

NIH External Researcher Conceptual Data Model v.1.0

Status of This Memo

This document specifies a standard for the National Institutes of Health (NIH) and requests discussion and suggestions for improvements. Distribution of this memo is unlimited.

Table of Contents

1	Executive Summary	3
2	Introduction	5
2.1	Purpose of the External Researcher CDM	5
2.2	Intended Audience	6
2.3	Scope of Standard	7
2.4	Limitations of External Researcher CDM	8
2.5	Assumptions.....	9
3	External Researcher CDM Overview	10
3.1	External Researcher Data Architecture Overview	10
3.2	External Researcher CDM Inputs	11
3.3	External Researcher CDM Content and Structure	11
3.4	Key External Researcher CDM Entities and Structures	13
4	External Researcher CDM.....	14
4.1	ORGANIZATION, PERSON AND ROLE.....	15
4.1.1	Introduction.....	15
4.1.2	Organization, Person and Role—Data Entities and Attributes.....	15
4.1.3	Organization, Person and Role—ORM Model.....	18
4.1.4	Organization, Person and Role—Relationships and Rules.....	19
4.2	ADDRESS AND CONTACT METHOD.....	21
4.2.1	Introduction.....	21
4.2.2	Address and Contact Method—Data Entities and Attributes	21
4.2.3	Address and Contact Method—ORM Model	23
4.2.4	Address and Contact Method—Relationships and Rules	24
4.3	PERSON NAME	27
4.3.1	Introduction.....	27
4.3.2	Person Name—Data Entities and Attributes.....	27
4.3.3	Person Name—ORM Model.....	29
4.3.4	Person Name—Relationships and Rules	30
5	References	33
6	Contact.....	34
7	Security Considerations.....	34
8	Changes	35
9	Authors' Address.....	35

Appendix A: Comprehensive External Researcher CDM—ORM Notation A-1
Appendix B: Comprehensive External Researcher CDM – ERD Notation B-1
Appendix C: Data Modeling Tutorial C-1
Appendix D: Glossary of Entities and Attributes used in External Researcher CDM..... D-1
Appendix E: EDU Person Object and Fact Types not used in External Researcher CDME-1

1 Executive Summary

The External Researcher Conceptual Data Model (CDM) provides a consistent and flexible data structure for the National Institutes of Health to store information about researchers external to NIH. An External Researcher is any person not working with NIH as an employee or contractor that takes part in research processes. NIH needs to provide permissions and allow access to processes for an External Researcher when doing business with NIH. A consistent approach to managing External Researcher data will facilitate tracking every NIH interaction over a person's career.

The External Researcher CDM considers the NIH's business goals and sets forth concepts that will unify, codify, and formalize the External Researcher data currently managed independently by NIH's various Institutes and Centers (ICs). This document identifies the major External Researcher data objects and critical business rules. The result of implementing these objects and rules will be better information systems that are flexible in meeting business needs.

Once the business processes to enroll an External Researcher in the electronic Research Administration (eRA) application and several extension systems were documented and understood, this data model was created to standardize data that can uniformly represent external researcher contact information and personnel role assignments. The model presented in this document incorporates several applicable Person standards. The EDU Person standard was selected as the foundation for this CDM because a significant number of the External Researchers involved with NIH are associated with universities and related institutions. The model was then further refined via numerous discussion sessions with Subject Matter Experts (SMEs).

Several concepts are expanded upon that allow the External Researcher CDM to address ongoing and emerging business needs:

- Assignment of a **Role** to a **Person** is made more dynamic in the following ways:
 - In an Organization, an individual **Person** can play multiple **Roles**, such as Principal Investigator and/or Signing Official. Depending upon the Roles assigned by an Organization, a Person can be granted specific permissions and responsibilities for accessing NIH computing resources.
 - **Role** assignments have both initiation and termination **Dates** specified, thus granting and revoking access at appropriate times without further intervention, as well as allowing retention of access rights for historical records.
- Augmentations made to the NIH Person Name NIHRFC:
 - Any External Researcher that travels for NIH must have an "**Official Government ID Name**". Most often there will be a "**Professional Name**" associated with a publication as a result of work performed with or for NIH. There is also a need for the "**Legal Name**" to be used when there are direct financial dealings with an External Researcher.

- Initiation and termination **Dates** are specified for a particular **Person Name**. When a name change occurs, this tracking allows the appropriate name to remain attached to historical records.
- Contact information for the External Researcher was established
 - For personal communication, treatment of **Address** is improved with the newly introduced concept of **Contact Method**. A **Contact Method** allows the storage of phone numbers, email addresses, and information for any other communication medium in a way that also distinguishes which method is most appropriate to contact someone based upon which organization they are representing and where they are located.

Adherence to the rules identified within this model will result in clearer representation of External Researcher information and role assignments as they pertain to NIH processes. For the technology staffs who implement information systems, the External Researcher CDM will provide the foundation to incorporate the key structural elements into the design of new or enhancement of the current information systems; and will provide a mechanism to map the detailed implementation activities to the External Researcher CDM. By providing the common language, definitions, rules, and relationships, the External Researcher CDM also serves as basis for naming conventions in information systems. As these current and future information systems are being developed, the lessons learned from their implementations will be incorporated into the External Researcher CDM to further refine the business rules and context. It is important to recognize that far-reaching change requires a comprehensive approach to educating stakeholders, understanding their needs and involving their participation to successfully adopt these concepts within the NIH.

2 Introduction¹

The External Researcher Conceptual Data Model (CDM) provides a description of the key data entities and relationships that support the National Institutes of Health (NIH's) External Researcher business processes.

The Federal Enterprise Architecture Data Reference Model (DRM)¹ defines a conceptual data model as follows:

A data model represents an abstract view of the real world; (ISO 11179-3) a higher-level data artifact that is often used to explore domain concepts with project stakeholders. Logical data models are often derived from conceptual data models. At this level, the data modeler attempts to identify the highest-level relationships among the different entities.

This model is intended to provide a basis for the way NIH Information Technology (IT) solutions will structure External Researcher data objects and observe related business rules. The business rules can then be used by information system designers and developers to ensure that data are represented consistently across NIH information systems, data can be effectively shared, and information systems meet business needs. The External Researcher CDM provides an overarching framework to organize detailed External Researcher data architecture efforts and provides a common taxonomy for describing data assets across the NIH.

An External Researcher is any person not working with NIH as an employee or contractor that takes part in research processes. NIH needs to provide permissions and allow access to processes for External Researchers when doing business with NIH. A consistent approach to managing External Researcher data will facilitate tracking every NIH interaction over a person's career.

2.1 Purpose of the External Researcher CDM

The purpose of the External Researcher CDM is to provide a conceptual view of the key data entities and relationships that support NIH's mission and scientific research.

The External Researcher CDM expands upon the Enterprise CDM² and complements the Person Name NIHRFC³ while focusing more specifically on information directly related to persons external to the NIH. NIH needs a CDM specific to External Researcher for a number of reasons:

- To provide a consistent basis for the development of the next generation of information systems that will utilize the External Researcher objects.
- To provide a common vocabulary for the discussion of key enterprise data elements used for External Researcher to allow for more consistent representation of requirements and design of information systems.

¹ All superscript numbers noted in the document are cited in Section 5 - References

- To effectively and consistently manage information about External Researcher in the future, and to provide an organizing framework for further development of External Researcher objects.
- To set the stage for federated authentication.
- To access future non-local sources of External Researcher data.

This high-level representation allows NIH management and stakeholders to effectively understand the plan for a future-state data architecture that will enhance NIH's ability to manage information about external people, and build more integrated, flexible information systems. As with all the data standards developed to date, this version of the External Researcher CDM represents an initial iteration and will be progressively refined through future updates as business processes and information systems evolve and as understanding of data requirements are further refined.

The External Researcher CDM is intended to provide the basis to build superior information systems that are more flexible to adapt to changing business needs including reporting. Current challenges in the interpretation of the definition of the term "External Researcher" and the flexible business rules have made it difficult for the current information systems to present the requisite information in a consistent and reliable manner. The External Researcher CDM tries to address these challenges by providing clear and concise business rules that can be applied in designing and developing future information systems.

The External Researcher CDM also supports the communication and outreach program among the various stakeholders within NIH by presenting a consistent and common vocabulary. The External Researcher CDM is a key artifact that can be used as a tool to map how the data are being used within an organization and understand the specific authoritative data source.

For the technology staffs who implement information systems, the External Researcher CDM will provide the foundation to incorporate the key structural elements into the design of new or enhancement of the current information systems; and will provide a mechanism to map the detailed implementation activities to the External Researcher CDM.

2.2 *Intended Audience*

This standard is available to the entire NIH community, but it is most relevant to the following NIH stakeholders:

- *Business Owners of Data*—The data entities described in this standard should be consistent with commonly used NIH business language, and the definitions of the entities should be understandable to business users.
- *Data Architects*—Those responsible for providing External Researcher-specific data architecture leadership at the NIH enterprise level and the Institutes and Centers (ICs) levels should use the External Researcher CDM as a reference and map their work products and data models to the conceptual level data entities identified within this CDM.
- *IT Leaders and Planners*—CIOs within NIH Institutes and Centers and other senior IT leadership should use the External Researcher CDM as the common taxonomy for identifying NIH External Researcher data assets in their strategy and IT planning documents.

- *Solution Architects*—Architects responsible for the overall design of a new solution or enhancement of existing information systems related to External Researchers will be a key audience of the External Researcher CDM. Information systems developed at NIH will often instantiate the high-level conceptual entities and relationships identified in the External Researcher CDM in their logical and physical database designs. Solution architects will need to align their data entities to the entities in the External Researcher CDM.
- *Information System Designers*—The External Researcher CDM will feed current state and future state information system designs by specifying the NIH standard objects and relationships for External Researchers.
- *IT Program and Project Managers*—The project managers of solution implementation efforts should be aware of the External Researcher CDM and other Enterprise Data Architecture artifacts and the alignment of their solutions with these artifacts. The Enterprise Performance Lifecycle (EPLC)⁴ and the NIH Capital Planning and Investment Control (CPIC)⁵ process will require mapping of External Researcher-specific solution development efforts to the External Researcher CDM.

2.3 Scope of Standard

The approved initial iteration of the External Researcher CDM is considered to be an NIH standard for all projects subject to alignment with the NIH Enterprise Architecture (EA) that maintain data about researcher objects. This document is intended to provide a high level and conceptual view of key data entities required by NIH's External Researcher data capture effort. In practical terms, this means that data architecture artifacts such as data dictionary, metamodels, etc., related to the External Researcher objects should:

- Provide a mapping of the artifact to the External Researcher CDM
- Align the naming of data elements with those used in the External Researcher CDM

The External Researcher CDM is to be used by all of NIH in the development of information systems. This ensures the implementation models of the future information systems support the shared core business processes. The model can also improve quality of design and save time in the development of data models and database designs in support of specific solutions.

