbers are not consecutive. An eight-digit number for a station, such as 03171000 includes the two-digit part number "03" plus the six-digit downstream-order number "171000." The part number is derived from Water Supply Paper (WSP) publication volumes or "Parts" and designates the major river basin; for example, part "03" is the Ohio, Cumberland, and Tennessee River Basins. In areas of high station density, more than eight digits may be used for a station number.* ***Wells and miscellaneous sites*** *The USGS well system is based on the grid system of latitude and longitude. The system provides the geographic location of the well and a unique number for each site. The number consists of 15 digits. The first 6 digits denote the degrees, minutes, and seconds of latitude, and the next 7 digits denote degrees, minutes, and seconds of longitude; the last 2 digits are a sequential number for wells within a 1-second grid. In the event that the latitude-longitude coordinates for a well are the same, a sequential number such as "01," "02," and so forth, would be assigned as one would for wells. This system may also be used for random water-quality miscellaneous sites. (A local well-numbering system is also used. It is a 2-part identifier, consisting of the abbreviation of county name and the serial number within the county.)* ***It cannot be emphasized too strongly that the site ID, once assigned, is used as a pure number and has no locational significance beyond representing the best location available at the time the site ID was assigned.*** *The latitude and longitude fields for the site should be used for the site's location.' [5]*	
Users	USGS	In use
Digital	Yes	Human and machine readable
Governance	None	Creation: based on convention (no associated system or service)
Scalability	Restricted to hydrological features. Has mechanisms to accommodate increasing numbers of sites.	
Restrictions	None.	
Suitability	None. Is a local system and uses conventions that are tied to geographic entities that are not in New Zealand.	

Landcare Research National Vegetation Survey Dataset Identifiers

Origin	Landcare Research	
Domain	Biology	
Type	System; Semantic; Persistent or Transient	
Structure/Example	`{region}-{sub-region}-{special-feature}-{year}`	`MOKAU RIVER-KAHIKATEA BEND-EXCLOSURE-1993`
Description	*The Landcare Research hosted National Vegetation Survey Databank is considering the definition of a naming convention for datasets, over and above the use of system GUIDs, DOIs, and other identifiers. It has been proposed to develop a convention for a delimited string based combining spatial extent, a description of a special feature of the data, the submitting organisation and the year. [6]*	
Users	Government, Research Institutes, Universities, Private Consultants	Proposed
Digital	Yes	Intended to be human and machine readable
Governance	None	Creation: not applicable
Scalability	To be determined	
Restrictions	To be determined	
Suitability	None. Not a formal standard.	

Land Use and Carbon Analysis System (LUCAS) Plot Network

Origin	Ministry for the Environment Land Use and Carbon Analysis System (LUCAS)	
Domain	Biology, Carbon	
Type	Opaque	
Structure/Example	String derived from a position on a x (east), y (north) 8km grid. The x axis is defined as an alphabetic sequence increasing to the east (A, B, ..., Z, AA, AB, ..., AZ, etc.) and a numeric sequence (1, 2, 3 ...) ascending to the south.	`Y150` `AB150`
Description	*'LUCAS is a national grid of permanent plots that systematically sample existing (pre-1990) natural forest. This plot network was specifically designed to monitor national biomass stocks and biomass stock change [...] and is based on 0.04-ha plots (20 × 20 m land surface area) located on an 8-km grid across New Zealand. The initial point (that is, the origin) was selected at random, and then the rest of the grid was derived in relation to that point, resulting in a set of evenly spaced randomly located plots. The 0.04 ha plot size was chosen to integrate with existing field protocols and vegetation survey data across New Zealand [...]' [7]*	
Users	Government, Research Institutes	2005 to present day
Digital	Yes	Human and machine readable
Governance	Ministry for Environment	Creation: based on convention (no associated system or service)
Scalability	2-dimensional geographic position only	
Restrictions	None	
Suitability	Limited. Specific to a location; cannot support identification of different sites within the same 8 km × 8 km area.	

