DRAFT - ICT Project Guidance

Longitudinal Sector Data Hub

Version:

0.1

Author:

Sky Sigal, Solution Architect

## Purpose

This document’s purpose is to describe the design of a [longitudinal](#Term_Longitudinal) [data hub](#Terms_DataHub) appropriate for the collection, management, and dissemination of structured, accurate and current education data between systems.

## Synopsis

A [data hub](#Terms_DataHub) must provide appropriate security and control of access to data, integrations to and from remote systems to collect and disseminate data, using data schemas that facilitate long term longitudinal data collection, while providing appropriate means to meet privacy and other regulation-based obligations.

## Contents

[Purpose 1](#_Toc163120645)

[Synopsis 1](#_Toc163120646)

[Contents 2](#_Toc163120647)

[Background 4](#_Toc163120648)

[Outcomes [Objectives] 4](#_Toc163120649)

[Options [Considered & Selected] 4](#_Toc163120650)

[Constraints 4](#_Toc163120651)

[Assumptions 4](#_Toc163120652)

[Dependencies 5](#_Toc163120653)

[Decisions 5](#_Toc163120654)

[Deliverables/Outputs 5](#_Toc163120655)

[Context View 6](#_Toc163120656)

[Information View 7](#_Toc163120657)

[Design Objectives 7](#_Toc163120658)

[Design Constraints 7](#_Toc163120659)

[Domains 8](#_Toc163120660)

[Social Domain Entities 9](#_Toc163120661)

[Information Domain Entities 13](#_Toc163120662)

[Education Domain Entities 14](#_Toc163120663)

[Health Entities 15](#_Toc163120664)

[Security Domain Entities 15](#_Toc163120665)

[Technical Domain Entities 16](#_Toc163120666)

[Functional View 18](#_Toc163120667)

[Social Domain Functionality 18](#_Toc163120668)

[Information (Record & Resource) Management 19](#_Toc163120669)

[Education Domain Functionality 20](#_Toc163120670)

[Security Profile Management 21](#_Toc163120671)

[Technical Domain Entities 21](#_Toc163120672)

[Interoperability View 23](#_Toc163120673)

[Universality 23](#_Toc163120674)

[APIs 23](#_Toc163120675)

[Security, Monitoring and Auditability 23](#_Toc163120676)

[Integration View 24](#_Toc163120677)

[APIs 25](#_Toc163120678)

[Service APIs 25](#_Toc163120679)

[3rd party APIs 25](#_Toc163120680)

[Other Integration Options 25](#_Toc163120681)

[Invoking Source APIs 25](#_Toc163120682)

[ETL 25](#_Toc163120683)

[Screen scraping 25](#_Toc163120684)

[Security View 27](#_Toc163120685)

[Discoverability 27](#_Toc163120686)

[WAF 27](#_Toc163120687)

[Data Handling 27](#_Toc163120688)

[Data At Rest 27](#_Toc163120689)

[Data In Transit 28](#_Toc163120690)

[Data delivered beyond the system 28](#_Toc163120691)

[Platform 28](#_Toc163120692)

[System Monitoring 28](#_Toc163120693)

[Privacy View 29](#_Toc163120694)

[Privacy Statements 29](#_Toc163120695)

[Anonymity 29](#_Toc163120696)

[Compliance View 30](#_Toc163120697)

[Regulation 30](#_Toc163120698)

[User Interface Obligations 30](#_Toc163120699)

[NZGAS 1.1 30](#_Toc163120700)

[NZWUS 1.3 30](#_Toc163120701)

[Information Obligations 30](#_Toc163120702)

[NZ Government Open Access and Licensing (NZGOAL) 30](#_Toc163120703)

[NZ Government Data & Information Management principles (DIMP) 31](#_Toc163120704)

[Security Obligations 31](#_Toc163120705)

[Functionality Obligations 31](#_Toc163120706)

[Support, Operations and Maintenance Obligations 31](#_Toc163120707)

[Supportability View 32](#_Toc163120708)

[Operability View 33](#_Toc163120709)

[Maintainability View 34](#_Toc163120710)

[Appendices 35](#_Toc163120711)

[Appendix A - Document Information 35](#_Toc163120712)

[Versions 35](#_Toc163120713)

[Images 35](#_Toc163120714)

[Tables 35](#_Toc163120715)

[References 35](#_Toc163120716)

[Review Distribution 35](#_Toc163120717)

[Audience 35](#_Toc163120718)

[Structure 35](#_Toc163120719)

[Diagrams 35](#_Toc163120720)

[Acronyms 36](#_Toc163120721)

[Terms 36](#_Toc163120722)

## Background

Te Rito’s key objective is to improve learning outcomes by having the education record of a child follow the child on their journey through multiple education providers. An education record permits teachers to utilise information gathered at other providers to inform their tailoring of an individual teaching approach.

## Outcomes

The desired outcome is a datahub that is regularly being updated with data from remote systems to which the ministry has permission and/or manages, then making it available to other permitted systems.

## Constraints

A key constraint of the Te Rito’s project is that an agreement was made that information shared by a provider on behalf of their consenting learners is only shared with other providers, and not the Ministry of Education.

The agreement is poorly scoped and has unintended impacts. By lack of scoping, it also includes the inability to share Ministry service information, whether directly sourced (from the source registries) or indirectly sourced (the same information, but via the providers’ SMSs).

The impact is that even though it is the only operational service that collates information from the various ministry managed service, the collated current and/or longitudinal data is unavailable to other ministry services.

Access to a single collated source would both increase business insight value and decrease ministry system integration costs in the future.

## Options

### Do nothing

An option is to change nothing and integrate from the Te Rito Learner Hub directly to ministry services. The integrations between the Te Rito Learner’s data hub and ministry services can be expected to proceed at the same limited pace. The data that is integrated will remain unusable by the ministry. Other services that require the data will have to organise their own integrations, becoming duplicative technical debt.

### Amend the agreement

An option is to change the agreement. It remains unclear if the vendor product used to develop the learner data hub can be configured to isolate the data from specific sources. It would remain unclear as to what would constitute an acceptable boundary as to what constitutes the ministry’s data, and the sector’s data. The option does not solve the integration needs of other services, while putting a ministry dependency on the learner’s data hub vendor, whereas right now it is only a dependency that impacts sector providers. Finally, there is a high risk that the sector misinforms itself as to how much data is being accessed has a high risk of causing schools to stop providing information to Te Rito, putting the whole project at risk.

### Develop a Sector Hub

An option is to develop a smaller hub specific to the needs of integrating sector data, without risk of ingesting learner data. The consolidated data would be accessible by other ministry and potentially external systems. The need to integrate with a single service would reduce the integration needs of other systems. The number of integrations the Learner Data hub has to develop and maintain is also reduced to one, simplifying the relationship with the vendor. Accessed primarily by systems, the user interface requirements are negligeable.

## Preferred Option

The development of a secondary, limited scope, data hub is the only option that meets both the integration and data objectives while not putting the project’s uptake at risk.

## Assumptions

Several assumptions have been made to support the proposed design.

The ministry will continue to maintain the existing systems of record, and if anything, add more as time progresses.

A common integration pattern can be developed that can correctly model existing data in way that is less expensive than connecting to existing systems and implementing their custom but relatively simple – due to having singular purposes -- APIs.

While design of SIF AU -- and by extension the minor variation that is SIF NZ -- is flawed and may impact implementability, maintainability and modularity, the [API](#Acronym_API) design is under the control of MOE and should be improvable over time. If it isn’t, MOE can at any time it wishes develop an alternate [API](#Acronym_API) with preferred characteristics.

While an economically challenging era, there will be sufficient skilled resources available in 2024/25 to develop the proposed service component.

## Dependencies

Until this system can provide integration points and remote systems are modified to use them, the delivery of the service is dependent on this service being integrated with legacy systems by whatever means available (e.g.: using a legacy datastore-based ETL approach, rather than a more current [API](#Acronym_API) based approach).

## Decisions

The internal data schema of the system will not be constrained to integration data schemas.

## Issues

The ministry currently does not have resources capable of developing new systems.

## Risks

Risks can be categories in two groups: Risks to the overall project, some of which this component could address, and risks to delivering this component.

### Project Risks

The following all-of-project risks are provided as background information.

#### Non-Delivery

After having already once been undelivered, reset and restarted, the non-delivery of Te Rito a second time would have impact on reputation of the ministry. More so than an inappropriate design, as once in operation the design may be amended --whereas non-delivery would have immediate negative effect. Based on these stark options Te Rito has been continued.

At present, at least until deployed, both risks remain. After deployed, the second risk remains.

