Te Rua Mahara o te Kāwanatanga
ARCHIVES
NEW ZEALAND

All-of-Government Ontology Options Paper

July 2021

Project team

Name	Organisation	Organisation role	Project role
Stephen Clarke	Archives NZ	Chief Archivist	Project Champion
Antony Moss	Archives NZ	Director Government Recordkeeping	Project Sponsor
Joanne Koreman	Archives NZ	Principal Advisor Government Information Management	Project Lead
Susan Jenkins	Archives NZ	Archivist/Archives Advisor	Disposal SME and Project Support
Hannah Mitchell-Shapiro	Archives NZ	Advisor Planning and Development	Project Support
Liz Wilson	Metataxis NZ Ltd	Information Architect	Author
Judi Vernau	Metataxis NZ Ltd	Ontologist	Author
Annabel Snow	Metataxis NZ Ltd	Ontologist	Contributor

Document control

Version	Date issued	Issued by	Comments
Draft v1.0	May 2021	Judi Vernau	For discussion
Draft v2.0	June 2021	Judi Vernau	Adding feedback
Draft v3.0	June 2021	Judi Vernau	Adding feedback
Final v1.0	June 2021	Judi Vernau	Issued
Final v1.1	July 2021	Judi Vernau	Minor change to nomenclature

1. Executive Summary

This options paper for an all-of-government ontology has been commissioned by Archives New Zealand, Te Rua Mahara o te Kawanatanga. Archives New Zealand is the agency that holds and provides access to government information of archival value. It establishes the regulatory framework for information and records management across the public sector, and the Public Records Act 2005 provides for the Chief Archivist to exercise a leadership role in this, setting standards for the creation, maintenance and efficient management of public records (including data). The purpose of this options paper is to explore the benefits of an all-of-government ontology in helping agencies to find, use, manage and share their information and data.

Ontologies provide common vocabularies, concept definitions and relationships between concepts. They support interoperability, the semantic analysis of content, autocategorisation of content and automation of business processes, knowledge representation and the use of artificial intelligence and machine-learning tools, widely thought to be the future of information management.

In developing the paper, we have explored options, benefits, demand, feasibility and practical implementation ideas for such an ontology with a wide range of stakeholders in the public, private and academic sectors. We appreciate the wide-ranging support from such a diverse community, particularly The Digital Public Service (Government Chief Digital Officer) and Statistics New Zealand (Government Chief Data Steward).

From an Archives business perspective, an all-of-government ontology will enable consistent categorisation of government information, and greatly improve access to holdings from multiple different points of view, both cultural and multilingual. But the benefits to such a semantic tool are arguably much wider across government, providing the ability to successfully evolve from our current siloed information and data architectures. An all-of-government ontology could support data sharing, data lineage, automatic categorisation and classification of content, and data analytics. There is already existing work, tools, standards and thinking in New Zealand which can be built upon to take this forward, and great enthusiasm expressed by interviewees: this is an exciting opportunity to look at how modern techniques and technologies can move Archives New Zealand forward in its aims, while also providing substantial benefit across government.

An all-of-government initiative such as an ontology necessarily requires different parts of the government information ecosystem to work together, and so we have explored options that propose joint efforts from key agencies for the initial stages, progressively including sector domains on a modular basis, but which may also include working with third parties.

The following are our options for an all-of-government ontology. Some can stand independently, some rely on other options:

1. Repository: to create hub or repository for government ontologies and taxonomies as a means to providing standards for terminology;
2. Foundation Ontology: to design and build a Foundation Ontology covering the key entities relevant to government to provide authoritative concepts (with URIs and definitions, and to include Māori concepts and te reo), and the fundamental building blocks for domain-specific ontologies;
3. Government organisational lineage: within the Foundational Ontology design, to prioritise the development of the classes, relationships and properties required to track the names and remits of Government departments through time;

4. Domain ontologies: to also prioritise the development of classes, relationships and properties for two specific domains - health and housing, since both are ongoing reorganisations, and senior leaders have expressed strong support for this work;
5. Autocategorisation and classification: to build on the Foundation Ontology to create terms, term properties and relationships which support autocategorisation and autoclassification of records when combined with the appropriate rulebase and algorithms;
6. Ontology service: to offer a service which incorporates the guardianship and publication of authoritative ontology models and associated taxonomies, and the provision of a suite of utilities for use in creating, using and disseminating them;
7. Third party platforms: to explore the potential of third party platforms such as the Azure cloud and associated technology stack (e.g. Syntex and Cortex, advanced artificial intelligence tools in M365) incorporating the Foundation Ontology (and potentially domain ontologies) to support processes for autocategorisation and the application of archival business processes.

From these options our recommendations are:

- To create a Foundation Ontology covering the key entities and relationships relevant to New Zealand government;
- To work alongside the development of the Foundation Ontology to create a Government organisational lineage;
- Within the developing Foundational Ontology work, to prioritise domain ontologies in health and housing to support the Ministry of Health and Kāinga Ora;
- To explore the potential for working with a third party such as Microsoft to integrate an all-of-government ontology into advanced artificial intelligence tools to support the information and data lifecycle.

There is also, in due course, the possibility of offering a full ontology service to agencies, but this is not part of the current recommendations.

Development of an all-of-government ontology will necessitate the creation of a centre of excellence, potentially within the Digital Public Service, which is responsible for building and managing the ontology and liaising with agencies.

2. Introduction

2.1 Background to the Options Paper

This Options Paper was commissioned by Archives NZ to allow them to understand and explore the potential for an all-of-government ontology (AOGO) and to test the idea across government with other functional leads. It supports alignment with Digital Public Service initiatives, in particular the Digital Public Service Strategy (DIA, 2020).

An ontology is a set of vocabularies which are formally categorised, defined, linked and structured to potentially support findability, interoperability, business efficiency and artificial intelligence. Ontologies (and other vocabularies) currently exist within the NZ government space but no work has been done to explore the potential of creating an all-of-government ontology. This paper is the first step towards that.

As part of the project, we talked to a range of potential stakeholders across New Zealand government, business and academia, as well as counterparts in Australia and the UK (the complete list of those consulted is in Appendix A). Desk research covered both national and international initiatives and standards. In order to support understanding and to test out some issues we also built a Demonstrator ontology covering a sample of NZ government agencies and functions together with some relevant events (the Demonstrator is described in more detail in section 7).

The Options Paper will:

- outline the necessary role ontology plays in supporting Archives New Zealand strategy and the outcomes of the Digital Public Service;
- show the high-value benefits which can be achieved in the public and private sectors;
- identify use cases where an ontology can support findability, efficiency and information sharing;
- demonstrate stakeholder support for potential features of the ontology;
- propose options for the scope and structure of an AOGO;
- outline the activities, resources and collaboration needed to create and maintain an AOGO and achieve benefits;
- suggest ongoing governance and management structures to support the AOGO in achieving desired outcomes and benefits.

2.2 Why now?

The AIIM report on opportunities for transforming information management and access (AIIM, 2021) describes the situation well[1]:

> 'There is a new awareness about the importance of information assets – document, content, records, [data,] and process leaders and practitioners need to seize this opportunity. Organisations cannot continue down the path of viewing information management decisions through a tactical cost-minimisation filter. In a digital age, the

[1] text in square brackets is ours, bold font is the author's

> everyday decisions that organisations [and governments] make about how they deal with information chaos become ***strategic business decisions*.**
>
> 'There is now a new realism about the gaps that exist in information management strategies, gaps that are reflected in what at first glance might seem to be a rather harsh grading of the current state of IM in most organisations … We need a higher level, more strategic discourse that addresses the bottom line for IM [and data management] investments. Budgeting for IM [and data] initiatives often takes a back seat to front line business process-driven budgeting in many organisations and [across government] even when the two areas should be connected.'

What are the reasons for this 'new realism'? The pace of technological advance has led us to an unmanageable system of data and documents that we fail to use to do our jobs. From the carefully curated world of the 1970's records office, clerks and paper files to the democratised world of do-it-yourself emails and typing and storing our own documents, our ability to control and use information has been constantly eroded: we are swamped by sheer volume and 'losing the battle against information chaos' (AIIM, 2021).

Government agencies and large corporates now routinely hold many copies of all their applications, data and documents in back-ups, business continuity and data centre redundancy architectures for real time failures; they also routinely fail to winnow out the rubbish from the reliable and useful**.** The pain of difficult filing, poor search and a hundred different ways to define 'customer' has led to huge inefficiencies and frustrations being 'priced in' to the cost of doing business.

Information technology's obsession with platforms and software tools and keeping the lights on has led to lack of consideration for the pain our information users suffer every day, and an ignorance of the semantic tools which can bring order out of chaos. In particular, metadata is the key to good findability, but agencies generally don't have the kinds of tools to help with automatic assignment of accurate metadata; similarly the disposal process for records is usually based on huge manual effort rather than being supported by the appropriate software.

2.3 So why this idea of ontology now?

Over the last two decades the power of ontology-based tools to help organise, find, manage and exploit content has been developed and proved in huge knowledge organisations across the world from Google to the BBC to finance companies, publishers and other information-rich companies. In the last decade these tools have been introduced into some NZ government information ecosystems, but failed to make a sufficiently impressive improvement to how we design these systems to be routinely a part of an IT department's infrastructure.

One major reason for our inability to utilise these tools is that technologists often have poor understanding or appreciation of the benefits of semantic tools; another is that where the benefits are understood, NZ government agencies have still been faced with needing to start the ontology and semantic designs from scratch. Business, innovators, researchers and cultural organisations have been similarly discouraged because the effort of using government data has been too onerous. However, advances in technologies and demands for more seamless remote working and collaboration have made the use and deployment of ontologies and semantic models more feasible.

API technology alone has enabled more cost-effective integration between semantic tools and data management systems. Artificial Intelligence, knowledge graphs and machine learning technologies have added a reiterative feedback loop to the maintenance and ongoing growth of concept definitions and vocabularies.

Despite this, the human-led work with subject and domain experts, stakeholders, users, intellectual property specialists, cultural guardians, etc remains as intensive as ever.

So what can New Zealand do to take advantage of this 'new realism'?

3. What are the problems we are trying to solve?

3.1 Data and information are 'locked away'

Two of the Archives New Zealand strategic focus areas in its Archives 2057 Strategy (ANZ, n.d.) are taking archives to the people, and upholding transparency. Currently it is hard to find archival – and other - government information unless you know specifically what you are looking for or what agency has created it. This impedes transparency and has a very immediate impact on people who are trying to find historical records but don't know which agency is or was responsible for what at any given time.

For many government departments, their own information and data has been collected and created over years and is held in uncontrolled or barely controlled repositories. So much of the information and data that staff need to do their jobs is locked away or hidden in documents, web pages, PDFs, and therefore not easily findable, usable or re-usable.

Our inability to file or control emails has led to the rise of email archives/vaults as another uncontrolled and inaccessible repository of information and data. Multiple mergers of government agencies and transfers of functions from one agency to another have magnified these pain points.

The inability to find information has many consequences, which include work being done again, time wasted looking for information and poor-quality work because relevant information was not found.

It can be hard to share or use data across government departments because it has been created for different purposes, uses different lenses, different language. An example of this is delivery-focussed departments, agencies and private organisations trying to use Stats NZ reference data.

Similarly, content that is useful and appropriate to the citizen, customer, client, researcher, policy maker, business and innovator should be readily available. Data analysts, researchers and policy writers need to be able to add semantic enrichment of concepts to data assets, data warehouses and data lakes, but frequently have no means to do this.

We need mechanisms which can link, interrogate, enrich and present the data and information that we have in order to exploit existing resources to the fullest extent. Being able to undertake fact and entity extraction from unstructured content such as documents, emails and social media and add that to a data environment will significantly add to the data available for analysis.

3.2 We need access to new forms of data and information

Demands for high quality data and the rise of automated processes of all kinds in the public sector and beyond has led to a huge proliferation of information and data systems being built by the public and private sector.

There is demonstrable demand for improvements and efficiencies in the design, documentation and re-usability of these new data and information systems and resources. However, the cost of extracting knowledge from documents, web pages or PDFs to populate usable data and knowledge systems is high and typically inefficient and low quality.

Te Tiriti partners and Te Ao Māori concepts must be included from the beginning of any system design, as they have not been included as a matter of course in the past. This has led to the current situation where data and information of interest to Māori are unable to be identified by agencies, and are therefore inaccessible to Māori.

Security and privacy concerns relating to information systems also need to be properly addressed to ensure that content is subject to the appropriate privacy and technical security requirements regarding access.

