Document Control

Project ID/Name: Government Enterprise Architecture of New Zealand
Author: Regine Deleu and John Fisher
Title: GEA-NZ v3.2 Data and Information Reference Taxonomy

Document Classification: UNCLASSIFIED

Document Approval

Department of Internal Affairs

The GEA-NZ Data and Information Reference Model is approved by Government Enterprise Architecture Group (GEAG) and Government Enterprise Architect, System Transformation, Service and System Transformation in-line with AoG policy, supporting the Government Chief Technology Officer (GCTO) and the Government Chief Information Officer (Gcio), Department of Internal Affairs, New Zealand.

Contact us: Enquiries regarding this document are welcome to Email: GEA@dia.govt.nz

Acknowledgements

The GEA-NZ v3.2 Data and Information Reference Model was developed by the Government Enterprise Architecture, System Transformation, Service and System Transformation, supporting the Government Chief Technology Officer (GCTO) and the Government Chief Information Officer (Gcio), Department of Internal Affairs, New Zealand.

The feedback and suggestions received from a number of experts from various agencies, and the collaboration working sessions, were greatly appreciated.

It was peer-reviewed and approved by the Government Enterprise Architecture Group (GEAG) members. Approval was on the 25th of May 2017.

GAE-NZ v3.2 Data and Information Reference Model
Revision history

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Description of changes</th>
<th>Author</th>
<th>Approved by</th>
</tr>
</thead>
<tbody>
<tr>
<td>v3.2</td>
<td>June 2017</td>
<td>Initial version of the Data and Information Reference Model document split of from the Data and Information Reference Taxonomy</td>
<td>Regine Deleu, John Fisher</td>
<td>GEAG</td>
</tr>
</tbody>
</table>

Table of Content

Document Control ........................................................................................................................................................................... 2
Document Approval ............................................................................................................................................................................... 2
Revision history .................................................................................................................................................................................... 3

Table of Content............................................................................................................................................................................... 3

Executive Summary .............................................................................................................................................................................. 4

Data and Information Reference Model Introduction .......................................................................................................................... 5
  GEA-NZ Framework Context ................................................................................................................................................................. 6
  Data and Information Reference Model Principles .......................................................................................................................... 8
  How to use .......................................................................................................................................................................................... 10
  Information Asset Catalogue .............................................................................................................................................................. 10
  Information Discovery ........................................................................................................................................................................... 10
  Data Governance ................................................................................................................................................................................. 10
  Categorisation .................................................................................................................................................................................. 11
  Data and Information Exchange ....................................................................................................................................................... 11
  Security and Privacy ............................................................................................................................................................................ 12

Abstract Data Architecture Model .......................................................................................................................................................... 13
  Overview .......................................................................................................................................................................................... 13
  Data Description ................................................................................................................................................................................. 15
  Data Context .................................................................................................................................................................................... 20
  Data Sharing ................................................................................................................................................................................... 24
Executive Summary

The purpose of this document is to provide NZ Government agencies with a core reference architecture model covering data and information, and the accompanying taxonomy or 'categorisation terminology', as part of the Government Enterprise Architecture for New Zealand (GEA-NZ v3.2) suite of reference models.

Data are simply facts or figures — bits of information, but not information itself. When data are processed, interpreted, organized, structured or presented so as to make them meaningful or useful, they are called information. Information provides the context for data.

The primary purpose of the GEA-NZ Data and Information Reference Model is to discover, describe, manage, protect, and to share and reuse information consistently within and across agencies and their business partners. The GEA-NZ Data and Information Reference Model is a flexible and standard-based framework to enable information sharing and reuse across agencies via a standard description and discovery of common data and information. It also promotes the uniform practice of data and information management.

Use of GEA-NZ reference models will:
- Provide common language to promote service, information, and system and technology interoperability.
- Support improved quality and speed of sharing information.
- Promote the identification and demand aggregation of sharable and common capabilities to improve the efficacy, utility and cost effectiveness of ICT across government.
- Foster traceability of features to meet the requirements of change.
- Support the re-use of solutions and services.
- Support the AoG Data Governance initiatives.

The intended audience of the Data and Information Reference Taxonomy is (but not limited to):

- Agencies’ Enterprise, Data, Information, and Solution Architects
- Business Service Management Providers
- Business partners involved in development and delivery of business, analytics, and technology solutions
- Business, Information, and IT Managers
- Business Subject Matter Experts
- Security Specialists for Certification & Accreditation activities (C&A)
- Suppliers to the New Zealand Government.

The key benefits to agencies, sectors and their business partners involved in the delivery of public services and joint capabilities:

- Drives sharing of information at the business layer to ease access to more, better, personalised and integrated services.
- Identification of opportunities to improve efficiencies and effectiveness of current capabilities. To guide change towards shared common services.
- For agencies’ Four Year Plan to help show what they will achieve and how it will be achieved aligning to Better Public Services and all-of-government shared services.
- This will ultimately result in increased collaboration between agencies, reduced risks, reduced number of incompatible systems across and within agencies, and it contributes to government-wide interoperability commitments in an affordable manner.
Introduction

The GEA-NZ Data and Information Reference Model describes and categorises data and information assets of the New Zealand Government across nine common subjects instead of a siloed, agency-by-agency view.

