

ED-FI WORKING DRAFT 5 - NEXT GENERATION DATA MODEL CONCEPTS

Ed-Fi Working Draft 5: Next Generation Data Model Concepts
Technical Suite: Suite 3
Status: [Active](#)

By: Ed-Fi Internationalization Work Group
Contact: eric.jansson@ed-fi.org
Publication Date: [December 16, 2020](#)

Synopsis

These materials propose major concepts for the next major generation of the Ed-Fi Unifying Data Model (UDM). The concepts as a whole attempt to align the UDM more around the central concerns and use cases for which the model has been used. Accordingly, increasing specificity and more complex patterns are proposed in some places, while other concepts propose increasing simplicity and flexibility. In addition, the proposed changes explore better ways to align the model with operational aspects and other standards and technology efforts within of the K12 ecosystem.

As a vehicle for communicating major changes, these materials do not account for all nuances around evolving the Suite 3 data model to the one described. Rather, the purpose is to propose larger, more impactful patterns and changes to the Ed-Fi UDM.

Contents

- [General Discussion](#)
- [Use Cases](#)
- [Working Draft Details](#)
 - [Entity Key Structure](#)
 - [Associations: Key Structure, Patterns and Naming](#)
 - [Refactoring of People Entities](#)
 - [More Flexible and Generic Organizational Structure](#)
 - [Revised Associations of People to Organizations](#)
 - [Flexible Grouping Mechanisms For People](#)
 - [Programs and Program Participation](#)
 - [Interaction with other K12 Education-focused Standards](#)
 - [More Specific Technology Bindings for the UDM](#)

General Discussion

The current Ed-Fi Unifying Data Model has proven a robust foundation for data exchange by the current Ed-Fi community. As with all data models, it must evolve continuously to help address new problems its user community wishes to solve and to adapt to changes in the current ecosystem.

This document proposes a set of data model concepts and patterns that would alter key data model domains or patterns for the next generation of the Ed-Fi Unifying Data Model. These changes would in many cases be breaking to systems conforming to downstream standards; they would result in changes that are not backwards-compatible, but rather require code changes and new deployments so as to not disrupt the ecosystem. It is this fact that makes these "next generation" concepts - one or more of these would trigger a generational change in the UDM.

These changes are generally addressing one of the following general issues or areas.

Increasing Domain Depth

Many new use cases and needs can be solved by tactical additions and small changes to the UDM. Over time, however, such changes begin to reveal a need for greater refactoring of one or more domain models so as to more efficiently address core use cases. This happens often in cases where community use cases began with more "superficial" data needs from a domain and over time increase in depth; correspondingly, the data model must evolve to match the complexity necessary for those new use cases. Such an evolution is in fact dictated by adherence to Domain-Driven Design, which is the foundational practice used to craft the Ed-Fi UDM.

Simplification

By contrast to increasing domain depth, in some places the data model may benefit from a move towards greater simplicity. This occurs where the data domain is more deeply articulated than the use cases require. Over-specification of some entities in fact can lead to an excessive reliance on extensibility in the field, which has arguably occurred in areas like the EducationOrganization domain models. Rather than ask all model users to bear complexity, this document proposes places where the UDM may benefit from increased simplification.

Internationalization

The core Ed-Fi Unifying Data Model was based on US K12 education, as that scope is the one most important to the Ed-Fi user community. This focus on US K12 will remain. However, there are organizations, and in particular many technology providers, who operate in other countries, and whose platforms are therefore deployed both the US K12 market and international market.

Accordingly, the concepts proposed here have considered how the data model could better serve users who operate both in and beyond the US K12 market. This stops short of trying to internationalize Ed-Fi beyond US K12 education, but does introduce the principle of considering the needs for how fully-internationalized applications and platforms can interoperate using Ed-Fi's standards.

Entity Naming Conventions

In some cases, current Ed-Fi naming conventions are producing entity or attribute names that are less than optimal for a variety of reasons. In a few places, this document and its data model propose changes.

Use Cases

The use cases for this data model are assumed to be ones identical to those for the current Suite 3 use cases. These are documented most clearly in the REST API standards documentation for Suite 3.

References

- [ED-FI RFC 24 - CORE STUDENT API](#)
- [ED-FI RFC 22 - ASSESSMENT OUTCOMES API](#)
- [ED-FI RFC 25 - SURVEY API](#)

Working Draft Details

The Working Draft contains a draft conceptual model in UML:

- [IWG-DataModelConcepts.vsd](#) (Visio format)
- [IWG-DataModelConcepts.pdf](#) (PDF format)

Specific concepts that are part of this draft model are discussed below.