Climate change data is a particularly compelling use case for these scenarios, as data which could be used has been created and stored over many years in many different formats and media. The content in this field is a data heavy and longitudinal example, as outlined in the report - *Focusing Aotearoa New Zealand's environmental reporting system* (Parliamentary Commissioner for the Environment, 2019).

> The ontology is essential to and at the heart of AI-driven technologies.'
> Seth Earley (Earley, 2020)

3.3 Compliant disposal of digital records is challenging to achieve

Government departments have a legal responsibility to delete, transfer or archive records according to agreed retention and disposal schedules. As volumes of information increase rapidly so do the challenges of identifying and sentencing digital records in a timely manner.

Keeping content for longer than required also increases costs of storage, the requirement to provide access, and privacy risks: timely disposal is therefore important to minimise those costs.

The extremely large and increasing volumes of content make it hard to find useful information, and make it much more likely that out-of-date content is used inadvertently. There can also be privacy and security risks associated with badly managed content. Government departments have a legal obligation to classify as either 'open access' or 'restricted access' information and records that have been in existence for 25 years or are to be transferred to Archives New Zealand or a local authority archives. This also can be difficult to achieve with the large volumes of digital content these organisations have responsibility for.

3.4 Lack of interoperability

Government agencies use different terminology and language which inhibits cross-government findability and usability. Inconsistent definitions across government information and data resources leads to inefficient and poor-quality data sharing and analysis.

> ‘If there's syntactic interoperability, what about semantic interoperability? Naming conventions are an issue. We need common lists of terms which are properly maintained. NZ place names are maintained by LINZ, and that works, but they have staff to work on that. We need similar master lists, with people who work with the relevant communities to keep them up to date. Putting all agency vocabularies under one hood would be beneficial.’
>
> Jochen Schmidt, Chief Scientist for Environmental Science, NIWA

Using different language and not having agreed identifiers and definitions means that data often has to be massaged in order to be usable by another agency or even by another business unit within one agency, causing extra time and effort, and the opportunity for new errors to creep in.

4. A closer look at some of the issues

In the previous section we reviewed some of the problems encountered by agencies trying to grapple with the difficulties of managing and making available ever-increasing volumes of information and data, while complying with their responsibilities. In this section we take a closer look at two sets of use cases and stories to highlight some of the issues. The areas chosen are:

- government administration and accountability, and
- innovation, data and resource sharing.

4.1 Government administration and accountability

Use cases
Poor search and findability leads to rework, loss of knowledge and insights, failure of accountability
Digital information of potential historical importance is sitting in vaults and cannot be assessed. Older records are hard or impossible to find, and government accountability is thereby lost because agencies change in name and remit over time, and there is no robust mechanism for tracking back. Examples: Recent Commissions of Inquiry have needed to search across multiple government agencies and back through time: • Public Inquiry into the Earthquake Commission (Christchurch earthquake) • Royal Commission of Inquiry into the terrorist attack on Christchurch mosques on 15 March 2019

- Royal Commission of Inquiry into Abuse in Care.

Individuals who are trying to locate personal information held by one or more government agency, or who are trying to find out who was responsible for what when, are unlikely to know where to look.

> 'Ontology and information design is crucial in linking the past to the current and the future' Rosemary McGrath, Stats NZ

Information and data are hidden in legacy systems

Migration of information and data from legacy environments to new information, knowledge and data systems is seldom successful or attempted, in a short-sighted effort to minimise the project costs.

Examples:

There are many examples of multiple document management systems, 'abandoned' shared drives, and onerous migration effort across government because of inadequate metadata tagging of documents and content.

The role of metadata in improving findability is not well understood

Making usable metadata schema and vocabularies easily implementable in Microsoft 365 is absolutely necessary because the platform is ubiquitous across agencies. Supporting semantic tools such as MS Cortex and SharePoint Syntex in the Microsoft 365 ecosystem can now significantly improve search and findability inside government agencies' document and knowledge environments. However, many agencies do not have a formal metadata scheme, let alone the kinds of vocabularies needed to take advantage of this.

Manual application of retention and disposal rules is no longer tenable

Agencies need to be able to use automated rules and algorithms to apply retention and disposal policies

- in formal repositories ie document management systems
- in non-formal repositories eg shared drives
- in formal but non-records-managed repositories eg emails.

But agencies often don't have the resources, skills or technology to pursue this. Where initiatives in this area are happening, there is overlap of effort and little sharing of experience and knowledge.

Stories: Archives New Zealand
Public sector organisations criticised for poor record-keeping, *Stuff* 7 January 2016 (Stuff, 2016)
'In an Archives New Zealand report, the chief archivist … said it was "disappointing" that barely half of the public offices audited by Archives New Zealand in 2014-15 had an appropriate level of record-keeping maturity, 10 years after the Public Records Act came into force. 'Low levels of record-keeping maturity indicate that some public offices are not effectively managing business risks or ensuring that records are created and maintained to enable government accountability. 'More than half of the public offices audited did not have regular monitoring or reporting on their record-keeping policies and procedures, while processes for disposing and transferring unimportant records were "underdeveloped" in nearly all of the audited offices. '"The [law] has now been in force for 10 years and most public offices are still not disposing of records appropriately…this is disappointing." '[The chief archivist] said over-retention of records could lead to higher storage costs, while records with long-term value were at risk of being accidentally destroyed or lost in systems "cluttered with lower-value records." 'Internal Affairs Minister, Peter Dunne said there was a "declining recognition" of the importance of keeping records in the digital age, which needed to be addressed. '"That type of mentality, Let's just move with today, that doesn't wash I'm afraid - I think there's got to be a much more profound responsibility on organisations to make sure that the records are maintained, and maintained in a way that they can be accessed at some time in the future."'
Survey finds worrying holes in management and accessibility of public records, *Stuff* January 28 2020 (Stuff, 2020)
'The Public Records Act (PRA) requires public offices to keep full and accurate records to enable the government to be held accountable. As [the chief archivist] put it, it's the flipside to the Official Information Act. You can't access information under the OIA if it was never created, if it's been filed in the wrong place, or if it's stored on ancient technology that has never been updated. Of the 189 complaints to the ombudsman in 2017-18, 82 related to information that did not exist, or could not be found…. '…[chief ombudsman Peter Boshier said] "I believe there is no excuse for poor record-keeping in the digital age. We have moved on from the days of officials placing sensitive files in the back of a basement filing cabinet in the hope they wouldn't be seen."… …'More than half the public servants replying to [a] survey said a lack of contextual information – think filing categories, keywords, metadata – risked preventing records from being discovered and interpreted.'
Record-keeping in public sector still worrying — Archives New Zealand report – *Stuff* 9 April 2021 (Stuff, 2021)
'Of 214 respondents in the new report – which includes ministries, departments, councils, district health boards, parliament offices and education entities – just 39 per cent had identified

information that they hold which was of importance to Māori. “It is a significant finding”, Chief Archivist Stephen Clarke said….

…’Other findings included 53 per cent of respondents saying they “definitely” or “possibly” had digital information which was inaccessible due to being stored in personal systems, or because of inadequate metadata and obsolete file formats….

…’Clarke said the public sector was still grappling with challenges of the digital environment, with many still “running to catch up”…and there were real-world consequences of inadequate record-keeping.’

4.2 Innovation, data and research sharing

Use cases
Inefficiency of new information systems design – common semantic and data standards are not available BEFORE people in separate agencies re-invent the wheel.
Examples: There are so many of these that it is typically the norm. Of high and urgent interest is the recently announced Health Reforms being overseen by the DPMC Health and Disability Review Transition Unit. The existing health data and administration environment is described as chaotic by senior information managers and architects, and an ontology supported semantic model will be essential as the sector gets to grip with potentially massive administrative changes (See DPMC, 2021).
Old knowledge and research of value get left behind or lost - especially in agencies merging/splitting and during technology replacements
Example: MPI Science Network - when multiple departments and systems were merged in 2012 many research data sets and documents were not migrated to new systems and now 9 years later many scientists cannot find their own earlier research, in PDFs in particular, as they are lost in the abandoned shared drives.

Story: Figure.nz
Figure.nz’s mission is to get the people of New Zealand using data to thrive. ‘We make it easy for everyone to find and use our country’s numbers for free, through our website, as well as helping people learn how to use those numbers in their day to day lives.’ Pātaka Raraunga is a Figure.nz website which helps Māori to access data on education, health, the economy and tourism, and includes a selection of tools and reports to help people put that information into context. Figure NZ typically download datasets from government websites including data.govt.nz and make them available using normalisation and visualisation tools. However, the lack of data definitions and standards means they typically spend several days renormalising and reformatting data sets before they can apply their own value-add.

Ngapera Riley, CEO of Figure NZ said:

> ‘Data is quite scary for some people, and it’s not just about pulling the data together, it’s about making it easier to find. There are 185 government agencies that collect data and most of them put it out in various sources, so we were trying to solve that problem with technology.’

But sorting ‘Māori data’ from ‘non-Māori data’ was not a straightforward process in the development of Pātaka Raraunga. The environment section, for example, rarely references statistics that directly involve Māori, but Riley explains this is an issue because the environment in Aotearoa is inherently a Māori issue.

> ‘How do we define what’s Māori data and what’s not? You’re so connected to the land and the water and the forests as Māori, so the environment was something we put in there but there was no Māori-specific data. Hopefully that’s something that can change.’

Story: Map of Agriculture

Andrew Cooke, founder of Rezare Systems and CTO of Map of Agriculture:

> ‘Map of Agriculture is a leading provider of insight and knowledge to the agri-food supply chain, delivering pioneering analytics, modelling and research. By working closely with farmers, we are able to gather valuable data, insights and opinions that benefit the agricultural industry as a whole.’

The three main wellsprings of information and data design have included:

- science models from academic research teams mapping science vocabularies to the farmer vocabularies from a very research-specific orientation;
- software providers including ex-farmers which are usually small start-ups and use their existing farms vocabularies;
- co-operatives owned by farmers, with a broad remit 'to help farmers' but the co-operatives’ vocabularies are typically their own, specific business vocabularies.

There are some specific issues which are prevalent in the NZ agriculture sector:

- These businesses keep people employed in rural areas.
- New Zealanders like to give it a go ourselves.
- They need luck to have the right people in the room, and getting agreement and conferring can mean expensive conferences, and are susceptible to our own parochial feedback loops.
- Poor internal knowledge management practices mean it isn’t uncommon to lose previous work, especially losing original data sets and having to rely on a precis in published papers.

> ‘Data standards for agricultural data were proposed back in the late 80s/early 90s and lack of data standards and models including ontologies, non-standard vocabularies, uncontrolled lists, have meant time and cost issues for farmers and data analysts.’
>
> Andrew Cooke, Map of Agriculture

5. How can Ontology help?

How does an Ontology fit into and support solving the problems and mitigating the risks outlined above?

Major data and information companies such as Google, Microsoft, Facebook and Amazon have long utilised ontologies to underpin their core business by supporting search, linking and surfacing data, and analytics. But the benefits for any organisation can include:

- semantic search which provides context to terms (results for a search on 'depression' can distinguish between the geographical, physical, psychological and historical contexts for the term);
- great improvement to the findability of information by storing synonyms, abbreviations, te reo versions, codes, acronyms, and other equivalences so that users can find content regardless of the labels they use for concepts;
- support for deep finding which enables people to navigate data concepts like they navigate the web using the relationships built into the ontology;
- support for faceted search and filters;
- ability to combine simple facts in order to infer new facts;
- provision of 'ready-made' models and vocabularies which can be human and machine-readable;
- auto-categorisation and metadata tagging supporting the automatic addition of tags to unstructured content to help users find content more easily and precisely;
- entity, fact and concept extraction to turn unstructured documents and content into data which can be queried;
- knowledge graphs which enhance a search engine's results with information gathered from a variety of sources;
- web design systems with publicly-available components for quick deployment;
- semantic enrichment of data in data warehouses or lakes for higher quality and analysis;
- auto-classification of content so that records can be identified as having privacy risks or for sentencing automatically;
- enabling of interoperability through building a common understanding of terms and a cross-government means of identifying them which can be extended to support sharing with the private sector as necessary.

An example of an MPI project run to test out some of these benefits is given in Appendix D.

6. So what is an Ontology?

In information science (as opposed to philosophy), an ontology is one of a number of knowledge organisation systems (KOS) used to support information discovery and sharing. Many people use the word to mean what would once have been known as a controlled vocabulary or authority file, a thesaurus, or more recently a taxonomy, and in fact these terms are part of a continuum from simple word lists to complex networks of concepts.

