Te Rito

Longitudinal Sector Data Hub Integration API  
DRAFT

Version:

0.2

Author:

Sky Sigal, Solution Architect

## Purpose

To describe the objective, outcomes and elements required of an API to transfer current and past educational state information in and out of a data hub.

## Synopsis

Integration development is often complex in breadth and depth, hence costly to implement. Strategies to reduce the cost include making it simpler, less time consuming and costly for others to implement the integrations. This is achievable by requiring only the simplest of widely used API protocols, removing the requirements to implement the full breadth of the APIs models before it can be used, increasing reuse of more abstract elements as opposed to requiring the use of case specific entities.

Related to that, the length of service of longitudinal data hubs implies being able to communicate with an ever-evolving set of systems, which in all probability cannot all be modified or evolved every time a new source system is integrated with it. To accommodate this reality, the data schema that requires development must be the developed with elements that are the abstracted, reusable, and flexible manner that can be reasonably designed at the time of need.

## Contents

[Description 1](#_Toc154927865)

[Synopsis 1](#_Toc154927866)

[Contents 2](#_Toc154927867)

[Purpose 3](#_Toc154927868)

[Background 3](#_Toc154927869)

[Outcomes [Objectives] 3](#_Toc154927870)

[Options [Considered & Selected] 3](#_Toc154927871)

[Constraints 3](#_Toc154927872)

[Assumptions 3](#_Toc154927873)

[Dependencies 3](#_Toc154927874)

[Decisions 3](#_Toc154927875)

[Deliverables/Outputs 3](#_Toc154927876)

[Heading Level 3 3](#_Toc154927877)

[Heading Level 4 4](#_Toc154927878)

[Appendices 5](#_Toc154927879)

[Appendix A - Document Information 5](#_Toc154927880)

[Versions 5](#_Toc154927881)

[Images 5](#_Toc154927882)

[Tables 5](#_Toc154927883)

[References 5](#_Toc154927884)

[Review Distribution 5](#_Toc154927885)

[Audience 5](#_Toc154927886)

[Structure 5](#_Toc154927887)

[Diagrams 6](#_Toc154927888)

[Terms 6](#_Toc154927889)

## Background

Integrations between systems is often costly to implement due to several factors.

Making it less costly is prudent to improve uptake.

## Outcomes [Objectives]

…

## Risks

API technology stack cannot be implemented by all 3rd parties at reasonable effort and cost.

API initially designed for primarily one domain are difficult to extend to integrate to accommodate the requirements of other business domains as time progresses.

API schema design managed by 3rd parties limit the evolution of the api to meet needs not in line with the strategies or needs of 3rd parties.

Lack of supporting information: for 3rd parties to develop clients around an API they require documentation that is current, accurate and clear. It is common to also supply an Software Development Kit (SDK) to demonstrate that the functionality exist.

## Options [Considered & Selected]

### Use an Already Accepted Integration API

One option is to use an integration standard already in use by most systems.

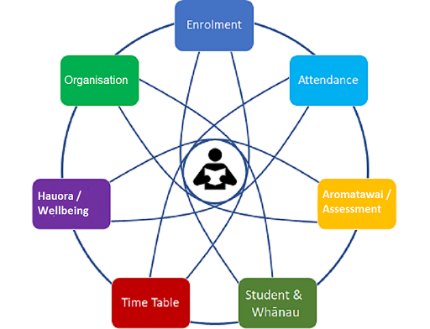
This is currently not feasible as there is no common standard already in use between sector systems. They vary per system, while also using different protocols, for different purposes.

Note:  
Internally, ELI, FIRST, NSI and ENROL all offers APIs that are specific to itself[[1]](#footnote-2).  
Externally, KAMAR, Edge, Hero, etc. all offer custom APIs[[2]](#footnote-3) while consuming to varying degrees the API end points made available by the just mentioned internal services.

### Use an Already developed Integration Standard

An option is to research for and use an integration standard that is recognised, appropriate, usable and maintained.

#### SIF



SIF is recognised.

However, SIF has poor market penetration and is barely used beyond mandated circumstances.