#### Vendor Managed Data and Service

There are differences of opinions as to whether there is risk inherent in the reliance on a SaaS as opposed to a SaaP for managing learner data. There is no instititutonal knowledge or precedent. Until now we have either owned the system and data (eg: NSI), or at least the data, on a service (eg: our documents on O365). In this case, we do not own or manage the data, or system.

### Vendor Relationship

While the relationship with the vendor is currently excellent, the very length of time data hubs are intended to be operational is a contributing factor to expecting this relationship to dis-align at some point during its service lifespan. Even if the learner data hub’s vendor was able to provide the ministry use of a consolidated view of its own information only, this service may be paused or terminated, affecting critical systems that were designed to rely on it. Putting in place options to not increase this exposure to risk and instead begin the process to mitigate this is simply prudent and a valuable investment.

### Component Risks

The following are risks specific to delivering the component described in this document.

#### Data Ownership

There is structural weaknesses in the ministry’s organisation in that departmental heads are of the opinion they own the operations, the data and the system. As two aspects – data and system -- are outside their experience, caution often leads them to prove difficult to obtain agreement to integrate with.

This perspective is out of line with other organisation’s where department heads own the delivery of the service, relying on use case specific *projections* of data extracted from data owned by the whole organisation, on systems owned by the whole organisation, but managed by an IT department.   
This project cannot solve this, and delivery will continue to be impacted by this state.

#### Integration Effort and Cost

The work, duration and budget required to integrated is almost always underestimated. Reasons include:

* Agreement is difficult to obtain between different parties.
* Different shapes of data requiring reshaping.
* A common shape cannot be obtained.
* Missing source attributes and/or records expected by the target system.
* Legacy systems use integration protocols which may be:
  + no longer best technical and/or security practice, requiring decisions as how to proceed,
  + inability to be removed due to impacts on service clients outside the influence of this organisation especially if no existing terms & conditions for the use of the older APIs, removing the option to mandate service clients invest in upgrading their services, or lose connectivity,
  + incorrect implementation of authorisation, response codes, etc. which has led to service clients to in turn be written incorrectly at great testing expense, leading to distrust of the system being implemented correctly this time, removing a desire to upgrade their services[[1]](#footnote-2).

#### Vendor Perception

The original designs for Te Rito were not followed due to a decision to stop work with the vendor originally commissioned for the learner data hub. Another vendor – CoreFour - stepped in and up to deliver some of the original RFP’s desired outcomes. This has included both the refinement of their datastore, and as a stopgap use of their integration technology to integrate with some of the ministry’s systems.

Due to this change in vendors, amongst other related changes, the limit to which the vendor is engaged for current and future work is unclear.

The ministry taking care of its own integration needs, to permit better consolidation, analysis of its own data and subsequent decisions, while also pursuing best practice of reducing unnecessary vendor reliance, risks being perceived by the vendor as a lost opportunity for instead continuing and expanding commercial activity.

# Context View

The service operates within an educational context of providers (early learning, primary, intermediate, secondary, tertiary), where a varied workforce (board of trustee members, principals, administrators, school maintenance, school nurses, teachers) provide education to learners enrolled by their caretakers or families.

The education process is based on following a curriculum that outlines expected learning outcomes towards gaining recognised achievements.

Teachers develop teaching plans for their cohorts of learners, making individualised plans as needed.

The progress of learners according to plan is tracked by observations, assessments, judgements, and recognised attainments.

## Roles

The service operates in a context where various roles participate in the education experience.

The service is not expected to define and/or have configured all possible roles but is expected to start with the following subset and be capable of adding more as and when required.

##### Principal

: the leader of a school, reporting to Board of Trustee members.

##### Administrator

: one or more individuals tasked with assisting the principal to operate & maintain the education provider’s resources, personnel, strategies.

##### Delegated Administrator

: one or more of the school administrators who have delegated permission to act on behalf of the school’s other personnel when dealing with ministry provided services.

##### Teacher

: a [person](#Entity_Person) providing education to learners according to plans developed to align with a national curriculum.

##### Nurse

: a [role](#Entity_Group_Role) that would require access to a learner’s [health profile](#Entity_Health_Profile).

##### Learner

: a [person](#Entity_Person) paying attention for the duration of their compulsory education, in return for being prepared according to a curriculum to be a young adult successfully participating constructively in society.

## Providers

The service operates in a context where presence, participation, learning and teaching traditionally has been provided at physical locations belonging to education providers. This will continue to evolve to incorporate distance, asynchronous and “blended” learning experiences.

Hence the system is expected to provide the functionality and [entities](#Term_Entity) necessary to correctly model a varied range of [provider types,](#Entity_GroupTypes) in both English and Kaupapa māori environments, from early learning centres, through primary, intermediate to secondary schools and in future beyond (vocational, tertiary, etc.)

Note:  
there are many more education provider types, but the service is not expected to need more at the start and remain capable of adding more as required.

## Enrolment

The system operates in a context where enrolment is required to reserve limited resources (e.g., teachers, spaces, educational resources, specialists, etc.) and provide information for funding decisions later.

Hence the system is expected to provide functionality and [entities](#Term_Entity) necessary to model the enrolment of learners at one or more of the above-described [providers](#Entity_Provider), whether enrolled in sequence or in parallel.

Note:  
the current ENROL system is unable to model parallel enrolment.

## Funding

The systems operate in a context where a combination of enrolment and evidenced individual student needs triggers funding being made available directly to schools.

The system is expected to provide the functionality and [entities](#Term_Entity) necessary to model the standard funding allocation[[2]](#footnote-3), progressing over time to addressing individual needs.

## Scheduling

The system operates in a context where reserved education network resources are allocated according to schedules developed on a per-provider basis.

The system is expected to provide the functionality and [entities](#Term_Entity) necessary to model the scheduling of teaching/learning events an on a per-provider basis.

Important:  
As a datahub, it is important to note that it does not need to provide the functionality to develop scheduling, but only import and model the scheduling developed by schools in their respective SMS systems, significantly simplifying the work scope.

Note:  
Early learning providers generally do not require scheduling, just participation, hence the ELI system being devoid of such functionality or modelling [entities](#Term_Entity) in its design.

## Participation

The system operates in a context where participation is considered a key desired quality of the education experience.

Hence the system is expected to provide the functionality and [entities](#Term_Entity) necessary to model participation of different roles, starting with teachers to learners, and in some cases expanding to incorporate parents and caregivers.

## Resources, Progress, Outcomes

The service is expected to operate in a context where education is compulsory.

The compulsory education experience is a process of following a curriculum that outlines a sequence to follow to transmit learnings from a teacher to a learner, making observations, collecting evidence, making assessments and/or judgements to determine progress.

Hence the system is required to provide the functionality and [entities](#Term_Entity) necessary to model the planning and progression through a curriculum.

## Systems

The system operates within an ecology of logically related but often disconnected digital systems.

Thus, the system is expected to provide the functionality and [entities](#Term_Entity) described earlier, along with the integration channels to the following systems to transfer information in from and later out to:

##### ESL

: the sector’s workforce identity provider (distinct from the DI4OL, the sector’s customer and parent’s workforce identity provider)

##### FIRST

: the system of record of providers.

##### NSI

: the system of record of past and current learners,

##### ENROL

: the system of record of the enrolment of learners at providers

##### OCH

: the system of record of the accepted curriculums and supporting materials for planning, teaching and assessing.

##### e-asTTle

: the system of record of standardised test-based evidence for assessments.

##### PaCT

: the system of record of assessments

##### Funding

: [to be investigated further].

##### EDW

: the education data warehouse.

Note:  
The above list is not final and may grow -- but it is sufficient for planning purposes for a year or more’s work.

### Isolated Funding and Systems

Due to the [AoG](#Acronym_AoG)’s – and thus the ministry’s – funding model, the above system’s have been developed as isolated systems, specific a business department’s needs, without need or desire to consolidate their system’s functionality and/or data with other systems. Project management of each system’s delivery has contributed to this outcome by ensuring that integration to other systems is outside the scope of delivery.

Hence the above systems have focused and limited usability and integration qualities, and the required integration has always been left for completion by others, later, using budgets that have not been initially planned for or requested. Consolidation has never been achieved making it impossible for whole of system information viewing, analysing, and decisions.

## Risks

The proposed option has some unique risks, and some that are shared by all the other discounted options.

# Information View

By their very nature, longitudinal systems are operational for a long time, over which it is it is to be expected that the systems that are available to source data from, and the schemas the dat.

An effect of the above is that the data specifications agreed upon at the start of service will require evolution over time.

To prepare for this eventuality, it is important that integration data schemes used to integrate with other systems be as abstract as possible, to be usable for as many use cases as possible. This observation is the basis of the choice and design of the following logical data elements.