DSIR Aquacodes

Origin	Department of Scientific and Industrial Research	
Domain	Hydrology	
Type	Local; Semantic; Persistent	
Structure/Example	*'An aquacode consists of an alphanumeric sequence of characters, generated from left to right. Digits and letters represent features as follows: Digits - represent complete river systems, river main channels, river tributaries or lakes; Letters - represent subdivisions (termed xones) of the coast, rivers, river tributaries or lake perimeters. Each zone can be further subdivided until the desired precision of location is obtained. The coding process is systematic, beginning at the coast and then analysing the river system layout.' [8]*	`DO7; DO7G8; DO75DSW`
Description	*A logical system for creating alpha-numeric identifiers (aquacodes) for lakes and waters (down to their component reaches) in New Zealand that was defined by the DSIR in 1990 [8]. It was intended to supersede existing conventions by introducing a topological constraint on the numbering that ensured that the structure and sequences also reflected the relative upstream or downstream position of river features. Monitoring sites could be identified using the aquacode of the water feature it coincided with.*	
Users	Research Institutes; Regional Councils; Utility Companies	Defunct
Digital	Yes (topological design intended to support computer modelling)	Human and machine readable
Governance	None	Creation: based on convention (no associated system or service)
Scalability	Focused on waterways and designed according to the interaction of parts of a hydrological network - limited scope for other domain features. Can be used to identify sampling stations on the network.	
Restrictions	None.	
Suitability	Limited. May provide a useful link into historical data, but too domain specific to be applied to, or be consistent with, conventions for other domains.	

5.2.2 International systems

What Three Words

Origin	what3words Ltd, United Kingdom	
Domain	Geography (location address)	
Type	System; Opaque; Persistent	
Structure/Example	`{word}.{word}.{word}`	`idea.speaks.scales` (cell within Westpac Stadium, Wellington)
Description	*what3words defines a global, unambiguous addressing system based on the division of the earth's land and sea surface into a 2-dimensional grid of 3 ×y 3 metre cells. Each of the 57 trillion cells is given a unique three-word combination to identify it, allowing any location to be addressed with a simple, memorable string. This allows addresses to be created and used in services such as satellite navigation systems even in developing nations with no formal addressing system. The words combinations are created using an algorithm that can be accessed online or be installed on devices for offline/remote area use. [9]*	
Users	Business; Government; Non-Government Organisations	in use
Digital	yes (text string, REST API call)	human and machine readable
Governance	Private: what3words Ltd, United Kingdom	Creation: mathematical algorithm (closed source)
Scalability	2-dimensional, fixed resolution geographic location only.	
Restrictions	Copyrighted. Subject to Terms and Conditions of Use. Free for individual use; tiered pricing policy for other users based on type of organisation and nation of origin.	
Suitability	Limited – solely intended as an identifier for a 2D cell on the earth's surface. A three-word location may have value as one of several location values for site to support discovery by users of the system.	

Life Science Identifier (LSID)

Origin	Interoperable Informatics Infrastructure Consortium (I3C, disbanded); Object Modelling Group	
Domain	Bioinformatics; Biodiversity	
Type	Global; Semantic (but intended to be opaque); Persistent	
Structure/Example	`URN:LSID:{naming-authority}:{namespace}:{object-id}:{revision}*` * optional	`URN:LSID:ebi.ac.uk:SWISS-PROT.accession:P34355:3`
Description	*'Life Science Identifiers (LSIDs) are persistent, location-independent, resource identifiers for uniquely naming biologically significant resources including but not limited to individual genes or proteins, or data objects that encode information about them. LSIDs are intended to be semantically opaque, in that the LSID assigned to a resource should not be counted on to describe the characteristics or attributes of the resource that the LSID refers to. The users of the LSIDs are permitted to use individual components (as specified elsewhere in this document) of LSIDs – although the LSID component parts themselves should be treated as opaque pieces of the identifier. LSIDs are expressed as a URN namespace [...].' [10]*	
Users	Businesses, research institutes, universities	2003 to present day
Digital	yes (text string, API call)	primarily machine readable (URN)
Governance	Unclear; specification: Object Modelling Group	Creation: LSID Assigning Service
Scalability	Intended application bioinformatics data (e.g. gene sequences, hybridizations, and compounds) and species.	
Restrictions	None	
Suitability	Limited. Can be used to identify biological features of interest in sampling activities, for example, water quality sampling. Soon to be defunct as their use is declining (A. Wilton, pers. comm., 2017).	