The reasons can be in part attributed to the technical protocols it officially uses (SOAP and XML) are both dated and more complex than more recent options (REST and JSON), its data schema is overly specific and prescribed, with both low extensibility and limited ability of being partially and incrementally implement. These qualities contribute to high costs to implement, while also being found inappropriate for use in most cases.

Attempts have been made to address these issues, leading to different incompatible versions of the SIF protocol (US, AU, GB,NZ) – which defeats the purpose of a common standard.

Note:  
At best, a fractured protocol that is not universally agreed upon or used or implemented provides material to consider when developing a new framework, and solid material to analyse as to why the protocol never gained universal or even wide traction.

### Other standard frameworks

Other frameworks have been developed and gained some level of market adoption. Common to all is keeping the scope of the protocol narrow, focusing on a single aspect of the business needs, leaving it to the systems at both ends of the integration channel to integrate this data with data sourced from other purposes.

In some cases, the protocols avoid defining transport, encoding, or even structure, to permit the widest adoption by whatever means available to each system. For example, one protocol only defines the attributes, leaving it up to implementation to decide in what format to transfer the data (CSV, SOAP, JSON, etc.).

### Development of a Fit for purpose API

The primary purpose is to enable an API for New Zealand’s Education needs, resilient

NZ’s education sector is uniquely different from those of other countries in many ways. It’s dual culture involves multiple curriculums as opposed to a single national ones, schools are managed in a devolved manner, where the national agencies provide support to schools, rather than demand accountability via information.

## Recommended Option

The TODO

## Decisions

## Design Objectives

Capable of mapping people, their relationships, their interactions, events, and the subscriptions to services, plans, assessments, improvements, and attainment of objectives, whether they occur in an educational or not context.

By being capable of mapping the above in a general sense, rather than specific education sense, the schema will provide the resilience necessary to remain useful through a long service lifespan, with the least breaking changes that would require service clients to have to update their code.

## Information Schema

A key objective is to reduce the number of education specific entities – or any other business domain for that matter – by keeping the entities as abstract as possible, reusable in as many future use case areas as possible.

## Domains

### Common

#### Attributes

##### Validity Range

: most relations between entities require additional metadata to define from when to when the relationship is valid. Best practice is the addition of FromUTC/ToUTC values to the table used to join the two entities.

Note:  
The advantages of validity ranges include:

* Decreasing security access risks by ensuring that by default relationships (eg: employment, roles, etc.) have a termination date.
* Increasing maintainability by permitting the development of relationships before they are needed, only making them available to end users at a predefined future date.

Remote Systems Identifiers

State

#### Entities

Remote Systems

[Group]