Note:  
An example of this preference for maintainable and evolvable abstraction is the use of a single [natural] [person](#Entity_Person), with different roles (Learner, Teacher, Parent) in different contexts – rather than using distinct [entities](#Term_Entity) (Learner, Teacher, Parent) with no relationship between them, increasing duplication and error while decreasing the quality of analysis, understanding and decision making.

## Design Objectives

An important objective when developing an information communication schema is to reduce the number of elements required to model the information. The reason is that each added entity requires more design, development, testing, monitoring, and maintenance, while limiting evolvability to meet unexpected integration scenarios.

A second objective is to normalise data such that it is not duplicated across multiple [entities](#Term_Entity). For example, it would be a design error to propose 5 disconnected and/or different [entities](#Term_Entity) (Teacher, Learner, Workforce, Parents, Community/Public [person](#Entity_Person)) to manage the same [person](#Entity_Person).

A third objective is to scope [entities](#Term_Entity) correctly to intended purpose, avoiding the common design error of modelling them on common norms. The way humans communicate, using a highly flexible language is often different, varied, flexible, but also often logically incorrect, compared what a computer can accept.

A fourth objective is to support the maximum amount of sharing of information for system wide use, without compromising personal privacy.

## Design Constraints

As stated earlier, a primary objective is to organize information effectively by dividing it into distinct profiles, each representing specific states or aspects. However, we must avoid the common pitfall of treating these profiles as mere subsets of a [person](#Entity_Person), as this can lead to complications. When profiles are nested within a [person](#Entity_Person), there's a risk of brilliant. unintentionally sharing sensitive information across different contexts, which could necessitate duplicating [person](#Entity_Person) [entities](#Term_Entity). Instead, we should elevate profiles to the top level, granting them independent status. This approach enables us to reuse the same [person](#Entity_Person) entity across various domains without compromising confidentiality or integrity, resulting in a more streamlined and maintainable information system. In simple terms, the following is an example of the comparison of a naïve and incorrect design, compared to a more correct, flexible, and maintainable schema design that supports the objective:

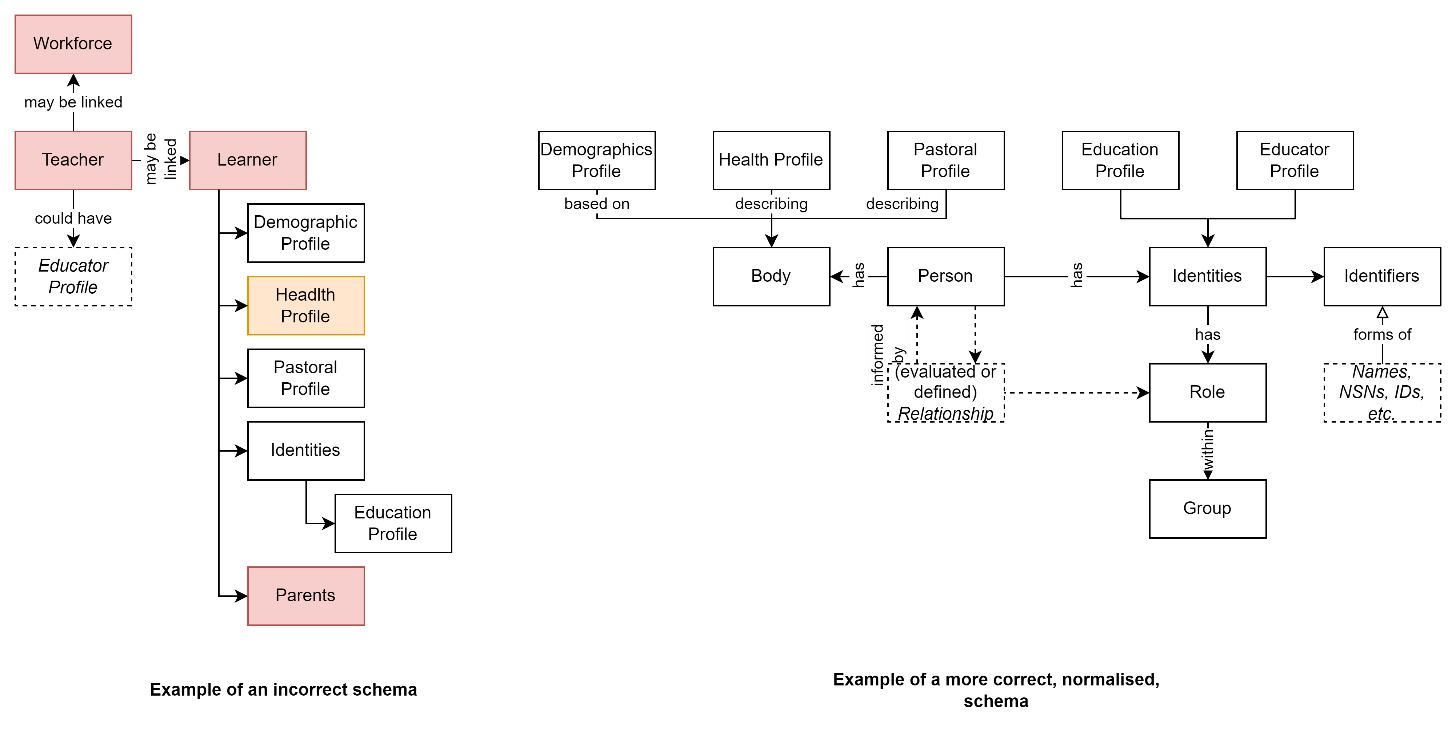


Figure 1: Comparison of normalised and denormalised schemas

Note:  
SIF-AU and therefore NZ-SIF as it currently is, is an example of an implementation of the denormalised approach that leads to well-understood information management issues, that become exposed over time in longitudinal data systems.

### Correct Analysis = expanded use case

HELP! I need the following to be made more understandable by lay persons.

A second aspect of the above approach to system information design is addressing the requirement for correct organisation of data [entities](#Term_Entity), based on prior correct analysis. As an example, it is common to talk about Learners as a top entity, with various profiles nested under it.

This results in an incorrect analysis. In the simplest terms, a person does not have a manila folder of their medical records attached to their body. They are different top [entities](#Term_Entity). There is a reference from the medical records to the body, not the other way around. This more correct two improved outcomes:

* permits an natural way to improve interoperability while not increasing the risk to privacy as now a user being shared across different systems without having to remove/filter out child properties (e.g., remove their medical records before the person is shared with another system).
* It permits consultants to share profiles to collaborate on outcomes without knowing the person it is associated with. For example, doctors can be provided a person’s medical profile and collaborate to make recommendations – even in a public p forum -- without access to a person’s [PII](#Acronym_PII). The same for pedagogical or pastoral records.

## Domains

The system requires [entities](#Term_Entity) to model multiple domains, including but not limited to:

* The social domain, covering individual [person](#Entity_Person)s and their inter-relationships within [group](#Entity_Group)s,
* The education domain, covering the aspect of educating individual [person](#Entity_Person)s,
* Separate but as important, are two Information, Communication and Technology ([ICT](#Acronym_ICT)) domains, covering [entities](#Term_Entity) required to implement the automation of the state and operations of the other domains.
  + Information management domain, covering records and resources, their state (draft, rejection, approval, publishing, maintaining, removal) and roles (collaborators, approvers, etc.) and associated workflows.
  + Technology & communication domains, required to manage system configuration, integration, as well as information storage, communication, access control, monitoring and maintenance.

Important:  
to meet the ISO-25010 expectations of being able to evolve and be maintained it is essential that the boundaries of these domains are correctly understood to ensure development of a solution does not merge into a monolithic and unmaintainable and unmodifiable liability rather than an asset.

## Social Domain Entities

Based on the outlined objectives and constraints, the following is a catalogue of the key social domain [entities](#Term_Entity) required to made available.

##### Persons

In a longitudinal system, a natural person[[3]](#footnote-4) would be expected to appear in the data at multiple times over the service lifespan, often as different roles in different contexts.   
An example of this natural progression over time might be a person being a known Infant before entering school, then a Learner while progressing through different providers of Compulsory Education, before perhaps studying to be a Teacher, then becoming one, possibly being a Caregiver to extended family members, before leaving the profession to become a full time Parent.

##### [Person] Identities

While a natural [person](#Entity_Person) has only a small set of characteristics that are set at birth (dob, biological sex) they may evolve multiple Identities over time. The identities contain more properties than a [person](#Entity_Person).