Herbarium Information Standards and Protocols for Interchange of Data (HISPID)

Origin	Council of Heads of Australasian Herbaria, Herbarium Information Systems Committee (HISCOM)	
Domain	Biology	
Type	System; Semantic; Persistent	
Structure/Example	Natural key combining the Herbarium Institution ID and the Herbarium's accession ID. [11] `{insid} + {accid}`	`insid: "NSW", accid: "390839"`
Description	*'Australian herbaria have been using the Australian standard for specimen data, known as Herbarium Information Standards and Protocols for Interchange of Data (HISPID), for almost two decades (Croft 1989; Whalen 1993; Conn 1996). Its use continues in Australian herbaria because databases in these institutions have been developed in concert with HISPID, and thus reports exist in most institutions to export data in this format. The current version of Australia's Virtual Herbarium, first prototyped in 1998, uses HISPID 3 (Conn 1996). It became TDWG's specimen data interchange standard until it was superseded by ABCD [Access to Biological Collection Data schema] in 2004. [HISPID 5 is an extension of ABCD.]' [12]*	
Users	Australasian Herbaria	in use
Digital	yes	human and machine readable
Governance	HISCOM	Creation: based on convention (no associated system or service)
Scalability	Intended application restricted to specimens held in herbaria.	
Restrictions	None.	
Suitability	Limited. The intended scope, physical samples, is narrower than that required by a monitoring site network.	

International GeoSample Number (IGSN)

Origin	System for Earth Sample Registration (SESAR)	
Domain	Geology, Pedology	
Type	System; Opaque; Persistent (based on URN syntax)	
Structure/Example	`ISGN:{NAMESPACE}{CODE}`	`IGSN:ICDP5054EEW1001`
Description	*'The IGSN is a persistent unique identifier for physical samples and specimens that eliminates the problems associated with the ambiguous naming of samples. The IGSN registration service helps discover, access, and share samples, ensure preservation of and access to sample data, aid identification of samples in the literature, and advance the exchange of digital sample data among interoperable data systems, thus maximizing the utility of samples for research, education, and society.' [13]*	
Users	Research Institutes, Universities	2011 to present day
Digital	yes (text string, resolvable Handle.net URL)	human and machine readable
Governance	IGSN e.V., Germany (http://www.igsn.org/)	Creation: requested from IGSN Registration Service hosted by an operating allocating agent
Scalability	17576 namespaces containing 1.55 billion samples. Scope restricted to physical samples and stored specimens.	
Restrictions	None.	
Suitability	Limited. The intended scope, physical samples, is narrower than that required by a monitoring site network. Allowing the association of IGSN values to samples taken at monitoring sites would help discovery, re-use, and integration of data.	

WMO Integrated Global Observing System (WIGOS)

Origin	World Meteorological Organisation (WMO)	
Domain	Meteorology, hydrology	
Type	System or Global (extended version); Opaque; Persistent	
Structure/Example	`{series}-{issuer}-{issue-number}-{local-identifier}`	`0-513-215-5678`

	Extended version (optional, for use outside the WIGOS):
	`int.wmo.wigos-{series}-{issuer}-{issue-number}-{local-identifier}-{supplementary-identifier}*`
	* optional
		`int.wmo.wigos-0-513-215-5678`
		or
		`int.wmo.wigos-0-513-215-5678-some.local.id`

Description	*'The WMO Integrated Global Observing System shall be a framework for all WMO observing systems and for WMO contributions to co-sponsored observing systems in support of all WMO Programmes and activities.' [14]* *'WIGOS station identifiers provide an essential link between observations and the metadata describing the environment in which they were made. Using the WIGOS station identifiers avoids the need to transfer all the station metadata alongside an observed value.' [15]*	
Users	WMO Member Organisations	2007 to present day
Digital	Yes. Intended for use in systems such as the WMO Information Resource (WIR).	human and machine readable
Governance	WMO and Member Organisations	Creation: assigned by providers (subject to some constraints on form) using series and issuer values allocated by the WMO.
Scalability	Currently only used to identify observing stations (series = 0); however, the convention is intended to be flexible enough to identify other entities (for example, instruments). No obvious limitation on the number of identifiable features.	
Restrictions	Only to be used by the WMO and member organisations.	
Suitability	Limited due to the restrictions on use but as the scope is the same, i.e. monitoring stations, the design principles are likely to be useful in definition of a local convention.	

6 Discussion

The requirements workshop and literature search uncovered many identifier schemes but very few were formally defined, considered for use in a distributed or shared system, or for application to other types of feature beyond their target scope. In other cases, for example, the NEII Site Register, consideration was given to the aggregation and indexing of sites across agencies and systems, but did not define a formal convention for identifying items in the register.