Government Enterprise Architecture for New Zealand

The overall consistency and cohesiveness of cross government services, shared services and common capabilities, can be improved when government entities apply common reference taxonomies to deliver consistent and aligned views of commonly required operational and technological services.

The objective of a Reference Model is to provide widely accepted core taxonomy, and an appropriate visual representation of that taxonomy. A Reference Taxonomy defines the terminology, and provides a useful, coherent, consistent, and structured description of the components of an Enterprise Architecture. The need for reference architectures and associated taxonomies as part of a Government Enterprise Architecture is described in the GEA-NZ Context Document.

There are eight architecturally significant dimensions within the Government Enterprise Architecture for New Zealand v3.1 framework. Each of these dimensions has a Reference Model which comprises:

- Structure and description of the dimension
- Context within the GEA-NZ
- Reference taxonomy, where appropriate
- Description of artefacts and relationships with other artefacts

GEA-NZ v3.1 uses reference taxonomies to provide categorisation terms to describe the architecture of capabilities for use across All of Government (AoG). Reference taxonomies reduce complexity by abstracting, organising and simplifying complex information sets.

The usage of the GEA-NZ reference models and taxonomies, at government, agency and sector level, will help drive ICT efficiencies and transformation programmes through identification of opportunities for development and reuse of shared solutions. This will enable the implementation of the Government ICT Strategy and Better Public Services: Results for New Zealanders.
GEA-NZ Framework Context

The GEA-NZ v3.2 Data and Information Reference Taxonomy is part of the GEA-NZ Data and Information Dimension (red rectangle in the diagram below) that is an integral part of the Government Enterprise Architecture for New Zealand (GEA-NZ) Framework. It is one of the 4 core dimensions of GEA-NZ (blue rectangle in the diagram below). The other 4 dimensions are the control dimensions.
The following table shows the GEA-NZ Data and Information Reference Taxonomy in context with the other GEA-NZ dimensions.

<table>
<thead>
<tr>
<th>Data and Information</th>
<th>Provides</th>
<th>Taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a framework for trusted data and information that can be used for strategic decision making</td>
<td>Strategy, Investment, and Policy</td>
</tr>
<tr>
<td></td>
<td>a framework for trusted data and information that can be used for governance and business performance management</td>
<td>Performance and Governance</td>
</tr>
<tr>
<td></td>
<td>the data and information structures that support business services, processes, capabilities, information sharing, and reuse</td>
<td>Business</td>
</tr>
<tr>
<td></td>
<td>authoritative data and information structures to be used by application and software services</td>
<td>Application and Software Services</td>
</tr>
<tr>
<td></td>
<td>the data and information requirements for technology and infrastructure services</td>
<td>Infrastructure</td>
</tr>
<tr>
<td></td>
<td>the data and information requirements and models needed for identity, privacy, and security</td>
<td>Identity, Privacy, and Security</td>
</tr>
<tr>
<td></td>
<td>sets the data and information requirements that drive development and scope of corresponding standards</td>
<td>Standards</td>
</tr>
</tbody>
</table>
Data and Information Reference Model Principles

The principle RB1D is specific to the Data and Information reference taxonomy; and the others (RT2-RT5) are generic to all GEA-NZ Reference Taxonomies.

Principle RT1D

Description
Principle Reference Taxonomy 1 Data and Information (RT1D); Elements that make up Data and Information Domains, Subjects, and Topics should be recognised terms, or NZ government terms.

Rational
This is good practice.
1. Using globally agreed terms makes the taxonomy more useful to both vendors and agencies.
2. Using unusual terms reduces usability and relevance of the taxonomy.

Exceptions
We may include elements for emerging and possibly experimental concepts for which we may want to provide guidance or direction. For example; this may occur where an element is specified to support the strategic direction of Government.

Principle RT2

Description
Reference Taxonomy Principle 2 (RT2) is that elements should be useful at the Category level (3rd level).

Rational
This is good practice. It assists the users apply the taxonomy to their architecture.
1. Use open language like "Category X includes;" rather than a list. The list should serve as examples.
2. An alternative to using level 4 sub-categories is summarising the "Narrower Terms:" that may belong to a category instead of creating lower level 4 sub-categories. This approach will be used to simplify the taxonomy for GEA-NZ v3.2.
3. There shall be NO AoG level 5.

Exceptions
Where level 4 is needed it is permitted and may be used where particular focus on a category supports the strategic direction of Government.

Principle RT3

Description
Reference Taxonomy Principle 3 (RT3) is that the source of a term, or the definition used SHOULD be noted in the notes as "Source: The_Source".