##### [Identity] Identifiers

An [identity](#Entity_Identity) has multiple identities of different types for different circumstances. The most common identifier for an [identity](#Entity_Identity) is a Name(s).[[4]](#footnote-5) An [identity](#Entity_Identity) may have multiple Names either at the same time or sequentially. For example, sequentially, a [person](#Entity_Person) may take their partner’s surname when they get married. Chinese people may have 4 or more official names at the same time, in different characters (either Latin and Chinese script). Additionally, a [person](#Entity_Person) may have different nicknames, some of which may be officially acceptable on official records. For example, a Chinese [person](#Entity_Person) may have been gifted the nickname ‘Ben’ in their first English course and this has become their name used in all conversations with English speakers. A [person](#Entity_Person) may have other system identifiers as well. For example, a [person](#Entity_Person) may have tax identifiers’s in multiple countries, as well as education sector identifiers in multiple countries.

##### [Identity] Identifier Types

Identifies the context and type of identifier as being social (name), tax, education, army, etc.

Note:  
While unfortunately quite common, it remains a logical design error to develop different [entities](#Term_Entity) for different roles, such as one for Learner, one for Administrator, one for Teacher, another for Parent, another for Other/Whanau.

##### [Person] Groups

People collaborate as members of logical [group](#Entity_Group)s, irrespective of Location. A [group](#Entity_Group) may have many members (e.g.: a class) or only one (e.g., a [person](#Entity_Person)’s estate, to which others can be invited, such as attorneys) whose membership change over time (e.g.: inheritance).

Note:   
While many systems [group](#Entity_Group) [system user](#Entity_SystemUsers)s, this system groups [person](#Entity_Person)s, who may not be [system user](#Entity_SystemUsers)s yet. This permits the ability to model families, caretakers or any other group including their security profiles, before all [person](#Entity_Person)s (e.g., grandparents) are [system user](#Entity_SystemUsers)s.

People don’t just use different name identifiers but share different identities as the basis of their role in different [group](#Entity_Group)s (“mom” in a Home group, “Ms. Lavery” in a group put together to teach and learn, etc.). As stated above, identifiers can change (Ms. Lavery got married to Mr. Smith in spring and (currently…) is preferring to being referred to as Mrs. Smith.

[Group](#Entity_Group)s are nestable (e.g., CoLs, Schools, Courses and Cohort). People’s Identities can belong to multiple groups (e.g., Cohort, Age and Course, as well as Schools if dual enrolled).

##### Group Types

[Group](#Entity_Group)s may be classified as different types - estate, scholastic, family, work, etc. – that in turn defines conventions (names, permissions and processes) specific to that type of [group](#Entity_Group).

[Group](#Entity_Group)s can be labelled as per their activity (Provider, Administrative, Teaching, Family, Ad hoc, etc.).

##### [Person or Group] Channels

Both a [person](#Entity_Person) or [group](#Entity_Group) may be contacted by multiple private channels, using legacy non-digital protocols (Postal Address) or more modern digital ones (Telephone, Mobile, Telegram, Instagram, etc.), or public non-private channels (e.g.: X, Mastodon, etc.).

Note:  
It is important to permit the list of channel types available to evolve over time as new ones become available all the time. Note also that postal is the only channel that requires multiple fields (street, city, etc.) whereas all electronic channels require only one (telephone, email, twitter, etc.).

##### [Person or Group] Locations

Locations define specific coordinate locations, usually GPS based.

A [group](#Entity_Group) may be assigned a fixed location in addition to being associated to communication channels.

A [person](#Entity_Person)’s location if tracked[[5]](#footnote-6) will be transient, periodically updatable. For example, a [group](#Entity_Group) of type School would be given a fixed Location. A classroom as well.

Note:  
A location is NOT to be confused with a communication channel of type Postal, although they can be linked if they are not a post office box.

##### [Person] Group Role

[Persons](#Entity_Person) have Roles within the [person groups](#Entity_Group). The type of role is often RASCI in nature, although the actual term used will vary by domain. For example, while classifiable as Responsible and/or Accountable, they may be colloquially referred as “Teacher” in a school context, or “Mother” in a family context.

##### Resources

Note:  
Resources – records – and Resource Roles are managed separately, under Information Domain [entities](#Term_Entity), described next.

##### [Person] Body

A natural [person](#Entity_Person) has a physical [body](#Entity_Body) shared by all its [identities](#Entity_Identity). It is this [body](#Entity_Body) that has physical traits inherited at birth that are not part of a [person](#Entity_Person)’s identities. This [body](#Entity_Body) is also the basis of the [person](#Entity_Person)’s medical and mental health records.

Note:  
while Mind is different than [body](#Entity_Body), we accept until proven to be needed otherwise that it shared across all identities and hence is a subset of [[body](#Entity_Body)](#Entity_Body). There is an argument for making it distinct, such that a [Pastoral Profile](#Entity_Pastoral_Profile) can refer to it in isolation from non-related aspects, including the [body](#Entity_Body).

##### Issues, Assessments, Assessment Types, Recommendations

A [person](#Entity_Person)’s [body](#Entity_Body) is assessed by a medical professional who records issues, physical assessments and recommendations. At school this may be a nurse, or a specialist invited to school premises (e.g.: a dentist in a caravan).

##### Demographic Profile

A [person](#Entity_Person)’s race and citizenship, based on attributes at birth, is shared across all Identities, therefore is part of a profile associated to their shared [person](#Entity_Body).

Note:  
From a purely normalise logical perspective, note that a [person](#Entity_Person)’s demographics should not be merged with a [person](#Entity_Person) or their [body](#Entity_Body), as the information is contextual, dependent on external classifications (the definitions of a denomination, etc.) as opposed to factual (dob, etc.)

##### Consents

A [person](#Entity_Person)’s records – especially their [body](#Entity_Body)’s Health and [pastoral](#Entity_Pastoral_Profile) records – may or may not be consented to be shared with other [group](#Entity_Group)s or [person](#Entity_Person)s.

Note:  
A [person](#Entity_Person)’s records are private and cannot be shared unless both the “viewee” has Consented and the viewer is Permitted.

##### Group/Resource Role Applications, Invitations, Acceptances

A [person](#Entity_Person) may apply to a [group](#Entity_Group) or [resource](#Entity_Resources)’s managers’ roles to be invited to a [role](#Entity_Group_Role) within the group or resource. If an *invitation* is issued, it can then be *accepted*, at which point the [identity](#Entity_Identity) can be associated to the [resource](#Entity_Resources) or [group](#Entity_Group).

Note:  
When importing [group](#Entity_Group) relationships from other systems, it is often the case that there is simply not enough information available to correctly model role invitation and acceptance. In such cases only it is acceptable to Assign the role to the [person](#Entity_Person).

Important:  
It is important that all Roles are time limited to short periods (e.g.: a year), to remove the possibility that a [person](#Entity_Person) retains full teacher right to a [group](#Entity_Group) long after they have left the organisation and/or retired.

## Health and Wellbeing Entities

While not a requirement of all [person](#Entity_Person)’s , a Person’s Education is affected by presence, particisiption, attention which is affected by a [person](#Entity_Person)’s wellbeing, hence the need for [entities](#Term_Entity) in this domain.

##### Health Profile

A health profile describes observations, assessments, recommendations, plans, etc. related to a [person](#Entity_Person)’s [body](#Entity_Body). In a scholastic environment, it is common for school’s to participate in the monitoring of a learner’s hearing, sight, teeth, etc.

The health of the mind is – in a pedagogical environment – is traditionally managed separately, under a pastoral profile.

##### Pastoral Profile

In a pedagogical learning environment a pastoral profile is used to manage the general wellbeing of a learner.

Note:  
in contrast to the compulsory public education sector, in a religious environment – as opposed to a pedagogical context -- the term “pastoral” is also used to manage the well-being of a person, but is more focused on the wellbeing in relationship to religion. Faith based providers use a pastoral profile to manage both aspects of wellbeing.

A [person](#Entity_Person)’s personal wellbeing, family, faith, political inclinations, etc. is accepted as shared across all [identities](#Entity_Identity), therefore a profile is associated to their shared [person](#Entity_Person).

## Education Domain Entities

Reusing both the social and information domain [entities](#Term_Entity), the system requires few education domain specific extension [entities](#Term_Entity).

##### [Education] Provider

An education [provider](#Entity_Provider) is essentially just a specific *type* of logical [group](#Entity_Group) of [persons](#Entity_Person) and [resources](#Entity_Resources).

As with any group, a [provider](#Entity_Provider) may have multiple communication [channels](#Entity_Channel) to get in touch with it, but does not necessarily have a [location](#Entity_Locations) (e.g.: Correspondence Schools could be an example of this).

##### Education Profile

A [person](#Entity_Person)’s [identity](#Entity_Identity) enrols in one or more education (e.g.: class) [group](#Entity_Group)s within an education [provider](#Entity_Provider)’s [group](#Entity_Group) to improve their understanding and application of their abilities to improve their skills and experience.