[Group] Available Roles

[Group] Role

[Group Membership] Application

[Group Role] Invitation

Acceptance

### Person Domain

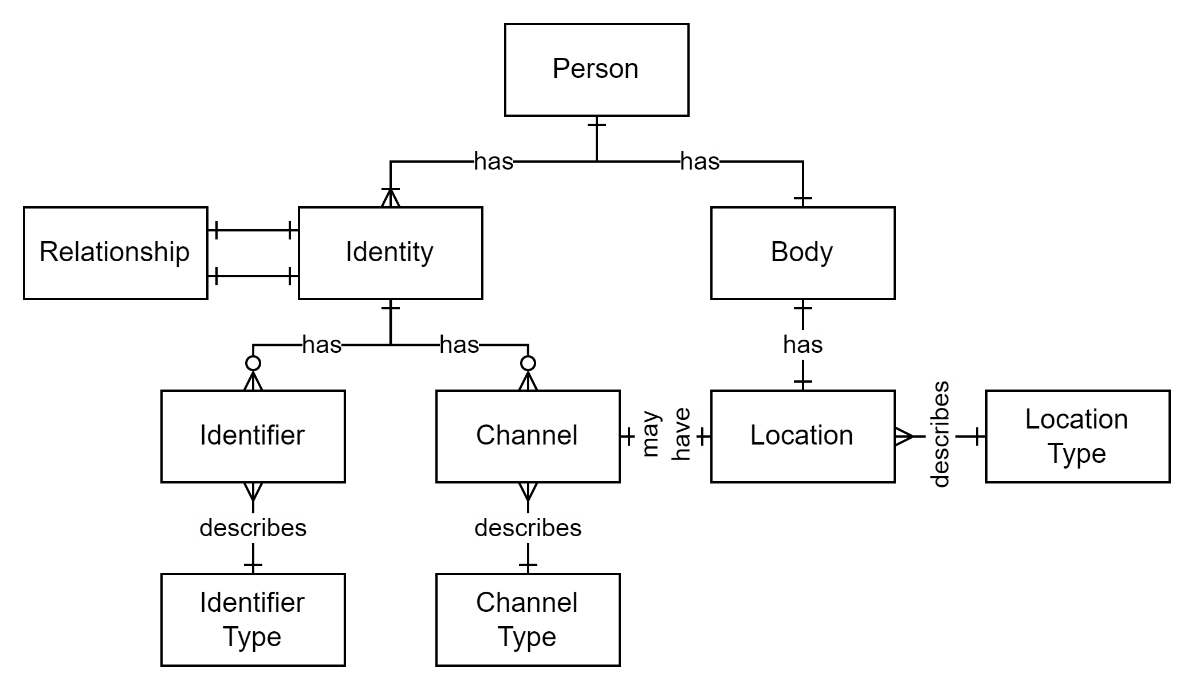


Figure : Person Domain

##### Person

: a [person](#Entity_Person) is a entity that has a single [body](#Entity_Body) and multiple [identities](#Entity_Identity).

##### Body

: a [body](#Entity_Body) gains immutable factual (as opposed to contextual) attributes at birth (sex, dob, genetics, and similar).

Note:   
Birth facts can later be used to support the development of a person’s health profile or demographic profile, but are not to be confused with these profiles.

##### Identities

: a [person](#Entity_Person) can develop multiple identities over time. A well-known example is when a person has an identity prior to marriage, and then another when married, changing one of their [identifiers](#Entity_Identifier) (their name). Another example is when a [person](#Entity_Person) transitions gender. Some identities are used as the basis of legal identification, others are just transitional experimentations.

##### Identifiers

: a person’s [identity](#Entity_Identity) can have one or more identifiers associated to it. An Identifier is a combination of [IdentityType](#Entity_IdentityType) and value(s). In the case of an identity type being ‘Name’, the value(s) will be their name(s). If the [IdentityType](#Entity_IdentityType) were – for example – “NSN”, the value would be the identity’s national student identifier.

Note:  
Due to many reasons[[3]](#footnote-4) is possible for an identity to have multiple values of the same type, but one should be defined as the preferred identifier of that type.

##### Identity Type

: an [identity](#Entity_Identity) is identified by one or more [types](#Entity_IdentityType) of identifiers. The most common is a person’s name (a social identifier). But it is important to note that names come in many formats, and that it is highly common for people to have many names – from nicknames, to known aliases, to multiple legally recognised names in different jurisdictions in different scripts (a Chinese person may have 4 or more legally recognised names).

In addition to social *names*, an [identity](#Entity_Identity) has other forms of [identifiers](#Entity_Identifier), such as system ids. For example: education, army, health system ids.

They are all valid -- but associated to different use cases.

The [identifier type](#Entity_IdentityType) defines what attributes are collected. For example, a student identifier only requires persisting one value (e.g.: the [National Student Number](#Term_NSN) ([NSN](#Acronym_NSN))), whereas a full name may require persisting multiple attributes (e.g.: the Given, Middle, Surname).

Note: the persisting of data in different fields provides no significant advantage and should not be a design deciding factor[[4]](#footnote-5).

Note:  
an [identity identifier](#Entity_Identifier) is distinct from an entity identifier, described elsewhere within this View.

##### Channels

: beyond face to face person to prson communication, one communicates with an [identity](#Entity_Identity) via channels they make available. A channel is defined by its [ChannelType](#Entity_ChannelType) (e.g., Postal, Email, Fax, Mobile, LinkedIn, X, etc.) and Value(s), which is a identifier unique to the channel type: i.e., an email address, a phone number, a LinkedIn or Twitter/X id, etc.