Rational
This is best practice. It assists the user of the taxonomy check for updated information, and provides useful context.
1. All new or revised terms used in the v3.2 Taxonomy MUST have the source shown in the description.
2. All terms SHOULD have a source.
Exceptions
Where elements / terms have not been revised, or are not new the "Source" MAY be missing in the v3.2 version of the taxonomy due to time constraints.

NOTE: In general the sources were FoNZ/NZGLS, the US FEAF v2, the AGA v3.0, augmented by Wikipedia.

Principle RT4

Description
Reference Taxonomy Principle 4 (RT4) is that related terms SHOULD be listed under "Related Terms:" in the element description.

Rational
This is best practice. It assists term discovery when alternative terms are searched for in the taxonomy.

1. "Related Terms:" MUST be used if a name changes, for instance as per principle RT1 (RT1D)
2. "Related Terms:" SHOULD be used if there is an alternative industry term.

Exceptions
Where terms have not been revised, or are not new, "Related Terms:" MAY be missing in the v3.2 version of the taxonomy due to time constraints.

Principle RT5

Description
Reference Taxonomy Principle 5 (RT5) is that element GUID’s, and Element Taxonomy ID’s, SHOULD be consistent across versions of the taxonomy.

Rational
This is good practice. It assists the user manage change between versions.

1. In Spark EA models the GUIDS for an element MUST not be changed where the element is otherwise the same, even if the name has changed, but has the same meaning, or the ID has had to change due to levelling. This means elements in models can be updated to reflect new version names and descriptions while preserving any relationships established by the user.
2. ID’s MUST NOT be reused for another element, when an element is removed from the taxonomy, or moved to another part of the taxonomy for levelling purposes. This avoids confusion where comparing across different versions.

Exceptions
1. Elements ID’s may change where we need to re-level a taxonomy. However element ID’s will not be re-used from version 3.2 onwards.
2. Element ID’s MAY be re-used to correct past errors in the taxonomy where elements may be incorrectly placed due to an ID error.
How to use

Use of the GEA-NZ Data and Information Reference Model will:

- Facilitate CIOs in establishing common information language to promote services, information, systems, and technology interoperability.
- Enable conversations to reach cross-agency agreements around: governance, data architecture and information sharing.
- Promote the identification and demand aggregation of sharable and common capabilities to improve the efficacy, utility and cost effectiveness of ICT across government.
- Foster traceability of features to meet the requirements of change.
- Support the re-use of solutions and services.
- Support the AoG Data Governance initiatives.

Ultimately, the aim is to increase collaboration between agencies, reduce risk, reduce number of incompatible systems across and within agencies, and contribute to government-wide interoperability commitments in an affordable manner.

Information Asset Catalogue

The GEA-NZ Data and Information Reference Model and Taxonomy identifies data and information assets, regardless of usage context. Used in combination with the GEA-NZ Business, Application & Software Services, and Infrastructure and the Information Asset Catalogue Template, agencies can identify and catalogue their data and information assets and manage them according to goals, objectives or business context.

Information Discovery

The GEA-NZ Data and Information Reference Taxonomy can be used to perform information discovery for programmes, projects and business processes. It is a simple but very efficient and effective process, for instance used during the workshop in the Discovery phase of the Accelerated Delivery Methodology. It helps to understand the context of a scenario or subject as early as possible in a project, programme, or scenario.

Data Governance

Data & Information is at the core of everything the Government does, so we need a strong Data Governance strategy to guide the creation, transformation, and sharing of data & information. Data Governance is a set of processes, around Data and Information:

- Ensuring that key Information is formally managed throughout the Government/Enterprise.
- Ensuring that the Information can be trusted and that people can be made accountable so that the Information is fit for purpose and that the value of those assets are fully realised.
- Altering the way of thinking around Information, how to handle information so it can be used by the entire organisation/Government/Country
Categorisation

The GEA-NZ Data and Information Reference Model and Taxonomy is a flexible and standard-based framework to enable information sharing and reuse across agencies via a standard description and discovery of common data and information. It also promotes the uniform practice of data and information management.

The GEA-NZ Data and Information Reference Model and Taxonomy includes three information pillars and nine category groups that can be used as a common language to categorise New Zealand government information consistently at a high level. The information pillars are:

- **Motivators**: Information relating to authority or governance such as plans, controls, contracts. They contain information in the form of potential, imagined or desired states. For example, risk management information is concerned with potential hazards, their likelihood and consequences.
- **Entities**: Information relating to instances of entities or things such as parties, places, and items. Entities are tangible items that the business must know about and for which different instances can be uniquely identified or distinguished from one another.
- **Activities**: Information needed to track or monitor moments, periods, events and cases that occur over time. This type of information is focused on occurrences that must be tracked for business reasons or represent a specific point in the evolution of ‘The Business’.

The nine information subjects within these three pillars will be described in more detail further in this document.