For the purposes of this standard, data architecture is defined as efforts to identify cross-system and/or cross-organization strategies for data management that may include a current state, target state and transition strategy. This may also include documents addressing the governance of NIH data assets. The future External Researcher data architecture efforts within NIH will be coordinated with The Office of the Architect. Additional extensions to the External Researcher data architecture artifacts may be developed by the design teams within the ICs due to their unique business rules.

Further, solution development efforts subject to alignment with the NIH EA should:

- Develop detailed data models that align with the defined entities and reflect business rules in the External Researcher CDM. (Note: The data structures or objects in the specific solutions may differ in order to support information system performance or other design/implementation constraints)
- Utilize the External Researcher CDM for mapping and sharing data with other systems.

2.4 Limitations of External Researcher CDM

The scope of this standard is the high-level conceptual view of the data entities and their key relationships in support of the management of the External Researcher objects and associated attributes. The model addresses the interactions between the External Researcher and various functional areas including research projects, loans, publications, membership in groups, etc. The scope of the External Researcher CDM is narrow in focus in that it addresses the names, roles, and addresses/contact information in these functional areas. However, most of the objects with minimal changes to the business rules can be expanded to accommodate additional functional areas.

An individual business process within a functional area may utilize attributes that are considered mandatory, optional, or excluded. A different business process might regard the same attribute differently. For example, Social Security Number may be required by one business process but not for another. Depicting this level of thorough analysis is outside the scope of this NIHRFC. However, an effort has been made in Table 1 below to capture the relevant objects used for each functional area. This list is intended as a starting point and guide for more detailed analysis.

Table 1 – Relevant Objects involved in Functional Areas

Current System/Use	Functional Area	Relevant Objects	Justification
eRA	Manage Grants	<ul style="list-style-type: none"> • Work Address and Contact Method • Professional Name • Allow access via Roles • Organization 	Access to NIH computing resources and sites.
	NIH Sponsored Travel	<ul style="list-style-type: none"> • Official Government ID Name • Home Address • Home Contact Method 	Reviewer travel supported by NIH
	NIH Site Access	<ul style="list-style-type: none"> • Official Government ID Name 	Required for access
	Reimbursement by Government	<ul style="list-style-type: none"> • Social Security Number • Home Address • Home Contact Method 	Sending funds
iEdison	People independent of organization	<ul style="list-style-type: none"> • Home Address • Home Contact Method • Legal Name • Organization 	Organization relinquishes interest in invention to PI or others
	Payment by Government	<ul style="list-style-type: none"> • Social Security Number • Home Address • Legal Name 	Sending funds, sending 1099
Loan Repayment	Reimbursement Government	<ul style="list-style-type: none"> • Social Security Number • Home Address • Home Contact Method 	Sending 1099 for loan interest
National Library of Medicine	Access Medical Knowledge	<ul style="list-style-type: none"> • Professional Name • Organizational affiliation 	Publication search

Current System/Use	Functional Area	Relevant Objects	Justification
External (Universities of Florida & Texas)	Establish Collaboration Environment	<ul style="list-style-type: none"> • Professional Name (historically) • Organizational affiliation 	Facilitating networking for science professionals

Detailed descriptions and development of External Researcher objects – logical and physical data models – are left for subsequent data architecture efforts or specific implementation efforts. It is recognized that this model represents a subset of the data entities required to fully model an NIH External Researcher, and that most follow-on architecture and implementation efforts will need to create additional data entities that are not addressed in this version of the CDM.

2.5 Assumptions

The following assumptions have been made in the development of the External Researcher CDM:

1. This model contains overlaps between a person playing the ‘External Researcher’ Role and the general Person object, but the primary focus has been on a person playing the “External Researcher” Role and its applicable business rules.
2. Work address information is captured for an External Researcher and not for an Organization.
3. Name, Address and Contact categorization must be available wherever the data is stored (e.g. Address - home, work).

3 External Researcher CDM Overview

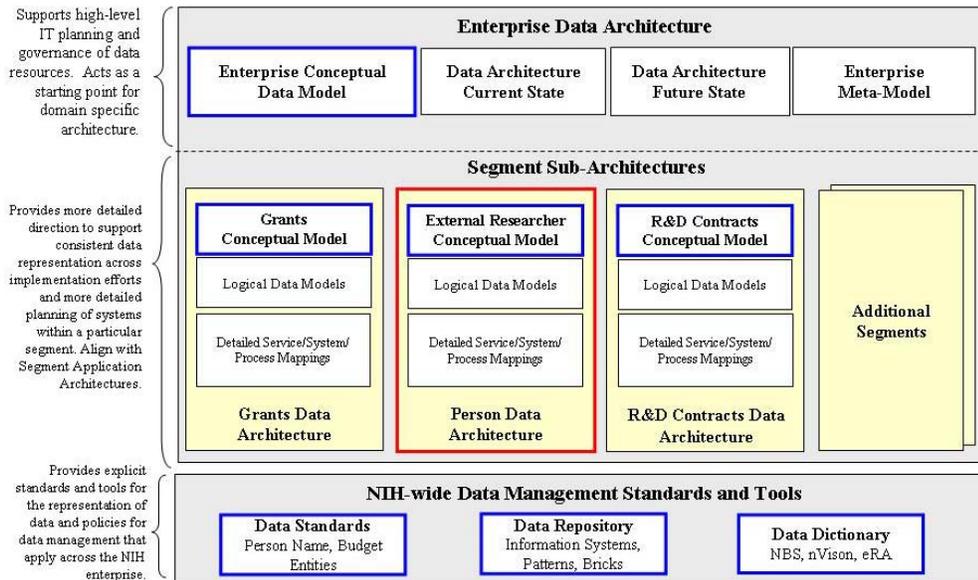
The External Researcher CDM defines the information requirements associated with business processes that involve an external researcher. The External Researcher CDM supports the notion of more effective sharing of External Researcher information across functional areas in order to facilitate decision making with respect to:

- Providing access to and permissions within NIH computer resources
- Allowing seamless integration between systems supporting External Researchers
- Tracking of an External Researcher through multiple types of interactions with NIH over time

3.1 External Researcher Data Architecture Overview

The External Researcher CDM is another component of the segment sub-architecture framework that will be developed in order to better align NIH’s information systems with NIH’s mission – to further science and medicine. Figure 1 shows the decomposition of data architecture and data standard components that may be used to manage NIH’s strategic data assets.

Figure 1 – Data Architecture Framework



Note: Items in blue are artifacts that have been developed by The Office of the Architect.

Enterprise Data Architecture is driven by the business strategy and alignment to NIH’s mission. This level supports the high level IT planning efforts to ensure goals and objectives are met across the organization. The Enterprise Data Architecture identifies those key subject areas and

entities that are shared across NIH including current and future state environment supporting data management.

The Enterprise CDM provides the model of the core data entities and relationships that support NIH.

By contrast, a segment architecture provides the detailed information for the individual elements of the enterprise describing core mission areas and the common/shared business processes and enterprise services, and is intended to deliver faster results. The Segment architecture is also driven by the business and delivers detailed and results-oriented artifacts specific to the core mission area. The Segment Sub-Architectures are more detailed. The CDMs within individual segments allow for detailed planning in a specific business area. Also included in the segment sub-architecture are the logical data models and data and process mappings, which will not be addressed in this standard. The conceptual data model described in this document is part of the External Researcher segment sub-architecture.

Segment architectures are closely related to EA through three tenets: reuse, structure and alignment. Segment architecture **reuses** important assets across all the layers defined at the enterprise level including, data, business, information systems and technology. Segment architectures **inherit the framework** used by the EA, though it may be extended or specialized to meet the needs of the core mission area or common/shared service. Segment architecture **aligns** with all the goals and objectives, drivers and strategies defined at the enterprise level.

3.2 External Researcher CDM Inputs

The External Researcher CDM was developed based on a number of key inputs:

- EDU Person standard⁶—The basis for the creation of this standard is the EDU Person standard. A significant portion of the external researchers involved with NIH are associated with universities.
- External Researcher Process Model (Current State)⁷—The Office of the Architect has developed a draft as-is External Researcher Process Model that documents process steps from the registration of an organization to the assignment of external persons in the eRA application.
- Discussions with key stakeholders—The business rules and context for the entities and relationships were also vetted with business stakeholders to ensure the accuracy of the current and future direction of the External Researcher processes and future concepts.
- Standards Reviewed⁸⁻¹⁹ —The Office of the Architect also reviewed and analyzed external standards efforts, internal NIH standards, and data used in multiple NIH applications that contain Person information. Similar Person data and alternate labels for attributes were identified and analyzed.

3.3 External Researcher CDM Content and Structure

The External Researcher CDM is comprised of: Entities, Relationships and Attributes. These components align with widely accepted nomenclature for the elements of a data model and are consistent with the approach to data description described in the Federal Enterprise Architecture Data Reference Model (DRM). These are defined in Table 2

Table 2 – External Researcher CDM Components

Component	Description
Entity	An abstraction for a person, place, object, event or concept described (or characterized) by common Attributes. For example, “Person” and “Organization” are Entities. An <i>instance</i> of an Entity represents one particular occurrence of the Entity, such as a specific person or a specific agency.
Relationship	Describes the association between two Entities. Relationships may also be described as business rules that specify the nature of the interaction between two Entities.
Attribute	A characteristic of an Entity whose value may be used to help distinguish one instance of an Entity from other instances of the same Entity. For example, an Attribute of an “Organization” Entity may be “DUNS Number.”

The External Researcher CDM will be presented in multiple notations to accommodate the wide variety of stakeholders: Object Role Modeling (ORM)²⁰ notation, sentences or fact types, and Entity-Relationship (ER) diagrams. The sentence structure is included as part of each focus area and is intended to make the CDM easier to readers without a technical background. All these sentences are normative in the order of precedence of the ORM, ERD and XML Schema Definition (XSD) models.

All entities and attributes that are specific to a focus area (the topics into which the CDM has been divided for clarity) have been presented in a tabular form in the document. A readable version of the comprehensive External Researcher conceptual data model (ORM version) has been presented in Appendix A – Comprehensive External Researcher CDM – ORM Notation. The table structure in an ER diagram is presented in Appendix B – Comprehensive External Researcher CDM – ERD Notation. An explanation for reading and evaluating the ORM model and sentences is provided in Appendix C – Data Modeling Tutorial. An entire glossary of all entities and attributes and the areas where they are used is listed in Appendix D – Glossary of Entities and Attributes used in External Researcher CDM. A glossary of those attributes defined in the EDU Person Model that have not been appropriated for use by NIH are available in Appendix E - EDU Person Object and Fact Types not used in External Researcher CDM.

The External Researcher CDM inherits the applicable and related entities, attributes and relationships from the Person Name and Enterprise standards. By not including all related and inherited entities, attributes and relationships in this model, the External Researcher CDM is able to focus exclusively on the External Researcher business area. This allows for the business area-specific entities, attributes and relationships to be captured and easily understood.

