The Role of the Data Model

Rutgers University has made this article freely available. Please share how this access benefits you. Your story matters. [https://rucore.libraries.rutgers.edu/rutgers-lib/49510/story/]

Citation to Publisher
Version: No citation available.

Citation to this Version: Agnew, Grace. The Role of the Data Model, 2015. Retrieved from doi:10.7282/T31R6SKD.
The Role of the Data Model

I keep six honest serving men
(They taught me all I knew)
Their names are What and Why and When
And How and Where and Who

Rudyard Kipling

Rudyard Kipling started his professional life as a journalist and knew the importance of understanding the crucial elements of any occurrence. Who did it, what did they do, when did they do it, how did they do it, and why?

We begin our work with metadata by asking the same questions about the data to be described by metadata and by looking at a methodology to structure that understanding into a data model, so that we have shared meaning—among the team members who create and manage metadata for an organization and, most importantly, with the data creator(s) and the current and future users of the resources described by the metadata, who will find and reuse the resources—today, tomorrow, and many years into the future.

The data model provides the conceptual underpinning for organizing, describing and managing resources through metadata. The data model demonstrates the nature of the information you are describing and the user needs that it serves. A data model is a way of capturing the essence of information and sharing it with others in an understandable manner.

There are many types of data models, serving multiple purposes, but our focus is on a basic conceptual data model, the ER model. This data model identifies the entities that have a role in the intellectual content or purpose of the resource(s) to be described, as well as the relationships between those entities. A good data model should provide the context for the resource—why was it created, by whom, when, where and for what purpose.
Data models can be expressed through a simple graphic or through standard modeling notations such as the Entity-Relationship (ER) model. The University of Texas, in its excellent overview of ER models, defines the entity as “the principal data object about which information is to be collected. Entities are usually recognizable concepts, either concrete or abstract, such as person, places, things, or events…” (U of Texas at Austin, 2004).

Entities relate to other entities in a data model. Information is all about relationships, because the way people make discoveries or innovate is often through creating relationships between two or more disparate pieces of information to discover something new, or to demonstrate cause and effect. To create new information generally involves creating relationships between different existing information resources. Newton related the apple’s weight to its falling to discover gravity, for example. The physician, John Snow, linked the victims of the 1832 cholera outbreak to a single water pump in London, and from that connection made novel discoveries about how disease is transmitted.

It is often fairly easy, particularly for librarians, to categorize all the characteristics of something. This is the heart of traditional cataloging. What metadata has added to traditional cataloging is the understanding that the relationships between concepts are as important as the concepts themselves.

The data model is critical to understanding information, and to documenting that understanding, because it captures both the concepts of information (“the entities”) and the relationships between them, both of which are described in metadata.

The data model is at the highest level of abstraction for the design of metadata. For this reason, it is often the hardest of the metadata tools to understand. In fact, you may be thinking, I didn’t have to learn to use a data model to catalog resources in MARC, or even in Dublin Core. Why do I have to learn this now?

The answer is that the data model lies at the heart of all metadata design. Understanding the model behind different metadata schemas will help you understand their purpose, what they describe (and what they don’t), so that you can effectively select and apply a metadata schema to describe resources.
More importantly, being able to design a data model will ensure that you document all the entities and relationships that are important to your organization and the resources it collects. Before you develop a repository, with a metadata implementation at its heart, you will want to develop an overarching data model that reflects the resources you will describe.

**Tools of Metadata**

So when will you use a data model? Four common uses for metadata design are:

1. When you are evaluating metadata standard to serve as the underlying schema for your organization or repository. Most metadata standards are designed to an underlying data model.

2. When you are developing an application profile, to tailor a metadata standard to serve the unique needs of your organization. You will use the data model to make sure you stretch and shape your metadata schema—be it Dublin Core,
MODS, or something else, to accommodate the who, what, when, where and why of your data to be described.