Entity Key Structure

Ed-Fi UDM for Suite 1 and 2 used a very strong natural key system for entity identity and – as a result – field projects often struggled with key volatility, particularly as data exchanges under those suites became more and more real-time or near-real time. If any field in the entity key was subject to change, this would result in issues with field projects and fixing the issue would lead to breaking changes in API standards and other products derivative of the UDM.

The UDM for Suite 3 addressed the most common places where key volatility occurred, opting in many places to use a "partial key surrogate": bringing the source system (and often) surrogate ID into the Ed-Fi entity key. This change solved most of the issues with key volatility, but some still remained.

The use of composite natural keys by the Ed-Fi UDM unquestionably has provided many benefits, particularly in terms of improving data quality. It has also forestalled the creation of new "Ed-Fi" keys that – once released into the ecosystem – would further fragment and confuse an ecosystem that already in some cases already has too many systems competing for entity identity.

Accordingly, the entity key pattern proposed for the next generation data model is as follows:

- For entities whose IDs are known to be managed as unique by education agencies, the unique ID is used. Generally Ed-Fi labels these as **[entity name]UniqueId** if the system is generally a unique ID system or **[entity name]Id** if the uniqueness is manually managed. e.g. **Student.StudentUniqueId** and **School.SchoolId**
- For entities that have a strong source system identity and where that identity is already being adopted outside that source system, the source system entity identity is used. Generally the naming pattern here is **[entity name]Identifier** - e.g. **Section.SectionIdentifier**

- In cases where there is not clearly a strong source system ID and/or there are clear and natural identifying properties to an entity and where references to that are not likely to propagate, the UDM will use those identifying properties - e.g. **StudentSectionAttendance** whose key is **AttendanceEventType** and **EventDate**

In the attendance domain model, you can see examples of each of these keys.

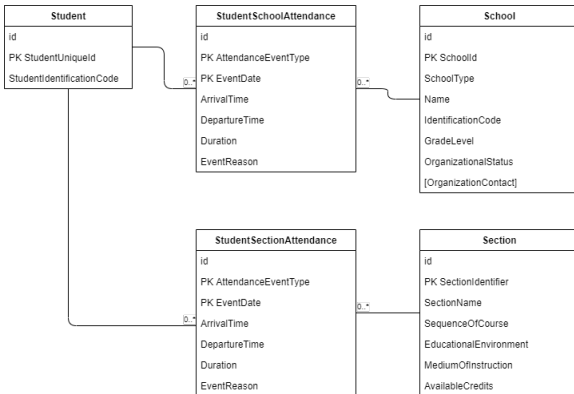


Figure 1: Attendance domain model, showing the different key structure patterns. Click to expand.

Associations: Key Structure, Patterns and Naming

In current Ed-Fi UDM, associations in the data model are named with the pattern **[entity name][entity name]Association** and they key is composed of the natural keys of both entities and sometimes other fields that are part of the association. This led to two problems.

1. First, names were often minimally semantic. As an example, **StudentSchoolAssociation** clarifies that this is an association, but does not clarify that this is a school enrollment record.
2. Second, the overuse of natural attributes for inclusion in the key led to the key volatility discussed above. As an example, use of dates to identify entities often caused issues as dates are easy to mis-enter into systems or forms.

The proposed changes are:

- Move naming to a new pattern: **[entity name][entity name][association modifier]**. In the example below, **StudentSchoolAssociation** becomes **StudentSchoolEnrollment**. This clarifies the semantics, while also allowing for identification of the entity as an association through use of the two associated entity names.
- Form the key of associations as the key of the elements being associated. As this will not always generate a unique key, consider additional properties as follows, using these strategies in the order presented
 - If the association has a strong identifier (one likely to already be in use outside the source system), introduce that into the key.
 - If the association has a natural status (e.g. is associated with a workflow) use that status in the key. For example this strategy could work with associations of students to programs, where the status ("evaluated", "enrolled", "exited" can describe the current relationship)
 - Use the best possible (i.e. least volatile, always available, and most reliable) entity field.

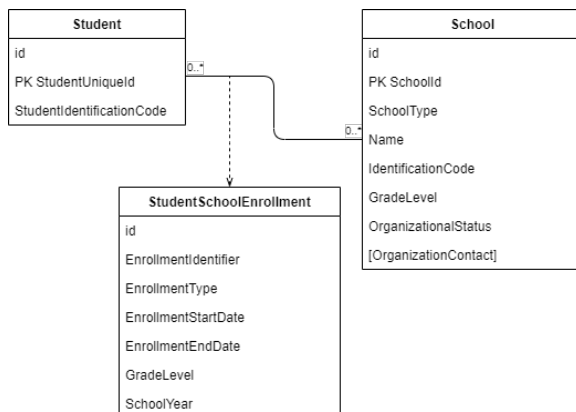


Figure 2: Example of proposed structure for associations. Click to expand.