The External Researcher CDM also will have specific business rules that will allow for unique identifiers for all entities including party, person and organization. This iteration of the External Researcher CDM captures some time dimensions and historical information.

3.4 Key External Researcher CDM Entities and Structures

There are a few key concepts associated with the External Researcher CDM that are integral to providing the flexibility that will be needed by NIH in its management of External Researchers in the future.

These core elements of the model are summarized in this section and are useful in understanding the remaining sections of this document. The definitions may differ slightly from current common usage at NIH or may be somewhat limited. These deviations must be addressed because precise definitions are needed to ensure that each object has a single, clear meaning within the context of the CDM.

Some of the key new entities represented in the model include:

- *Address Instance*—A description of the physical location at which a person may be found.
- *Contact Method*—Location-specific contact information for a person utilizing a specified medium.
- *Person Name Instance*—One of the many possible names that a person may use dependent upon the circumstances in which it is applied.

The External Researcher CDM also addresses a number of ongoing and emerging business needs which include: a government issued id name for travel, limiting role assignment to defined date ranges, tracking of name changes by their applicable dates, and the storage of multiple location-based contact methods.

4 External Researcher CDM

This section describes the External Researcher CDM in detail, including definitions of all the Entities, Key Attributes and Relationships that comprise the model. For those readers that might appreciate a data modeling refresher, a tutorial has been included in Appendix C.

The CDM has been decomposed into the following focus areas to support greater clarity:

1. Organization, Person and Role
2. Address and Contact Method
3. Person Name

A subsection for each of these focus areas presents an Object Role Model (ORM) model for the area, a definition for each entity and attribute, and a sentence about relationships amongst the entities and attributes. The constraint on each relationship is abbreviated as 1:1 (one-to-one), M:1 (many-to-one), 1:M (one-to-many), or M:M (many-to-many) when the sentence is read from left to right.

Critical objects relevant to the External Researcher CDM have been bolded in each introductory section. The presentation of the object names is structured. The primary objects are bolded and left justified and any subtype of the primary object is bolded and indented. All attributes of an object are in italics and right justified.

A complete list of entities and attributes has been defined in Appendix D - Glossary of Entities and Attributes in the External Researcher CDM.

4.1 ORGANIZATION, PERSON AND ROLE

4.1.1 Introduction

This section describes the business rules used to manage organizations and people. It also discusses the roles played by the people in those organizations. The effective management of this information is essential to meeting NIH's business needs for External Researchers.

An **Organization** can assign an individual **Researcher** multiple **Roles**, such as Principal Investigator and/or Signing Official. Depending upon the Roles assigned by an Organization, a Researcher can be granted specific permissions and responsibilities for accessing NIH computing resources. The same person can be assigned different roles by different organizations. For example, a Principal Investigator can be assigned the Role of Reviewer by NIH.

This model allows for continued flexibility and expansion of the number of roles that NIH manages about the External Researcher, which allows for flexibility in business processes over time. By allowing for each role played by a person to be tracked, NIH will be able to track the history of the persons involved in various business activities, such as principal investigator, loan recipient, etc.

4.1.2 Organization, Person and Role—Data Entities and Attributes

These are the high-level objects that represent the relationships between Party, Person and Organization and the Roles they play in the External Researcher business area as shown in

Table 3.

All of these entities have been defined as part of the External Researcher CDM as related to the Organization, Person and Role. These entities represent the core business data that need to be collected and managed throughout the NIH. For each of the entities, the following information is provided:

- *Entity Name*: The name used to refer to the entity. (**Bolded** within this standard.)
- *Attribute Name*: The name used to refer to a key attribute. (*Italicized* within this standard.)
- *Definition*: A description of the entity in plain English, consistent with the understood common usage within NIH whenever possible.
- *Source*: The point of origin for the definitions identified within this standard.

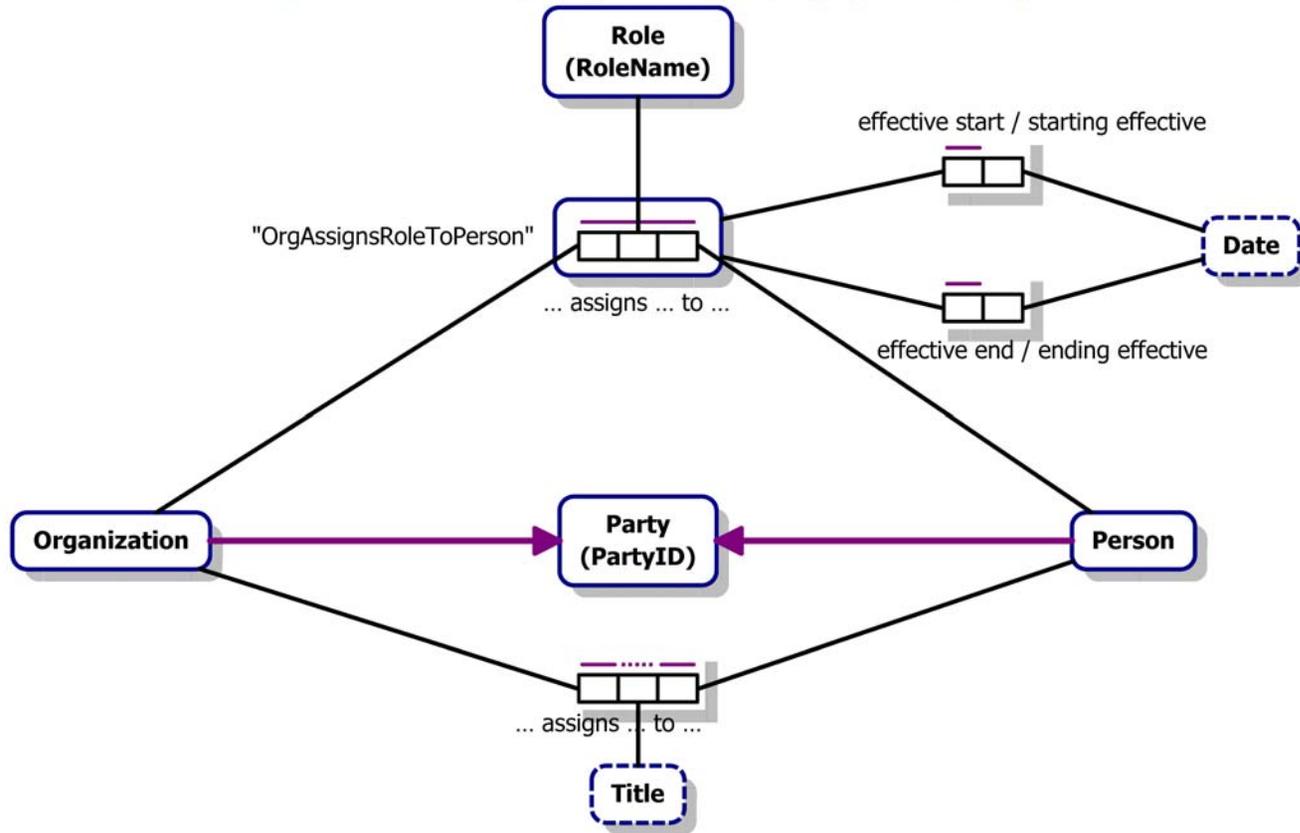
Table 3—Organization, Person and Role— CDM Entities and Key Attributes

#	Name	Definition	Source
1	Party	Information about people, organizations, and other actors in NIH processes, and their roles. Highest level entity that can participate in roles and be assigned permissions; includes person, organization and intelligent agent.	NIH Enterprise Conceptual Data Model NIHRFC0025/STD0012
2	<i>Party ID</i>	<i>A unique identifier of a Party.</i>	<i>NIH Grants Conceptual Data Model NIHRFC0026</i>
3	Organization	A formal grouping of people and/or business units coordinated to perform a specific purpose or obtain a specified objective.	NIH Enterprise Conceptual Data Model NIHRFC0025/STD0012
4	Person	Any individual of interest to the NIH for whom the NIH maintains information.	NIH Enterprise Conceptual Data Model NIHRFC0025/STD0012
5	Role	A named object assigned to a person that is given permissions and responsibilities to some resource or set of resources. Roles are defined by Organizations. A Role can be defined as part of another role.	Proposed by The Office of the Architect
6	<i>Role Name</i>	<i>A short description specifying the role.</i>	<i>NIH Grants Conceptual Data Model NIHRFC0026</i>
7	Person Playing Role	A compound object representing the assignment of a particular role to a person.	Proposed by The Office of the Architect
8	<i>Date</i>	<i>A particular period of time at which something happened or existed, or is expected to happen.</i>	<i>NIH R&D Contracts Conceptual Data Model NIHRFC0047</i>
9	<i>Title</i>	<i>The name of the position that a person holds in an organization</i>	<i>Proposed by The Office of the Architect</i>

4.1.3 Organization, Person and Role—ORM Model

Figure 2: Organization, Person and Role

{'External Researcher', 'Principal Investigator', 'Signing Official', 'etc.'}



4.1.4 Organization, Person and Role—Relationships and Rules

In addition to the key entities defined as part of the External Researcher CDM, there are relationships between entities that can be expressed as statements of business rules. Instances allow for a better understanding of the rule that is being expressed. Where applicable, population constraints are provided. Additional assumptions and comments may also be provided about the relationship.

Table 4—Organization, Person and Role—CDM Relationships and Rules

#	Relationship/Rule	Example	Constraint	Assumption	Comment
1	Party is identified by party id <PartyID>.	Party is identified by party id 7890000001 .	ID is unique		
2	Organization is a subtype of Party with party id <PartyID>.				
3	Person is a subtype of Party with party id <PartyID>.				
4	Organization with party id <PartyID> assigns the Title <Title> to Person with party id <PartyID>.	Organization with party id 1230000009 assigns the Title Director of Research to Person with party id 7890000001 .	M:1		
5	Role is identified by Role Name <RoleName>.	Role is identified by Role Name Principal Investigator .	RoleName is unique		Aligned with “Scoped Affiliation” in the EDU Person standard
6	Organization with party id <PartyID> assigns the Role of <RoleName>.to Person with party id <PartyID>.	Organization with party id 1230000009 assigns the Role of Principal Investigator to Person with party id 7890000001 .	M:M		

#	Relationship/Rule	Example	Constraint	Assumption	Comment
7	Organization with party id <PartyID> assigns the Role of <RoleName>.to Person with party id <PartyID> with effective start date of <EffectiveStartDate>.	Organization with party id 1230000009 assigns the Role of Principal Investigator to Person with party id 7890000001 with effective start date of 1/1/2010 .	M:1		
8	Organization with party id <PartyID> assigns the Role of <RoleName>.to Person with party id <PartyID> with effective end date of <EffectiveEndDate>.	Organization with party id 1230000009 assigns the Role of Principal Investigator to Person with party id 7890000001 with effective end date of 12/31/2010 .	M:1		

4.2 ADDRESS AND CONTACT METHOD

4.2.1 Introduction

An External Researcher can be associated with multiple Organizations that have working relationships with NIH. Within each Organization, an External Researcher can have multiple addresses and contact methods. All address and contact method data are ascribed to an External Researcher; no address or contact method data are maintained for the Organization.