Note:  
in current times, most channels are digital and only require one value -- whereas historical, non-digital ones, such as postal, are composites of multiple values (unity, street, neighbourhood, city, code, country).

##### Channel Type

: the type of [channel](#Entity_Channel) (e.g., Postal, Email, Fax, Mobile, LinkedIn, X, etc.)

##### Location

: locations are the combination of [LocationType](#Entity_LocationType) and GPS coordinates which can be associated to an organisation’s assets and addresses – but can be associated to a person’s body[[5]](#footnote-6).

Note:   
It is a logical error to define Addresses as locations. PO boxes addresses demonstrate this most evidently: they remain a communication [Channel](#Entity_Channel), although they can be associated to a Location.

##### Location Type

: the customisable type of location (e.g.: “primary”, “secondary”, “home”, ”work”, etc.).

## Social Domain

Often poorly distinguished from a Person Domain, a social domain[[6]](#footnote-7) extends the individual domain, providing the entities to model the relationships of individuals to other individuals and groups and/or organisations.

Individuals belong to multiple groups at one time, where a group can be a family, a school, an organisation, or simply an ad-hoc gathering or circle of friends. Within each group individuals have a role to play. In a family group, they can have a member or responsible role. In a classroom group, they can the role of learner or teacher.

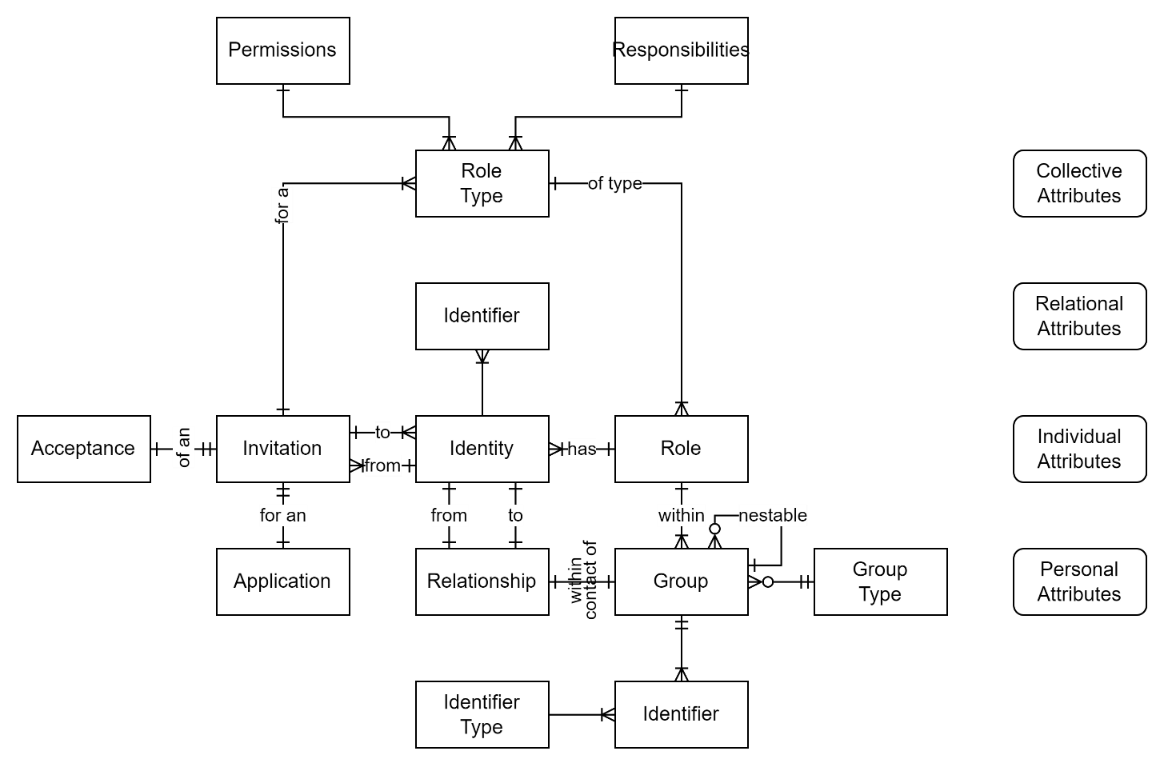


Figure : Social Domain Entities & Relationships

##### Group

: a group is a collection of Roles associated to Identities of a Person. Each identity has a Role following [RASCI](#Acronym_RASCI) characteristics (Accountable, Responsible, Informed, Consulted, Ignored, Excluded) even if they have localised labels (Parent, Cousin, restricted access).