The World Meteorological Organisation convention for station identifiers is the most promising candidate for an environmental monitoring network in New Zealand. While its use is restricted to the WMO and its member organisations, its design principles are directly applicable. If implemented in a properly governed system such an identifier scheme would meet the functional and encoding requirements outlined in section 5.1.3 and would:

1. provide a system-wide identifier with scope for global distribution in other contexts
2. create a value that could be incorporated into common digital identifiers like URIs
3. support a sufficient range of values to support the identification of many features, and also has scope to be applied to many types of feature beyond sites
4. be a sufficiently logical structure to be parsed for meaning, but not in such a way that long-term stability is compromised
5. be centrally governed, but allow for values to be created 'off-line' when data are created by participating agencies.

The workshop identified useful information to be carried with the identifier, for example, location or the type of sensors deployed (Table 1). Initial proposals that they become part of the semantics of the identifier were problematic as:

1. too many facets can make an identifier long and difficult to transmit, present as text or transcribe
2. facets based on changeable values like agencies (names change, authorities merge) or locations (positions may become more accurate or new projections may be adopted) can make an identifier unstable – there is a temptation to update an identifier to match the new data, such as the use of a new location datum, or the addition or removal of sensors
3. meaningful labels can confuse identity with search, meaning a user may 'hack' an identifier to find a feature based on expected facet values and wrongly assume missing data when no result is returned. Conversely, the ID can wrongly convey out-of-date data, capturing the state of a site when it was established, but not as it may be later.

That said, the data expected in meaningful labels are very useful, especially as they will support the documentation and indexing and discovery of sites. Also, meaningful labels have value when presenting data and should be provided as non-authoritative names or labels where useful. A registry of similar scope to the Australian NEII Environmental Monitoring Site Register would mean that the identifiers could be associated with the meta-data record and labels. Furthermore, the registry could be used to document identifiers assigned by other parties or under other conventions (for example the IGSN, or the NIWA Hydrological Site Index) to further promote the integration of both complimentary digital data and historical hard-copy reports.

7 Recommendations

7.1 Proposal

We propose the adoption of a national site identifier convention based on the WMO WIGOS convention. The identifiers will be assigned by the agencies that establish the site, but subject to the constraints and approval (confirmation the value is valid and unique) of an accepted naming authority. Once the convention is agreed and in use, a web-based national monitoring site register should be defined and built. This register will support the governance activities of the naming authority and also the discovery and description of environmental monitoring sites in New Zealand. The register will not hold or publish the actual monitoring data, instead it will direct users to the appropriate data source – whether or not a user can access the data will be determined by the data source, not by the register.

To implement a national monitoring site identifier the participating regional councils will need to:

1. establish a NEMS Naming Authority to govern their allocation and use of identifiers
2. define a convention for the creation of site identifiers
3. implement a simple on-line registry (with data services) to:
    a. define and allocate shared identifier components (e.g. feature type codes and authority identifiers)
    b. keep track of identifiers that have been allocated, and
    c. check new identifiers for uniqueness.

To implement a national monitoring site register the participating regional councils will need to:

1. create a NEMS Monitoring Site Registration Committee (this will incorporate the previously establish naming authority)
2. define minimum meta-data required for site registration
3. extend the online name registry to a full site meta-data registry that:
    a. deploys complimentary web services for automated registration and discovery of sites; and
    b. provides links to actual data (historic reports, web sites, web services etc.) but defers management of access to the host organisation.

7.2 Example implementation

7.2.1 Identifier convention

A structured identifier using tokens designed according to the WMO WIGOS Site Identifier is proposed. It shall have three forms (Table 2):

1. A system identifier for use in 'authoritative' systems supporting the monitoring network (e.g. the site register)
2. An essentially globally unique extended identifier to allow the string to be used outside the system and as text in documents
3. An HTTP URI for use in a Linked Data based distributed information system.

The identifier shall be case insensitive and conform to the structure and value constraints presented in Table 2 and Table 3. Numeric values shall be preferred over alpha-numeric or alpha values. In other words, alpha-numeric values shall only be introduced when all numeric options in a sequence are exhausted. For example when a two-digit sequence beginning with 00 reaches 99 then the next sequence shall be 0a, 0b, ..., 0z, 1a, 1b and so on.

For the purposes of this example the system shall be known as the New Zealand Environmental Monitoring Network (NZEMN).