The presentation of the Address and Contact Method fact types are as flexible as possible so that new types of Contact Method can be added to the current set. This general presentation of data at the conceptual level may be modified to a more structured environment for implementation, but the appropriate typing of the data must be made available as the data are shared across NIH.

4.2.2 Address and Contact Method—Data Entities and Attributes

Table 5 shows the data entities that have been defined as part of the External Researcher CDM as related to Address and Contact Method. These entities represent the core business data that need to be collected and managed throughout the NIH. For each of the entities, the following information is provided:

- *Entity Name*: The name used to refer to the entity. (**Bolded** within this standard.)
- *Attribute Name*: The name used to refer to an attribute. (*Italicized* within this standard.)
- *Definition*: A description of the entity in plain English, consistent with the understood common usage within NIH.
- *Source*: The point of origin for the definitions identified within this standard.

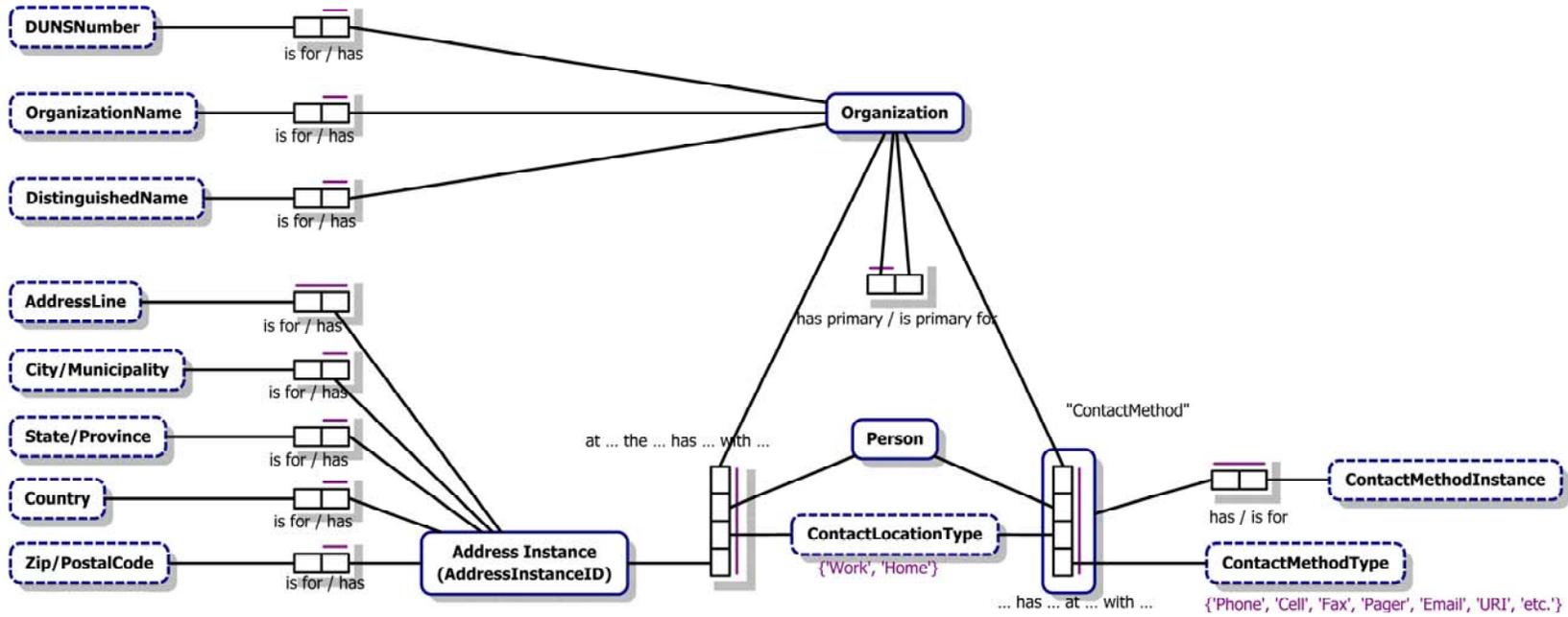
Table 5—Address and Contact Method— CDM Entities and Key Attributes

#	Name	Definition	Source
1	Organization	A formal grouping of people and/or business units coordinated to perform a specific purpose or obtain a specified objective.	NIH Enterprise Conceptual Data Model NIHRFC0025/STD0012
2	<i>DUNS Number</i>	<i>The DUNS number is a unique nine-digit number assigned by Dun and Bradstreet Information Services. It is recognized as the universal standard for identifying and keeping track of more than 92 million businesses worldwide. This is also known as Data Universal Numbering System (DUNS)</i>	<i>NIH Research Portfolio Online Reporting Tool (RePORT)²¹</i>
3	<i>Organization Name</i>	<i>The title by which the organization is referred to within NIH.</i>	<i>Proposed by The Office of the Architect</i>

#	Name	Definition	Source
4	<i>Distinguished Name</i>	<i>A unique name for the organization, usually the main domain name used for online communication.</i>	<i>Proposed by The Office of the Architect</i>
5	<i>Person</i>	<i>Any individual of interest to the NIH for whom the NIH maintains information.</i>	<i>NIH Enterprise Conceptual Data Model NIHRFC0025/STD0012</i>
6	<i>Address Instance</i>	<i>A description of the location at which a person may be found.</i>	<i>NIH R&D Contracts Conceptual Data Model NIHRFC0047</i>
7	<i>Address Instance ID</i>	<i>A unique identifier for an Address Instance</i>	<i>Proposed by The Office of the Architect</i>
8	<i>Address Line</i>	<i>A free text area that usually will contain street number or other location-specific information.</i>	<i>Proposed by The Office of the Architect</i>
9	<i>City/Municipality</i>	<i>The community in which an address is located.</i>	<i>Proposed by The Office of the Architect</i>
10	<i>State/Province</i>	<i>A sub-unit of a country with limited sovereignty in a federally organized country.</i>	<i>ISO/DIS 21090²²</i>
11	<i>Country</i>	<i>Locality of isolation indicated in terms of political names for nations, followed by regions and localities.</i>	<i>NCBI (adapted) http://www.ncbi.nlm.nih.gov/projects/collab/country.html</i>
12	<i>Zip/Postal Code</i>	<i>A postal code designating a region defined by the postal service.</i>	<i>ISO/DIS 21090</i>
13	<i>Contact Method</i>	<i>A compound object enabling electronic communications amongst people.</i>	<i>Proposed by The Office of the Architect</i>
14	<i>Contact Location Type</i>	<i>Type of place used in identifying a contact method, such as work or home.</i>	<i>Proposed by The Office of the Architect</i>
15	<i>Contact Method Type</i>	<i>The kind of medium to use for the contact method, such as a phone call or email.</i>	<i>Proposed by The Office of the Architect</i>
16	<i>Contact Method Instance</i>	<i>May contain specific contact details for the medium indicated by Contact Method Type such as a phone number or email address.</i>	<i>Proposed by The Office of the Architect</i>

4.2.3 Address and Contact Method—ORM Model

Figure 3—Address and Contact Method



4.2.4 Address and Contact Method—Relationships and Rules

In addition to the key entities defined as part of the External Researcher CDM, there are relationships between entities that can be expressed as statements of business rules. Instances allow for a better understanding of the rule that is being expressed. Where applicable, population constraints are provided. Additional assumptions and comments may also be provided about the relationship.

Table 6—Address and Contact Method—CDM Relationships and Rules

	Relationship/Rule	Example	Constraint	Assumption	Comment
1	Organization with party id <PartyID> has primary Organization with party id <PartyID>.	Organization with party id 8633718944 has primary Organization with party id 7833718944 .	M:1		
2	Organization with party id <PartyID> has DUNS Number <DUNSNumber>.	Organization with party id 7833718944 has DUNS Number 123456789 .	M:1		
3	Organization with party id <PartyID> has Organization Name <OrganizationName>.	Organization with party id 7833718944 has Organization Name University of Oklahoma .	M:1		
4	Organization with party id <PartyID> has Distinguished Name <DistinguishedName>.	Organization with party id 7833718944 has Distinguished Name ou.edu .	M:1		
5	Address Instance is identified by <AddressInstanceID>.	Address Instance is identified by 93726341 .	AddressInstanceID is unique		
6	At Organization with party id <PartyID>, the Person with party id <PartyID> has Contact Location Type of <ContactLocationType> with the Address Instance <AddressInstanceID>.	At Organization with party id 1230000009 , the Person with party id 7890000001 has Contact Location Type of Work with the Address Instance 92495650 .	M:M		Values for Contact Location Type: Work, Home

	Relationship/Rule	Example	Constraint	Assumption	Comment
7	Address Instance with id <AddressInstanceID> has Address Line of <AddressLine>.	Address Instance with id 93726341 has Address Line of 660 Parrington Oval .	M:1		
8	Address Instance with id <AddressInstanceID> has City/Municipality of <City>.	Address Instance with id 93726341 has City/Municipality of Norman .	M:1		
9	Address Instance with id <AddressInstanceID> has State/Province of <State>.	Address Instance with id 93726341 has State/Province of OK .	M:1		
10	Address Instance with id <AddressInstanceID> has Country of <Country>.	Address Instance with id 93726341 has Country of United States .	M:1		Country should be implemented using Country Code defined in ISO 3166. ²³
11	Address Instance with id <AddressInstanceID> has Zip/Postal Code of <Postal Code>.	Address Instance with id 93726341 has Zip/Postal Code of 73019 .	M:1		
12	Organization with party id <PartyID> has Person with party id <PartyID> at Contact Location Type of <ContactLocationType> with Contact Method Type of <ContactMethodType>.	Organization with party id 1230000009 has Person with party id 7890000001 at Contact Location Type of Work with Contact Method Type of Email .	M:M		Values for Contact Method Type: Phone, Cell, Fax, Pager, Email, URI

	Relationship/Rule	Example	Constraint	Assumption	Comment
13	Person with party id <PartyID> at Organization with party id <PartyID> for Contact Location Type of <ContactLocationType> with Contact Method Type of <ContactMethodType> has Contact Method Instance <ContactMethodInstance>.	Person with party id 7890000001 at Organization with party id 1230000009 for Contact Location Type of Work with Contact Method Type of Email has Contact Method Instance james.smith@ou.edu .	M:M		Values for Contact Method Type: Phone, Cell, Fax, Pager, Email, URI

4.3 PERSON NAME

4.3.1 Introduction

The Person Name attributes for External Researcher follow the Person Name standards. We are proposing three new name types. Any External Researcher that travels for NIH must have an “**Official Government ID Name**”. Most often there will be a “**Professional Name**” associated with the publications as a result of work performed with or for NIH. There is also a need for the “**Legal Name**” to be used when there are direct financial dealings with an External Researcher. If the Name data come from a trusted external source, then the name must be accompanied by a Name Type. Likewise, when the data are managed at NIH all shared names must be specified with a Name Type.