Data and Information Exchange

The GEA-NZ Data and Information Reference Model and Taxonomy provides a standard means by which data is described, categorised, and shared. It also provides guidance for data and information exchange within and across agencies and their business partners. Data and Information Exchange bases its definition and documentation on the following three components:

- **Data Context**: Data Context categorizes data and information assets into the three pillars of information as outlined above. Data Context will answer key questions about the data required within and across agencies and establishes the basis for data governance. Data Context also enables discovery of data, and can provides linkages to the other GEA-NZ Reference Models.
- **Data Description**: Data Description provides a means to uniformly capture the semantic and syntactic structure of data. This enables comparison of metadata (data about data) for purposes of harmonization, and supports the ability to respond to questions regarding what is available in terms of Data Descriptions (metadata).
- **Data Sharing**: The Data Sharing describes the access and exchange of data, where access consists of recurring requests (such as a query of a data asset), and exchange consists of fixed, recurring information exchanges between parties. Data sharing is enabled by the capabilities provided by both the Data Context and Data Description areas.

The interaction between these components is shown in the following diagram.
Security and Privacy

An institutional process that includes roles and responsibilities for data stewardship for each project or program in an agency has been defined as part of a policy that governs data Quality, Security, Privacy and Confidentiality.

The roles and responsibilities associated with data and information management are covered in detail in the AoG document “Data and Information Management – Roles and Responsibilities”.

There are a number of areas that should be addressed in building a Security, Privacy and Confidentiality Policy for an agency. These include:

- Constructing a policy that is compliant with legislation, Executive Orders and Standards.
- Addressing sensitivity of information that eliminates possible compromise of sources and methods of information collection and analysis.
- Establishing the practices of data stewardship.
- Addressing specific data access policies defined by the responsible steward; for example:
  - Data is available for open, unrestricted access
  - Data is accessible only to a group
  - Data access is a function of the person (his or her identity), data about that person (e.g., current position), and data about the environment (e.g., physical location)
  - Data is self-protecting through digital rights management or similar technology.

---

1 See the GEA-NZ v3.1 Data and Information Quality Management framework.
Abstract Data Architecture Model

Overview

The abstract model is an architectural pattern developed to optimize agency data architectures. It is abstract in that it allows multiple technical implementations. This architectural pattern is designed to optimize an agency’s data architecture for information integration, interoperability, discovery and sharing.

The pattern achieves this optimization by defining, arranging and relating the standard information concepts in an informal data model, specifying common attributes for each concept. Each component of the architecture, description, context and exchange, is described using a model diagram of their contained data concepts, depicted as boxes, coupled with descriptions and associated attributes.

The linkages between these three models are provided by shared, fundamental information concepts modelled in the following diagram. The linkage concepts are shown in the same colours as below on the three detailed models.

Four types of relationship are shown between the concept boxes namely:

- A formal UML association with a crowfoot to indicate “many”
- A formal UML “generalises” relationship where appropriate
- A solid arrow to indicate a general relationship
- A dotted arrow to indicate an optional general relationship

Before defining the information concepts in detail, it is important to understand the main features of the architecture model for each of the three components of the data architecture.

---

2 Some agencies use the DOD Discovery Metadata Specification (DDMS) for Digital Data Resource attributes while another agency may choose to use the Dublin Core elements; both their implementation maps to the abstract model.

GEA-NZ v3.2 Data and Information Reference Taxonomy
Data Description

The Data Description area focuses on understanding the data at two levels of abstraction: the metadata artefacts required to understand the data and how those metadata artefacts are aggregated into a managed Data Asset catalogue.

There are two basic types of metadata recommended for the Data Description section of the abstract model:

- Logical data models to describe Structured Data Resources
- Digital Data Resource metadata (such as Dublin Core elements) to describe Semi-Structured and Unstructured Data Resources.

The division of data into these two types is intended to support harmonization (via comparison of logical data models) and registration (via description of universal resource attributes).

Implementation of the Data Schema concept group would take the form of Entity-Relationship diagrams, class diagrams, etc. Implementation of the Digital Data Resource could be documents in a content management system or a metadata catalogue.

Data Context

The Data Context area focuses on management mechanisms to capture the context of data in an organization. Those mechanisms are the Data and Information Reference Model and Taxonomy and a Data Asset catalogue.

A Data Asset is a collection of Digital Data Resources that is managed by an organization, categorized for discovery, and governed by a data steward.

A key attribute of a Data Asset is whether it is authoritative and if so designated, authoritative on which Entity or Attribute of the logical data model (see Data Schema in the Data Description section of the abstract model). Implementation of a Data Asset catalogue is recorded in a metadata registry.

Data Sharing

The Data Sharing area focuses on how information is packaged for and/or exposed to the Consumer. The key concepts are Exchange Packages as containers for fixed messages and Query Points as descriptions of data access points.

---

3 It should be noted that the term “entity” here, and in subsequent Dublin Core attributes, does not have the same exact meaning as the “Entity” in the Information Categorisation.

4 Implementation of the taxonomy could take the form of extensible Markup Language (XML) Topic Maps, Web Ontology Language (OWL) hierarchies or ISO11179 Classification schemes.

5 Implementation of Exchange Packages could be standard XML messages or EDI transaction sets. Implementation of Query Points could be descriptions in a Universal Description, Discovery and Integration (UDDI) or ebXML registry of a data access Web service.