> There is no clear distinction between what is referred to as "vocabularies" and "ontologies".'
>
> (W3C, 2021)

According to ISO 25964, '……. an ontology is a formal explicit specification of a shared conceptualization ' (ISO 25964, 2011). This is a rather broad, often repeated, but rarely understood definition. A more helpful description is:

> 'Ontologies are formalized vocabularies of terms, often covering a specific domain and shared by a community of users. They specify the definitions of terms by describing their relationships with other terms in the ontology.' (W3C, 2012)

In other words, an ontology models the type of objects and concepts that exist within a given domain, together with their properties and relationships. It identifies the terminology within that domain and categorises the individual terms by assigning a class in order to create relationships between them. It is the formal categorisation, strict rules and relationships that distinguish an ontology from a thesaurus or taxonomy.

To illustrate this, in the following example of a taxonomy, individual terms are presented in a hierarchy, which can be useful for providing a browse structure and could be used as a loose vocabulary for tagging content, especially if synonyms are included behind the scenes (such as in a SharePoint term store). The user may know nothing more about each term other than that it is broadly related to the term above (parent) and the term below (child) in the hierarchy.

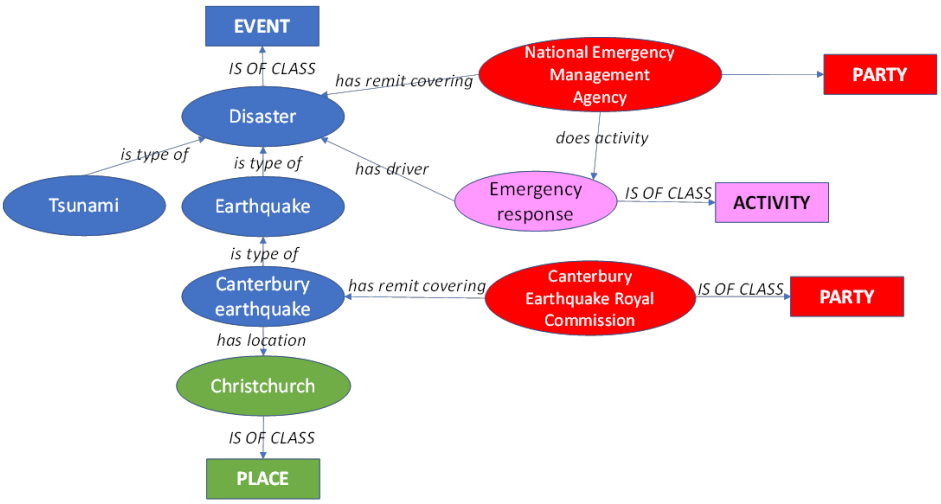
TAXONOMY
Events
- Disasters
  - Earthquakes
    - Canterbury earthquake
      - Canterbury Earthquake Royal Commission
  - Emergency response
  - Flooding
  - National Emergency Management Agency
  - Tsunamis

Places
- New Zealand
  - Canterbury
    - Ashburton
    - Christchurch
    - Kaikura
    - Timaru

In an ontology, the semantic unit (or unit of meaning) is the concept, and each concept has a preferred term, ie the term which is used as the primary label (like 'Flooding' and 'Canterbury' in the

above example). Concepts are grouped hierarchically in ‘concept schemes’ which are sets of concepts of the same type (like all of the place names above). Each ‘child’ term in the hierarchy must be a type of its ‘parent’ (so the parent of ‘Earthquake’ is ‘Disaster’, as an earthquake is a type of disaster).

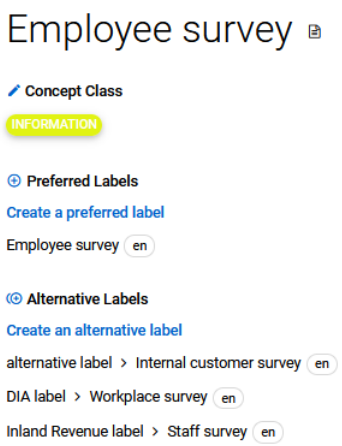
Each concept belongs to a formal category or class which supports the creation of rich relationships between concepts and the assignment of class-based properties. So again using our taxonomic example above, the ontology could contain:



EVENTS could have the relationship *has location* to link it to the PLACE where it occurred; PARTY could have the relationship *does activity* to link it to an ACTIVITY. All concepts within the class of EVENT could be assigned the property of 'Date of occurrence'; all concepts within the class PARTY could be assigned the property of 'Address'.

The ontology can also store synonyms, abbreviations, other language equivalents, definitions, and any other properties which are required for each concept to support findability, interoperability and reasoning.

An example of this, from a possible all-of-government domain of Human Resources Management, is shown below.



All of these aspects of an ontology can be seen in the AOGO Demonstrator (see section 7).

7. Ontology in action – the AOGO Demonstrator

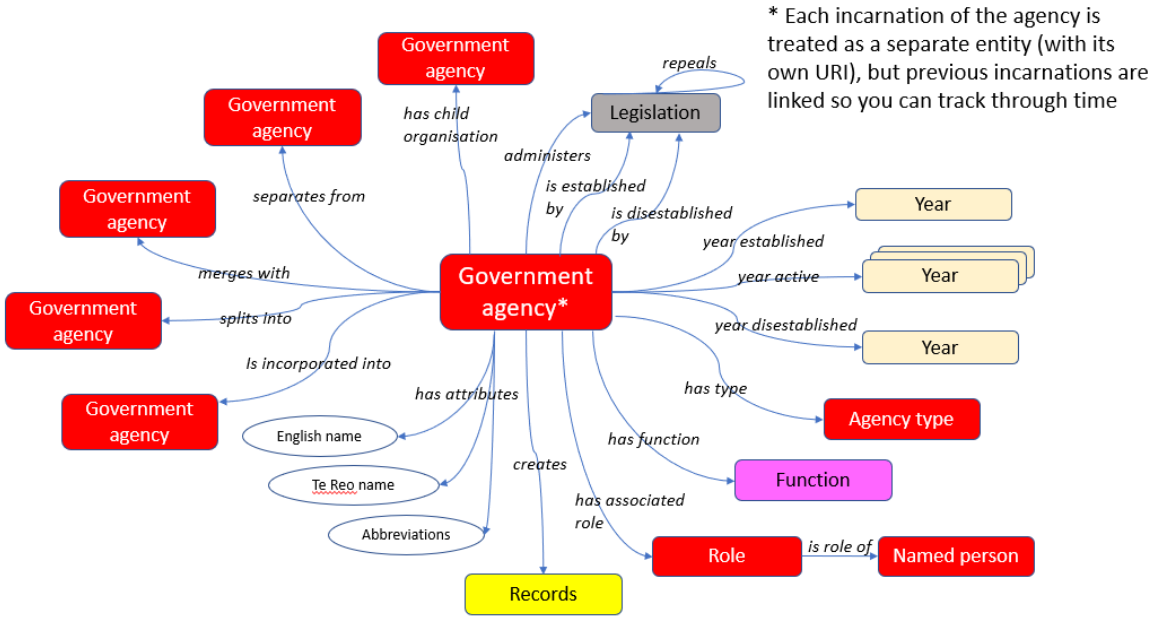
In order to support our thinking for this Options Paper and also to help others to understand the structure and content of an ontology, we built a 'Demonstrator' of 11 high level classes, 24 sub-classes, 45 relationships, 16 term properties and 750+ concepts, using Smartlogic's Semaphore.

The specific goals of the Demonstrator were to show:

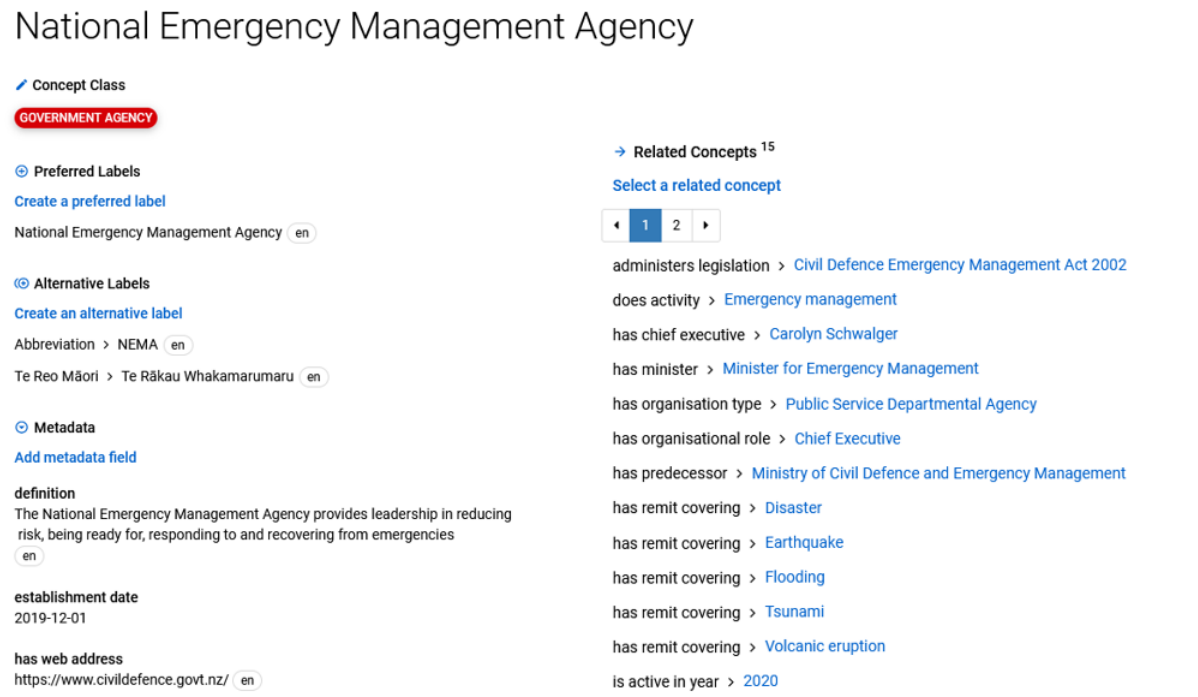
- what an AOGO could 'look like' (from an internal perspective, not as part of user experience design) and how it could be structured, including
  typical taxonomies, terms and relationships (hierarchical, associative, equivalence). In particular we aimed to incorporate the tracking of the names and remits of Government departments through time;
- what the properties of each term could be (eg definition, source);
- governance considerations (eg linked models);
- interoperability considerations (eg URIs, mapping to Agency terms);
- autocategorisation;
- autoclassification;
- inferencing;
- and finally a flavour of some of the semantic issues for an AOGO.

The design of the AOGO Demonstrator was purely for the above purposes: it was not intended or expected that an actual AOGO would use exactly the same structures or features. It was based mainly around the Christchurch Earthquake, as we could use that scenario to cover all of the above goals.

The diagram below shows the domain model used for demonstrating the Government organisational lineage:

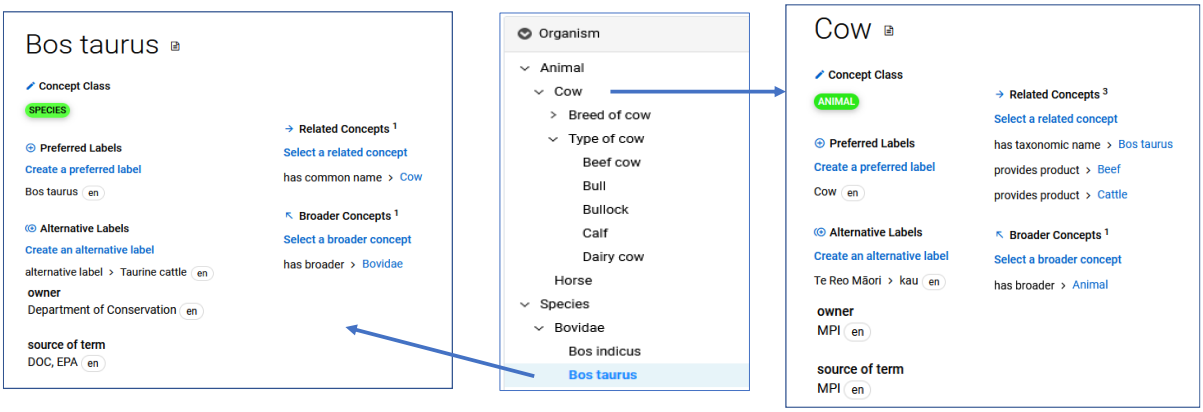


Below is a screenshot from the Demonstrator itself for the National Emergency Management Agency. This gives an indication of the kinds of properties and relationships that can be captured for each concept.