4.3.2 Person Name—Data Entities and Attributes

Table 7 shows the data entities that have been defined as part of the External Researcher CDM. These entities represent the Person Name data that need to be collected and managed throughout the NIH. For each of the entities, the following information is provided:

- *Entity Name*: The name used to refer to the entity. (**Bolded** within this standard).
- *Attribute Name*: The name used to refer to an attribute. (*Italicized* within this standard).
- *Definition*: A description of the entity in plain English, consistent with the understood common usage within NIH.
- *Source*: The point of origin for the definitions identified within this standard.

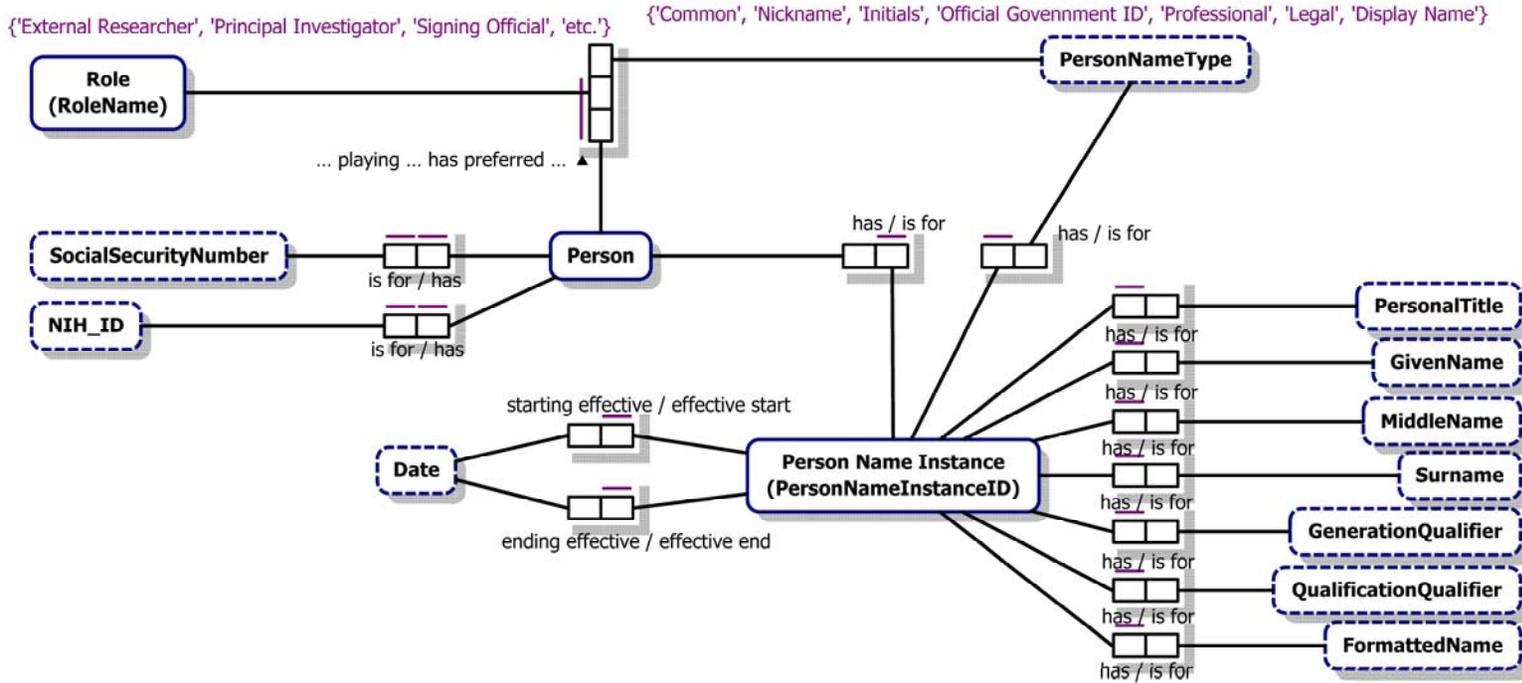
Table 7—Person Name—CDM Entities and Key Attributes

#	Name	Definition	Source
1	Person	Any individual of interest to the NIH for whom the NIH maintains information.	NIH Enterprise Conceptual Data Model NIHRFC0025/ST D0012
2	<i>Social Security Number</i>	<i>A unique number issued by the Social Security Administration that identifies a person.</i>	<i>Proposed by The Office of the Architect</i>
3	<i>NIH_ID</i>	<i>A unique number issued by NIH that identifies a person.</i>	<i>Proposed by The Office of the Architect</i>
4	Person Name Instance	One of a number of person names that refer to a person.	Person Name NIHRFC0002
5	<i>Person Name Instance ID</i>	<i>Unique identifier for a person name instance.</i>	<i>Person Name NIHRFC0002</i>
6	<i>Person Name Type</i>	<i>The types of names that are used to refer to a person at NIH.</i>	<i>Person Name NIHRFC0002</i>
7	<i>Personal Title</i>	<i>A person’s desired title.</i>	<i>Person Name NIHRFC0002</i>

#	Name	Definition	Source
8	<i>Given Name</i>	<i>First name of a person.</i>	<i>Person Name NIHRFC0002</i>
9	<i>Middle Name</i>	<i>Middle name or initial of a person.</i>	<i>Person Name NIHRFC0002</i>
10	<i>Surname</i>	<i>Last name of a person.</i>	<i>Person Name NIHRFC0002</i>
11	<i>Generation Qualifier</i>	<i>Denotes the generation of a person.</i>	<i>Person Name NIHRFC0002</i>
12	<i>Qualification Qualifier</i>	<i>The earned qualification of a person.</i>	<i>Person Name NIHRFC0002</i>
13	<i>Formatted Name (Display Name)</i>	<i>Complete name string derived via the rules for each name type.</i>	<i>Person Name NIHRFC0002</i>
14	<i>Date</i>	<i>A particular period of time at which something happened or existed, or is expected to happen.</i>	<i>NIH R&D Contracts Conceptual Data Model NIHRFC0047</i>
15	Role	A named object assigned to a person that is given permissions and responsibilities to some resource or set of resources. Roles are defined by Organizations. A Role can be defined as part of another role.	Proposed by The Office of the Architect
16	<i>RoleName</i>	<i>A short description specifying the role.</i>	<i>NIH Grants Conceptual Data Model NIHRFC0026</i>

4.3.3 Person Name—ORM Model

Figure 4—Person Name



4.3.4 Person Name—Relationships and Rules

In addition to the key entities defined as part of the External Researcher CDM, there are relationships between entities that can be expressed as statements of business rules. Instances allow for a better understanding of the rule that is being expressed. Where applicable, population constraints are provided. Additional assumptions and comments may also be provided about the relationship.

Table 8—Person Name—CDM Relationships and Rules

#	Relationship/Rule	Example	Constraint	Assumption	Comment
1	Person with party id <PartyID> has Social Security Number <SocialSecurityNumber>.	Person with party id 789000000 has Social Security Number 123-45-6789 .	1:1		
2	Person with party id <PartyID> has NIH_ID <NIHID>.	Person with party id 789000000 has NIH_ID 0011234567 .	1:1		This is aligned with “Uid” in the EDU Person standard
3	Person name instance is identified by <PersonNameInstanceID >.	Person name instance is identified by 000021 .	PersonNameInstanceID is unique		
4	Person name instance <PersonNameInstanceID > is for person with party id <PartyID>.	Person name instance 000021 is for person with party id 0019999222 .	M:1		
5	Person name instance <PersonNameInstanceID > has person name type of <PersonNameType>.	Person name instance 000021 has person name type of Professional .	M:1		

#	Relationship/Rule	Example	Constraint	Assumption	Comment
6	Person with party id <PartyID> playing role of <RoleName> has preferred person name type of <PersonNameType>	Person with party id 7890000000 playing role of Principal Investigator has preferred person name type of Professional .	M:1		
7	Person name instance <PersonNameInstanceID> has a personal title of <PersonalTitle>.	Person name instance 000021 has a personal title of Dr .	M:1		
8	Person name instance <PersonNameInstanceID> has a given name of <GivenName>.	Person name instance 000021 has a given name of James .	M:1		
9	Person name instance <PersonNameInstanceID> has a middle name of <MiddleName>.	Person name instance 000021 has a middle name of K .	M:1		
10	Person name instance <PersonNameInstanceID> has a surname of <Surname>.	Person name instance 000021 has a surname of Smith .	M:1		
11	Person name instance <PersonNameInstanceID> has a generation qualifier of <GenerationQualifier>.	Person name instance 000021 has a generation qualifier of III .	M:1		
12	Person name instance <PersonNameInstanceID> has a qualification suffix of <QualificationSuffix>.	Person name instance 000021 has a qualification suffix of M.D., Ph.D.	M:1		

#	Relationship/Rule	Example	Constraint	Assumption	Comment
13	Person name instance with <PersonNameInstanceID > has a formatted name of <FormattedName>.	Person name instance with 000021 has a formatted name of Dr. James K. Smith, III M.D., Ph.D.	M:1		This is aligned with "Display Name" in the EDU Person standard
14	Person name instance <PersonNameInstanceID > has an effective start date of <EffectiveStartDate>.	Person name instance 000021 has an effective start date of 1/21/2003.	M:1		
15	Person name instance <PersonNameInstanceID > has an effective end date of <EffectiveEndDate>.	Person name instance 000021 an effective end date of 6/12/2006.	M:1		

5 References

1. Federal Enterprise Architecture Program. "The Data Reference Model." Version 2.0. http://www.whitehouse.gov/OMB/egov/documents/DRM_2_0_Final.pdf. September, 2006.
2. NIH Enterprise Conceptual Data Model—NIHRFC 0025 (STD0012) version 1.0. <http://enterprisearchitecture.nih.gov/NR/rdonlyres/5D3017EA-22C1-4BCC-8E0F-79EB7B5C797A/0/NRFC0025.pdf>. January, 2007.
3. NIH Person Name—NIHRFC 0002 (STD0001) version. <http://enterprisearchitecture.nih.gov/NR/rdonlyres/B6CC7AE9-A6AA-44E7-8452-51D20B8E1AA1/0/NRFC0002.pdf>. January, 2007.
4. Enterprise Performance Life Cycle (EPLC). <http://www.cc.nih.gov/dcri/epcl.html>. January, 2009.
5. NIH Office of the Chief Information Officer (2006), Capital Planning and Investment Control (CPIC) Policy and Guidelines Version 3. <http://irm.cit.nih.gov/itmra/nihcpicpolicy.doc>. September, 2006.
6. internet2.edu Middleware.eduPerson Object Class Specification. <http://middleware.internet2.edu/eduperson/docs/internet2-mace-dir-eduperson-200806.html>. June, 2008.
7. Person Process model (current state, not yet released). <http://enterprisearchitecture.nih.gov/ArchLib/AT/BA/NIHExternalDirectory.htm>
8. HR-XML Consortium Person Name Recommendation. http://ns.hr-xml.org/2_5/HR-XML-2_5/CPO/PersonName.html. April, 2007.
9. UsPerson version 1.0. <http://www.parlato.com/docs/com/parlato/beans/UsPerson.html>. October, 1997.
10. Erika Rissi; Researcher Profile and Commons Personal Profile Analysis. www.nsf.gov, http://era.nih.gov/services_for_applicants/apply/personal_profile.cfm. April, 2009.
11. Collaborative Data Dictionary (CDD) Working Draft. <http://www.uta.edu/research/collaborate/cdd.php>. 2006.
12. University of Florida LDAP Schema Design. <http://open-systems.ufl.edu/node/221>. date unknown.
13. Medbiquitous ANSI /MEDBIQ PP.10.1-2008 Healthcare Professional Profile Specifications Document. http://www.medbiq.org/working_groups/professional_profile/NameSpecification.pdf. June, 2008.