GEA-NZ v3.2 Data and Information Reference Taxonomy
**Data Description**

The Data Description component of the abstract model depicts the concepts that provide the description of the data and the relationships between them. The Data Schema shown in this section of the abstract model is an aggregation of related concepts.

Following are descriptions for each of the concepts depicted on the model above.

**Data Schema**

Data Schema is a representation of metadata, often in the form of data artefacts such as logical or conceptual data models. The Data Schema concept group is comprised of those concepts pertaining to the representation of structured data. A Data Schema provides a means to provision data sharing services that is independent of the values of the data in the data resource that it describes.

Relationships:
- A Data Schema defines a Structured Data Resource
- A Data Schema describes a Structured Data Asset
**Entity**

An Entity is an abstraction for a person, place, object, event, or even a concept. In fact anything that can be identified and have a common set of Attributes. For example, “Person”, “City”, “Vehicle” and “Entity” are all Entities. An instance of an Entity represents one particular occurrence of the Entity, such as a specific person or a specific city.

Relationships:
- An Entity contains Attributes
- An Entity participates in Relationships with other Entities

**Attribute Domain**

An Attribute Domain defines a type of attribute by specifying its data type (e.g. "string" or "integer"), its length and its validation rules (which can be null for text). It is a constraint on the representation and values that an instance of an Attribute may hold. Data management is simplified by adopting a naming convention for Attributes such as Entity (qualifier) Domain.

Relationships:
- Could be used in Business Rules relationships

**Attribute**

An Attribute is a characteristic of an Entity. Their values may be used to distinguish one instance of an Entity from other instances of the same Entity, indicate relationships between Entities or simply to provide data about an Entity. For example, the Attributes of a Vehicle entity would include:
- Vehicle Registration Number to differentiate one vehicle from another
- Manufacturer Name to relate it to a vehicle manufacturer
- Vehicle Colour to provide vehicle information

An Attribute is defined by a Domain

**Relationship**

A Relationship describes the relationship\(^6\) between two Entities. A Relationship therefore has two Relationships, one with each of the Entities it relates to.

**Data Asset**

A Data Asset is a managed container for data detailed by the Data Resources it contains. In many cases, this will be a relational database; however, a Data Asset may also be a Web site, a document repository, directory or data service. Data Asset is further described under Data Context.

**Digital Data Resource**

A Digital Data Resource is a digital container of information. A Digital Data Resource may be one of three specific types of data resources namely “Structured Data Resource”, “Semi-Structured Data Resource”, and “Unstructured Data Resource”. It is a container for the metadata for the data resource. It relates to one of three specific Data Resources and is used to describe the composition of a Data Asset.

\(^6\) It should be noted that the term “relationship” is used in two ways here. The concept named “Relationship” participates in relationships with other concepts in the abstract model. It also defines the relationships between entities when used in the Data Concept descriptions.
• A Digital Data Resource describes a Structured Data Asset or a
• A Digital Data Resource describes a Semi-structured Data Asset or a
• A Digital Data Resource describes an Unstructured Data Asset

**Structured Data Resource**
A Digital Data Resource which contains structured data and typically corresponds to a Data Subject Area within a data model. A Data Subject Area is defined by a Data Schema and contains a set of closely related Entities which form a nameable cohesive data resource. This data can be accessed in a uniform manner, independent of data values, once the Data Schema is known.

Relationships:
• A Structured Data Resource is a type of Digital Data Resource
• A Structured Data Resource is defined by a Data Schema

**Semi-Structured Data Resource**
A Digital Data Resource which contains semi-structured data. This will generally consist in part of structured data and in part of unstructured data. This could literally be a document with text and tables. However there is an important type of data that fits this description and that is structured data that carries its own meta data with it. XML messages fall into this category as do things like Data Lakes.

Relationships:
• A Semi-Structured Data Resource is a type of Digital Data Resource
• A Semi-Structured Data Resource can be contained in a Document

**Unstructured Data Resource**
A Digital Data Resource which contains unstructured data. This is typically, but not necessarily, text. It can also carry meta data, such as Dublin Core, with itself. Unstructured data is collection of data values that are likely to be processed only by specialized application programs.

Relationships:
• An Unstructured Data Resource is a type of Digital Data Resource
• A Semi-Structured Data Resource can be contained in a Document

**Document**
A Document is a file containing Unstructured and/or Semi-Structured Data Resources.

Relationships:
• A Document may contain an Unstructured or Semi-Structured Data Resource
• A Document refers to an Entity

Note: While a Document can contain structured data, it normally has explanatory material included, which would therefore cause it to be considered semi-structured. It is for this reason that there is no “contains” relationship from Document to Structured Data Resource. It is important to separate Documents from Structured Data Resources because they are processed very differently.
### Data Description Attributes