Groups can be nested.

Groups are not limited to identity roles, and can contain other types of entities (e.g., folders, assets, resources) which the group has a relationship to (generally manages).

Groups can be logical in that they automatically include identities as Members that have certain qualities (e.g.: members of a school AND are Age 11 are Members of a cohort group).

Note:  
It is important to note that Identities are not members of a Group directly but each have a Role within a Group. In flexible systems, Roles can be assigned to both Identities and Organisations.

##### Group Type

: [Groups](#Entity_Group) cover a tremendous range of ways that [identities](#Entity_Identity) of [persons](#Entity_Person) can collaborate. They can be ad-hoc, family, or any one of many educational group types (e.g., ECL, school, tertiary, CoL, MOE Local office, industry training Organisation, etc.) as well as external groups (supplier, consumers, employers, consultants, etc.)

Note:   
It is a design consideration whether to permit different tenancies (e.g.: SMS) to refer to different types, but that have common Code Set values, or to force all to agree to the same type. We argue that flexibility of the first option is more prudent to improve take up by suppliers with the least amount of upfront configuration (the Code set references can always be added later).

##### Group Identifiers

: groups, like identities, can have multiple identifiers in different systems. For example, a Group can have a Business Number Identifier, education provider identifier, religious identifier, etc.

##### Group Type

: groups can be typed (e.g., family, school district, school, class, etc.). The type is used to indicate its purpose (family, class, etc.), but also logically constrains the type of individual and/role that can be associated to it, based on attributes associated to the [identity](#Entity_Identity) (e.g., qualifications).

##### Group Role

: an individual identity has a role in a group, even if labelled with a localised name (e.g.: “parent”, “teacher”) often aligning closely with the rights and responsibilities of traditional [RASCI](#Acronym_RASCI) roles. While informed by group roles, [relationships](#Entity_Relationship) remain distinct and supplemental to them.

##### Relationship

: a relationship is a single-direction link between a group member and the group, and/or between members of a group. For example, “brother” within a family group, or “teacher of” within a school’s class group. While relationships can cover both directions (“brothers”, “siblings”, etc.) they are in general less valuable than having two distinct relationships (each being a ‘brother’ to another individual) and their used should be avoided.

Note:  
Relationships are not expected to span outside of groups.

Profiles

Individuals have multiple profiles, for use in different domains. For example, individuals have demographic profiles, health profiles, and education profiles.

Education Profile

: an individual’s education profile is composed of gportfolio, assessments, reports, and qualifications.

Portfolio

: individuals can curate a presentation of work they believe best describes their capabilities.

If they are system users, they have a security profile. education profiles.

### Classification Domain

Classification, Categorisation.

Discovery and finding of elements in the system is improved by being able to organise Individuals and entities into logical classes based on intrinsic or applied attributes.

##### Tag

: the default way of applying attributes to entities. Tags can be nested (e.g.: “Teacher/Physics” can be under “Teacher”).

Tags can also be limited to being associated to specific types of entities (e.g., a [person](#Entity_Person) or [identity](#Entity_Identity) probably should not be tagged as a “Farm/Animal”).

Tag Type

: a tag can be typed. The praedicamenta provides the list to consider: is a, quantity, has a qualify/having, location, when.

Relationship Type:

A relationship can be developed between an active (affecting) and passive entity (affected). The effect is generation, destruction, increase, diminution, alteration,

##### Grouping

: logical grouping is already covered earlier, under Groups.

### Event Scheduling Domain

[Identities](#Entity_Identity) come together physically or virtually for a single event or a series of events.