In order to explore some of the issues, we also included something from the natural world as the terms provide a great example of how different agencies view what is potentially the same concept. Our example for this was:

which is labelled as ‘Bos taurus’ in the Department of Conservation, ‘Cow’ at MPI, ‘Cattle’ at Inland Revenue, etc. Are these concepts exactly the same, or is the scientific method of classifying species actually different to the classification of animals for regulatory or taxation purposes?



The ontology needs to offer a way to reconcile such different contexts by harmonising or mapping agency terms, which will not always be straightforward, but will be very important for information sharing and cross-agency searching. This is one example of the kinds of questions that the design of the ontology will have to tackle. Another is how the ontology incorporates the concept of time, since government departments change frequently in name and portfolios, and if we wish to track who was responsible for what when, the ontology must be able to incorporate this facet.

There is a script to accompany the Demonstrator that can be used to search and navigate its content and structure, and to see how it supports the stated goals.

8. What is an all-of-government ontology?

We have described above the typical components of an ontology (classes, relationships and concept schemes within a given domain), but what could this mean for an all-of-government ontology? By all-of-government, we mean containing classes and concepts that are relevant across government departments (at a minimum).

8.1 Previous NZ all-of-government work

Functions of New Zealand (FONZ) and Subjects of New Zealand (SONZ) are two all-of-government vocabularies developed in the early 2000s for use with the New Zealand Government Locator Service (NGLS) metadata standard. The vocabularies are in the form of thesauri and it was intended that the terms would be used for tagging and searching public-facing information. However the terms were not user-friendly (for example 'Ensuring public health' and 'Enforcing legal sanctions and remedies') and neither thesaurus is now used, to our knowledge.

The Government Enterprise Architecture for New Zealand (GEA-NZ) contains reference taxonomies that are also cross-government in scope. Their purpose is to provide a consistent view of government functions and common language, and thereby to support interoperability, findability and data lineage. Each taxonomy includes terms, IDs and descriptions, and has a fairly lose structure in that each section is a collection of terms relating to a specific topic, but the terms themselves represent different kinds of entities (for example under Defence we find 'Military law', 'NZ Defence Force', and 'Defence communications'). GEA-NZ would be an important source of vocabulary for an all-of-government ontology., but is not one in itself, according to our definition.

Other government initiatives have worked on authoritative sources of terminology for specific domains (for example the DPS Commitment 11 data set and the Archives NZ AIMS data model for government functions; the LINZ set of place names; Stats NZ standards and classifications). All of these will be key sources of authoritative terminology: the aim of an all-of-government ontology would be to provide rich relationships *between* the individual terms.

8.2 Relevant work outside of New Zealand

Outside of New Zealand, as far as our research has shown, the need for all-of-government vocabularies to aid the finding and sharing of official information has been generally understood and acted upon since the mid 1990s, obviously coinciding with the rise in use of the internet. Most activities in this area were focussed on helping citizens to locate information and government

services, and produced a variety of thesauri, taxonomies and pick-lists for use in tagging content on web sites.

More recent initiatives continue the focus on findability but are also aimed at information sharing. The following are examples of initiatives in the UK, Australia and the USA.

8.2.1 Government ontology work in the UK

In 2002 a paper by the UK Office of the e-Envoy (a part of the Cabinet office charged with developing an eGovernment Metadata Standard) noted that ontology 'is the new kid on the block' (UK, 2005). The earliest UK pan-government vocabularies (the Government Category List (GCL) in 2001 and its replacement, the Integrated Public Sector Vocabulary (IPSV), in 2005) were still taxonomies for use in tagging agency web content, although agencies were encouraged to use the IPSV for tagging internal content. The IPSV was no longer mandated by 2012 when its home moved to The National Archives (TNA). It is TNA who have been responsible for and active in government ontology work since then, creating the Digital Records Infrastructure (DRI) which includes an ontology covering the structure of records, the agencies which produce them, and the related government functions. The stated purpose is to support new forms of user engagement, participation, data re-use and research.

The National Archives also sponsored the creation of an organisation ontology to support the publication of information on organisations and organisational structures including governmental organisations in linked data. The Organisation Ontology has been developed and standardised within the W3C Government Linked Data working group and has become a full W3C recommendation (W3C, 2014).

It is worth noting another much earlier ontology initiative which was for a specific domain: the Common Basic Specification was developed between 1987 and the late 1990s. It was a generic model describing the functions the National Health Service (NHS) undertakes and the information required to carry them out, with the intention of supporting joined up information management and exchange across the NHS. What makes it noteworthy is the fact that it was a well-funded, major government project which allegedly floundered because (probably among other things) insufficient attention was paid to articulating the benefits, there was little information about how to use the model, it was developed using a particular world view which didn't fit all contexts, and it was based entirely on a top down (conceptual-based) rather than a bottom up (content-based) approach: a good ontology needs both.

8.2.2 Government ontology work in Australia

A similar path has been taken in Australia, in terms of modelling record-keeping and government structures and functions. There appear to have been a number of initiatives in this field, for example:

- the Commonwealth Records Series Thesaurus (CRST) was started by the National Archives of Australia (NAA) in 1999, to support the archiving of government records;
- the Australian Government's Interactive Functions Thesaurus (AGIFT) was developed by the NAA (last updated in 2016) and is a list of commonwealth, state and local government functions presented as a hierarchy. Its purpose is to support public access to government information and services by providing a standard list of terms for tagging
- the Commonwealth Records Series Ontology (CRSO) was adapted from the CRST in 2018 and intended to assist with the attribution of disposal classes and support search;

- the Longitudinal Spine of Government Functions was an output from a project run from 2018-19 which aimed to show how existing datasets could be linked by means of a semantic ‘spine’;
- the Classification of Functions of Government – Australia (COFOG-A) was issued by the Australian Bureau of Statistics and based on the scheme (COFOG) published by the United Nations Statistics Division (draft in 2019);
- the Australian Government Records Interoperability Framework Ontology (AGRIF Ontology, 2020) was developed by the Australian Department of Finance, using three people over nine months, but not currently used. It is thought that the 'command and control' approach to the management and use of the ontology has acted against take-up.

In addition, the Australian Government Linked Data Working Group (AGLDWG) was set up in 2012 as a 'community of Commonwealth Government experts and champions' working to communicate the benefits of Linked Data to individuals, government and business, and to help organisations to create and use Linked Data sets. Members are volunteers, but may be sponsored by their Departments. They have instituted a URI minting service and are currently working on a new all-of-government ontology.

8.2.3 Government ontology work in the USA

In the United States, the focus for government in terms of ontology was originally information sharing between agencies. A paper from the MITRE Corporation in 2004 (MITRE, 2004) gives a detailed review of current ontology models and posits what a pan-government upper ontology could look like.

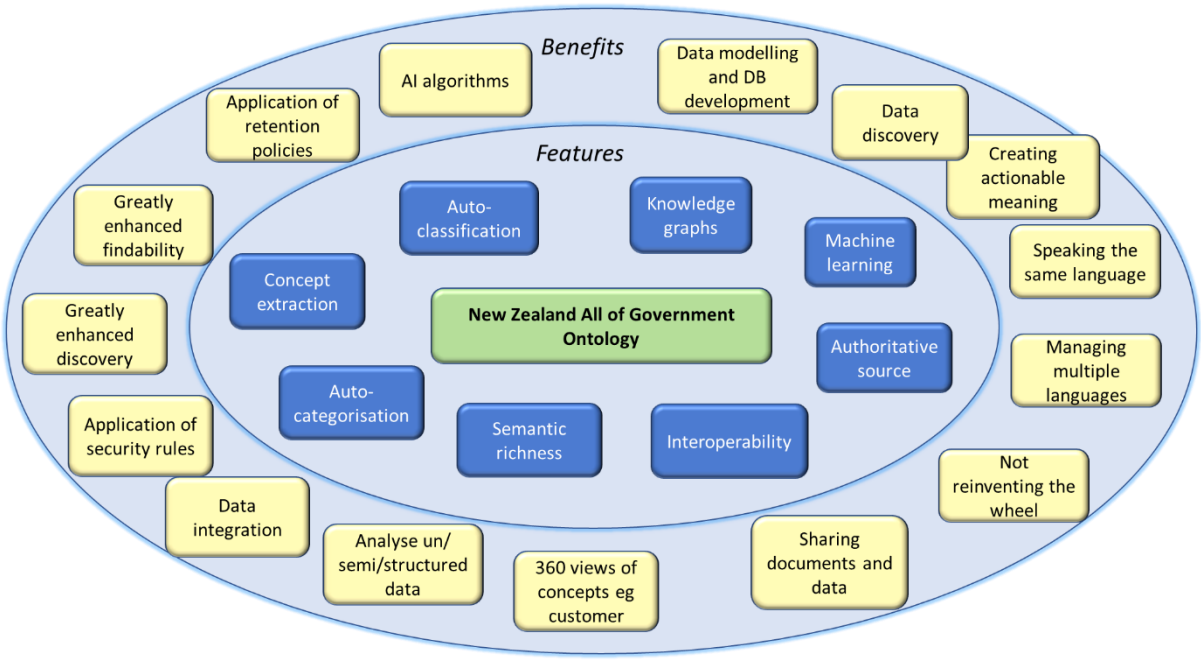
The following year the National Information Exchange Model was started and is used within the justice and security sectors to identify common entities. The fundamental building block of NIEM is the data component, a basic business data element that represents real-world objects and concepts such as people, places, events or material things. Components that are frequently and uniformly used by agencies are specified in NIEM and can then be reused for information exchange, regardless of the nature of the business or the operational context. To become a universal component, consensus by all domains is needed on its semantics and structure, and the baseline set is comparatively small.

A company, TopQuadrant, has developed a number of government-focussed ontologies, published under the brand OE-GOV (www.oe-gov.us). The set includes a Government Core Ontology which defines the basic organisational structure of government, and contains a description of each agency and its reporting line. It is not clear from our research whether a part of the US government commissioned these ontologies or whether they are effectively a set of commercial products. It is also not clear how they are being used.

9. What are the options for an all-of-government ontology?

We have explored some of the problems faced by government departments in finding and sharing information and data, and complying with the Public Records Act. We have also noted the need to make information and data available to non-government organisations in order to promote

transparency and to support research and innovation. The benefits and the features required to support them are summarized in the following diagram:



With this in mind, the following is a set of options for what an all-of-government ontology might be. They are not mutually-exclusive.

Option 1: A hub or repository for government ontologies and taxonomies

This is a single platform which holds existing ontology or taxonomy models and products and makes them available inside and outside of government.

Data.govt.nz already helps people discover, collect, manage, use, share, re-use data. It holds nearly 30,000 datasets and the largest contributors are LINZ and GNS Science, with the three largest categories of data being Land, Local and Regional Government, and Environment and Science. This fairly reflects the current need identified by agencies to provide in freely and widely shared government data in pursuit of the public good.

Senior Product Owner of data.govt.nz, Christian Linnell, perceives an opportunity for an all-of-government ontology to assist as an appraisal step, rather than just accepting all submitted datasets. The current system is largely concerned with privacy and accuracy rather than quality data and tagging which has led to poor findability.

The 'ontology' would therefore comprise a means of:

- categorising the data sets to improve findability;
- applying, or mandating, standards and syntax to improve quality;
- tracking provenance and currency.

'Poor quality data is often the result of the lack of good vocabularies.'

Jochen Schmidt, Chief Scientist, Environmental Information, NIWA

Option 2: A Foundation Ontology

Any ontology requires identification of the high-level classes and relationships that make up its basic structure. To that extent, a foundation (or core) ontology is a sine qua non, but the importance for an all-of-government ontology is that these classes and relationships must be relevant to and appropriate for all agencies (and their partners) to provide the fundamental conceptual building blocks on which more domain-specific ontologies and taxonomies can be built. It will therefore be important to get cross-agency agreement on what these components are.

> 'The core ontology allows the representation of the multi-faceted domain knowledge of varying sectors to highlight deeper understanding.'
>
> Rosemary McGrath, Chief Architect, Stats NZ

> 'Government needs world understanding or sense making. If we use data, then the world must be represented in the data.'
>
> Aaron Jordan, Chief Digital Officer, LINZ

The Foundation Ontology must therefore be as generic and objective as possible.

As well as providing the conceptual basis for domain or agency-specific ontologies, the Foundation Ontology should also support information sharing. It will therefore need to include:

- Unique Resource Identifiers;
- agreed entities;
- definitions wherever possible;
- provenance (where did this term come from and can I trust it?);
- owners, who are responsible for quality and maintenance;
- easily sharable formats with good documentation;
- ideally, mapping between different agency terms for the same concept.