Table 2 The structure of the three forms of environmental monitoring identifier

Form	Structure	Example
System Identifier	{series}-{issuer}-{issue-number}-{local-identifier}	10-01-00-654321
Extended Identifier	nzemn:{series}-{issuer}-{issue-number}-{local-identifier}	nzemn:10-01-00-654321
HTTP URI*	https://data.nems.org.nz/nzemn/id/{series}-{issuer}-{issue-number}-{local-identifier}	https://data.nems.org.nz/nzemn/id/10-01-00-654321

* The 'nzemn' prefix to the Extended Identifier can be used a well-known namespace prefix for the base of the HTTP URI allowing CURIEs to have the same form as their text representation.

Table 3 Identifier facet value constraints

Token	Form	Range	Authority	Comments
{series}	Alpha-numeric; ##	00 to zz	NEMS-NA	Number of available values: 1,296. The series can be broken down into sub-series ranges based on the type of feature, for example: 00 to 0z: system reserved 10 to 1z: sites or stations 20 to 2z: hydrological features 30 to 3z: soil features etc. Care must be taken not to make the sub-typing too complex and the number of types should be a fraction of the 36 available values in each range.
{issuer}	Alpha-numeric; ##	00 to zz	NEMS-NA	Number of available values: 1,296. Each issuer shall be allocated an issuer number that acts as a namespace for their identifier, for example 00: System reserved/shared features 01: Northland Regional Council 02: Auckland Council 03: Waikato Regional Council 04: Bay of Plenty Regional etc. It would be desirable for each number to map to a single ISO3166-2 code. A number can be retired or transferred to another issuer (for example when Auckland Council was created it could take over the Auckland Regional Council number while the Auckland City Council number could be retired). Only one agency should be allocated a number at any one time.
{issue-number}	Alpha-numeric; ##	00 to zz	ISSUER	Number of available values: 1,296. To be created at the discretion of the issuer to delegate the creation of identifiers to another party, for example a contractor, system or individual data capture device. Default value is '00'.
{local-identifier}	Alpha-numeric; ######	00000 to zzzzz	ISSUER	Number of available values: 2,176,782,336 To be created at the discretion of the issuer. Must conform to the required form. Must be unique in combination with the issuer and issue-number.
	Alphanumeric	Max length: 6	ISSUER	An alternative is to follow the IGSN practice and allow the provider to use any value (say a local system id), subject to some restrictions on length and assurance of permanence once created.

7.2.2 Name register

The initial implementation of a name register is a matter of process supported by a functionally simple web site and web-service to keep track of identifiers and to help test that an identifier has not already been used. The main task will to establish the naming authority – ideally a NEMS committee – and provide light-weight governance of the identifier scheme.

This proposed NEMS Naming Authority's responsibility shall:

1. ratify a convention for monitoring site identifiers
2. oversee work to establish conventions for the identification of related features, for example, lakes and waterways
3. administer an online database that:
    a. manages the allocation of identifier tokens (series and issuer)
    b. (optional) create a complete new identifier for an issue on request
    c. checks new tokens for uniqueness and conformance to the convention, and
    d. stores a record of identifiers that have been used by each issuer.

7.2.3 Site register

The Site Register will fulfil the role of the naming authority's on-line database and augment it with the ability to store meta-data relevant to the indexing and discovery of sites and links to data collected at the sites (subject to appropriate restrictions on access). The specifics of the register are beyond the scope of this report, but one could be set-up using existing models for registration and description of sampling and observation data:

1. Governance of the registry could confirm to ISO19135 [16]
2. The structure and management of the registry and its content could be according to ISO19135 and the UK Government Linked Data Working Group [17]
3. Site metadata could be captured according to the specifications of the NZ Environmental Observation Data Profile of ISO19156 [18]
4. The register itself would be a web tool built on web services that conform to the standards of the Open Geospatial Consortium (spatial data) and World Wide Web Consortium (Linked Data).

7.3 Future work

A national convention can be realized in two steps: first, the establishment of the naming authority and supporting tools; and second, the implementation of a national site register. In the right circumstances both could be completed at once. We recommend the following steps:

1. Establish the Name Register as per section 0:
    a. Convene the naming authority as a NEMS, or NEMS-endorsed, committee
    b. Build a simple on-line database to support the creation and administration of the identifiers.
2. Implement a National Monitoring Site Register as per section 7.2.3:
    a. Establish the terms of reference for a governing body
    b. Define the requirements for a site register, including, but not limited to, its functionality, data standards, security model, access constraints and licensing conditions
    c. Build a set of open-standards based web services to manage the registry, administer identifiers and manage the resolution of HTTP URI versions of identifiers to appropriate locations
    d. Build a client interface to these web services for shared use.