In an educational context, this is for a single course (e.g., an evening PLD presentation) or series of courses (e.g., a semester’s classes).

The scheduling and resourcing of events may be complex or simple. For example, early learning centres only need to offer one teacher to a room for the whole day, whereas secondary schools will have more complex and fine-grained scheduling needs.

People who come together can be from the same organisation (a school’s learners and teachers) or from various organisations (e.g.: from COLs, for regional PLD, etc.).

Note:  
A data hub may only be required to record events that occurred, rather than what may happen (requiring full scheduling). If only the first case, the number of tables required by the hub are fewer, but it is our professional advice to develop the entities required to capture planned events as well. Below are listed entities required of a more mature system that can be used to record both past and planned records.

##### Event

: describes a fixed or recurring set of scheduled occurred or to occur events.

##### Event Type

: Events can be organised for a myriad of reasons. In an educational context, the events can be for Education (e.g.: classes), Informing (e.g.: parent-teacher meetings), etc. The Type can be used to inform Scheduling.

##### Event Schedule

: the set of parameter that describe the algorithm type and associated set of parameters. Includes but is not limited to recording whether the event is a Fixed (single date) or Recurring event, the start date, and if recurring, the nullable end date, the recurrence type (Daily, Weekly, Monthly, Yearly, etc.) and – if not a Daily event – the offset from the start (e.g., 3rd day of the week or month), the desired number of iterations, instructions on how to handle iterations that conflict with omissions (skip, choose the next nearest period, etc.), whether to align to Periods or skip straight to time.

The scheduled iterations are checked against [Bounds](#Entity_EventBoundary).

##### Boundaries

: date ranges which can be defined to either typed as being one that contain (e.g., school terms) or excludes (e.g.: mid-term holidays) scheduled iterations.

##### Boundary Type

: a categorisation of a Boundary as being inclusive or exclusive boundary.

##### Holidays

: a specific type of [Boundary](#Entity_EventBoundary) is a list of known national (e.g.: Mother’s Day) or other holidays that happen regularly, and whether they impact event scheduling. Most national holidays are single days, although not all (e.g.: Statutory days between Christmas and New Year’s). Holidays can be national (specific to Samoa versus New Zealand) as well as regional (e.g., Wellington Anniversary Day versus Auckland’s Anniversary Day). Effort is required to provide the capability of including or excluding Anniversaries from a schedule that it could impact.

##### Prerequisites

: events can only be scheduled if they can be offered when an appropriate quantity of specific resources are available. For example, a teaching event will not be feasible without a teacher, overhead projector, microscopes, etc.).

##### Periods

: slices of time that may or may not align with hours. Schools use periods for classes (e.g.: a school may have 40-minute class duration, with 5 minutes in between for moving between classes, with a larger gap left for lunch). A Period is described by its [period type](#Entity_PeriodType). [Period](#Entity_Period)s are collected together as [Period Set](#Entity_PeriodSet)s.

##### Period Type

: the type of Periods: e.g., may be for classes, or for moving around between classes.

##### Periods Sets

: sets of [Period](#Entity_Period)s used to develop a whole duration (e.g.: a series of [Period](#Entity_Period)s of type “Learning”, followed by a period of type “lunch break”, and then resuming with more Periods of type “learning”. The whole set would be a set that could be titled “Teaching Day”.

Note:  
Resources are described in further detail in the next section.

Event Resource Type

Resource Pools

##### Prerequisites

: the list of prerequisites is dependent on the resources needed. For example, laptop resources may need prior charging. These activities can be allocated to Events for other people (technicians).

##### Assignment

: the concretisation of the reservation of a resource from a resource pool.

### Resource Availability Domain

Events are used to reserve resources.

##### Resource Type

: Resources can be locations, people, devices, media, etc., which in turn can be further subdivided (e.g.: classrooms, gyms, playing fields, etc., teachers, learners, parents, etc., laptops, projectors, etc., books, etc.)

##### Event Resource Pools

: an event requires resources of specific type.

While there might be a preference for a resource, it may not be available (e.g., either already taken by another event or is temporarily out of commission due to a leak). To accommodate flexibility, Events are scheduled against Pools of possible resources, which are reserved based on preference.