Progression of an [identity](#Entity_Identity) is mapped using [evidence](#Entity_Evidence), [assessments](#Entity_Assessment) and [achievement](#Entity_Achievement) awards.

##### Enrolment

A domain specific term for reserving limited resources.

Note:  
See Subscription in the information domain listed later.

##### Scheduling

Scheduling is an aspect of reserving resources to facilitate teaching learners. It involves reserving time (e.g., teaching period), location (e.g., classroom), material (e.g., computers, projectors), personnel (e.g., teachers) and consumers (e.g., learners).

Note:   
scheduling is not a requirement for most early learning providers but is essential to primary, intermediate, and secondary providers.

##### Attendance

A domain specific term for a general concept, [Attendance](#Entity_Attendance) used to be limited to tracking physical presence, and has evolved be include virtual presence, at scheduled times, or asynchronously, later at different times.

##### Portfolio

Learner output continues to be a mix of physical and digital output, produced both in class or virtually, requiring persisting as digital records (e.g.: test scores) and digital records of media (e.g.: scanned artwork).

Note:  
The increasing size of media requires ensuring the system has sufficient resources to persist them for many years while remaining responsive and available – as well as inexpensive.

##### Evidence

The material in [portfolios](#Entity_Portfolio) as well as non-digital or physical activities become the basis of evidence to support judgements and assessment. These can be developed as a system record, or a record of summarised media (e.g.: a scan of a test).

##### Assessments

Assessments of ability may be made based on collected evidence, informing an understanding of progress.

##### Achievements

Achievements are the records of achieving small tasks, on their progression towards attainments, and can be the result of judgement or standards-based test.

##### Attainments

Attainments are the records of being assessed as passing a specified level,

**Important:**  
The terminology used in the education domain varies over time and location. It is a design error of both internal and communication [entities](#Term_Entity) to map [entities](#Term_Entity) to terminology that is external domain vernacular terms that may change over time. It is far more important - even crucial - to strive to develop [entities](#Term_Entity) that are well defined (i.e., bound) and logically named. Leave evolving nomenclature to be limited to a presentation layer, rather than letting it infect multiple system layers, right down to the data storage layer.

## Information Domain Entities

Individuals interact to create and or consume resources.

While social domain [entities](#Term_Entity) are required to map how individuals come together, they are insufficient to map the development, maintenance and access to resources created.

##### Resources

The term “Resources” cover both records in their own right (assessment records, etc), and records of uploaded digital media (evidence, medical records, etc.).

##### Resource Metadata

To facilitate searches, [resources](#Entity_Resources) are saved with metadata (created when/by, title, purpose, etc.) to facilitate later discovery.

Note:   
consider the inclusion -- but not limitation -- to the use of Dublin Core metadata to categorise resources to support efficient discovery later.

##### Resource State

Simple resources may have none, but more important Resources may be collaborated on through a variety of states (Draft, for Review, Rejected, Approved, Released, Replaced, Retracted, Retired, etc.).

##### Records

In addition to the above core concepts additional business context records are expected to be required.

The above records are sufficient to map most conditions. For example, an adult Learner at night school, who is a mother to a child and carers to others may be relatively easily mapped to a [person](#Entity_Person), with multiple Identities associated to [roles](#Entity_Group_Role) in multiple [groups](#Entity_Group), with medical records associated to their [person](#Entity_Person)’s [body](#Entity_Body), and scholastic progress records associated to at least one of their [person](#Entity_Person)’s [identities](#Entity_Identity). If one needs to contact a learner, one can use one of the available communication channels associated to the [identity](#Entity_Identity) associated to the Learner role at the school’s [group](#Entity_Group). They will be assigned a role to homework resources assigned by their teacher.

##### Resource Roles

Users have roles that relate to [resources](#Entity_Resources). They may be the original creator of the resource or be invited (see Invitations above) to become collaborators, reviewers, approvers, maintainers, or consumers of the resource, which come with Responsibilities which they can Accept or not in order before their role allocation Approved.

Note:  
it is an impactful design error to design/offer a service that only provides for the means of unilaterally allocating a role.

## Security Domain Entities

A certain number of [entities](#Term_Entity) are required to provide privacy, consent, auditability.

[Consent](#Entity_Consents) has already been covered in the social and information domain [entities](#Term_Entity) listed above, and Auditability is addressable via standard technical domain [entities](#Term_Entity) (specifically Session Operation records) described next.

Privacy has been addressed in the social domain by isolating identifiers from other objects making them easier to control access in a more granular manner, permitting wider sharing of non-[PII](#Acronym_PII) information while reducing risk of exposing Personal Identifying information ([PII](#Acronym_PII)).

## Technical Domain Entities

The entities outlined in the above domains (i.e., social, education) are reliant on the system having sufficient [entities](#Term_Entity) defined to develop automation to facilitate users. As such, the system is also expected to have many of the following [entities](#Term_Entity) as well.

Note:   
The following entries are standard to mature systems and there is sufficient computer science[[6]](#footnote-7) literature readily available and do not require in depth descriptions here.

##### Confidential System Integration Service Keys

##### System Configuration Settings

##### Diagnostic Trace log entries

##### Error Report entries

##### Tenancies

: a logical separation of user [group](#Entity_Group)s.

**Important:**while resources can be logically separated, it is a legacy design error from the 90’s to physically separate users (it disallows collaboration and/or dual enrolment across organisations, schools, etc.)

##### Resource Routes

: route management becomes a requirement when resource versioning is introduced, such that links to older versions can suggest later editions, etc.

##### User Sessions

##### User Operation Audit records

: permanent records of activity initiated by [system users](#Entity_SystemUsers).

##### System Users

: [persons](#Entity_Person) who are using 3rd party digital identities they manage, to identify themselves to this system and use its functionality.

##### Permissions

: system permissions are bundled as system access controls that mirror roles.

##### Access Levels

: the technical implementation of social domain roles.

##### Accounts

: accounts are used to manage tenancies (e.g.: per department, per school).

##### Subscriptions

Note: a subscription is a business domain neutral name for enrolment.

##### User Digital Identities

: a [system user](#Entity_SystemUsers) may be identified based on one or more 3rd party digital identities (MS, Google, etc.)

##### User [System Settings] Profiles

: [system users](#Entity_SystemUsers)’ collection of personal (as opposed to all-user) system settings.

##### User Security Profiles:

The collection of access roles and/or permissions associated to a [system user](#Entity_SystemUsers).

## Design Considerations

A key consideration applicable to all [entities](#Term_Entity) that are transmitted between multiple systems is the need for [entities](#Term_Entity) to have multiple identifiers, each associated to a different systems.

For example, school A may have a person record describing Learner “Billy T”, with an id of “abc124” that is shared with the datahub, where it is saved as with an Id of “89874”, and shared with another SMS, where “abc124” has already been issued, and therefore requires another id of “fec458”.

As all items can -- at some point in the future -- expect to be round-tripped back to the SMS, each record will require storing at least all 3 of the above ids, each one linked to information about the various systems.

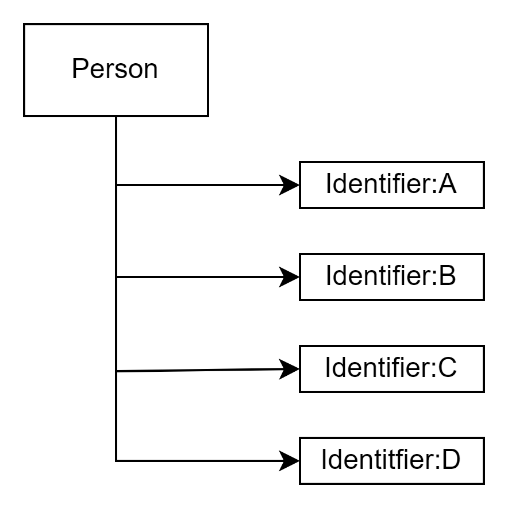


Figure 2: Multiple Identifiers per object.

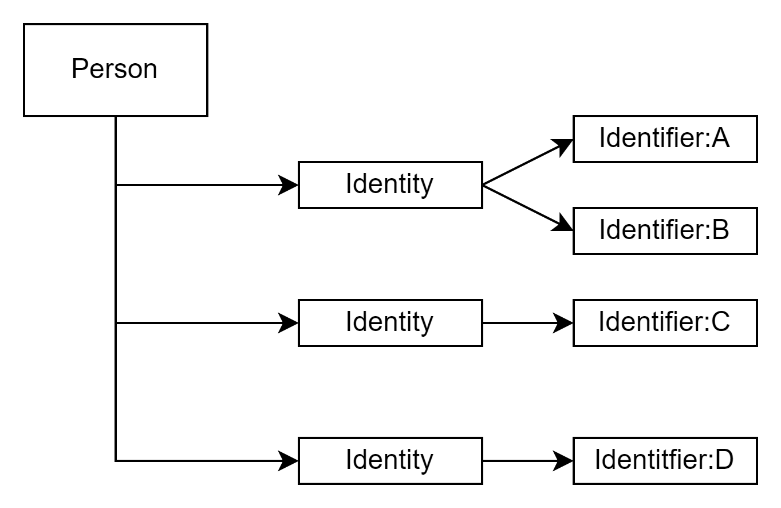


Figure 3: Multiple Identifiers

# Functional View

A [data hub](#Terms_DataHub) requires the exposing of its functionality primarily to service clients via [API](#Acronym_API)s, as opposed to end users, via User Interfaces ([UI](#Acronym_UI)).