3. When you are designing your own metadata schema, because the available standards are simply not adequate to your purpose.

4. When you are describing a new information format or genre, and one for which there is not a lot of guidance, such as research data.

Before we get started, let me supply two basic definitions.

A **metadata schema** is metadata that is structured to support consistency of expression, generally in an XML or an RDF (Resource Description Framework) schema. A **metadata standard** is a metadata schema that is made publicly available for reuse and is maintained according to generally accepted policies and practices, such as periodic review and revision and making those revisions available for public comment. A standards body maintains a standard, such as the Library of Congress or NISO (National Information Standards Organization). It may seem I am using the terms interchangeably, but in fact, I use “standard” for those metadata schemas, such as Dublin Core and MODS that are maintained as international standards and I use “schema” for locally designed and maintained schemas, such as the metadata schemas you will design in this course. Even if a schema is offered publicly for reuse by others, it is not a standard unless it is developed and maintained as an international standard, according to accepted policy and practice.

So, now let’s get started....
Data models are expressed in standard forms of notation, so that they can be shared and understood within a community. We will work with a standard notation, the **Entity-Relationship** or **ER notation**.

Brydon notes two critical issues about the data model:

1. Data modeling is first and foremost *a tool for communication*. There is no single “right” model. Instead, a valuable model highlights tricky issues, allows users, designers, and implementers to discuss the issues using the same vocabulary, and leads to better design decisions.

2. The modeling process is inherently iterative: you create a model, check its assumptions with users, make the necessary changes, and repeat the cycle until you are sure you understand the critical issues. (Brydon, 2001)

In Entity-Relationship Modeling, entities are related to each other through a line and notation indicating **cardinality**, which is simply, how many instances of an entity are related to each other and **obligation**, notation that indicates whether the relationship is **required** (e.g., at least one instance of an entity must be related to at least one instance of another entity).

An entity may relate to another entity in a **one-to-one** relationship, for example, a data model may document that each resource will have one, and only one, instance of metadata. It is important to understand that this is not an **absolute** relationship, but rather a relationship relative to the data model.

The data model for your library’s catalog might document that each resource will have only one metadata instance, because obviously it is not in the user’s best interests to have multiple catalog records for the same resource. In reality, there may be many records available for a single resource in the broader information space. Every institution owning that resource has a record and there may be multiple records to select from in OCLC. However, the data model for the library’s catalog would document that each resource in the library’s collection has one and only one catalog record, which serves two purposes—serving as the ideal toward which the library catalog aims and documenting how the catalog record and the resource interrelate,
which forms the basis for all the workflow and decision making that the catalogers who maintain the library catalog are engaged in.

The data model is thus a very high level, abstract representation that nonetheless has tremendous impact on policy making and the daily decision making that are required to construct a library catalog.

A relationship may be **one to many** (1:N) For entity A, there may be one or many instances of entity B, but for entity B, there is only one instance of entity A. The author *L. Frank Baum* (entity A) has written many books, but each book (entity B), (for example, *The Wizard of Oz*) has only one author, L. Frank Baum.

A **many-to-many** relationship states that for each instance of entity A, there are one or many instances of entity B and for each instance of entity B, one or many instances of entity A. If the notation indicates that an occurrence of an entity is optional, there may be zero, one or many instances of that entity. For example, a resource may have zero, one or many permissions associated with its use (e.g., display, copy, print, etc.). Each permission may apply to zero, one or many resources.

The way these relationships are documented is known as **notation**. A very common notation for data modeling is ER, or Entity Relationship notation. There are some variations in ER notation, but the principles of:

- documenting **entities**,
- describing **relationships**,
- identifying **cardinality** (number of entity instances allowed)
- identifying **obligation** (whether the entity is required and this whether the relationship must occur)
- identifying **attributes** or properties of entities

are common to all ER notations.
A standard notation form is for ER modeling, which we will use to construct our data model assignment, is provided below:

Let’s consider a basic ER diagram from the University of Texas at Austin tutorial.