Refactoring of People Entities

People entities in the UDM have tended to grow in size and complexity over time, as people are central to the data model use cases. To address this, the proposal is to keep the domain models for capturing data on persons largely intact, but refactor for increased clarity and to allow for more options in downstream standards, such as Ed-Fi REST API standards.

While there was some discussion of adding an inheritance model from a core **Person** entity, this approach was rejected. The core issue with this is that the K12 industry and the organizations that operate within it generally do not identify unique individuals, but identify individuals in roles: they are "person-roles." The model will retain its core focus on "person-roles": **Students, Staff and Guardians** (this last is a renaming of the current **Parent** entity to follow real-world domain conventions), and others as needed.

- **Person-roles and Person.** The **Person** entity would remain as it has been modeled in early access versions of the current Suite 3 data model: an optional entity to capture cases where person-roles overlap (e.g. where a **Staff** is also a **Guardian**). Other person-roles do not inherit from **Person**. References elsewhere in the model would also favor references to person-roles over references to **Person**, reflecting that in most K12 use cases, person-roles are either generally mutually exclusive, not highly relevant to the uses cases supported, or not able to be reliably matched with individuals. In a few use cases (e.g. in a future survey domain) use of **Person** references may be used to simplify the model and reflect the generic nature of person-roles (e.g. survey respondents can be any person-role).
- **Segmentation of person attributes.** The main change to this domain model from the current model is to break up the various sets of user attributes into separate domain entities: **Contact, Identification, and Demographics**. This change will enhance clarity and keep entity size smaller and focused on similar concerns. Per the design introduced in Suite 3, these sets of user attributes would be scoped by **EducationOrganization**, reflecting that the values for these entities are often determined via discrete and separate organizational processes and therefore can differ between organizations.

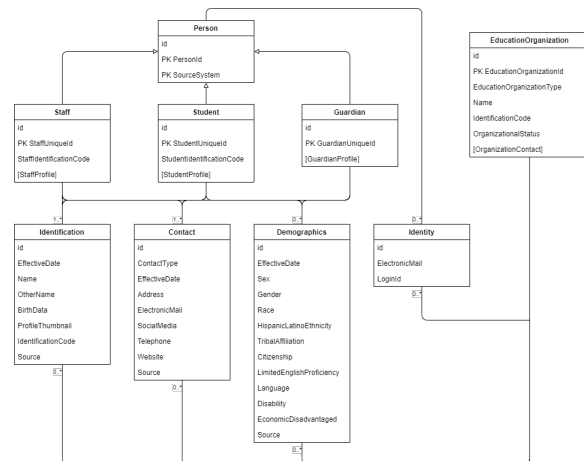


Figure 3: Proposed domain design for modeling person entities. [Click to expand.](#)

More Flexible and Generic Organizational Structure

The current UDM has classes that represent each different class of organization: **School, LocalEducationAgency, StateEducationAgency, CommunityProvider, and PostsecondaryInstitution**. This model has worked, but had its limitations:

1. First, it has introduced unnecessary domain complexity, complexity not helpful for solving primary community use cases. The current model forces the creation of new organization entities as the model expands, and the theoretical quantity of such entities is quite large (e.g. JuvenileServiceCenter, EarlyChildhoodProvider, PrivateTutoringCenter, etc.). While that is OK in theory, the precise attributes of these organizations are rarely relevant to the actual use cases. Given that Ed-Fi focuses on student success, a very detailed domain-driven design for organizational entities has simply not proven critically valuable.
2. Second, downstream implementations have often struggled with issues relating to the subclassing required to make this work. Subclassing has been found in field work to be a powerful modeling technique, but one that introduces significant complexity when applied to areas like REST APIs. In general, subclassing can work, but needs to be reserved to places where there is high value to its use. For the same reason as above – lack of need for use cases – this additional complexity has had low utility.

Instead, the proposed concept is to have 2 entities: **EducationOrganization** and **School**. Instead of subclasses, **EducationOrganizations** simply have a type. **School** is retained as it has stronger and more relevant domain attributes, such as a type of its own, grade levels it is chartered to serve, and so on.

EducationOrganizations can reference other **EducationOrganizations** to describe hierarchies or networks of related organizations.