14. ISO/IEC (2003). ISO/IEC 11179-3, Information Technology Metadata Registries (MDR), Part 3: Registry Metamodel and Basic Attributes. <http://jtc1sc32.org/doc/N1151-1200/32N1168-ISO-IEC11179-3-2003COR1.zip>. September, 2006.
15. Integration Services Center - Integration Description: NIH Federation Target Applications. <http://enterprisearchitecture.nih.gov/NR/rdonlyres/69B19CDA-F07E-489E-B99C-244AC54D09E5/0/DomainTeamApplicationIntegration20040726.pdf>. July, 2004.
16. NIH Unique Identifier for People, the NIH ID Number v1.0 NRFC0005. <http://enterprisearchitecture.nih.gov/NR/rdonlyres/8FEA4894-D234-4305-92CD-14C99A1D3E96/0/NRFC0005.pdf>. January, 2006.
17. NIH Organizational Person Schema. <http://nedinfo.nih.gov/amgtech/docs/schema/current.html>. December, 2009.
18. NIH Grants Conceptual Data Model – NIHRFC 0026, version 1.0. <http://enterprisearchitecture.nih.gov/NR/rdonlyres/37CF0ABE-D835-497E-AF27-9FB354799EE6/0/NIHRFC0026NIHGrantsConceptualDataModel.pdf>. April, 2008.
19. NIH R&D Contracts Conceptual Data Model NIHRFC0047. <http://enterprisearchitecture.nih.gov/YourPart/Collaborate/ApprovedNRFCs.htm>. <http://enterprisearchitecture.nih.gov/YourPart/Collaborate/NRFCDiscussionArea.htm>. Draft January, 2010.
20. Terry Halpin “Object Role Modeling.” <http://ww.orm.net>. September, 2006.
21. NIH Research Portfolio Online Reporting Tool (RePORT). <http://report.nih.gov/glossary.aspx?filter=P>. January, 2010.
22. DRAFT Health Informatics – Harmonized data types for informatics exchange, ISO/DIS 21090, April 2008
23. English Country Names ISO 3166 http://www.iso.org/iso/english_country_names_and_code_elements
24. For more information on the NLM notation and interpreting NLM models see <http://www.sharpinformatics.com>.

6 Contact

To contact the NIHRFC Editor, send an e-mail message to EnterpriseArchitecture@mail.nih.gov.

7 Security Considerations

This NIHRFC raises no security issues.

8 Changes

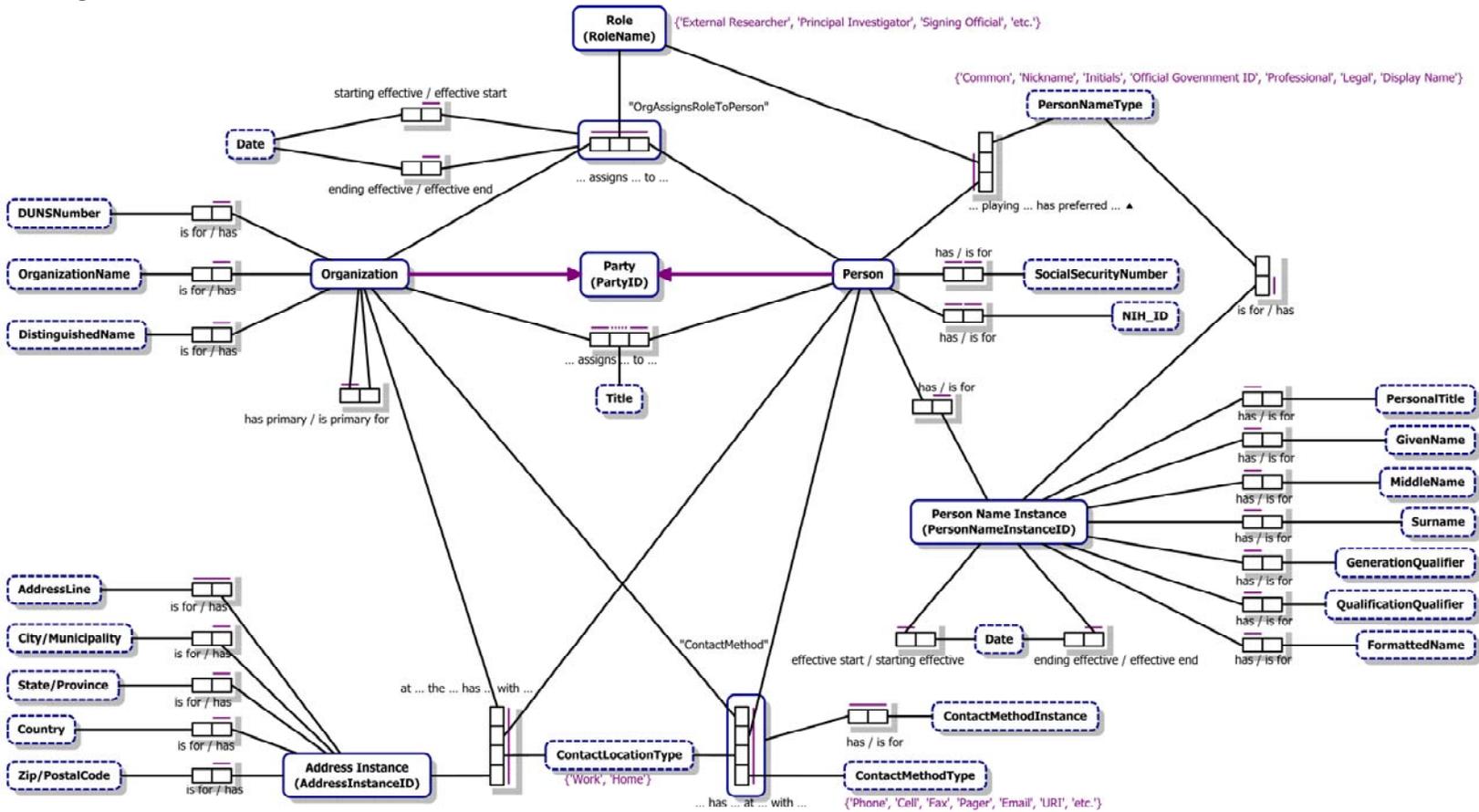
Version	Date	Change	Authority	Author of Change
0.1	Mar, 2010	Original Draft	NIHRFC0001	Eric Mechenbier, Taruna Reddy, John Sharp
0.2	May, 2010	Revised per Comments: Included international address; revised definition of Role to be person specific; explained mandatory/optional requirements for use cases; changed External Researcher from an object to a Role that an individual can play; included fact type for preferred name type.	NIHRFC0001	Eric Mechenbier, Taruna Reddy, John Sharp
0.3	June 2010	Changed the title to External Researcher for better alignment with the target audience. Removed the delegation-of-Role fact type because this rule is more appropriate for specific information systems and not as an NIH standard.	NIHRFC0001	Eric Mechenbier, Taruna Reddy, John Sharp
1.0	June 2010	Approved by the ARB	NIHRFC0001	Kiley Ohlson

9 Authors' Address

Eric Mechenbier, Taruna Reddy, John Sharp
National Institutes of Health
10401 Fernwood Road
MSC 4806
Bethesda, Maryland 20817
Phone: 505.243.1498
E-mail: sharp@sharpinformatics.com

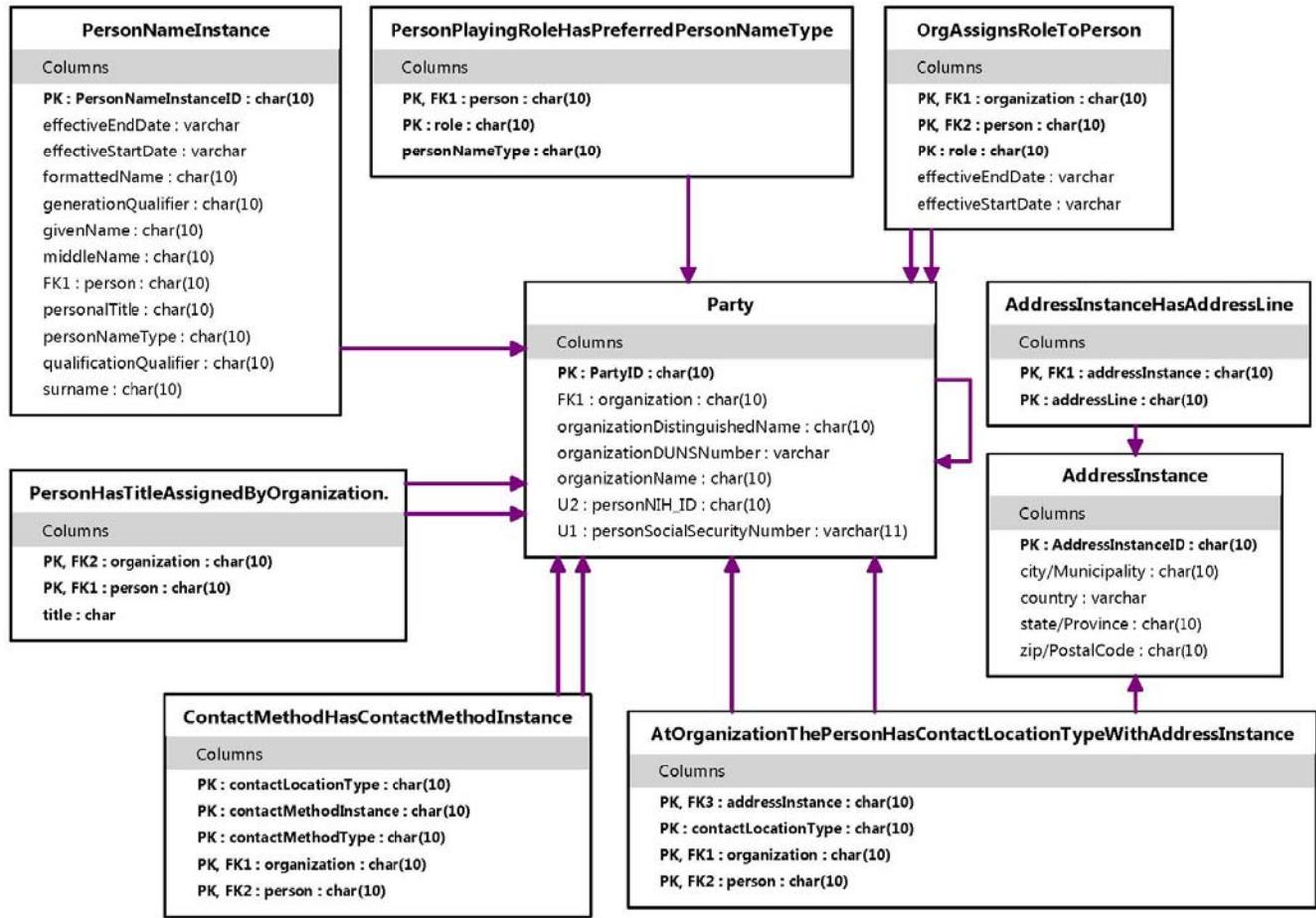
Appendix A: Comprehensive External Researcher CDM—ORM Notation