##### Resources

: metadata to describe other records (Persons) or external resources (school spaces).

##### Resource Identifiers

: as with persons, a Resource may have multiple identities. The same photo or artefact used as a teaching resource may be described with a M`aori name, as well as an English name. The identifier/name may evolve over time. A teaching resource that is a book may also have ISBN numbers. It may be on shared with another organisation, and therefore have two distinct identifiers.

##### Classifiers

: whereas identifiers are generally unique to a single resource, multiple resources can be classified by 3rd party identifiers (e.g.: classified by an external curriculum part’s id).

##### Event Resource Limits

: resources have limits in terms of available times (e.g., business opening hours), quantity of people (e.g., microscopes and laptops can only be used by one or two people at a time, rooms can only accommodate a certain number of people without breaking fire regulations, etc.).

### Enrolment and Attendance Domain

Enrolment is used to reach a minimum to reserve resources. Attendance is a measure of the reserved resources being successfully utilised.

##### Enrolment

: similar to a “subscription” in that one is associated, implying possible service consumption, but making no statement of actual consumption (e.g., attendance). Used as a basis of reserving resources, employment contracts, funding required, etc.

##### Attendance/Absentee Type

: a type to identify whether an identity is still pending attendance, attended, virtually or not, or was absent.

##### Absentee Reason

: if absent, why, The Reason can be submitted beforehand (eg: Apologies or School passes) or afterwards.

### Education Domain

The hub’s baseline purpose is to facilitate the collection and transferring between service providers information that facilitates improving learning outcomes for a learner. A key entity managed within the hub is an education profile, associated to a Person’s Identity.

Note:  
the profile is intentionally separate from other profiles (health, etc.), and intentionally not placed as a subset of a Person or identity.

Appendices

Appendix A - Document Information

### Versions

* 1. Initial Draft

### Images

[Figure 1: ERD Crow's Feet Notation 15](#_Toc166490768)

[Figure 2: Example of ERD Crow's Feet Notation Use 15](#_Toc166490769)

### Tables

### References

**There are no sources in the current document.**

### Review Distribution

The document was distributed for review as below:

|  |  |
| --- | --- |
| Identity | Notes |
|  |  |
|  |  |
|  |  |

### Audience

The document is technical in nature, but parts are expected to be read and/or validated by a non-technical audience.

### Structure

Where possible, the document structure is guided by either ISO-\* standards or best practice.

### Diagrams

Diagrams are developed for a wide audience. Unless specifically for a technical audience, where the use of industry standard diagram types (ArchiMate, UML, C4, ERD), is appropriate, diagrams are developed as simple “box & line” monochrome diagrams.

#### Entity Relation Diagrams

Entity Relational Diagrams were developed using Crow’s Feet Notation:

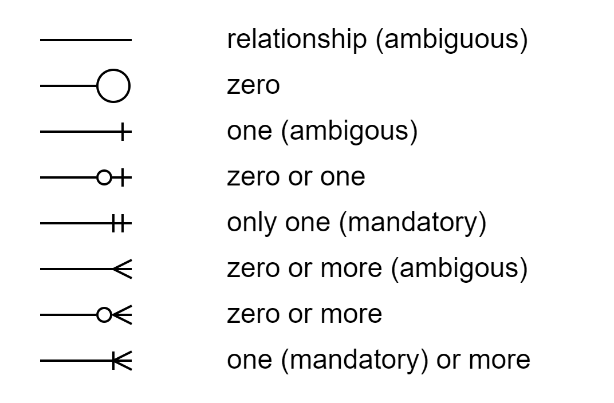


Figure : ERD Crow's Feet Notation

An example of how the above notation is used is demonstrated below:

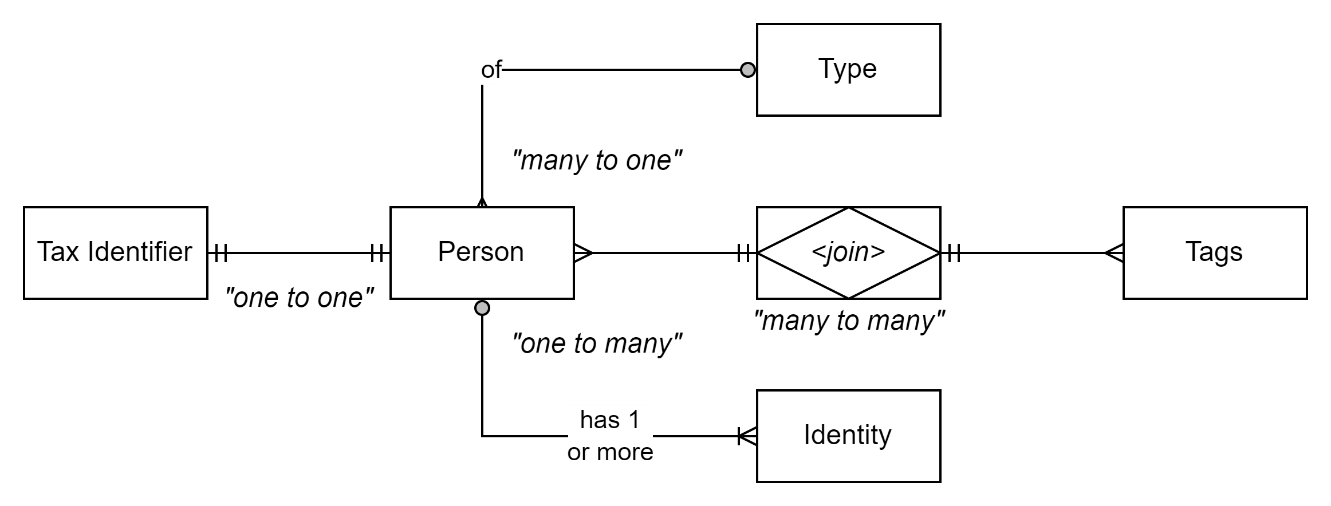


Figure : Example of ERD Crow's Feet Notation Use

### Acronyms

Refer to the project’s Glossary.

##### IT

: acronym for Information, using Technology to automate and facilitate its management.

##### ICT

: acronym for Information & Communication Technology, the domain of defining Information elements and using technology to automate their communication between entities. IT is a subset of ICT.

##### NSN

: the acronym for the [National Student Number](#Term_NSN).

##### RASCII

: acronym for **R**esponsible (to do the work), **A**ccountable (allocating the work and reporting on its progress to **I**nformed staked holders), **S**upporting (the work) **C**onsulting (on the requirements underpinning the work) stakeholders, **I**nformed (kept abreast of progress) stakeholders, or **I**gnored/Excluded others.

### Terms

##### Entity

: database and system objects that possess inherent [system identifiers](#Term_SystemIdentifier), as opposed to [value objects](#Term_ValueObject) which don’t.

##### National Student Number

: a numeric system identifier associated to a person’s [identity](#Entity_Identity) in the NZ compulsory education system.

##### System Identifier

: a unique identity within a system set (e.g. a database table). The identity is not used by outside the system.

##### Value Object

: an object that is identified purely by its properties, as opposed to a [system identifier](#Term_SystemIdentifier) (see [Entity](#Term_Entities)).

1. FIRST’s APIs are REST, augmented by ODATA, ENROL is an unsupported early version SOAP used to wrap simpler CSV payloads, NSI’s externally accessible APIs are REST based. [↑](#footnote-ref-2)
2. Using a wide range of protocols (custom xml feeds, REST, SOAP). [↑](#footnote-ref-3)
3. It is totally possible that a person is provided multiple names or even NSNs, simply due to human error and later correction. [↑](#footnote-ref-4)
4. I.e., consider avoiding making database fields for first, middle, last, etc which are really only but a western civilisation concept as opposed to universal attributes. [↑](#footnote-ref-5)
5. For example tracking learners transit to school in a school chartered bus. [↑](#footnote-ref-6)
6. [individual self.pdf (southampton.ac.uk)](https://www.southampton.ac.uk/~crsi/individual%20self.pdf) [↑](#footnote-ref-7)