This section will expand on the concepts presented above to include the attributes that are associated with each concept in the Data Description section of the abstract model. A description will be provided for each attribute, along with an example where necessary for clarity. All Unstructured Data Resource attributes and their descriptions are taken from the [Dublin Core Metadata Initiative (DCMI), Version 1.1](https://dublincore.org/documents/dcmi-terms/). All references to “resource” within descriptions of Unstructured Data Resource should therefore be interpreted as “Unstructured Data Resource”. The above URL provides additional information on attribute descriptions and usage.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Attribute</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity</td>
<td>Identifier</td>
<td>A unique string associated with an Entity for identification purposes.</td>
<td>“D205XKA”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Entity</td>
<td>“Person”</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Entity</td>
<td></td>
</tr>
<tr>
<td>Attribute Domain</td>
<td>Name</td>
<td>The name of the Domain</td>
<td>“Age”</td>
</tr>
<tr>
<td></td>
<td>Data Type</td>
<td>The name of the Data Type</td>
<td>“Integer”</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>The validation to be applied</td>
<td>“1&lt;120”</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Domain</td>
<td></td>
</tr>
<tr>
<td>Attribute</td>
<td>Name</td>
<td>The name of the Attribute</td>
<td>“Person Birth Date”</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Attribute</td>
<td></td>
</tr>
<tr>
<td>Relationship</td>
<td>Name</td>
<td>The name of the Relationship</td>
<td>“works for”</td>
</tr>
<tr>
<td></td>
<td>Origin</td>
<td>Name of the concept that is the origin (i.e. the “from” concept”) of the Relationship</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Destination</td>
<td>Name of the concept that is the destination (i.e. the “to” concept”) of the Relationship</td>
<td></td>
</tr>
<tr>
<td>Structured Data Resource</td>
<td></td>
<td>See all concepts within “Data Schema” group.</td>
<td></td>
</tr>
<tr>
<td>Semi-Structured Data Resource</td>
<td></td>
<td>See “Structured Data Resource” and “Unstructured Data Resource”</td>
<td></td>
</tr>
<tr>
<td>Unstructured Data Resource (Based on Dublin Core)</td>
<td>Title</td>
<td>A name given to the resource</td>
<td>“Information Exchange Report – June 2015”</td>
</tr>
<tr>
<td></td>
<td>Resource Identifier</td>
<td>An unambiguous reference to the resource within a given context.</td>
<td>“D205XKA”</td>
</tr>
</tbody>
</table>

---

7. It should be noted that the term “attribute” is used here in the same way as the concept named “Attribute”. Here “attribute” is used to describe characteristics of each of a concept “entity” in the abstract model.

8. The “Identifier” attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementation based on the GEA-NZ Data and Information Reference Model will introduce such aspects as needed according to their requirements.

9. As shown in the abstract model, a Digital Data Resource may be one of these three specific types of data resources.

---

*GEA-NZ v3.2 Data and Information Reference Taxonomy Page 18 of 27*
| **Date** | A date of an event in the lifecycle of the resource. Will typically be associated with the creation or availability of the resource |
| **Creator** | An entity (“Party”, see Information Categorisation) primarily responsible for making the content of the resource |
| **Format** | The physical or digital manifestation of the resource. Typically, format may include the media-type or dimensions of the resource. “text/plain” |
| **Description** | An account of the content of the resource |
| **Source** | A reference to a resource from which the present resource is derived. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system. |
| **Subject** | A topic of the content of the resource |
| **Resource Type** | The nature or genre of the content of the resource. “Service” |
| **Publisher** | An entity (“Party”, see Information Categorisation) responsible for making the resource available |
| **Contributor** | An entity (“Party”, see Information Categorisation) responsible for making contributions to the content of the resource |
| **Language** | A language of the intellectual content of the resource. “eng” |
| **Relation** | A reference to a related resource. “521LXF” |
| **Coverage** | The extent or scope of the content of the resource. “Wellington” |
| **Rights Management** | Information about rights held in and over the resource. “Public domain” |

*Document* See “Structured Data Resource” and “Unstructured Data Resource”
Data Context

The Data Context section of the abstract model depicts the concepts that comprise the Data Context area and the relationships between them.

Following are descriptions for each of the concepts depicted on the model above.

Taxonomy

A Taxonomy is a collection of controlled vocabulary terms organized into a hierarchical structure. Taxonomies provide a means for categorizing or classifying information within a reasonably well-defined associative structure. In a Taxonomy each term is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy. A common example of a taxonomy is the hierarchical structure used to classify living things within the biological sciences from Carols Linnaeus\(^1\), as shown below:

\(^{10}\) Carl Linnaeus was a Swedish botanist, physician, and zoologist, who laid the foundations for the modern biological naming scheme of binomial nomenclature. He is known as the father of modern taxonomy, and is also considered one of the fathers of modern ecology.
<table>
<thead>
<tr>
<th>Category</th>
<th>Value for humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kingdom</td>
<td>Animalia</td>
</tr>
<tr>
<td>Phylum</td>
<td>Chordata</td>
</tr>
<tr>
<td>Class</td>
<td>Aves</td>
</tr>
<tr>
<td>Order</td>
<td>Anseriformes</td>
</tr>
<tr>
<td>Family</td>
<td>Anatidae</td>
</tr>
<tr>
<td>Genus</td>
<td>Tadorninae</td>
</tr>
<tr>
<td>Species</td>
<td>Hymenolaimus</td>
</tr>
</tbody>
</table>