(U of Texas at Austin, 2004).
In this notation, the two entities in the model are each documented by a rectangle. The relationship is documented by a linking line and text that describes the relationship.

A department, which is one and only one, manages many projects, all of which are optional. The attributes of each entity are documented within the rectangle.

Many ER diagrams eliminate the “obligation” notation, but I like to include it because obligation is a very critical concept.

So, let’s look more closely at each individual notation of **ER Notation**. There are variations on this notation. What is important is to select a form of notation and be consistent in its use, so that the meaning of the data model can be understood and shared.

**Diagram 1.**

**Data Model ER Notation**

- **Entity**
- **Relationship**
- **One to One**
- **Each must have one**

Number of entity instances = **degree**

Whether an instance is required = **obligation**
Each entity is documented in a rectangle. The line between entities indicates that each entity has a relationship to the other. The relationship is described in text below the line.

In the second set of boxes, each entity has one mandatory instance, so the entities are related to each other in a required one to one (1:1) relationship.

**Diagram 2**

Data Model ER Notation

A vertical line on the relationship bar indicates that the entity is required. A single line indicates that the entity will have one required instance. Some notations use double vertical lines --||-- to emphasize “one and only one.” You can choose in your data model to use a single line or the double line, when one and only one entity relates to another entity, for clarity or emphasis.
In this diagram, we see that at least one instance of each entity is required. There is only one instance of entity 1, but there is at least one, but potentially may be more, instances of the second entity. For example, a depositor must have deposited at least one resource in the repository, but may have deposited many resources. Each resource (second, or right hand entity) has only one depositor.

As you look at the data model in Diagram 3 and apply it to my example, you may be thinking, but what about co-authored articles, won’t they have many depositors? Remember that the reality of the data model is contextual, not absolute. For the purposes of your repository design, you might decide that each article will have only one depositor, although it might have many authors. As long as each article is only deposited once, with a single depositor, you are not violating your data model. Another repository might want multiple depositors, for each author of an article. This is also perfectly valid, as long as the repository workflow never violates the model.
In this diagram, we see how to designate that an entity is **optional**. For example, a resource may have zero or one depositors but a depositor will always have at least one, but potentially many, resources. This enables a repository to describe resources that many be artifacts (e.g., a rock or mineral) and biofacts (e.g., a fossil or a pressed leaf) that have no depositor. However, a depositor, to be documented as such in the repository metadata, will always be associated with at least one resource that (s)he has deposited.
In our final diagram, we examine the relationship between two optional entities. In this diagram, the first entity has one or zero instances and is related to an entity with zero or many instances. For example, a rights metadata schema may have zero or more copyright holders, who may provide zero or many use permissions for the resource.

So, now you are probably saying, I understand how entities relate to each other in terms of cardinality and obligation but what constitutes an entity? Does an entity always correspond to a metadata data element, such as author and title, or will it document something at a higher and less abstract level, such as the repository itself and the resources it contains?

The answer is, both. And sometimes neither. The data model is contextual and will document whatever is necessary to support your information task. If you are designing a repository or portal, you will probably develop a data model at a high level with an entity for the repository itself, because you want to be sure that you design the repository to support all the necessary relationships, such as the
relationship of the repository to depositors, as well as ensuring that you adequately describe the resources within the repository and that the way resources relate to the repository are supported by the metadata. So if you are designing a metadata implementation in tandem with the repository that maintains the metadata, you would start with the major entities for the repository—the repository itself, the contributors to the repository, and the resources within the repository. You would document attributes for each entity, which would be used to ensure that the administrative metadata needed to manage relationships between entities is captured, as well as the descriptive metadata to describe authors and resources. A data model at this level also helps establish policies and procedures to manage the repository.