> 'To be effective agencies need to connect across an ecosystem of government organisations, third parties, iwi and trusted intermediaries.'
>
> Colin Holden, Digital Public Service, DIA

> 'Shared language and shared understanding would be a good start.'
>
> Honiana Love, Chief Executive, Nga Taonga Sound & Vision

Option 3: The Government organisational lineage

A Government 'organisational lineage', structured as an ontology, which shows which agencies were and are responsible for what over time would be hugely beneficial to support the finding of records and information of importance to the searcher. This has been a key requirement for some time, but has become particularly important for citizens trying to locate information about events that have seriously impacted their lives.

We believe that the modelling for a Government lineage would be relatively simple as there are a number of useful models around the world and in New Zealand to help with the design, and some agencies have collected data that can be used as input. In particular the AIMS model at Archives NZ, the Government A-Z (https://www.govt.nz/organisations/) and the OGP Commitment 11 model and dataset will be very important inputs, as well as the W3C standard for organisations (W3C, 2014) and the Australian AGRIF-O model.

The AOGO Demonstrator shows some potential classes, relationships and sample data. While the actual classes of the Government organisational lineage would need to easily fit with any wider model (see above: Foundation ontology), the fact that this is a clearly defined domain of agency, function and portfolio which, as stated, has been modelled before, should make this relatively straightforward.

The Government organisational lineage would need to be built in a tool which supports designing and building the model and associated taxonomies, and making them available for human and machine consumption.

Option 4: A set of domain-specific ontologies

Designing and building domain-specific ontologies will be important for improving findability and sharing between agencies within a specific sector; these should be built within the framework of the Foundation Ontology in order to ensure wider interoperability and findability.

Good examples of particular domains with urgent needs are health and housing. As mentioned earlier, health reforms planning is underway via the DPMC Health and Disability Review Transition Unit (DPMC Transition Unit).

> 'Initial work should include prototypes to involve particularly challenging "unstructured or semi-structured" data domains such as GIS project files and data architecture artefacts, which are deeply embedded in specific technology and restricted in use by software licence.'
>
> Liz Kolster, Senior Advisor, Statistics New Zealand

We suggest an immediate outreach to DPMC Transition Unit to discuss a Digital Public Service/Archives NZ-led approach to foundation and health domain knowledge and ontology modelling as a pilot project. Martin Frauenstein, Principal Architect at the Ministry of Health and Alastair Kenworthy, Chief Standards Advisor at the Ministry of Health are key supporters of this approach. Housing is another area where reorganisations and pressing social need make this an ideal time to look at introducing semantic tools that will support efficiency and findability. If at all possible, this second domain pilot should be undertaken alongside health.

Option 5: A starter kit for automatic categorisation and classification

Combining an ontology with a rulebase can support the automatic application of subject tags and disposal classes to documents and data sets.

All government agencies have support functions such as human relations management, financial management and IT management. While there will be some differences in how these functions are referred to, the actual activities will be broadly the same and as such fall under the same General Disposal Authorities (currently GDA6 and GDA7). It is therefore likely that a GDA6 'starter kit' for automatically assigning support activities, and therefore assigning GDA6 classes could be built (indeed some work has already been done in this area by a third party).

The 'starter kit' could then be built on by individual agencies to add terms and properties relevant to their particular business functions and disposal authorities.

Option 6: Ontology Service

An ontology service could be a hub or repository for more than just government ontologies (or *the* government ontology). It would identify a range of utilities and services which can be used by government ontology builders, contributors and customers, including third parties. These utilities might typically include auto-categorisation engines, text crawlers, knowledge graph engines, visualisation tools, ontology standards, and analytic tools. A brief survey of common, relevant and widely available semantic technologies covering ontology management tools, knowledge graphs and auto-categorisation tools is given in Appendix E. Dr Amin Haller of the Australian National University has also drawn our attention to a suite of tools that were built or acquired for use with the AGRIF ontology, and which could be made available for an NZ all-of-government ontology.

Where agencies cannot or will not make use of the semantic technologies to comply with Archives NZ requirements, an autoclassification service could be available to apply semantic rules and retention policies to triage documents and data sets for historic and archival value**.**

> 'Integrated services are enabled through digital foundations that can be used across the public service, making it possible to reuse data, rules and transactions, as well as government-wide standards and frameworks.'
>
> Strategy for a Digital Public Service (DIA, 2020)

Option 7: Third party platforms

The Azure cloud and associated technology stack (e.g. Syntex and Cortex, advanced artificial intelligence tools in M365) have the potential to give the New Zealand government an unprecedented opportunity to explore applying information lifecycle and analytical tools across government information, based on incorporating the Foundation Ontology (and potentially domain ontologies). This could be explored to look at enabling and automating processes for autocategorisation and application of archival business processes.

10. Recommendations for an NZ all-of-government ontology

New Zealand is a relatively small country with a centralised government. Although the scope of government activities is as broad as almost any other country, the potential for co-operation and shared approaches is higher than for most. So much good work has been done to model particular domains, create vocabularies and ontologies, and make data available; tools and standards to support findability and interoperability are increasingly available; there is expertise at home and abroad to advise and contribute; agencies have increasing need for sophisticated information discovery tools; but few agencies have the skills and resources to develop and maintain these tools themselves. In short, the time is right to build the kind of semantic eco-system that an all-of-government ontology (and related services) would provide.

Our recommendations are therefore:

- to design and build a Foundation Ontology covering the key entities relevant to government;
- within that design to prioritise the development of the classes, relationships and properties required by a Government organisational lineage, so that work on the concept schemes can begin as soon as possible;
- to also prioritise the development of classes, relationships and properties for two specific domains: health and housing;
- to build on the Foundation Ontology to create terms, term properties and relationships which support autocategorisation and autoclassification when combined with the appropriate rulebase and algorithms;
- to create a business unit responsible for creating and managing the all-of-government ontology and related tools (see further below);
- in due course, to offer a service which incorporates the guardianship and publication of authoritative models and taxonomies, and the provision of a suite of utilities for use in creating, using and disseminating taxonomies.

> 'The core ontology and [lineage] are key enablers of the broader Government Enterprise Architecture – New Zealand, Data, Information and Analytics Domain, specifically the authoritative and reliable multi-contextual models and vocabularies.'
>
> Rosemary McGrath, Chief Architect, Stats NZ

In light of some of our recommendations and some of the hindrances and constraints noted in the following section, a phased approach to build capability may dictate the sequence for development of the recommended components.

Section 12: All-of-Government Ontology Roadmap outlines in which order we suggest that capability be addressed.

11. A Government ontology capability

> 'A national vision gives permission in agencies to do the work, but you need to

keep it modern and not get stuck in old school thinking.'
Doug Lambert, Data Strategy and Governance Lead, Inland Revenue

11.1 AOG Ontology Office

In order to plan, design, build, maintain and govern an all-of-government ontology (and associated services) it will be crucial to put the appropriate organisational structures and capabilities in place. As stated above, our recommendation is therefore to set up an Office or Unit which will provide the required leadership and organisational and secretariat support, as well as playing a coordinating role with government agencies and the community.

During the discussions and research for this Options Paper, we considered where such an Ontology Office could be sited. The Digital Public Service appears to be a natural and useful home for the Office secretariat and is well placed to lead cross-agency efforts. It has several groups which have some specific overlaps of skills and technologies relevant to ontology design as well as motivated and interested staff.

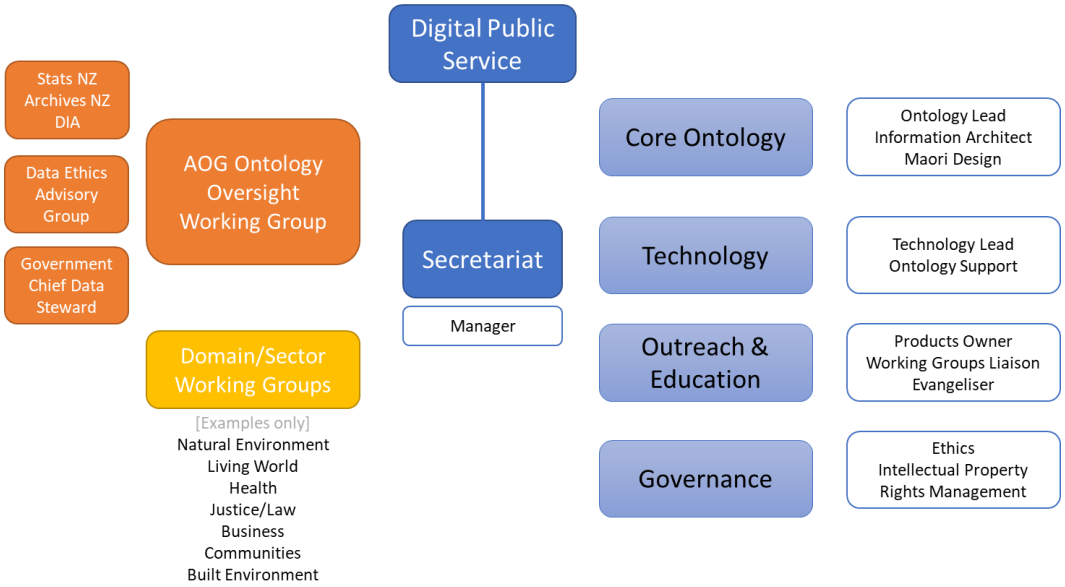
The functions within the Office should include:

- Ontology design and basic build
- Technology and standards
- Outreach and education
- Governance

Representation of te Ao Maori should be integral to all areas.

'The digital transformation of government will be culturally inclusive — particularly in reference to te ao Māori and tikanga concepts.'
Strategy for a Digital Public Service (DIA, 2020)

The AOG Ontology Office would be comprised of a Secretariat to manage the day-to-day running, an Oversight Working Group, and working groups in each domain or sector.



11.2 Key All-of-Government Ontology Office roles

Role	*Description*
Manager	Administrative lead for the Office. Investment process support for Technology Platform. Needs good management skills and a good understanding of all aspects of benefits realisation and ontology development.
Ontology Lead	Leading on the information design, standards, ontology standards etc. Mentor to other information architecture specialists. Contributes to publications and outreach. Needs in depth experience of ontology design and development, and good knowledge of relevant standards.
Technology Lead	Leading on technology relationships with Microsoft, ontology platforms, APIs and other publishing platforms. Needs good understanding and experience of integrating semantic tools into technology stacks.
Technology Support	Ongoing support and management of the ontology platforms.
Information Architect (Semantics)	Supporting Ontology Lead with ontology outputs e.g. Government organisational lineage. Liaises with domain/sector efforts. Builds and ingests ontologies into technology platform. Needs extensive experience of designing and building ontologies.
Information Architect (Technical)	Building schemas and queries, rule-base development, wrangling data, advising other government departments. Needs extensive experience of XML, graph databases, query languages. Ideally has experience of Ontology Language for the Web (OWL).
Data Governance Lead	Developing Governance frameworks. Creating and managing IP agreements with contributors and users.

Maori Data Design and Governance	Leading on inclusion of Te Ao Māori concepts and relationships with Māori information practitioners.
Champion and Educator	Leading on publications, speaking engagements (nationally and internationally). Liaising with government agencies re uptake of ontologies and standards. Developing and leading information architecture training and education for students, researchers, academics, information technology students and practitioners. Needs in depth understanding of the structure and benefits of an ontology, and the patience and enthusiasm to help others in this.

11.3 Ontology oversight working group

A working group comprising experts from Stats NZ, Department for Internal Affairs and Archives NZ should be available to advise on priorities. This could be a new working group, or part of the responsibilities of an existing senior group. Governance oversight would be provided by the Digital Public Service and additional input from the Government Chief Data Steward and the Data Ethics Advisory Group should be sought or available as and when needed.

11.4 Domain/sector working groups and owners

Each domain or area of the ontology will need a working group of concept stewards who are subject matter experts with some level of authority within their area. These domain groups will be responsible for identifying the appropriate owners of concept schemes (for example owners for the lists of medical conditions, offences or place names), and for harmonizing concepts and resolving any conflicts of terminology.

There are some clear candidates among government agencies to be responsible for particular concepts – here's an example of 'Person' being a core concept, but with the Ministry of Health being the concept steward for 'Patient'. The following are some suggestions for stewards for particular concepts.