The taxonomies which are used within GEA-NZ are:
- Business Reference Taxonomy
- Data and Information Reference Taxonomy
- Application and Software Services Reference Taxonomy
- Infrastructure Reference Taxonomy
- Government Business Capability Reference Taxonomy

Other taxonomies can be used for specific purposes like the Carols Linnaeus taxonomy within Department of Conservation (DoC), Geospatial taxonomy within Land Information New Zealand (LINZ), etc.

Relationships:
- A Taxonomy contains a Topic
- A Taxonomy is represented as a Structured Data Resource

**Structured Data Resource**

This concept is described in Data Description.

**Topic**

A Topic is an information category within a Taxonomy. Topic is the central concept for applying context to data. For example, an agency may have a Taxonomy that represents their organizational structure. In such a Taxonomy, each role in the organizational structure (e.g. CIO) represents a Topic.

Relationships:
- A Topic categorizes a Data Asset
- A Topic may categorize a Digital Data Resource
- A Topic may categorize a Query Point
- A Topic may categorize an Exchange Package
- A Topic participates in a Relationship with another Topic
Digital Data Resource
This concept is described in Data Description.

Query Point
This concept is described in Data Sharing.

Exchange Package
This concept is described in Data Sharing.

Topic Relationship
Topic Relationship describes the relationship between two Topics, for example, a relationship between a Party and an Item.

Relationships:
- A Relationship relates a Topic one with another.

Data Asset
Data Asset is a managed container for data. In many cases, this will be a relational database; however, a Data Asset may also be a Web site, a document repository, directory or data service.

Data resources held within a Data Asset will usually have meta data associated with them. Digital Data Resources will have a Data Schema. Documents stored and managed within a document repository will have its management context provided for it through the metadata that is associated with it. Such metadata may include the Dublin Core attributes as described in the Data Description chapter.

Relationships:
- A Data Asset provides management context for a Digital Data Resource(s)
- A Data Asset may be defined by Data Schema(s)
- A Data Asset may be managed by a Data Steward

Data Steward
A Data Steward is the person responsible for managing a Data Asset.

- A Data Asset may be managed by a Data Steward

GEA-NZ Reference Model
GEA-NZ Reference Model represents the four GEA-NZ v3.2 reference models (Business, Data and Information, Application and ICT Services, and Infrastructure). Its purpose is to provide a linkage to these reference models, which are themselves Taxonomies. These are depicted as a special kind of Taxonomy due to their importance in overall categorisation of information.

Relationships:
- The GEA-NZ Reference Models are types of Taxonomies
## Data Context Attributes

This section will expand on the concepts presented above to include attributes that are associated with each concept in the Data Context section of the abstract model. A description will be provided for each attribute, along with an example where necessary for clarity.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Attribute</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxonomy</td>
<td>Identifier</td>
<td>A unique string associated with a Taxonomy for identification purposes.</td>
<td>“A1.02.03”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Taxonomy</td>
<td>“Application and ICT Services”</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Taxonomy</td>
<td></td>
</tr>
<tr>
<td>Topic</td>
<td>Name</td>
<td>The name of the Topic</td>
<td>“Accounts Payable”</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Topic</td>
<td></td>
</tr>
<tr>
<td>Topic Relationship</td>
<td>Name</td>
<td>The name of the Relationship</td>
<td>“part of”</td>
</tr>
<tr>
<td></td>
<td>Origin</td>
<td>Name of the concept that is the origin (i.e. the “from” concept”) of the Relationship</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Destination</td>
<td>Name of the concept that is the destination (i.e. the “to” concept”) of the Relationship</td>
<td></td>
</tr>
<tr>
<td>Data Asset</td>
<td>Identifier</td>
<td>A unique string associated with a Data Asset for identification purposes.</td>
<td>“SchDB1”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Data Asset</td>
<td>“Schools Database”</td>
</tr>
<tr>
<td></td>
<td>Type</td>
<td>Type of the Data Asset – e.g. database, Web site, registry, directory, data service, etc.</td>
<td>“database”</td>
</tr>
<tr>
<td></td>
<td>Geospatial Enabled</td>
<td>Designates whether or not the Data Asset supports or provides Geospatial data.</td>
<td>“no”</td>
</tr>
<tr>
<td>Data Steward</td>
<td>Employee ID</td>
<td>Data Steward’s employee ID.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>Data Steward’s Name</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Department</td>
<td>Department for which Data Steward works.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Initial Date</td>
<td>The date that Data Steward became associated with the Data Asset.</td>
<td></td>
</tr>
<tr>
<td>GEA-NZ Reference Model</td>
<td>Acronym</td>
<td>Reference Model acronym</td>
<td>“ARM”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>Reference Model name</td>
<td>“Application and ICT Services Reference Model”</td>
</tr>
</tbody>
</table>

---

11 The “Identifier” attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementation based on the GEA-NZ Data and Information Reference Model will introduce such aspects as needed according to their requirements.
Data Sharing

The Data Sharing section of the abstract model covers two primary aspects of data sharing:

- **Data Exchange**: Fixed, recurring transactions between parties, such as the regular exchange of environmental testing data among national, local, and regional entities. These exchanges are implemented with Data Exchange services.
- **Data Access**: Requests for data services, such as a query of a Data Asset. These requests are supported by Data Access Services.