The work to establish a site register could be undertaken in an appropriately funded EnviroLink Tools project. How this would be sustained once it has been deployed is unclear. If no agency or consortium is willing to take responsibility for ongoing maintenance, and governance, then the value of a short-lived register is limited.

8 Conclusions

Well-defined and stable identifiers are the structural fabric of any distributed information system. When in place, all involved can be confident they are dealing with the same objects, making it much easier to integrate, compare, and analyse data stored in multiple places and systems over long periods of time. Defining and using such an identifier will greatly enhance the exchange and integration of environmental monitoring data in New Zealand.

A national convention for identifying environmental monitoring sites and their features of interest is conceptually and technically feasible. A convention, using the design of the WMO WIGOS Station Identifier, is proposed, initially as a site identifier to be managed by a formal naming authority, but ultimately as the kernel of a national site register that provides additional site meta-data and discovery services.

The only impediments to implementation are social and financial. No agency can be coerced into using such a convention, so national support is essential for widespread adoption. More important, however, the specification, administration and implementation require financial support. This will be in-kind, where participating agencies fund and share the retro-fitting of systems to use the identifiers, funding the groups responsible for governing the system and the tools required to support it; tools that, if properly designed and implemented, should relieve the burden of implementation on participating agencies.

9 Acknowledgements

The authors thank the following for their support, technical feedback and reviews of this report: Sean Hodges and Brent Watson (Horizons Regional Council); Nicholas Holwerda and Grant Lawrence (Auckland Council); Jochen Schmidt (NIWA); David Medyckyj-Scott (Landcare Research). We gratefully acknowledge the financial support of the MBIE Envirolink scheme and the sponsorship of Jeff Watson, Manager – Catchment Data, Horizons Regional Council.

10 Glossary of terms and abbreviations

For consistency, NEMS terms and definitions are favoured by this document [19]. All definitions provided below are taken verbatim from their source.

authority: the organisation responsible for creating an identifier or data object.

commissioning agency: The agency that initiates data collection. [19]

CURIE: Compact URI, e.g. IGSN:ICDP5054EEW1001, that can map directly on to a URI, e.g. http://igsn.org/ICDP5054EEW1001. [20]

custodian: the agency responsible for ensuring the preservation and dissemination of data. [19]

domain feature: feature of a type defined within a particular application domain (for example, hydrology). [2]

feature: abstraction of real-world phenomena. [1]

feature of interest: feature whose properties are of interest in the investigation of which the observation is a part. [2]

feature type: class of features having common characteristics. [1]

Linked Data: a method of publishing structured data on the World Wide Web as a set of interlinked data.

metadata: a set of data that describes and gives information about other data. It may describe the content, quality, condition, location or other characteristics of the data, and operations on or modifications to that data. [19]

observation: act of measuring or otherwise determining the value of a property. [2]

procedure: method, algorithm or instrument, or system of these, which may be used in making an observation. [2]

register: set of files containing identifiers assigned to items with descriptions of the associated items. [16]

registry: an information system on which a register is maintained. [16]

sampling feature: feature that is involved in making observations concerning a domain feature. [2]

sensor: a device that detects or measures a physical property and records, indicates or otherwise responds to it. [19]

site: the geographical location where monitoring takes place. (See also ‘station’.) [19]

site number: a unique numeric station identifier usually derived from a national numbering system such as Catchments of New Zealand a map grid, e.g. site number 75207, where 752 = Clutha River, and 01 = Balclutha Station, or modified from a MetService Network Number; e.g. 941301 for C94131, Tarata. [19]

site identifier: a unique label for a site, which may be numeric or alphanumeric. [19]

station: the collective term for sensors at a particular site. [19]

URI: Uniform Resource Identifier

URL: Uniform Resource Location

URN: Uniform Resource Name

WS16: a standard form developed by the Water and Soil Division of the Department of Scientific and Industrial Research (DSIR) used to record station history metadata. Earlier versions of this form were known as SCC16's or Form 16's. [19]

11 References

[1] ISO 19109 2015. Geographic information – Rules for application schema. Geneva, Switzerland, International Organization for Standardization. Last accessed at https://www.iso.org/standard/59193.html on 2017-06-14.