That said, the system also requires a [UI](#Acronym_UI) to permit operating and managing the system itself, such as updating reference data, its use, as well as developing reports on the data, whether it be for current state or comparisons of state over time.

The functionality required is the browsing & listing, retrieval, editing, adding and logical deletion (i.e., logical state flag changes) of records and resource metadata records.

The functionality is described below, again broken down into distinct domains to improve maintainability over time while remaining practical to use by [system user](#Entity_SystemUsers)s.

## Social Domain Functionality

Irrespective of a specific business domain, functionality is required to manage the [entities](#Term_Entity) used to map the interactions of individuals within [group](#Entity_Group)s.

##### Person Management

: functionality to create, retrieve, update, and logically delete [person](#Entity_Person)s, who may or may not be [system user](#Entity_SystemUsers)s.

##### Person Body Management

: functionality to manage a [person](#Entity_Person)’s birthdate defined attributes, irrespective of the *identities* they may project in different environments or [group](#Entity_Group)s.

##### Person Identity Management

: the management of one or more identities that a [person](#Entity_Person) may project in different environments or [group](#Entity_Group)s.

##### Person Group Management

: the management of [group](#Entity_Group)s, their nested structure (schools, courses, etc.), and typing (provider, class, family/caregivers, ad-hoc, etc.).

##### Person Group Role Application, Invitation, and Acceptance Management

: the means for [group](#Entity_Group)s to self-administer the association of [person](#Entity_Person)s to groups closest to source of knowledge, rather than using a central processing bottleneck (i.e. by allocation, centrally).

##### Demographic Profile Management

: functionality to manage Demographic Profiles of [person](#Entity_Person)s. This includes but is not limited to race and citizenship.

Important:  
For multiple reasons – both logical and technical -the education profile is not a sub-entity of a [person](#Entity_Person)’s [identity](#Entity_Identity), but a distinct domain entity record that links to a [person](#Entity_Person)’s [identity](#Entity_Identity). A key reason is that it permits managing demographic information without necessarily making information attached to it be available to users who request information about the [person](#Entity_Person).  
The inversion permits managing of [person](#Entity_Person)s who are not to be tracked for their demographics (e.g., teachers, workforce, caretakers, community, etc.) and do not require having or exposing a demographic profile.

## Health Domain Functionality

##### Health Profile Management

: functionality to manage health information about a [person](#Entity_Person)’s [body](#Entity_Body) and mind. This includes attaching Observations, Evidence, Assessments, Plans, Events, Progressions toward.

Note:  
the functionality is similar in many respects to the operations required to manage an Education Profile.

##### Pastoral Profile Management

: functionality to manage information about a person’s wellbeing. Again, a profile is used to collect and manage Observations, Evidence, Assessments, Plans, Events, Progress.

## Education Domain Functionality

##### Education Profile Management

: functionality required to manage education records, associated to a [person](#Entity_Person)’s [identity](#Entity_Identity).

Important:  
For multiple reasons – both logical and technical -the education profile is not a sub-entity of a [person](#Entity_Person)’s [identity](#Entity_Identity), but a distinct domain entity record that links to a [person](#Entity_Person)’s [identity](#Entity_Identity). A key reason is that it permits managing an education without necessarily making information attached to it be available to users who request information about the [person](#Entity_Person).  
The inversion permits managing of [person](#Entity_Person)s who are not Learners (i.e., teachers, workforce, caretakers, community, etc.) and do not require having or exposing an education profile.

##### Enrolment Record Management

: functionality required to manage the enrolment status of an Education Profile with an Education Provider.

##### Attendance Record Management

: functionality to manage the attendance of an enrolled Education Profile to teaching events organised by an Education Provider’s.

## Information (Record & Resource) Management

Well-designed, mature, systems provide functionality to meet information management needs irrespective of the business purpose of the information.

Note:   
When this baseline common functionality is put in place first, it provides the elements to permit being profitably leveraged for business specific needs – e.g., Education Domain Functionality, described earlier – with the least duplication and delivery risk.

##### R&R Metadata Management

: the management of metadata about media uploaded by [system user](#Entity_SystemUsers)s, such as media mime type, size, original filename, current name(s), etc.

##### R&R Tag Management

: the management of tags able to be associated to resources and records.

##### R&R State Management

: the management of resources and records through different states (e.g.: Draft, For Review, Rejected, Approved, For Publishing, Published, Merged, Removed, Restored) is required to manage the development of R&Rs through multi-stakeholder workflows.

##### R&R Collection Management

: the management of the allocation of resources to nestable groups of resources.

Note:   
an R&R group is often referred to as a ‘folder’.

##### Record & Resource Version Management

: the management of versions of records is required to keep records to provide insights as to change over years within a longitudinal datahub.

Note:  
The use of Temporal Tables – available in SQL Server, Oracle and Postgres -- is an option to use to provide this functionality.

##### Resource Access-Control/Role Management

: the management of the access control (‘roles’) to develop resources.

Note:   
Common roles include Creator, Collaborator, Approved, Publisher/Manager, Endorser, Consumer, etc.

##### Resource Group Role Application, Invitation, and Acceptance Management

In simpler systems, the resource permissions granted to users can be based on [group](#Entity_Group) roles, but when required to import information from remote systems – and more importantly, round-trip them --it is prudent to implement a system that can more accurately map more complete flows so as to not lose required information.

##### Copyright Management

: the hub is currently scoped to include sector data. While most resources are records (sourced from FIRST, NSI, etc.) there is no certainty that the data remains constrained to records that exclude media. In which case, to adhere to DIMP constraints, copyright information is required to be associated as a type of [resource metadata](#Entity_Resource_Metadata).

**Important:**copyright – including copyleft – is a western concept that does not correctly map Maori expectations, which more closely relates to Consent. It is a design error to provide a data hub for use in New Zealand that only provides functionality that adheres to copyright.

## Security Profile Management

##### Security Profile Management

: functionality to manage the associations of system Permissions to Users, who are in turn associated to an [identity](#Entity_Identity) that belongs to a [person](#Entity_Person).

Important:

A system Security Profile can be a sub entity of a [system user](#Entity_SystemUsers), but there are some benefits to consider implementing the Security Profile as a parent entity that in turn links to a [system user](#Entity_SystemUsers).   
The inversion would permit [system user](#Entity_SystemUsers)s be transmitted across the wire without including Security aspects specific to only one of the two systems in question.

##### System Group Role Management

: functionality required to develop User Roles for Groups, as logical groupings of System Permissions.

##### System Group Role Application, Invitation, Acceptance, Allocation Management

: functionality to manage the application of a [person](#Entity_Person) for a specific role, that – if accepted – leads to the issuance of an Invitation for a specific [person](#Entity_Person) to a specific role within a specific [group](#Entity_Group), that can be Accepted or not when received, which would trigger a role allocation step.

Note:  
It is important to model the full role association steps rather than just the Allocation stage. It is an important component that permits the improving and facilitating of the users onboarding -- to the right [group](#Entity_Group) (s) – process, later.

## Technical Domain Entities

The system is to be used, monitored, and managed by [people](#Entity_Person) that are [system user](#Entity_SystemUsers)s.

Based on their roles within relevant [group](#Entity_Group)s, [system user](#Entity_SystemUsers)s are provided system permissions.

##### System User’s Identity Identifier Management

: the management and association of 3rd party external digital identities to a [system user](#Entity_SystemUsers). Note that a user is associated to a [person](#Entity_Person) (see above).

##### System User’s Identity System Preferences Profile Management

: the management of the personal preferences of a [system user](#Entity_SystemUsers). This may include their avatar image, their display name, their styling preferences (colours, sizes, etc.), their preferred language.

##### System User’s Security Profile Management

: the management of the permissions and access levels granted to a [system user](#Entity_SystemUsers).

##### System Permission Management

: the management of system defined permissions that may be bundled together as an Access Level (commonly referred to as a System Role)

##### Access Level Management

: the management of system permissions bundled together as logical [group](#Entity_Group)s, intended to facilitate the assignment of permissions to [system user](#Entity_SystemUsers)s.