Ontology Concept	**Potential Concept Steward**	**Working Party Members**
Foundation Ontology Classes	AOG Ontology Office	
Government Agency	Archives NZ	
Government functions	Public Service Commission	
Business	MBIE	
Person	AOG Ontology Office	
Patient	Ministry of Health	HNZ, ACC, MHA, Health & Disability Commissioner, Medical Council, MOJ, GP College etc.
Taxpayer	Inland Revenue	MSD
Citizen	DIA	
Resident	Immigration	Inland Revenue
Worker	MBIE	
Student	Ministry of Education	
Authority		
Legislative and Regulatory	NZ Legislation (Parliamentary Counsel Office)	
Location	LINZ	
Medical condition	Ministry of Health	MSD, Inland Revenue, ACC, Medical Council etc

11.5 DPS capabilities for an all-of-government ontology

The Digital Public Service (DPS) branch of the Department of Internal Affairs works with and through agencies to drive and deliver customer-centred digital government. Its role is to support and enable agencies to progress a modern and adaptive public service that can respond to New Zealanders' expectations of faster, more agile, accessible, inclusive and integrated government services. It does this by focusing on building agency capability to work together as a public service, identifying opportunities to design joined-up, responsive and robust digital services that accelerate our shift to a unified public service.

The branch supports the role of the Government Chief Digital Officer (GCDO) in its broader context of driving a more unified digital public service, framed in the Strategy for a Digital Public Service.

The DPS supports the role of the GCDO in particular in setting digital policy and standards as identified in the Strategy for a Digital Public Service (https://www.digital.govt.nz/digital-government/strategy/strategy-summary/strategy-for-a-digital-public-service/).

> 'The public service is well on the way to helping people find the information and services they want and need from government. However, the current system is geared towards delivery by individual agencies rather than as part of an integrated system.'
>
> *Strategy for a Digital Public Service* (DIA, 2020)

11.6 Reviewing standards and existing work

Initially the Office will need to take cognizance of existing initiatives in NZ and around the world, in particular the importance of standards and not reinventing the wheel. A proper review of what is out there already would be an essential first step. Just some of the primary candidates would be:

Existing New Zealand ontologies, for example:

- Inland Revenue ontology
- Ministry of Justice ontology
- NZTA ontology
- MPI farming and pest control ontologies
- Department of Conservation ontology
- MBIE buildings ontology
- Local Government NZ ontology project

Other New Zealand models, initiatives and standards, for example:

- Stats NZ models and vocabularies
- GEA-NZ
- DPS work on the Commitment 11 dataset and other initiatives
- the Archives NZ Maori Metadata project
- Maori Subject Headings
- Health Information Standards

- Land Air and Water Aotearoa (LAWA) work
- Geospatial Data Standards
- Archives NZ AIMS model

International standards, for example:

- Organization Ontology
- Friend of a Friend (FOAF ontology)
- Common Data Model
- PROV-O ontology for provenance
- W3C Time Ontology

Non-NZ government initiatives, for example:

- UK Digital Records Initiative
- AGRIF-Ontology (Australia)
- The suite of ontology tools built to support the development and maintenance of AGRIF.

11.7 Ongoing AOGO governance and education

Clearly ontologies are, like the world they represent, constantly growing and adapting. Although New Zealand is a small country, we are as complex as any other and the New Zealand Government is interested in nearly every facet of our society, economy and environment.

Once the Foundation Ontology classes are developed, they should not be considered to be complete and 'put on the shelf'. Not only are the technology and standards ecosystems essential for take-up and use, but the ongoing governance and development of ontologies and semantic designs must occur to maintain relevance and usefulness. It is therefore vital that sufficient resource is allocated to these activities.

Another important part of the work will be to provide ongoing education and upskilling of agencies to support their use of the ontology and related tools (See also Section 13: Hindrances and constraints.)

12. All-of-government ontology roadmap

We suggest the following roadmap for development of the all-of-government ontology:

- AOGO Office set-up
- Review of existing work and standards
- Technology selection and implementation
- Foundation Ontology design and build
- Government organisational lineage design and build
- Prioritise domain-specific initiatives i.e. Health and Housing
- Investigation of joint ventures utilising innovation and private sector funding to be undertaken at the same time as other activities.

13. Hindrances and constraints

Although, as discussed elsewhere, ontologies and information architecture has been a known concept in New Zealand since the early 2010s, several issues have either hindered or constrained its uptake, and will need to be considered in order to move forward.

- Education and professional development in information and data architecture for both ontology practitioners and in technology education has been distinctly lacking. An approach to VUW Masters in Information Management was initially welcomed but the project management module was prioritised. A two-day VUW Professional Short Course is run two or three times a year by Judi Vernau and is currently of an introductory nature only.

- There is no mention or availability of ontology or conceptual or semantic design in the University of Auckland Masters in Data Management.

- Ontology expertise and practitioners are rare in New Zealand. Those information and data architects with these skills are typically employed in large corporates and focused on data warehousing. In recent years there has been an increase in the large government departments seeking to recruit information architects, but resources are scarce.

- IT departments lack understanding and commitment to well-designed data and content ecosystems. The poor semantic designs used in new systems cause problems over years and decades for end users which the IT departments are seldom around to see.

- Programme and project management methodologies lack understanding and commitment to information architecture supporting efficient data system design and semantically interoperable data systems.

Appendix A – People consulted for this paper

Name	Organisation	Role
Aaron Jordan	LINZ (Land Information NZ)	Chief Digital Officer
Aimee Whitcroft	NZTA (NZ Transport Agency)	Open Data Lead, Data and Information
Airin Alamsjah	NZTA (NZ Transport Agency)	Principal Analyst
Alan Bell	DIA, Digital Public Service	Executive Director Digital Identity Programme and member of DPS SLT
Alastair Kenworthy	MoH (Ministry of Health)	Chief Standards Advisor
Andrea McIntosh	Whangarei District Council	Team Leader, Information Architect
Andrew Cooke	Rezare Systems/Map of Agriculture	Chief Technology Officer
Anne Marie Cavanagh	DIA, Digital Public Service	Deputy Government Chief Digital Officer and member of DPS SLT
Armin Haller	Australian National University	Associate Professor, Research School of Management and Research School of Computer Science
Arron Judson	Scion Research	General Manager
Ben Joyce	Cyma Enterprise Architecture	Recruitment Consultant
Caleb Moses	Dragonfly Data Science	Data Scientist
Carol Feuerriegel	Inland Revenue	General Manager, Information & Knowledge
Ceara Owen	MBIE	Sector Partnerships - Standards NZ
Christian Linnell	DIA, Digital Public Service	Senior Product Owner
Colin Holden	DIA, Digital Public Service	General Manager, System Strategy and Initiatives
Dale Cousens	DIA, National Library	Team Leader Collection Development (Acquisitions)
Deborah Brunning	Stats NZ	Manager, Data Leadership and Capability
Dennis Lim	MPI (Ministry of Primary Industries)	Information Architect

Name	Organisation	Role
Dion O'Neale	University of Auckland	Principal Investigator, Complex Systems and Network Science
Doug Lambert	Inland Revenue	Data Strategy and Governance Lead
Elynoel Cristoria	DPS Agency Standards and Integration	Standards Publishing Advisor
Genna Grant	NZTA (NZ Transport Agency)	Manager, Corporate Information & Records
Gerard Rooijakkers	Auckland Transport	Corporate Information Manager
Gordon Morris	Te Tumu Paeroa/ MLC/ Te Puni Kokiri	Kaitohutohu Mātauranga Papa Whenua – Geospatial Knowledge Advisor
Honiana Love	Nga Taonga Sound & Vision	Tumu Whakarae - Chief Executive
Ian Cowan	DIA, Technology Services and Solutions	IT Business Partner and member of DPS SLT
James Collier	DIA, Digital Public Service	Director System Settings and Design
Jamie Lowe	NZTA (NZ Transport Agency)	Manager, Operational Policy and Advice
Jane Kennedy	DIA, Digital Public Service	General Manager, All of Government Services Delivery
Jason Krause	Stats NZ	Integration Architect
Jeff Kennedy	eResearch Centre, University of Auckland	Enterprise Architect
Jennifer Andrews	Organisational Capability and Services	Manager Communications and member of DPS SLT
Jim Clendon	DIA, Digital Public Service	Architecture Modelling Consultant (All of Government Enterprise Architect)
Joanne Knight	DIA, Digital Public Service	Identification Consultant
Joanne Koreman	DIA, Archives NZ	Principal Advisor Government Information Management (AOG ontology)
Jochen Schmidt	NIWA	Chief Scientist, Environmental Information
John Machin	Australia Government Dept of Finance	Information Architect, Digital Reporting

Name	Organisation	Role
Jone Garmendia	The National Archives, UK	Head of Cataloguing
Kate McDonald	World Economic Forum	Regulation for AI Representative
Kaye Maree Dunn	Ahau NZ	Founder
Kerri Siatiras	Siatiras Consulting Limited	ALGIM trainer and consultant
Keith Chung	Stats NZ	
Kim Gutchlag	National Library	Manager, Collection Development and Description
Leigh Vollans	DIA, Digital Public Service	Senior Operational Policy Analyst - Digital Identity Programme
Lesley Officer	ALGIM	ALGIM IM Project Lead
Liliana Grace	Figure Foundation	
Lisa Austin	DIA, Archives NZ	Product Owner - AIMS programme
Liz Kolster	Stats NZ	Senior Advisor
Marcus Gustafsson	eResearch Centre, University of Auckland	eResearch Operations Manager
Mark Gahegan	eResearch Centre, University of Auckland	Director
Martin Frauenstein	MoH (Ministry of Health)	Principal Architect
Michelle Edgerley	DIA, Digital Public Service	Strategic Advisor Level 1
Ngapera Riley	Figure.nz	Chief Executive Officer
Nick Cater	Microsoft NZ	Data and AI, Solution Specialist
Noni Oldfield	Inland Revenue	Senior Enterprise Domain Architect
Paddy Power	Stats NZ	Senior Manager - Data and Information Management
Paul Hassall	NZTA (NZ Transport Agency)	Data Manager
Paul Norrie	Callaghan Innovation	Business Innovation Advisor
Ritesh Anand	DIA, Digital Public Service	Manager Agency Standards and Integration
Rosemary McGrath	Stats NZ	Chief Architect
Russell Cook	DIA, Digital Public Service	General Manager Agency Partnerships and Capability & Government Chief Privacy Officer

Name	Organisation	Role
Sara Knight	MPI (Ministry of Primary Industries)	Lead Information Lifecycle Analyst
Shane Simpson	DIA, Digital Public Service	Strategic Advisor Level 2
Stella Dextre Clarke	ISKO - International Society for Knowledge Organisation	Taxonomist (Retired)
Stella Ward	DIA, Digital Public Service	Executive Director Cloud Programme and member of DPS SLT
Stephen Clarke	DIA, Archives NZ	Chief Archivist
Talei Masters	Archives NZ	Senior Archivist
Te Aroha Grace	Figure Group	Chief Relationship Officer
Trevor Himona	DIA, Digital Public Service	Te Pou Matihiko and member of DPS SLT
Vanita Parbhu	Microsoft NZ	Technical Specialist
Vernon Wybrow	DIA, Archives NZ	Manager, Disposal & Acquisition
Victoria Wray	DPS Agency Standards and Integration	Digital Service Design Consultant
Vikash Kumar	NZTA (NZ Transport Agency)	Manager, Data Science
Yvette Wharton	eResearch Centre, University of Auckland	eResearch Solutions Lead

Appendix B – Glossary of Terms and Concepts

API
An application programming interface is an interface or communication protocol between different parts of a computer program intended to simplify the implementation and maintenance of software. An API may be for a web-based system, operating system, database system, computer hardware, or software library.
(https://en.wikipedia.org/wiki/Application_programming_interface)

Autocategorization
The automatic assignment of metadata tags to information objects.

Autoclassification
The automatic assignment of classifications such as retention and disposal classes or security classifications to information objects.

Concept scheme
A set of terms forming a taxonomy within an ontology class, for example all of the place names, or all names of organisations.