Figure 5—External Researcher CDM—ORM Model



Appendix B: Comprehensive External Researcher CDM – ERD Notation

Figure 6—External Researcher CDM—ERD Model



Appendix C: Data Modeling Tutorial

What is data modeling and why is it used?

Data modeling is the process of exploring and representing data in a structured manner within a knowledge or subject area. Data models identify the data elements that the business uses and how they relate to one another. This is represented by entities (or kinds of significant things) about which an organization wishes to know, collect, and maintain information about, the attributes (characteristics of the entities), and the relationships among the entities. In addition to defining and organizing the data, data modeling imposes constraints or limitations (implicitly or explicitly) on how that data are placed within a structure.

Data models typically address only structured data and do not describe any unstructured data such as e-mail messages, graphics, pictures, etc.

Data models can be one or more of three kinds: conceptual, logical and physical data models:

- A *conceptual data model*, sometimes called domain models, typically are used to explore the domain concepts at a high level with stakeholders who assist in specifying the entities, attributes and the relationships among them.
- A *logical data model* describes the tables and columns.
- A *physical data model* describes the physical and internal mechanisms within database depicting the data columns of the tables and the relationships between the tables.

The conceptual data model is typically devoid of detailed implementation information such as database vendor, how the physical database will be built (i.e., relational, objected oriented or other dimensional information), etc. The entities and relationships can be depicted pictorially to allow stakeholders and users to easily view the information structure as shown in Figure 7. Data instances represent the real world occurrence of data as seen in Figure 8. This is one of the mechanisms to test the validity of the concepts and relationships within the data model.

Figure 7: Entity and Relationships

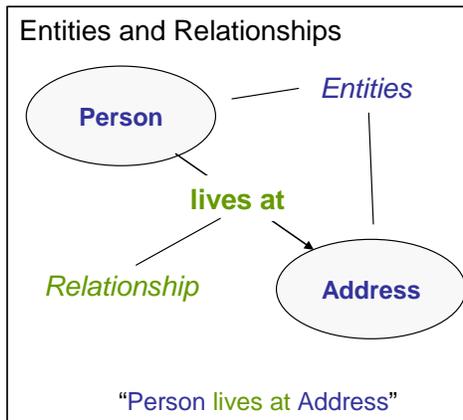
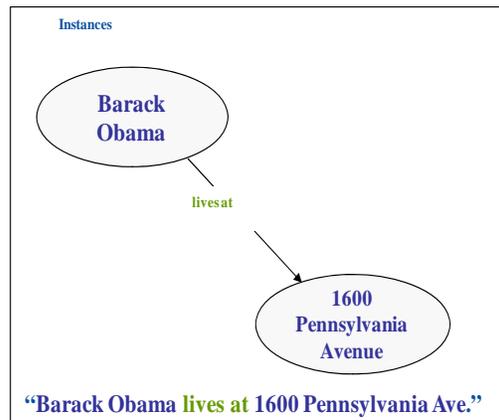
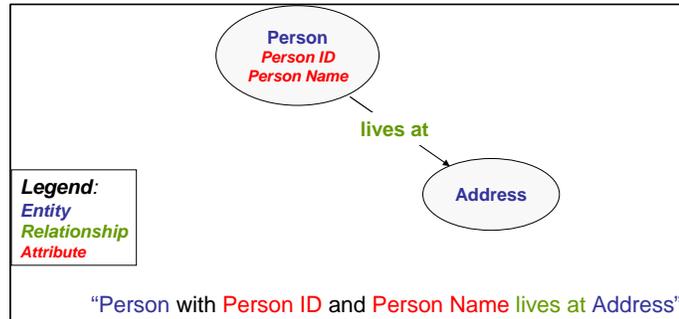


Figure 8: Instances of Data



In addition to the entities and relationships within the model, we also define characteristics of the entities known as attributes as shown in Figure 9.

Figure 9: Entity, Relationships and Attributes



From the conceptual data model, further decomposition of the model provides the detailed information (logical and physical information) and structures as to how the model will be physically built, along with the information systems that will use the models.

Models provide a formal, rigorous way of representing the world by providing an unambiguous input to the design and development of IT solutions, and another mechanism to communicate about data. Good and consistent models allow reliable data to be shared across information systems and also help facilitate the evolution of information systems changes.

Categories of Data Models

Table 9 shows the definitions and purpose of different types of data models that may be created as part of NIH's Enterprise Data Architecture or in support of specific solution implementations.

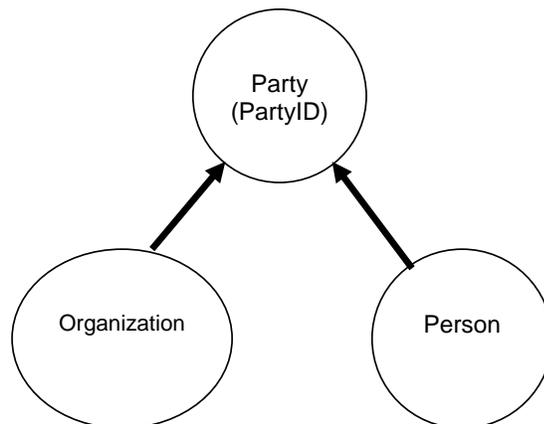
Table 9—Categories of Data Models

	Description	Purpose
Conceptual Data Model	A CDM represents the overall logical structure of a database, which is <u>independent</u> of any software or data storage structure. A conceptual model often contains data objects not yet implemented in the physical databases. It gives a formal representation of the data needed to run an enterprise or a business activity.	Suitable as an enterprise level artifact to provide an overall set of key data entities to facilitate management of enterprise data resources and support effective information sharing.
Logical Data Model	A logical view of the conceptual data model. Data Architecture theories such as "normalization" are applied to transform the conceptual data model into the logical data model that moves the data modeling further towards the ultimate prescription for the data architecture to be implemented. Relationships get absorbed as "attributes" known as foreign keys or pointers within appropriate logical model entities. This may be explicit or implied in the logical data model. As long as the resulting physical data model includes the necessary foreign key columns and joins, the inclusion of foreign-keys in the logical data model is a matter of convenience. Logical Data Model does not have any specific restrictions and/or requirements imposed by the Database Management System (DBMS) to be used for creating the actual database.	Suitable for representing the detailed business rules governing the structure of data elements and their relationships. Provides a more detailed view of the data and is more suitable for specific systems designs of for the description of enterprise data standards.
Physical Data Model	The mapping of conceptual or logical database design data groupings into physical database areas, files, records, elements, fields , and keys while adhering to the physical constraints of the hardware, DBMS software, and communications network to provide physical data integrity while meeting the performance and security constraints of the services to be performed against the database.	Suitable for the design of specific implementations of a data model. Generally not suitable for enterprise standards or architecture specification

Subtypes and Supertypes of Objects

The External Researcher CDM model has special notations for describing Subtypes and Supertypes. A supertype is a high level object that has widespread use across the model. In this model, subtypes are smaller sets of the supertype that share common attributes and they allow for the specification of more precise rules.

Figure 10—Party Supertype with Organization and Person Subtypes

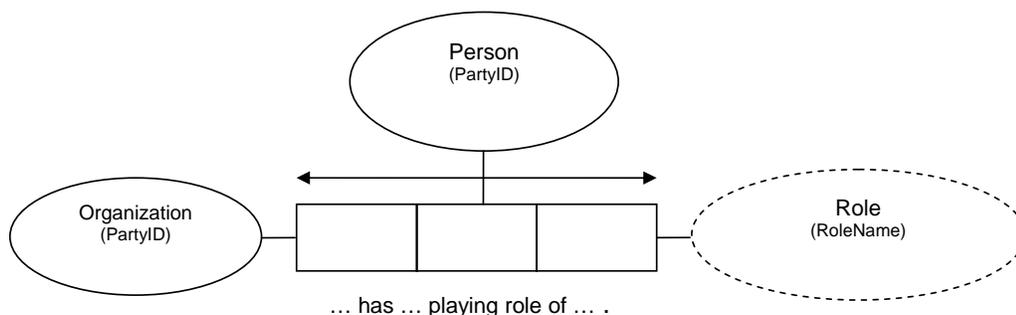


In the example above, Party (the supertype) can be used to designate rules where either an instance of a Person or Organization (the subtypes) can be valid. For example, a Party could be a grant applicant that is either a person requesting the grant for a fellowship or an organization requesting the grant for medical research. Other rules only apply to groups of Organizations and the rules are written so that Subject Matter Experts can understand and validate the rules. For example, an Offeror (subtype of Organization), can be the awardee of an R&D Contract.

ORM Diagrams and Natural Language Modeling

Object Role Modeling (ORM) is a graphical modeling technique that precisely displays fact types and business rules. Natural Language Modeling (NLM)²³ is completely sentence-based and provides subject matter experts with the ability to establish and validate fact types and business rules without becoming proficient in reading graphical models. Both of these focus on the establishment of fact types and rules that are sentence-based.

Figure 11—ORM Fact Type



The fact type reading for this ORM diagram is:

Organization with <PartyID> has Person <PartyID> playing the Role of <RoleName>.

The arrow over the Organization, Person and Role objects means that there is a many-to-many (M:M) relationship between the combined objects.

Natural Language Modeling allows these rules to be validated by only asking questions about the objects.