Resource level metadata might need its own data model, if you are designing a schema or an extensive application profile, particularly if you are bringing a complex or important data collection to an existing repository. A resource level data model also helps you evaluate existing resource description schemas, to find one that best suits your data.

Often, a higher level data model is all that is needed, and the attributes within each entity are adequate to define the metadata that describes each entity.

Let’s work through a couple of examples, for large data projects at the Rutgers University Libraries that I designed, which will hopefully make data modeling clearer.
EXAMPLE ONE: WAAND (Women Artists Archive National Database) http://waand.rutgers.edu

Let’s re-construct the data model for a digital library project at the Rutgers University Libraries.

The Women Artists Archives National Database (WAAND, http://waand.rutgers.edu) documents the papers and other archival materials of women artists active in the United States since 1945. Each archive participating in WAAND will have at least one, but may have more, collections to describe in the archive. Each collection will have at least one but may have more artist’s papers documented in the collection. WAAND provides a registration form to archives, who are able, once approved, to enter their own metadata about their collections, through an online form.

So let’s build an entity relationship diagram documenting the entities of WAAND (the directory), archive, collection and artist’s papers.
First we have the entity, "WAAND"

A repository will not always be an entity in a model, but in this case, since I was designing the repository, I made it an entity.

WAAND has a relationship with the many archives that contribute to it.
WAAND Data Model Example

WAAND has a relationship with the many archives that contribute to it.

There is one and only one WAAND Database. Obviously WAAND would not exist as a directory database without at least one archive, but potentially many more, contributing to it.

There is one and only one directory WAAND but many archives that contribute to WAAND. It is mandatory that at least one archive contributes to WAAND, otherwise it is not a functioning directory.
Each Archive has a relationship to the collection(s) it is contributing to WAAND.

**WAAND Data Model Example**

One or more archives may contain one or more collections. Collections may be owned by more than one archive.

There must be at least one collection for each Archive contributing to WAAND, but there may be more than one collection.
Through our discussions with archivists managing collections of women artists, we discovered that the papers of a prominent artist, such as Georgia O’Keeffe, are often located in several collections. Thus, each collection includes one or more artist’s papers, and each artist’s papers are included in one or more collections. We also learned that a collection may be titled, “Latina Artists Collective” or “Feminist Artists’ Collective” and include the papers of many artists. Therefore, the relationship between collections and individual artists’ papers is many to many.

Each collection must contain the papers of at least one Artist. However, a collection, such as the “Latina Artists’ collective” collection may include many artists, too many to separate into individual collections by artist. Also, a prominent artist, such as Grandma Moses, may be included in multiple collections. So the relation is one to many / one to many, meaning that at least one collection must be related to at least one artist, and vice versa.
Next we document the relationship between the artist’s papers and WAAND. WAAND is organized around the artist and her papers, so this is the dominant entity relationship. Through our research, we discovered that this is how researchers will primarily search the directory. The name of the collection where the artist papers reside is documented and is critical for actually using the artists’ papers, but is not the organizing principle for WAAND.
Now we document the relationships that are still outstanding. All entities in a data model will generally relate to each other entity. This isn’t required, but it is generally the case for a well-constructed data model. If an entity isn’t related to all other entities, it may be that it doesn’t rise to the level of an entity, but is actually an attribute of an entity, as explained below.

Each archive makes accessible the Artist’s papers. The whole point of providing information about the Archive is to enable researchers to actually obtain access to the papers of the artists, which are otherwise very difficult to discover.
The Final entity relationship is between WAAND and the collection. The collection is the organizing principle at each archive for managing the artist’s papers. But as we indicated earlier, the collection and the artist’s papers are not one and the same. Truthfully, when we began designing the data model, we thought that they would be. But we discovered in testing the data model with participating archives, that there were many collections, such as the Latina artists’ collection, that included papers for many artists, and those artists’ papers could not be located except through the collections in which they were stored.

WAAND contains records for individual artists’ papers, but it identifies the collections which own them, which is critical for locating the individual artists’ papers in the archives that own them.