Findability
Findability is the ease with which information contained on a website or in a repository can be found, both from outside the site (using search engines and the like) and by users already in the repository.
(https://en.wikipedia.org/wiki/Findability)

Foundation Ontology
A Foundation Ontology (also known as an upper ontology, core ontology, top-level ontology, upper model, or foundation ontology) is an ontology (in the sense used in information science) which consists of very general terms (such as "object", "property", "relation") that are common across all domains. An important function of an upper ontology is to support broad semantic interoperability among a large number of domain-specific ontologies by providing a common starting point for the formulation of definitions. Terms in the domain ontology are ranked under the terms in the upper ontology, e.g., the upper ontology classes are superclasses or supersets of all the classes in the domain ontologies.
(https://en.wikipedia.org/wiki/Upper_ontology)

Example (from the Department of Conservation)

Knowledge Graph
A Knowledge Graph is a model of a knowledge domain created by subject-matter experts with the help of intelligent machine learning algorithms. It provides a structure and common interface for all of your data and enables the creation of smart multilateral relations throughout your databases. Structured as an additional virtual data layer, the Knowledge Graph lies on top of your existing databases or data sets to link all your data together at scale – be it structured or unstructured.
(www.poolparty.biz)

Linked Data
Linked data is structured data which is interlinked with other data so it becomes more useful through semantic queries.
(https://en.wikipedia.org/wiki/Linked_data)

Resource Description Framework (RDF)
The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. It has come to be used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax notations and data serialisation formats. It is also used in knowledge management applications.
(https://en.wikipedia.org/wiki/Resource_Description_Framework)

Triplestore

A triplestore or RDF store is a purpose-built database for the storage and retrieval of triples through semantic queries. A triple is a data entity composed of subject-predicate-object, like "Bob is 35" or "Bob knows Fred". Much like a relational database, information in a triplestore is stored and retrieved via a query language. Unlike a relational database, a triplestore is optimized for the storage and retrieval of triples. In addition to queries, triples can usually be imported and exported using Resource Description Framework (RDF) and other formats.
(https://en.wikipedia.org/wiki/Triplestore)

Appendix C – Options Paper Authors

Judi Vernau

Judi has worked in the field of information architecture for more that 40 years, and was co-founder of Metataxis Ltd in the UK in 2002. Her particular expertise is in content structuring, taxonomy and ontology, and she has worked for clients across the public, private and third sectors, in the UK, Europe, USA , Hong Kong, New Zealand and Australia.

Judi came to New Zealand in 2013 to develop an ontology for the Department of Conservation, and has since designed ontologies for Inland Revenue, Ministry of Justice, Ministry for Primary Industries and Foodstuffs. She was a visiting lecturer at London City University and currently runs training courses in Information Architecture at Victoria, University of Wellington.

Liz Wilson

Liz trained as an archivist in the 1980s and has gradually made her way through records management, document management, information governance to information architecture in both NZ and the UK private and public sectors.

Since returning to NZ in 2012, she has worked in Information Architecture for the Department of Conservation, Foodstuffs North Island and Ministry for Primary Industries. She is currently Principal Advisor Information for Kainga Ora.

Appendix D – An ontology in action – an MPI proof of concept

In 2019 the Ministry for Primary Industries developed a proof-of-concept to:

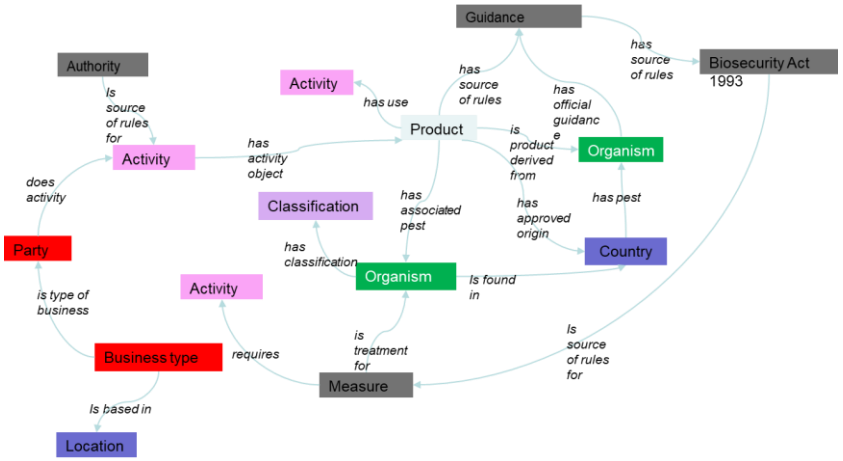
- translate complex information into searchable ideas;
- demonstrate how fact extraction can identify data and complex relationships contained in the Import Health Standards (IHS) documents;
- demonstrate how the extracted data can be combined with a concept model to drive a user interface.

The scenario was: 'I want to import coconuts from Tuvalu – what do I need to be aware of?'

Feedback from importers and MPI advisors was that the IHS were hard to find, read and interpret.

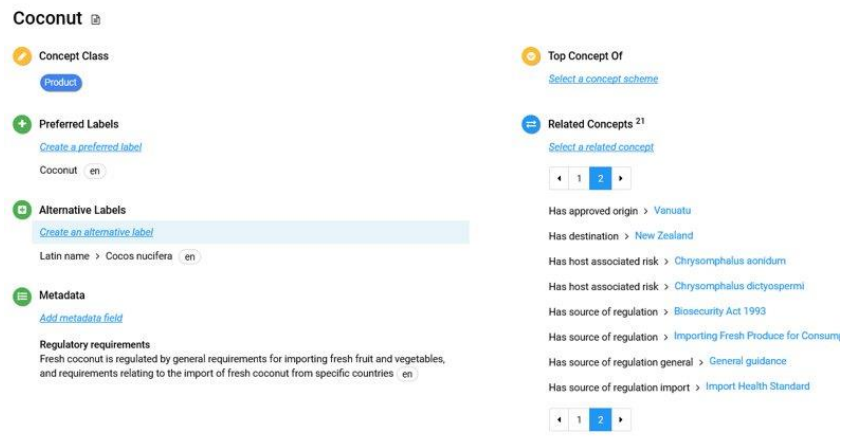
All the information was contained in the IHS in document form, and did not exist in a query-able data structure or knowledge base. The manual effort to create a data structure was considered significant, and a method for automated data extraction was sought. The solution generated a graph database that could be queried and created a user interface that presented this data dynamically and allowed exploration of the data that is searchable.

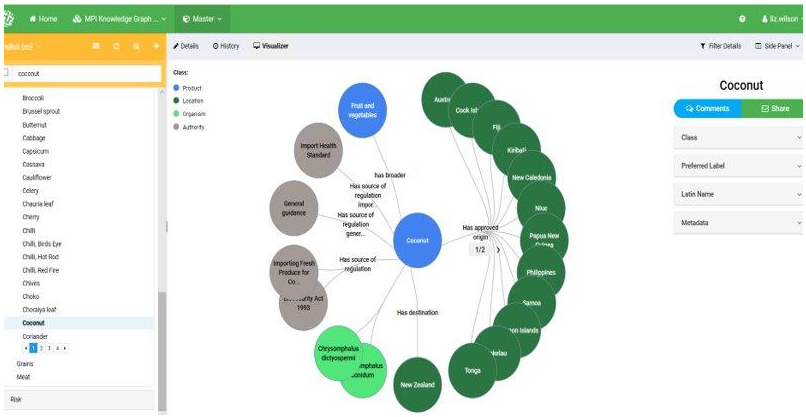
A model was created to help figure out the concepts within the IHSs, including the relationships between pests, importing, organisms and locations .



The resulting ontology was captured in Semaphore Ontology editor.

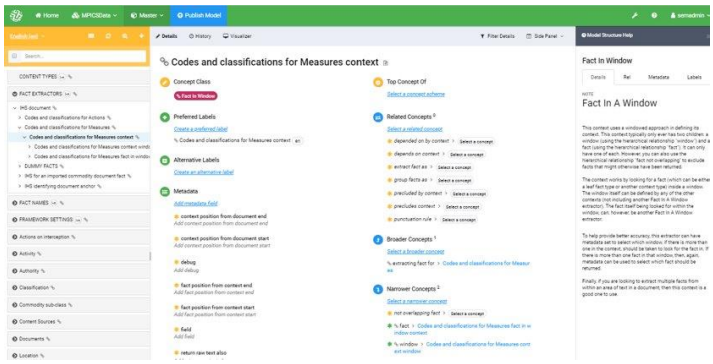
Concept Model for Coconut-Tuvalu





Data and relationships were extracted from the IHS.

Fact extraction rules were created to extract pest, measures and actions from IHS documents

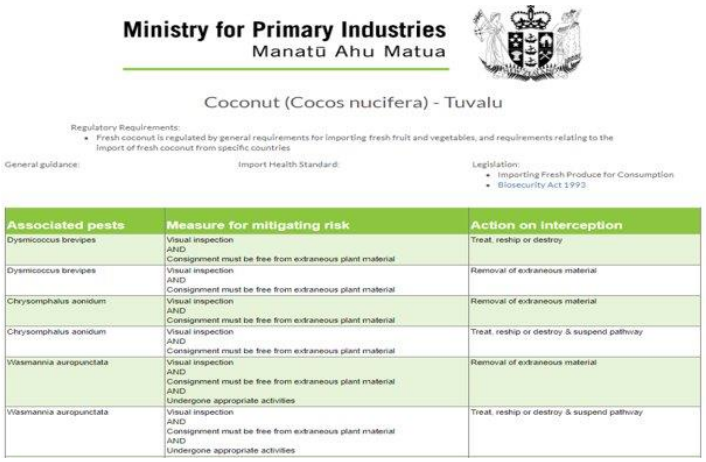


Ontology Editor stored the data in a graph database in triples.

The facts extracted from the IHS documents were loaded into the Ontology Editor graph database

A special SES API called Concept Mapping extracts concepts from natural language, and provides the 'did you mean' type query functionality

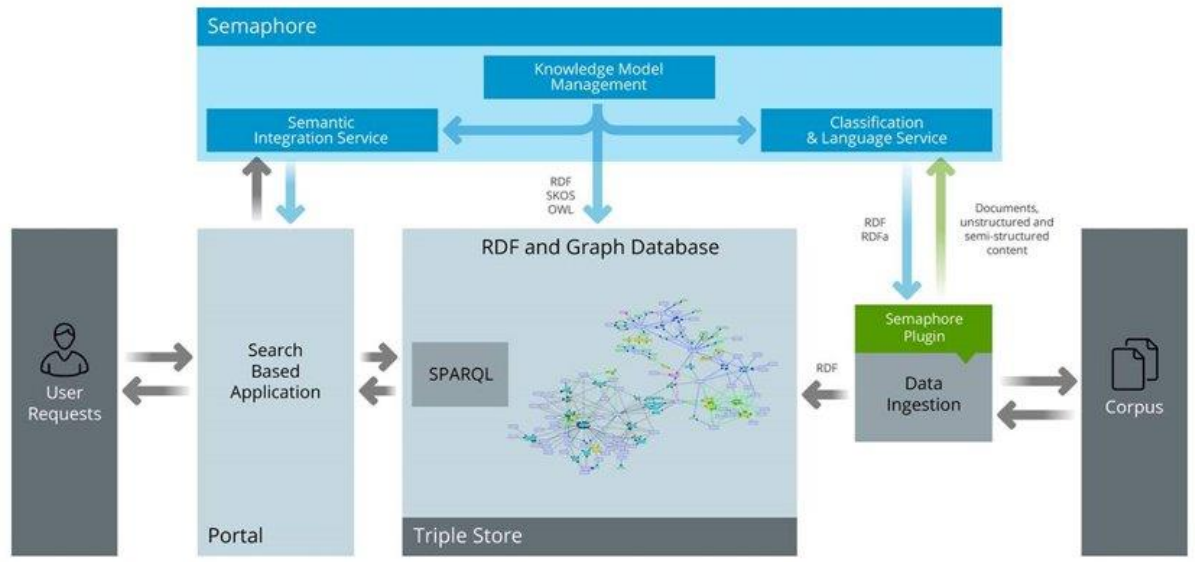
When submitting an enquiry about Coconut and Tuvalu, relevant pests found in the commodity and from that country, as well as measures at shipment and actions on interception, can be displayed.



Actions on Interception

Associated pests	Action on interception
Dysmicoccus brevipes	Treat, reship or destroy AND/OR Removal of extraneous material
Chrysomphalus aonidum	Removal of extraneous material AND/OR Treat, reship or destroy & suspend pathway
Wasmannia auropunctata	Removal of extraneous material AND/OR Treat, reship or destroy & suspend pathway
Paratrechina vaga Forel	Treat, reship or destroy AND/OR Removal of extraneous material
Paratrechina bourbonica	Treat, reship or destroy AND/OR Removal of extraneous material
Tetramorium simillimum	Treat, reship or destroy AND/OR Removal of extraneous material
Neotermes rainbowi	Treat, reship or destroy AND/OR Removal of extraneous material
Pseudoepicoccum cocos	Treat, reship or destroy AND/OR Removal of extraneous material
Graeffea crouanii	Treat, reship or destroy AND/OR

The following architecture underpinned this application:



Appendix E – Brief survey of semantic tools

The table below gives a brief overview of currently available semantic tools which could be used to create, present and use an ontology.