It is possible that **School** could be the unique subclass of **EducationOrganization** in order to allow the model to retain the current practice of allowing flexibility in references to EducationOrganizations (e.g. a **CourseTranscript** can reference an **EducationOrganization**, which can be one of any possible subclasses of **EducationOrganization**)

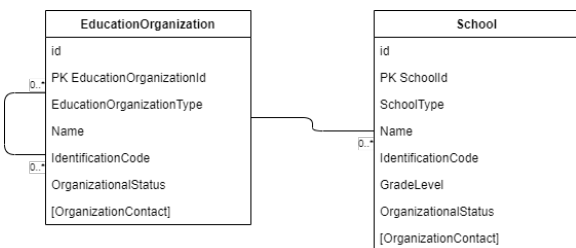


Figure 4: Proposed simplified model for organizations. Click to expand.

Revised Associations of People to Organizations

In a number of places the model proposes changes to the associations of people to organizations. As with some of the items listed above, one important goal of these changes is to increase domain specificity where the additional details provide value, and offer more generic structures in places where it does not. The attempt is to follow the principles of Domain Driven Design, in which the technical domain entities should mirror the business domain entities, yet not over-articulate elements that are not core to the model's purpose.

As noted above, associations now are named with the associated entity names plus a modifier rather than the general term "Association." Accordingly, there are a number of changes to the relationships of people to organizations. These include:

Ed-Fi Suite 1-3	Proposed
StudentSchoolAssociation	StudentSchoolEnrollment
StudentSectionAssociaton	StudentSectionEnrollment
StaffSchoolAssociation	StaffSchoolAssignment
StaffSectionAssociation	StaffSectionAssignment

These changes should enhance the usability of the domain model and make the entity names more clearly semantic.

In other places where the association semantics have proven less important to community use case, it is proposed that the model provide for increased flexibility via the provision of a more generic "typed" associations, as with **StaffEducationOrganizationAffiliation**. This model can be used to define any number of possible staff-organization relationships without needing to add new entities or extend the model. For example, it is possible to denote concepts like "applicant", "visiting faculty" or "in service teacher" using this without having to create full domain entities.

Of course, over time if the concepts of "visiting faculty" or "in service teacher" become core to the use cases which the model serves – and therefore a deeper understanding of these entities is needed – then these can be added as domain entities.

To help drive standardization, **StaffEducationOrganizationAffiliation.AffiliationType** can be a controlled vocabulary.

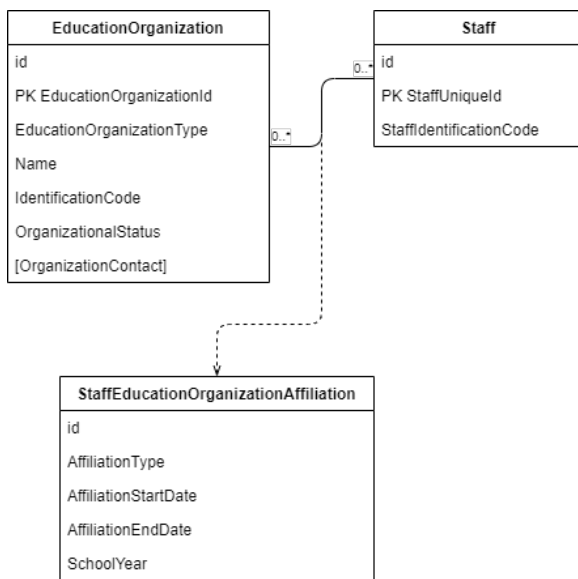


Figure 5: Example of proposed patterns for more flexible associations of people to organizations. Click to expand.

Flexible Grouping Mechanisms For People

Similarly to the above approach to organizational relationships to people, in some places there is a need to designate more informal groups of people. In schools for example, there are clubs, sports teams, honor groups, service organizations, student council, etc. These groups are rarely the focus of community use cases, but occasionally are a consideration or are of secondary interest.

In Ed-Fi Suite 1-3, these were served through the **Cohort** model. The proposal is to make a few changes to the **Cohort** model and adopt a new name: **StudentGroup**. As with **Cohorts**, these groups will retain a type but also introduce a type (a controlled vocabulary) that describes the nature of the students or staff participation. Associations built from these (or similar structures) would also default to using the generic relationship modifier "Affiliation", as in **StudentStudentGroupAffiliation**