# Interoperability View

As a data hub a key concern of the service component is ensuring data be able to be imported into the system and made available later to other systems.

Information must be usable by remaining accurate and complete, to a point sufficient to model the problem on which action or a decision is required.

Note:  
Interoperability – an information concern -- is distinct from Integration -- a technical concern -- discussed later.

As such a certain number of outcomes are required to be delivered.

## Universality

The system must persist and transmit data using universal types and formats where such exist. The use of UTF-8, UTC, UUID are required.

## APIs

The information is required to be available and manageable via access controlled, audited, [API](#Acronym_API)s. [API](#Acronym_API)s can generally be categorised as Resource or Process based.

Note:  
Current best practice is to use Resource-based [API](#Acronym_API)s (i.e. REST) for communication between systems in different networks by different vendors, reserving Process (e.g.: RPC) based calls to same-network system communication.

*All* of the functionality described in the Functional View should be exposed via [API](#Acronym_API). This not only permits 3rd parties contributing information to the service, but also the development of independent user interfaces systems, rather than embedding interface development in the server (a very common, but lower long-term value, design and development approach).

## Security, Monitoring and Auditability

Communication is encrypted, and access to the [API](#Acronym_API)s is controlled and audited, and this is covered in the Security View.

# Integration View

Data Hubs are dependent on integration being established between multiple systems.

It is a well-established design principle to improve resilience and availability by diminishing the number of dependencies a system has. To do this, the preferred outcome is for a datahub to provide [API](#Acronym_API)s for other systems to use so that the hub does not rely on the remote services.

For a number of reasons, that approach is not always possible:

* many Legacy systems may not be technically capable or have sufficient funding to provide for the development of automation to invoke the above-described [API](#Acronym_API)s.
* the service provider may not wish to make the effort to facilitate access to their data, whether it by PULL or PUSH.

Therefore, the decision tree as to which approach can be used will be approximately as follows:

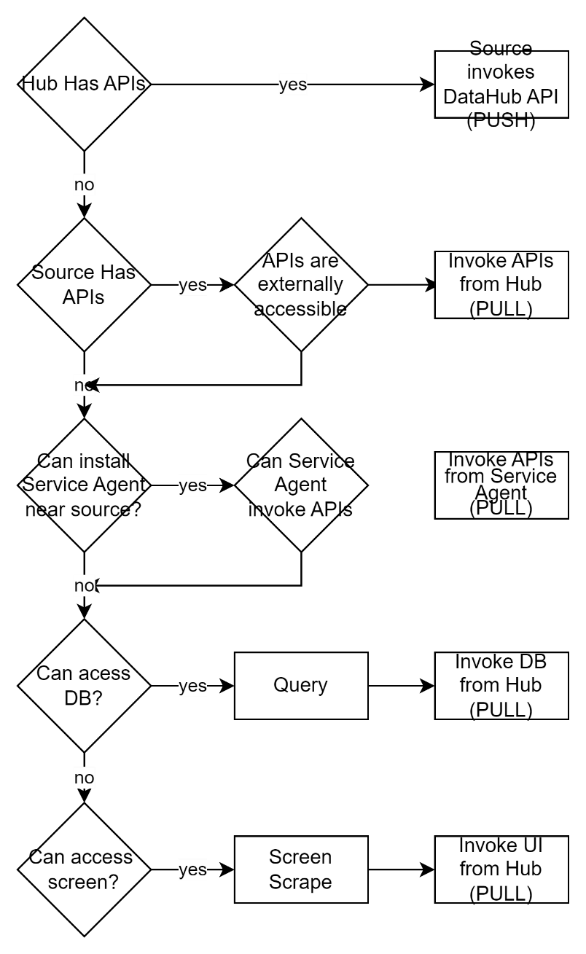


Figure 4: Integration approach decision flow

## SAPIs

When discussing integration by [API](#Acronym_API)s it is important to disambiguate which service is being discussed.

#### Service APIs

The data hub is required to provide [API](#Acronym_API)s to expose the capabilities and functionality outlined in the Functionality View. The [API](#Acronym_API)s are for invocation by 3rd party systems to transmit to and/or query information from the data hub.

#### 3rd party APIs

Many legacy systems may not be technically capable or have sufficient funding to provide for the development of automation to invoke the above-described [API](#Acronym_API)s.

In such cases, the list of viable alternative integration options is few, and are discussed in the next section.

## Other Integration Options

### Invoking Source APIs

If the source system is not modifiable, but does present an integration [API](#Acronym_API), then it recommended that the server invoke them to extract information when needed, often via a timer based regular poll, whether it be initiated from the data hub or an agent on its behalf.

**IMPORTANT**:  
It is essential to understand that [API](#Acronym_API)s can be categorised as one of two kinds: User Interface support [API](#Acronym_API)s, and Integration [API](#Acronym_API)s. User Interface [API](#Acronym_API)s are generally limited to accepting queries about a single tenancy at a time and do no lend themselves to being used for integration purposes and another approach is required.

### ETL

When the [API](#Acronym_API)s are absent or not of the right kind, If permitted, a remotely controlled Service Agent can be installed on a device within the same network as the target system, such that it can be instructed to connect to the system’s datastore, query it directly, and send back the results to the Service Agent’s controller (i.e., the datahub).

Important:  
The approach should be avoided if [API](#Acronym_API)s available, as there are numerous prerequisites and issues to consider.

### Screen scraping

If a remotely or trigger controlled service agent cannot be installed in the same network as the service to query, but the system is remotely viewable, then the only option left is performing screen scraping.

**Important:**  
The approach should be avoided if there are [API](#Acronym_API)s available, as there are numerous prerequisites and issues to consider, and can be expected to deliver poor quality data.

# Infrastructure View

The two key components of the system are:

* a data hub, comprised of a web service backed by a relational datastore, and
* a service agent, installable as required to PULL data from remote systems.

## Service Deployment

The system’s server component is deployed to organisation-managed cloud infrastructure, preferring fully managed (e.g., PaaS) to self-managed (e.g., IaaS) services.

To host There are no technical reasons for requiring more than a PaaS platform, which is sufficient to meet the security and infrastructure obligations summarised in the Compliance View.

## Service Agent Deployment

The system’s service agent can be deployed as required to devices within the same network as source remote systems.

As the source systems are all on organisation managed infrastructure the deployments are done by pipeline automation, as opposed to having to develop an installation package.

Note:   
The similarity to EdsbyLink is to be expected: the integration process is a well-known design pattern that has been used to develop several Azure products (e.g., SQL Server Agent, data pipelines agent, etc.)

**Important:**A key difference between the service agents used for the learner hub (i.e., EdsbyLink) and this hub’s service agents is that they are not expected to be accepted and/or installed on within school networks.

# Security View

## Discoverability

The system is deployed using the organisation’s on a domain that follows the sponsor organisation’s DNS naming strategy (e.g.: *service*.education.govt.nz) to provide a first line of trustability to service consumers.

## WAF

The service is protected behind a Web Access Firewall.

The WAF requires configuration to adhere to UN embargo obligations as well as more traditional controls (Denial of Service attacks, SQL and command injection attacks, etc).

## Data Handling

Delivery team members undergo security training, with exposure to the following points:

* Production Data is never to be made available outside of a Production Data environment, whether in cleartext, encrypted or a scrambled manner.
* No access to production data of any kind (configuration, operational, reporting or backup) is permitted in any environment that is not audited for both access and change.
* Test data is developed for purpose and is *not* taken from production data.

## Data At Rest

Dedicated security stores are used for persisting credentials necessary to set up integrations between devices and services.

Consider using multiple data stores: a reason being to separate Personal Identifying Information ([PII](#Acronym_PII)) from Operational Data.

Note:  
When separated, if a nefarious actor gains access to a database they can see that [person](#Entity_Person)s have or have done things – but cannot know whom did them.   
  
Using separate datastores does complicate – but only slightly -- the development of Browsable or Reportable lists of [person](#Entity_Person)s.

All data stores are encrypted where technically feasible.

## Data In Transit

Data is transferred between networks through secured encrypted channels where technically feasible (e.g.: HTTP/S).

Unencrypted (i.e., HTTP) traffic is redirected to encrypted (i.e., HTTP/S) equivalent endpoints.

Unless technically impactful, Encryption algorithms used are maintained to the current latest versions.

Functional mementos (i.e., session, language, etc. Web Browser Cookies) are marked as only transmittable over secured channels.

Server-side issued mementos (i.e., Web Browser Cookies) are marked as being closed and inaccessible via client-side script to disallow tamper ability. The server checks returned cookies that they are still marked as secure and http-only.