[2] ISO 19156 2011. Geographic information – Observations and measurements. Geneva, Switzerland, International Organization for Standardization. Last accessed at https://www.iso.org/standard/32574.html on 2017-06-14.

[3] Sollins K, Masinter L 1994. Functional requirements for uniform resource names. Internet Engineering Task Force Request for Comments 1737. Last accessed at https://tools.ietf.org/rfc/rfc1737 on 2017-06-14.

[4] Bureau of Meteorology 2016. The National Environmental Monitoring Sites Register. Last accessed at http://www.neii.gov.au/nemsr on 2017-06-14.

[5] US Geological Survey 2011. Do site numbers have any particular meaning? Last accessed at https://help.waterdata.usgs.gov/faq/sites/do-station-numbers-have-any-particular-meaning on 2017-06-14.

[6] Landcare Research 2014. NVS Project naming conventions. Unpublished discussion document.

[7] Holdaway RJ, Easdale TA, Carswell FE, Richardson SJ, Peltzer DA, Mason NWH, Brandon AM, Coomes DA 2016. Nationally representative plot network reveals contrasting drivers of net biomass change in secondary and old-growth forests. Ecosystems. doi:10.1007/s10021-016-0084-x.

[8] Pearson CP, Phuah PL, Thompson WR 1990. An illustration of "Aquacoding" New Zealand waterways. New Zealand Journal of Hydrology 29(2): 63–74.

[9] What3Words. Last accessed at https://what3words.com/ on 2017-06-14.

[10] Object Management Group 2004. Life sciences identifiers specification. Last accessed at http://www.omg.org/cgi-bin/doc?dtc/04-05-01 on 2017-06-14.

[11] Conn RJ ed, 1996. HISPID3. Herbarium information standards and protocols for interchange of data, Version 3. Last accessed at http://plantnet.rbgsyd.nsw.gov.au/HISCOM/HISPID/HISPID3/H3.html on 2017-06-14.

[12] Herbarium Information Systems Committee, 2007. HISPID/ABCD Workshop Executive Summary. Last accessed at http://hiscom.rbg.vic.gov.au/wiki/HISPID/ABCD_Workshop_Executive_Summary on 2017-06-14.

[13] IGSN e.V. IGSN Documentation. Last accessed at http://igsn.github.io/ on 2017-06-14.

[14] World Meteorological Organisation 2015. Manual on the WMO Integrated Global Observing System. Annex VIII to the WMO Technical Regulations. WMO Report

Number 1160. Last accessed at https://library.wmo.int/pmb_ged/wmo_1160_en.pdf on 2017-06-14.

[15] World Meteorological Organisation 2016. 5.2 WIGOS Regulatory Material complemented with guidance material – 5.2(2) WIGOS Station Identifiers. WMO Document Number ICG-WIGOS-5/Doc.5.2(2). Last accessed at https://www.wmo.int/pages/prog/www/WIGOS-WIS/meetings/ICG-WIGOS-5/ICG-WIGOS-5_Doc-5.2(2)_WIGOS-IDs_Guidance.docx on 2017-06-14.

[16] ISO19135-1 2015. Geographic information – Procedures for item registration – Part 1: Fundamentals. Geneva, Switzerland, International Organization for Standardization. Last accessed at https://www.iso.org/standard/54721.html on 2017-06-14.

[17] UK Government Linked Data Working Group 2017. Registry core: principles and concepts. Last accessed at https://github.com/UKGovLD/registry-core/wiki/Principles-and-concepts on 2017-06-14.

[18] Ritchie ABH ed. 2014. New Zealand environmental observation data profile. Last accessed at https://github.com/EODP-NZ/eodp-dev/raw/master/doc/spec-eodp-core-1.0.pdf on 2017-06-14.

[19] National Environmental Monitoring Standards 2016. National environmental monitoring standard glossary - terms, definitions and symbols. Version 1.0. Last accessed at https://www.lawa.org.nz/media/2982084/NEMS-Glossary.pdf on 2017-06-14.

[20] Birbeck M, McCarron S eds 2009. CURIE Syntax 1.0 – a syntax for expressing Compact URIs. W3C Candidate Recommendation 16 January 2009. Last accessed at https://www.w3.org/TR/2009/CR-curie-20090116/ on 2017-06-14.