OM = Ontology management
AC = Autocategorisation/autoclassification
KG = Knowledge graph capabilities

Tool	Overview	OM	AC	KG
Amazon Neptune -	https://aws.amazon.com/neptune/ "Amazon Neptune is a fast, reliable, fully managed graph database service that makes it easy to build and run applications that work with highly connected datasets." • Commercial product • W3C ingestible and can build in SPARQL, OWL and RDF amongst others • Can import directly or integrate from relational databases • Has a HTTP REST endpoint for connecting on to other visualisation tools if required • Highly scalable • Big community of users and support information			x

Tool	Overview	OM	AC	KG
Cognitum Fluent Editor	https://www.cognitum.eu/semantics/FluentEditor/ "Fluent Editor, an ontology editor, is an Award Winning comprehensive tool for editing and manipulating complex ontologies that uses Controlled Natural Language. Fluent Editor provides one with a more suitable for human users alternative to XML-based OWL editors. " • Fluent Editor is free for individual developers, open source projects, academic research, education, and small professional teams. • W3C standards, lots of plugins • Has diagrams similar to Microsoft Visio • Interoperability with Protégé • Supported by Cognitum AI (company) • Collaborates with Protégé- could provide ongoing ways for all industries to contribute and amend/improve/extend	x		
Data Harmony MAIstro suite	https://www.accessinn.com/data-harmony/ "Data Harmony (DH) 4.0 is a fully customizable suite of software products designed to maximize precise, efficient information management and retrieval. Our core and extension modules focus on everything from taxonomy construction to automatic indexing, database records management, information retrieval, and more." • Commercial product • Thesaurus based with hierarchies and conditions for using them. • Java based. • Rules based. • Indexing of existing records at a rate of 6-10 records per hour. • Has existing APIs with Microsoft SharePoint 2010 and 2007, MarkLogic, OpenText, MuseGlobal, Oracle, and SAP. • TCP/IP for transmission of 1,500 bytes each. Can be accessed by users with password access from anywhere, and is scalable.		x	
Ellie	https://ellie.ai "Ellie is a cloud-based intuitive visual diagramming tool with enterprise-level data modeling and information architecture features." • Commercial product • Clear and easy to get started • Narrative in glossary easy for anyone to maintain • Works in a browser so no installs required • Can import and export • Has APIs • Integrates with data warehousing tools • Can share the data model/s across the whole organisation • Slick visualisation, auto look up for existing terms	x		

Tool	Overview	OM	AC	KG
Grakn	https://grakn.ai/grakn-core Recently rebranded as Vaticle TypeDB Knowledge graph modelling language and associated query language (Graql, now TypeQL) • Open source • Has inference, analytics, rules, strong GitHub presence • Need to build it directly in Grakn • Enterprise version also available with costs associated but also lots more modules • Can be deployed on premises or in cloud – Amazon Web Services and Google Cloud already set up for deployment • Can text mine straight to knowledge graphs, ie like a clever Wordle • Unclear if the enterprise version can import ontologies or if they need to be built directly inside the software			x
Microsoft Project Cortex	https://techcommunity.microsoft.com/t5/microsoft-365-blog/introducing-project-cortex/ba-p/966091 https://resources.techcommunity.microsoft.com/knowledge-content-services/ • Commercial product • Microsoft's approach to knowledge management • Azure Information Protection scanner searches for all content on premises • Autoclassify content as it loads, using machine learning models previously trained/taught (by whom?) • Can ingest from table storage and many other connectors eg java, python, .net, node, rest and GO • Enables SharePoint Syntex using advanced AI and machine teaching to amplify human expertise, automate content processing, and transform content into knowledge. • With SharePoint Syntex, you can create AI models that capture expertise to classify and extract information and automatically apply metadata. Automate the capture, ingestion, and categorisation of content and streamline content-centric processes. Connect and manage content to improve security and compliance.		x	x
Neo4j	https://neo4j.com/ "Neo4j is a native graph database platform, built from the ground up to leverage not only data but also data relationships. Neo4j has a flexible structure defined by stored relationships between data records." • Has a free community version for starting up then enterprise licence needed for larger scale • Can integrate with existing BI platforms • Decisions based reasoning based on the information in context • Has a load of existing APIs, googledrive,			x

Tool	Overview	OM	AC	KG
	• Requires programming in the software but seems fairly standard • Lots of support material including a book in the "for dummies" series • Very active support community • Scalable			
NeOn	http://neon-toolkit.org/wiki/Main_Page.html "The NeOn Toolkit is [an] ontology engineering environment originally developed as part of the NeOn Project and now supported, together with other technologies from NeOn, by the NeOn Foundation." • Basic version is open source supported by the NeOn Foundation • Basic is based on the IBM Eclipse Platform, has 45 plugins. Cannot see any screenshots, many links on wiki no longer working. Latest downloadable version is from 2011. Level of community participation is unknown.	x		
Netwrix Data Classific-ation	https://www.netwrix.com/data_classification_software.html Previously known as ConceptSearching's ConceptClassifier – the company has been bought by Netwrix • Identifies and classifies sensitive data • Categorisation and classification • Integrates with SharePoint, Oracle, EMC and GoogleDrive • Commercial product		x	
OWLGrEd	http://owlgred.lumii.lv/get_started A graphic editor for OWL • Commercial product • Visualisation and editing • Export as SVG or diagram • Visuals very like MS Visio, simple colouring • Free to trial • Up to date community • W3C standards	x		
PoolParty Semantic Suite	https://www.poolparty.biz/ PoolParty is a platform to implement Enterprise Knowledge Graphs to suppport data analysis, text mining and knowledge discovery • Commercial product • W3C compliant • Can cope with structured and unstructured data • Can export • Named by Gartner in 2019 as a magic quadrant software • Used by Australian government and MBIE	x	x	x

Tool	Overview	OM	AC	KG
Protégé	• Automated reasoning keeps the data in context • Can identify trends along with visualising links between data • Full suite of product options to work with • Modern visualisation • Comes with some existing libraries that you can then extend or load in your own or create your own Has API https://protege.stanford.edu/ "A free, open-source ontology editor and framework for building intelligent systems" • Open source supported by Stanford University with established community • Open source doesn't mean truly free, but is a good option if there is a java developer available • Based on Java • Can use Stanford hosted solution or install on own server • Web Protégé: classic visualisation - a little old fashioned but serves the purpose and easy to input terms, no visualisation, just folder structure • Desktop Protégé still has folder structure plus a better visualisation • All changes logged and you can go back to prior revisions if needed • Users can request changes to terms within the tool so the developer/sysadmin can see and respond as needed • Configurable GUI, API • W3C standards supported • Autocompletion during upload, not full auto-categorisation but minimal input • Can graph using add-on widgets	x		
RAVN-Tech HyperTag	https://github.com/Ravn-Tech/HyperTag https://blog.neotree.uber.space/posts/hypertag-file-organization-made-for-humans "Knowledge Management CLI for Humans using Machine Learning & Tags" • Open source		x	x
SAS Text Miner	https://www.sas.com/text-analytics/ • Commercial product • Can suggest terms to add through interactive GUIs to easily identify relevance, modify algorithms, document assignments and group materials into meaningful aggregates. • Extends text mining beyond basic start-and-stop lists by using custom entities and term trend discovery to refine automatically generated rules and topics.		x	

Tool	Overview	OM	AC	KG
	• Can visualise key topics and related phrases to see how terms change over time and amend the thesaurus/ontology. Shows rule performance. • Can visualise concept links.			
Smartlogic Semaphore	https://www.smartlogic.com/semaphore "Semaphore - a modular software platform - provides the semantic layer in your digital ecosystem so you can manage knowledge models, automatically extract and classify the context and meaning from structured and unstructured information, and generate rich semantic metadata." • Commercial product • Used by several government agencies in New Zealand • Reasonable visualisation • W3C standards • Can auto ingest/autocategorise • Has existing APIs with Microsoft365, Oracle, OpenText, MarkLogic.	x	x	x
TopQuadrant TopBraid Enterprise Data Governance (EDG)	https://www.topquadrant.com/ "A suite of programs. The Vocabulary Management package has the ontology aspects where you can include, extend, enhance and connect existing vocabularies; develop with unlimited hierarchy levels; define classes, attributes, relationship and rules; publish different views for different audiences." • Commercial product • Visualisation fairly standard, folder tree plus process map • Import and crosswalk (join two ontologies) via Excel • Good client list of big and complex companies • Has a technology partner in Australia - Surround Australia, that focuses on the government market • W3C standards	x	x	x
VocBench3	http://vocbench.uniroma2.it/ "VocBench is a web-based, multilingual, collaborative development platform for managing OWL ontologies, SKOS(/XL) thesauri, Ontolex-lemon lexicons and generic RDF datasets." • Open source but funded to be maintained by EU funding • Uses Semantic Turkey - open source knowledge acquisition and management - for business and data access layers realisation • VOC3 released 2017 • Role based access control • Import by spreadsheet	x		

Tool	Overview	OM	AC	KG
	• Uses Apache Maven as software project management and comprehension tool • Requires development skills to keep up to date and initialise • W3C standards			
Vitro	https://github.com/vivo-project/Vitro "Vitro is a "full stack" framework for building semantic web applications. It is not domain specific." • Web based editor • Java web app running on a Tomcat servlet container • Can create or load ontologies in OWL, edit instances and relationships, build a public web site to display data • Created by Cornell University • Big GitHub community • Requires development skills to initialise and keep up to date • W3C Standard	x		

Appendix F – Citations

AGRIF-O (2019) *AGRIF Ontology*. See: https://catalogue.linked.data.gov.au/resource/121, accessed in May 2021.

AIIM (2021) *AIIM State of the intelligent information management industry.* See https://info.aiim.org/state-of-the-intelligent-information-management-industry-2021, accessed in May 2021.

ANZ (n.d.) *Archives 2057 Strategy.* See: www.archives.govt.nz/publications/archives-2057-strategy, accessed in June 2021.

DIA (2020) *Strategy for a digital public service.* See https://www.digital.govt.nz/digital-government/strategy/strategy-summary/strategy-for-a-digital-public-service/, accessed in May 2021.

DPMC (2021) *Health and Disability Review Transition Unit.* See: https://dpmc.govt.nz/our-business-units/transition-unit, accessed in May 2021.

Earley (2020) *The AI-powered enterprise*, by Seth Earley. LifeTree Media Ltd, 2020.

Harris (2004) *Ontology Review 1. The NHS Common Basic Specification. Why top level Ontologies don't work.* See: http://virtualtravelog.net/2004/01/ontology-review-1-the-nhs-common-basic-specification-why-top-level-ontologies-dont-work/, accessed in May 2021.

ISO 25964 (2011) *ISO/DIS 25964-2:2011 Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies*, International Standards Organization, 2011.

MITRE (2004) *Toward the use of an upper ontology for U.S. government and U.S. military domains: an evaluation*, by S.K. Semy, M.K. Pulvermacher and L.J. Obrst. MITRE Corporation, 2004.

Parliamentary Commissioner for the Environment (2019) *Focusing Aotearoa New Zealand's environmental reporting system,* 2019. See: https://www.pce.parliament.nz/publications/focusing-aotearoa-new-zealand-s-environmental-reporting-system, accessed in May 2021.

Stuff (2016) *Public sector organisations criticised for poor record-keeping,* Stuff 7 January 2016 See: https://www.stuff.co.nz/national/politics/75647486/public-sector-organisations-criticised-for-poor-record-keeping, accessed in May 2021.

Stuff (2020) *Survey finds worrying holes in management and accessibility of public records*, Stuff January 28 2020. See: https://www.stuff.co.nz/national/118797967/survey-finds-worrying-holes-in-management-and-accessibility-of-public-records, accessed in May 2021.

Stuff (2021) *Record-keeping in public sector still worrying — Archives New Zealand report*, Stuff 9 April 2021. See: https://www.stuff.co.nz/national/124789079/recordkeeping-in-public-sector-still-worrying--archives-new-zealand-report, accessed in May 2021.

UK (2005) *Tomatoes are not the only fruit,* UK Cabinet Office e-Government Unit, 2005. See: https://nanopdf.com/download/tomatoes-are-not-the-only-fruit_pdf, accessed in May 2021.

W3C (2012) *OWL2 Web Ontology Language Document Overview 2012.* See: http://www.w3.org /TR/0wl2.overview/, accessed in May 2021.

W3C (2014) *The Organization Ontology: W3C Recommendation 16 January 2014*. See: https://www.w3.org/TR/vocab-org/, accessed in May 2021.

W3C (2021) *Vocabularies.* See https://www.w3.org/standards/semanticweb/ontology, accessed in May 2021.