The Data Sharing area is supported by the Data Description and Data Context areas in the following ways:

- **Data Description**: Uniform definition of Exchange Packages and Query Points supports the capability to effectively share them within and between agencies.
- **Data Context**: Categorization of Exchange Packages and Query Points supports their discovery, and their subsequent use in data access and data exchange.

The Data Sharing section of the abstract model depicts the Data Sharing concepts for the abstract model and the relationships between them.

**Exchange Package**

An Exchange Package is a description of a specific recurring data exchange between a Supplier and a Consumer. An Exchange Package contains information (metadata) relating to the exchange (such as Supplier ID, Consumer ID, validity period for data, etc.), as well as a reference to the Payload (message content) for the exchange.

An Exchange Package can also be used to define the result format for a query that is accepted and processed by a Query Point in a data sharing scenario.

**Relationships:**

- An Exchange Package refers to an Entity
- An Exchange Package is disseminated to a Consumer
• An Exchange Package queries a Query Point
• An Exchange Package refers to a Payload Definition

**Entity**
This concept is described in Data Description.

**Attribute**
This concept is described in Data Description.

**Supplier**
A Supplier is a Party (person or organization) that supplies data to a Consumer.

Relationship:
• A Supplier produces an Exchange Package

**Consumer**
A Consumer is a Party (person or organization) that consumes data that is supplied by a Supplier.

Relationship:
• A Consumer receives an Exchange Package

**Payload Definition**
Payload Definition is an electronic definition that defines the requirements for the Payload (data) that is exchanged between a Supplier and a Consumer, e.g. XMI.

Relationship:
• A Payload definition is used by an Exchange Package
• A Payload definition references Attributes

**Query Point**
Query Point is an endpoint that provides an interface for accessing and querying a Data Asset. A concrete representation of a Query Point may be a specific URL at which a query Web Service may be invoked.

Relationship:
• A Query Point accesses a Data Asset

**Data Asset**
This concept is described in Data Context.
### Data Sharing Attributes

This section will expand on the concepts presented above to include attributes that are associated with each concept. A description will be provided for each attribute, along with an example where necessary for clarity.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Attribute</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange Package</td>
<td>Identifier</td>
<td>A unique string associated with an Exchange Package for identification purposes.</td>
<td>“PAJ937”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Exchange Package.</td>
<td>“Payables Child Support”</td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Exchange Package.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classification</td>
<td>The security classification for the Exchange Package</td>
<td>“R” (Restricted)</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>The frequency at which the exchange occurs</td>
<td>“Monthly”</td>
</tr>
<tr>
<td>Supplier</td>
<td>Identifier</td>
<td>A unique string associated with a Supplier for identification purposes.</td>
<td>“051246985”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Supplier.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary Contact</td>
<td>Name and contact information of the Supplier’s primary contact for this particular exchange.</td>
<td></td>
</tr>
<tr>
<td>Consumer</td>
<td>Identifier</td>
<td>A unique string associated with a Consumer for identification purposes.</td>
<td>“98546321”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Consumer.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary Contact</td>
<td>Name and contact information of the Consumer’s primary contact for this particular exchange.</td>
<td></td>
</tr>
<tr>
<td>Payload Definition</td>
<td>Identifier</td>
<td>A unique string associated with a Payload Definition for identification purposes.</td>
<td>“PCS4152-985-2014”</td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Payload Definition.</td>
<td></td>
</tr>
<tr>
<td>Query Point</td>
<td>Identifier</td>
<td>A unique string associated with a Query Point for identification purposes.</td>
<td><a href="http://www.thisisanexample.nz/querypoint">http://www.thisisanexample.nz/querypoint</a></td>
</tr>
<tr>
<td></td>
<td>Name</td>
<td>The name of the Query Point.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Description</td>
<td>A description of the Query Point</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Query Language</td>
<td>A stipulation of the query languages that are supported by the Query Point (e.g. SQL-92, CQL (239.50), XQuery, HTTP GET, etc.).</td>
<td>“SQL-92”</td>
</tr>
</tbody>
</table>

The "Identifier" attribute is described at an abstract level in order to be consistent with the abstract nature of the reference model. Therefore, there are no references to aspects such as identifier uniqueness, representation format, or similar. Implementation based on the GEA-NZ Data and Information Reference Model will introduce such aspects as needed according to their requirements.