Each data model will generally have a primary entity, for which one instance of the entity exists in relation to the other entities. When we are designing metadata, this is generally the entity that is the focus of
the metadata record. In this case, the primary entity was the WAAND database because it is uniting information about three equally critical entities, the archive, which physically owns the papers that the researcher needs, the collection, which organizes the papers and must be identified in order to find the papers and the artists’ papers, which are the ultimate goal of the researcher.

In most cases, the primary entity of the data model will be entity described in metadata. In WAAND, no one entity described predominates, so WAAND itself, and the way it integrates the information from the other three entities to enable a researcher to discover, locate and obtain the information she needs, is the primary entity. The concept of primary entity will become clearer over time as you work with data models.

**The Final Piece: Attributes or properties**

Many data models also document the attributes or properties of the entities to be documented in metadata. Documenting these attributes ensures metadata that documents the core or critical information that each entity, and the relationships between entities, require.

So what might the primary attributes be? The attributes will uniquely identify each entity and may also enable relationships between entities. For example, the contact information for an archive enables the archive to contribute to WAAND. The name of the archive enables the archive to be discovered in WAAND. Information about contacting or visiting the archive enables accessibility of the artists’ papers, which is the core business model behind WAAND. Some core attributes are thus:

- Name of the archive
- Contact information for the archive
- URL for the archive
- Restrictions or permissions to use the archive

For collections, core attributes might be:

- Name of the collection
- URL for the collection
- Cataloging status of collection
- Finding aid for collection
- Access restrictions on collection
Since the collection is the organizing principle behind the artist’s papers, it is likely that the collection itself is cataloged, even if individual artist’s papers are not. Information about the collection thus provides a way to discover the artist’s papers/

Some attributes might document the relationship between entities:

Name(s) of artists contained in the collection

In the case of the artist papers, attributes support discovery and use of the papers:

- Name of the artist
- Types of materials contained in papers
- Extent of the collection (e.g., number of items, footage)
- Time period the papers cover
- Geographic areas represented in the papers (e.g., the Santa Fe years, for Georgia O’Keeffe)

Obviously, there are many attributes, and data models that include attributes usually select the most salient attributes, which reflect the context and purpose of the data. There are many things that can be documented about archives, but the attributes documented for the WAAND directory are specific to the use of collections and artists’ papers documented in the WAAND directory.

A final note about this Getty-Foundation funded project. Before we finalized the data model, and began designing the directory metadata schema, we hosted a meeting of women artists and archivists of women artists’ papers, to understand the landscape and their needs. We designed a draft data model and metadata schema, and created a draft directory, to test and refine the model, before we finalized the data model, metadata schema, and directory application, which is now almost ten years old and continues to prove very usable and useful.
EXAMPLE TWO – Video Mosaic Collaborative
http://videomosaic.org

The Video Mosaic Collaborative (videomosaic.org) is an NSF-funded collection of digital videos from multiple research projects.

In these interventions, education faculty, graduate students, and teachers ("mediators"), use tools such as manipulatives (Unifix cubes, Cuisenaire rods, etc.) and other objects (graph paper, magic markers, etc.) to elicit mathematics reasoning and learning in students. Techniques such as active listening and guided prompts are also used to encourage students to make their own discoveries about how abstract concepts such as fractions and probability, work in real world problems, such as placing toppings on a pizza or determining the shortest route for a taxi cab to travel from one location to another.

The heart of the collection is a unique longitudinal study that follows the same cohort of students from 2nd grade through the conclusion of high school.

In multiple discussions with the researchers about the significant contributions that the collection can make to education researchers and to practicing teachers, we agreed that the major contributions were twofold: identifying the student reasoning elicited in each
intervention, so that effective teaching strategies can be replicated in current classrooms, and following students across settings to study the individual development of each student in the longitudinal study.