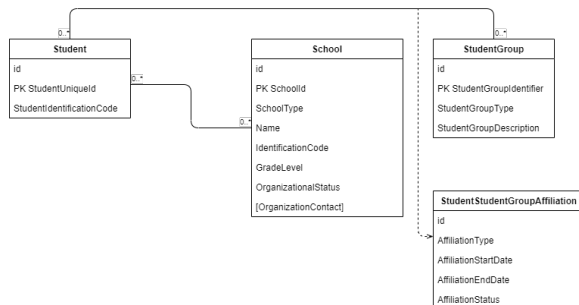


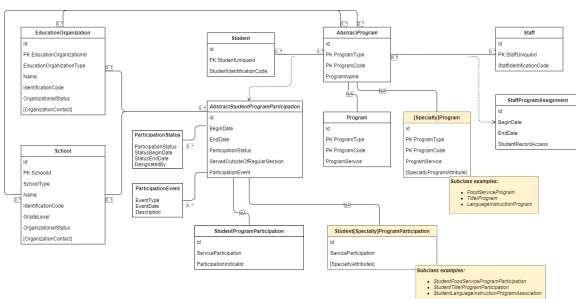
Figure 6: Example of proposed pattern for flexible grouping of individuals below the organization structure. Click to expand.

Programs and Program Participation

The proposed program model would remain largely parallel in concept and structure as today, retaining the use of abstract classes and inheritance. Other models that would remove the use of subclassing and inheritance were considered, but they had various drawbacks, such as relying on convention for domain consistency, or exploding numbers of classes. The use of inheritance as a central strategy of this domain seems justified by the value it provides and by how central this domain is to data usage and analysis of student performance and needs.

The main proposed change is to opt for some decomposition into more entities to reduce the size and complexity of some key entities, particularly **AbstractStudentProgramParticipation** (a renaming of **GeneralStudentProgramParticipation** today). For that entity, elements such as **ParticipationStatus** and **ParticipationEvent** would be split off and used to shrink the size and complexity of larger entities, as well as offer more options for coordination of disparate systems as the domain model is used to create downstream specifications, such as REST API bindings.

In addition, allowing for **Program** to be abstract and subclassed was another key change. In the Ed-Fi use cases, an increasing number of local programs are being captured in the model and these may benefit from the ability to record local program features.



Interaction with other K12 Education-focused Standards

Significant time was spent analyzing overlaps with other major standards in the US K12 space and discussing strategies for interoperating with those standards. Detailed comparisons with other standards known to have operational presence in the K12 space – particularly presence in use cases close to classroom and building-level analysis of student performance – were generated and consulted in the course of the investigations of this proposed data model.

The general question researched was the extent to which the Ed-Fi model could take on structures, definitions and nomenclatures of those other standards, or vice versa - if they could adopt those of the Ed-Fi UDM.

In a number of places the data model changes proposed in this document factored in the structures found in the most prominent and used standards today. For example:

- Many entity name revisions that clarified more common usage of other vernacular – such as "Parent" to "Guardian" or the restructured naming of association entities in the Ed-Fi model – were adopted in this proposed version.
- Likewise, the key structure proposed here represents a norming to patterns more common in those other standards that would enhance interoperability (for example, mapping a composite natural key to single identity string – the most common pattern in other standards – would be difficult to stabilize in field work).

However, this research and analysis also revealed the difficulty of coordinating across specific standards. In retrospect, this was to be expected: different standards serve different use cases and concerns, and to expect that we can get to a single standard from which all others can derive or map proved impractical. The work to map standards that was undertaken as part of the Internationalization Work Group revealed each time that changes would be painful for the current user community of that domain structure, while at the same time the utility of such change was often ambiguous.

What emerged instead from these investigations was a vision that we live in a "multi-lingual" standards environment, with different standards serving different use cases: conceptualized, structured and defined to solve sets of discrete use cases.

This is not to say that standards should not coordinate. Rather the main conclusion was that the focus of such coordination should not be to attempt to collapse standards together wholesale, but to examine the specific use cases which are most impactful to users who rely on those standards and focus coordination on those specific intersections.

It was also noted that there are other practical issues to coordination of standards that were beyond the intent of the workgroup to investigate or propose answers to. These issues related to issues such as ecosystem maturity and demand generation, and the organizational context in which different standards are governed.

More Specific Technology Bindings for the UDM

One concept that was raised was to expand the number of downstream bindings of the Ed-Fi UDM to include new bindings adapted for different analytical methods or other use cases.

An example would be a "data lake" binding. Such binding could be the place to start to aggregate data model information such as key structure, rather than have key structure derive from the UDM itself. To continue with the key structure example, this was attractive as a concept because the current data model key structure has prioritized data quality over easy flow of data. That prioritization has been critical to the Ed-Fi community, but we are starting to see cases where community users would prefer easier flow of data and resolution of data quality issues downstream. Conceptualizing a richer set of bindings could help expand utility while keeping the overall community collaborating on a rich semantic data model.