## Data delivered beyond the system

Information generated for consumption outside the system’s ability to audit access – i.e., diagnostics logs, electronic mail – are cleansed of confidential information ([PI](#Acronym_PI), passwords, integration configuration and credentials, etc.).

Information transmitted to users outside the system ability to audit access – i.e., emails, email attachments -- must be limited to notifications that information is available on the system, where access can be audited.

Additionally, the message headers are required to describe the security profile of the message (CONFIDENTIAL, etc.)

## Platform

The service is deployed to an ISO-27001 Level2 compliant environment that provides a 99.95% or better availability, that is monitored for abnormal activity, and audits access using sponsor services where they are available.

It is outside scope for the service provider is not expected to develop a monitoring solution for the sponsor.

## System Monitoring

The system must be capable of the use of industry standard integrations patterns to be connected to the sponsor’s organisation monitoring systems for inappropriate performance, resource consumption, operation outcomes, and user access.

# Privacy View

## Privacy Statements

Data is provided to the system primarily by systems as opposed to [person](#Entity_Person)s. Their privacy statements must be sighted to cover the sharing of data with this system’s sponsor organisation.

Users of this system’s User Interface must be provided a privacy statement that covers the information being shared with the data hub’s service clients that integrate with it.

## Anonymity

[Person](#Entity_Person)s may request that information retained about them be forgotten. The EU’s General Data Protection Rights (GDPR) has clarified that the removal does not necessitate the physical removal of data, but only requires data anonymity be delivered.

[Anonymisation and pseudonymisation | Data Protection Commission](https://www.dataprotection.ie/en/dpc-guidance/anonymisation-pseudonymisation#:~:text=Fully%20%27anonymised%27%20data%20does%20not,individuals%20are%20no%20longer%20identifiable.)

# Compliance View

### Regulation

Regardless of the quality of Quality Requirements developed to manage expectations, a number of regulations are required to be followed before the service can be made available for use.

A part of the Change Advisory Board (CAB) obligations to verify compliance to the following

Note:  
CABs obligations is to also verify for Security, Supportability, Operability and Maintainability, the expectations covered in the applicable views.

### User Interface Obligations

#### NZGAS 1.1

Since 2019, the NZ Government Web Accessibility Standard mandates that all external and internal websites meet WCAG 2.1+, AA+.

   supporting meeting 

#### NZWUS 1.3

Since 2019 NZ Govt Web Usability Standard mandates that all external and internal websites meet a common set of usability specifications covering the following:

* all of government logo & link
* Contact information
* General Copyright statement
* Privacy statement
* Link to non-html files
* printable web pages that non-content aspects, including navigation, etc.

### Information Obligations

*A number of frameworks are regulated as expected outcomes of financing by treasury.*

#### NZ Government Open Access and Licensing (NZGOAL)

Although anything specific to a Learner or other [person](#Entity_Person) is not public information unless granted so, the hub’s content includes information of value to the public.

Such resources must meet both the country’s and all-of government copyright obligations – specifically ensuring that copyright is definable, and made available via interfaces (including [API](#Acronym_API)s).

Additionally, Information is to be made available to authenticated and audited end users.

Note:   
It is a design error to mistake openness to [API](#Acronym_API)s available to unauthenticated users, as it leads to an elevated risk of service unavailability and use unaccountability.

#### NZ Government Data & Information Management principles (DIMP)

Management of information is expected to align with DIMP principles. The functionality provided under the Information Management Domain of the Functionality View – specifically, metadata, versioning, and state workflows -- provide the alignment required.

### Security Obligations

The design of the system must adhere to NZISM’s guidance, to protect system and information, which forms part of the PSR process required of all government managed services.

The system requires a data classification being assigned to the data managed by the system.

Additionally, the system requires being certified and accredited as being sufficient to protect information of this classification. The controls required are addressed within the Security View.

### Functionality Obligations

A number of Digital Services and Digital System design principles are expected to be followed. Being a data hub, essentially headless/user interface-less, for access primarily by remote systems as opposed to being directly accessible by end-users, the principles –are not expected to follow where applicable, while not being obligations.

### Infrastructure Obligations

All government agencies are obligated to select cloud infrastructure unless a compelling reason to do otherwise is presented.

### Data Obligations

In addition to DIMP obligations mentioned above, Data Sovereignty agreements direct the use of NZ hosted cloud infrastructure when it becomes available and accredited, while remaining comparable in functionality[[7]](#footnote-8).

### Support, Operations and Maintenance Obligations

While not regulated, Ministry policies, whether reflected in Quality Requirements or not, dictate that the service is delivered with sufficient documentation to adequately support customers of the service, operate it, and maintain it in a secure, available and cost-effective state. To this effect, an Application Support Guide (ASG) is supplied, as well as operation guidance and maintenance.

Most importantly, the service’s infrastructure and service is delivered via automation using ministry systems, and documentation is provided in electronic format, again on provided ministry services.

# Qualities View

The solution’s deliverability is measured against ISO-25010 quality objectives of

* Functional Completeness
* Security
* Resource Performance
* Compatibility
* Integration
* Reliability
* Security
* Maintainability  
  Flexibility
* Safety

These objectives are defined in finer detail in the Quality Requirements associated to the project.

# Supportability View

As per the Compliance View, a system specific Application Support Guide (ASG) will be delivered to the satisfaction of ministry’s Support Services at CAB.

# Operability View

As per the Compliance View, the solution’s environment, system, and data provisioning will be delivered by automation using Ministry services.

System specific documentation on their execution and maintenance will be provided in electronic format on Ministry systems.

# Maintainability View

The any custom code of the system – whether it be for the whole system or just integration or deployment or operation strings -- will be provided to the ministry on ministry managed services (either its ADO or GitHub Enterprise account subscription).

Appendices

Appendix A - Document Information

### Versions

* 1. Initial Draft
  2. More
  3. Added Business and System context information.

### Images

[Figure 1: Comparison of normalised and denormalised schemas 13](#_Toc163463773)

[Figure 2: Multiple Identifiers per object. 24](#_Toc163463774)

[Figure 3: Multiple Identifiers 25](#_Toc163463775)

[Figure 4: Integration approach decision flow 32](#_Toc163463776)

### Tables

**No table of figures entries found.**

### References

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

### Review Distribution

The document was distributed for review as below:

|  |  |
| --- | --- |
| Identity | Notes |
| Sandy Britain, Enterprise Architect |  |
| Russell Campbell, Project Manager |  |
| Amy Orr, Data Domain Architect |  |
| Marg Mcleod, Strategic ICT Business Advisor |  |

### 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), is appropriate, diagrams are developed as simple “box & line” monochrome diagrams.

### Acronyms

Refer to the project’s Glossary.

##### API

: Application Programming Interface

##### AoG

: All of Government. A vernacular term for specifying that something affects all government entities, including this one.

##### ASG

: Application Support Guide

##### ETL

: acronym for Extract, Transform, Load – an integration approach that can involve [API](#Acronym_API)s but more often than not involves some form of direct privileged access to a system’s operational datastores.

##### 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.

##### PI

: Personal Information.

##### PII

: Personal Identifying Information. A subset of PI.

##### UI

: User Interface

### Terms

##### Data Hub

: …

##### Entity

: is a term with different meanings depending on context. At a high level, an entity, in information technology, is a thing with distinct properties and independent existence. A [person](#Entity_Person) is an entity. An invoice is an entity. Entities can link to or be linked to by other independent [entities](#Term_Entity). A Group is an example of such.   
In contrast, limbs are not discussed in isolation to the [body](#Entity_Body) they belong to, nor is a line item a thing that can live independently to a parent invoice. These sub-parts of a parent entity are referred to information theory as [value-objects](#Term_ValueObject).   
When the term is used in a business or social domain, it often is taken to refer to an business or government sector organisation.

When the term is used in a system information context the term is generally referring to a model of a single record within a system database table.

##### Longitudinal

: …

##### Value Objects

: see [entity](#Term_Entity) above.

1. This is a factor with NSI, where the original analysis, and subsequent response codes, authentication, threading, and batching was incorrectly implemented on directions from the customer. [↑](#footnote-ref-2)
2. See ENROL [↑](#footnote-ref-3)
3. As opposed to “juridical person”. [↑](#footnote-ref-4)
4. While often discussed as a “person’s name”, it is actually an “*identity’s* name”. [↑](#footnote-ref-5)
5. School bus transits audit records are an example of the transient nature of a location while in responsibility of an education system. [↑](#footnote-ref-6)
6. Computer Science is a well understood and defined *science*, although one wouldn’t know it from the way most ICT projects are run (more as an unbound *art* at adds tremendous risk to the delivery of expectations and value). [↑](#footnote-ref-7)
7. Catalyst, while NZ based, offers a lower capability and is not accredited by MOE. [↑](#footnote-ref-8)