So, armed with this understanding, let’s get started with the data model design. We identified four major entities—the student(s) participating in an intervention, the setting of the intervention (critical for both considering how setting might have an impact on the intervention and for tracking individual students across settings), the intervention itself, and the form(s) of reasoning, strategies and heuristics displayed by students in the intervention.

In the interest of brevity, each entity is presented with the major attributes already defined. In reality, these took months of design and negotiation with the grant investigators in the Graduate School of Education. There are four major entities, but the organizing principle is the intervention, where the student(s) and the setting intersect and where reasoning, and thus learning, occurs.
Before we get started exploring this data model, let’s stop and compare to WAAND. Why wasn’t the database containing this information an entity, as was the case with WAAND? Remember, that I said there are no right or wrong data models, just what makes contextual sense for your data and can thus be validated from exemplars or test cases from the data.

For WAAND, we were designing a directory structure where none existed intended to enable researchers to find the papers of women artists in existing physical archives. We discovered these papers might be their own collections, for prominent artists such as Faith Ringold, or they might be hidden in a larger collection, such as women Latina artists. The directory itself needed to be designed to enable researchers to not only find that these artist’s papers existed, but to locate where they existed in order to obtain access. The directory was the primary entity for enabling the three goals of identify, locate and access. In the case of the Video Mosaic, we were working with a very prominent research collection, which we would be placing in our existing repository, RUcore (Rutgers University community repository), which has a very sophisticated and extensible MODS implementation. So we were designing a metadata application profile intended to provide full support for a unique and important research collection. If you take the follow on course, Metadata Implementation, you will be designing an application profile for a research imaging collection, much as I designed the Video Mosaic Collaborative application profile.

**Let’s get started...**

So, let’s look at the first two entities, **student** and **setting**. Some interventions involve only one student, but all interventions involve at least one student, so intervention will include at least one student in at least one setting. The collection was developed over 20 years through many NSF grants. Many of the students participated in the longitudinal study, from 2\textsuperscript{nd} grade through 12\textsuperscript{th} grade. Those students will be located in more than one setting. Other students will be located in only one setting. Every intervention will have at least one student in one setting.

So the relationship is **one to many** students in **one to many** settings.
Now, let’s add the **Intervention** entity into the model. Each metadata record actually documents an intervention, where one or more students participate in a learning event, located in a single setting.

An intervention requires at least one student, or the intervention cannot occur. Each student also must participate in at least one intervention to be documented in metadata. If a student was enrolled in a class but missed class the day of the intervention, or sat in the back and did not participate in the intervention, obviously there would be no point in documenting that student in metadata.

Since the intervention is the **primary entity**, meaning it is the thing documented in metadata record, the primary entity occurs once, with a single setting and one or more students, in each instance of the metadata data model. This is where a metadata data model may differ from other data models. The intent is to design a metadata record that represents an instance of the data model, so just as a catalog record describes one book, a metadata model is an instance of one entity but contains the other entities. Does this make sense?
So now, let’s bring in the final entity, **Form of reasoning**, which is the transformative entity, i.e., the innovative aspect of the research, or the proof of the research hypothesis. The crux of the research is that, as a result of the intervention, the participating student(s) will exhibit one or more **Form of reasoning**.

*An important question to ask right now, is why wouldn’t Form of reasoning be an attribute of intervention?*

Obviously, the students exhibit forms of reasoning through the intervention that takes place, where skilled mediators pose a problem and use tools to elicit forms of reasoning in a mathematics topical area, or strand, such as counting or fractions.

The answer, again, is contextual. These videos are research data, meaning they were conducted to test a hypothesis, which is that forms of reasoning could be elicited in students from a well-designed intervention. In working with the researchers, I asked them first, who their audience was, and they identified two key audiences: other education researchers who would reuse the videos and teachers, who could use the interventions to learn to teach more effectively. I then asked what **impact** they wanted the videos to have. The response was they wanted both audiences to see that students did display specific forms of reasoning, and they had gone to great trouble identifying the forms of reasoning displayed. Forms of reasoning is thus the “gold” and deserves to be treated as an entity. Does this make sense.

The final entity, **Forms of Reasoning**, has only one attribute, “Name,” but since a vocabulary for this entity could not be discovered, this data element took months to develop. We will discuss vocabulary development in the final lesson of this course.

For the final entity, one or more forms of reasoning, such as “guessing and checking” or “recursive reasoning” will occur as the result of an intervention and obviously students might display a form of reasoning such recursive reasoning in more than one intervention. In an
intervention, one or more students displays one or more forms of reasoning, and each intervention produces one or more forms of reasoning. Each form of reasoning will occur in one or more interventions.

When we put all the entities and relationships together, we have a metadata model that takes the intervention itself as the primary entity, that emphasizes that one or more forms of reasoning are elicited from an intervention, but that also supports tracing a child’s development in this unique longitudinal study across the different settings of grade and school, as well as within interventions, which occur where students, location and the intervention itself intersect to elicit or produce forms of reasoning.

This is a really important collection which provides unique opportunities to test the use of manipulatives and mediator strategies in different settings such as grade level and school classroom or after school informal learning to elicit the exhibition of reasoning skills. Each intervention stands on its own, but because the same students were followed from first grade to 12th grade, the collection can also be used to follow and study the development of individual students. You would not understand the richness and complexity of the collection by browsing the videos. Nor would you understand the wealth it provides if we simply documented where each intervention occurs, the first names (for privacy reasons) of the students, grade level, etc. No, you only see the power of the collection and really make the best use of it—whether you are an experienced faculty researcher or a teacher looking for guidance—when all the entities of the data model come together in a description. Can you see this? It is important to understand, because this is why we create data models.
This collection is used all over the world, and much of its impact is because people searching the collection understand it—what happened and why. Why is critical for research data. Researchers are only successful if they can answer why they did something. Great art often has meaning if we understand the why of it. It is often important, and never more than for research, to document the why.
An interesting discussion that occurred over the months we were developing the data model was whether the "mediator"—the researcher(s) who led the interventions, was an entity or an attribute. In examining the research documents and talking to the Principal Investigator, who participated in every single project that produced video for the collection, it was clear that the interventions were designed separately from their occurrence, according to the hypotheses and parameters of each research project. While there can be some minor variation in the way an intervention occurs, based on the mediator, this proved to actually be rare. The researchers were well trained and consistent in presenting the intervention in the classroom, so the identity of the mediator(s) had little or no impact on the forms of reasoning displayed, and the same intervention design, conducted in different settings, transpired in the same manner. Variations in forms of reasoning were due to the student and to the setting (a 4\textsuperscript{th} grader exhibits different reasoning skills than an 8\textsuperscript{th} grader), but not to the well trained mediators. So mediator became an attribute, because you might want to find the mediator in an intervention, but not an entity. These are the sorts of interesting discussions that occur when you develop a data model for a unique and complex research data collection!

Another issue for discussion was whether or not a Form of reasoning was exhibited in each intervention. None of the researchers was aware of any intervention captured that didn’t display at least one form of reasoning, and this occurrence was viewed as highly improbable, unless the participating student(s) simply sat silent, which no one believed had ever occurred, so the model documents one or many instances of forms of reasoning occurring in each intervention, through the participation of one or more students. As you can see, the guiding principle for data model design for a research project is to represent the information view of the researchers. The metadata designer asks lots of questions, but the final arbiter of each model entity is the researcher.

So now, you may be feeling a bit overwhelmed, or you may be raring to get started on your own data model. Either way, let’s jump in!
Assignment 1A (the first of two assignments for lesson one) is to design a small data model. Please proceed to Assignment 1A.

Sources


Brydon, Michael. An Introduction to Data Modeling. Available at course website.