Given the populated fact type: “Organization **435543231** has Person **101342872** playing the Role of Principal Investigator.” **is true:**

Q1.1. Can Organization **678543325** have Person **101342872** playing the Role of Principal Investigator?—**Yes**

Q1.2. Can Organization **435543231** have Person **781344432** playing the Role of Principal Investigator?—**Yes**

Q1.3. Can Organization **435543231** have Person **101342872** playing the Role of Signing Official?—**Yes**

The results of this analysis can be expressed as a matrix:

Organization <PartyID> has Person <PartyID> playing the Role of Principal Investigator.

	Organization <PartyID>	Person <PartyID>	Role <RoleName>	
Instance	435543231	101342872	Principal Investigator	
				Allowed?
Q1.1	another	101342872	Principal Investigator	Yes
Q1.2	435543231	Another	Principal Investigator	Yes
Q1.3	435543231	101342872	another	Yes

A “yes” answer means that each variable is independent of the others. A “no” answer would mean that the variable in question is dependent on one or more of the other variables.

The Q1 matrix is the first step in the NLM procedure. Because the answers are all “Yes”, a second Question is asked about this fact type.

Q2: Does instance: “Organization **43543231** have Person **101342872** playing the Role of Principal Investigator.” exactly identify an object? **Yes**

The Q2 Yes answer means that this is an Identification Fact Type which is commonly called a key. Other Q1 answer patterns would similarly go through a set of steps resulting in the creation of the needed fact types for the data model. The resulting Table would be:

OrganizationPersonRole

Organization <PartyID>	Person <PartyID>	Role <RoleName>
---	-----------------------------------	----------------------------------

Appendix D: Glossary of Entities and Attributes used in External Researcher CDM**Table 10— Glossary of Entities and Attributes used in External Researcher CDM**

#	Name	Definition	Source	Where Used
1	Party	Information about people, organizations, and other actors in NIH processes, and their roles. Highest level entity that can participate in roles and be assigned permissions; includes person, organization and intelligent agents.	NIH Enterprise Conceptual Data Model NIHRFC0025/ST D0012	Organization, Person, and Role
2	<i>Party ID</i>	<i>A unique identifier of a Party.</i>	<i>NIH Grants Conceptual Data Model NIHRFC0026</i>	Organization, Person, and Role
3	Organization	A formal grouping of people and/or business units coordinated to perform a specific purpose or obtain a specified objective.	NIH Enterprise Conceptual Data Model NIHRFC0025/ST D0012	Organization, Person, and Role Address and Contact Method
4	Person	Any individual of interest to the NIH for whom the NIH maintains information.	NIH Enterprise Conceptual Data Model NIHRFC0025/ST D0012	Organization, Person, and Role Address and Contact Method Person Name
5	Role	A named object assigned to a person that is given permissions and responsibilities to some resource or set of resources. Roles are defined by Organizations. A Role can be defined as part of another role.	Proposed by The Office of the Architect	Organization, Person, and Role Person Name
6	<i>Role Name</i>	<i>A short description specifying the role.</i>	<i>NIH Grants Conceptual Data Model NIHRFC0026</i>	Organization, Person, and Role Person Name

Formatted: Portuguese (Brazil)

Formatted: Portuguese (Brazil)

Formatted: Portuguese (Brazil)

#	Name	Definition	Source	Where Used
7	Person Playing Role	A compound object representing the assignment of a particular role to a person.	Proposed by The Office of the Architect	Organization, Person, and Role
8	<i>Date</i>	A particular period of time at which something happened or existed, or is expected to happen.	NIH R&D Contracts Conceptual Data Model NIHRFC0047	Organization, Person, and Role Person Name
9	<i>Title</i>	<i>The name of the position that a person holds in an organization</i>	Proposed by The Office of the Architect	Organization, Person, and Role
10	<i>DUNS Number</i>	<i>The DUNS number is a unique nine-digit number assigned by Dun and Bradstreet Information Services. It is recognized as the universal standard for identifying and keeping track of more than 92 million businesses worldwide. This is also known as Data Universal Numbering System (DUNS)</i>	<i>NIH Research Portfolio Online Reporting Tool (RePORT)</i>	Address and Contact Method
11	<i>Organization Name</i>	<i>The title by which the organization is referred to within NIH.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
12	<i>Distinguished Name</i>	<i>A unique name for the organization, usually the main domain name used for online communication.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
13	Address Instance	A description of the location at which a person may be found.	NIH R&D Contracts Conceptual Data Model NIHRFC0047	Address and Contact Method
14	<i>Address Instance ID</i>	<i>A unique identifier for an Address Instance</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method

Formatted: Portuguese (Brazil)

#	Name	Definition	Source	Where Used
15	<i>Address Line</i>	<i>A free text area that usually will contain street number or other location-specific information.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
16	<i>City/Municipality</i>	<i>The community in which an address is located.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
17	<i>State/Province</i>	<i>A sub-unit of a country with limited sovereignty in a federally organized country</i>	<i>ISO/DIS 21090</i>	Address and Contact Method
18	<i>Country</i>	<i>Locality of isolation indicated in terms of political names for nations, followed by regions and localities</i>	<i>NCBI (adapted)</i> <i>http://www.ncbi.nlm.nih.gov/projects/collab/country.html</i>	Address and Contact Method
19	<i>Zip/Postal Code</i>	<i>A postal code designating a region defined by the postal service.</i>	<i>ISO/DIS 21090</i>	Address and Contact Method
20	Contact Method	A compound object enabling electronic communications amongst people.	Proposed by The Office of the Architect	Address and Contact Method
21	<i>Contact Location Type</i>	<i>Type of place used in identifying a contact method, such as work or home.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
22	<i>Contact Method Type</i>	<i>The kind of medium to use for the contact method, such as a phone call or email.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
23	<i>Contact Method Instance</i>	<i>May contain specific contact details for the medium indicated by Contact Method Type such as a phone number or email address.</i>	<i>Proposed by The Office of the Architect</i>	Address and Contact Method
24	<i>Social Security Number</i>	<i>A unique number issued by the Social Security Administration that identifies a person.</i>	<i>Proposed by The Office of the Architect</i>	Person Name

#	Name	Definition	Source	Where Used
25	<i>NIH_ID</i>	<i>A unique number issued by NIH that identifies a person.</i>	<i>Proposed by The Office of the Architect</i>	Person Name
26	Person Name Instance	<i>One of a number of person names that refer to a person.</i>	<i>Person Name NIHRFC0002</i>	Person Name
27	<i>Person Name Instance ID</i>	<i>Unique identifier for a person name instance.</i>	<i>Person Name NIHRFC0002</i>	Person Name
28	<i>Person Name Type</i>	<i>The types of names that are used to refer to a person at NIH.</i>	<i>Person Name NIHRFC0002</i>	Person Name
29	<i>Personal Title</i>	<i>A person's desired title.</i>	<i>Person Name NIHRFC0002</i>	Person Name
30	<i>Given Name</i>	<i>First name of a person.</i>	<i>Person Name NIHRFC0002</i>	Person Name
31	<i>Middle Name</i>	<i>Middle name or initial of a person.</i>	<i>Person Name NIHRFC0002</i>	Person Name
32	<i>Surname</i>	<i>Last name of a person.</i>	<i>Person Name NIHRFC0002</i>	Person Name
33	<i>Generation Qualifier</i>	<i>Denotes the generation of a person.</i>	<i>Person Name NIHRFC0002</i>	Person Name
34	<i>Qualification Qualifier</i>	<i>The earned qualification of a person.</i>	<i>Person Name NIHRFC0002</i>	Person Name
35	<i>Formatted Name (Display Name)</i>	<i>Complete name string derived via the rules for each name type.</i>	<i>Person Name NIHRFC0002</i>	Person Name

Appendix E: EDU Person Object and Fact Types not used in External Researcher CDM**Table 11—EDU Person CDM Entities and Key Attributes**

#	Name	Definition	Source
1	Affiliation	Specifies the person's relationship(s) to the institution in broad categories such as student, faculty, staff, alum, etc.	eduPerson Object Class Specification (200806)
2	Primary Affiliation	Specifies the person's PRIMARY relationship to the institution in broad categories such as student, faculty, staff, alum, etc.	eduPerson Object Class Specification (200806)
3	<i>Audio</i>	<i>Avoid. Not clearly defined, no de facto standard.</i>	<i>eduPerson Object Class Specification (200806)</i>
4	<i>Description</i>	<i>Text which describes the person object.</i>	<i>NIH Organizational Person Schema (adapted)</i>
5	<i>Jpeg Photo</i>	<i>JPEG photograph of person.</i>	<i>NIH Organizational Person Schema (adapted)</i>
6	<i>Language</i>	<i>Preferred written or spoken language for a person.</i>	<i>eduPerson Object Class Specification (200806)</i>
7	<i>Person Assurance</i>	<i>Set of URIs that assert compliance with specific standards for identity assurance.</i>	<i>eduPerson Object Class Specification (200806)</i>
8	<i>Person Entitlement</i>	<i>URI (either URN or URL) that indicates a set of rights to specific resources.</i>	<i>eduPerson Object Class Specification (200806)</i>
9	<i>See Also</i>	<i>The 'seeAlso' attribute type contains the distinguished names of objects that are related to the subject object. Each related object name is one value of this multi-valued attribute.</i>	<i>eduPerson Object Class Specification (200806)</i>

#	Name	Definition	Source
10	<i>Targeted ID</i>	<i>A persistent, non-reassigned, privacy-preserving identifier for a principal shared between a pair of coordinating entities, denoted by the SAML 2 architectural overview [1] as identity provider and service provider (or a group of service providers). An identity provider uses the appropriate value of this attribute when communicating with a particular service provider or group of service providers, and does not reveal that value to any other service provider except in limited circumstances.</i>	<i>eduPerson Object Class Specification (200806)</i>
11	<i>Unique Identifier</i>	<i>The 'uniqueIdentifier' attribute specifies a unique identifier for an object represented in the Directory. The domain within which the identifier is unique and the exact semantics of the identifier are for local definition. For a person, this might be an institution- wide payroll number. For an organizational unit, it might be a department code.</i>	<i>eduPerson Object Class Specification (200806)</i>
12	<i>User Certificate</i>	<i>A user's X.509 certificate.</i>	<i>eduPerson Object Class Specification (200806)</i>
13	<i>User Password</i>	<i>This attribute identifies the entry's password and encryption method in the following format</i>	<i>eduPerson Object Class Specification (200806)</i>
14	<i>User SMIME Certificate</i>	<i>An X.509 certificate specifically for use in S/MIME applications</i>	<i>eduPerson Object Class Specification (200806)</i>

#	Name	Definition	Source
15	X500 Unique Identifier	Avoid. X500UniqueIdentifier syntax is specified as bit string, and that is not likely to be a good fit for many of the institutional attribute value choices, especially as part of the DN.	eduPerson Object Class Specification (200806)
16	P.O. Box	For those addresses that require it, this contains the number of the Post Office Box.	Proposed by The Office of the Architect

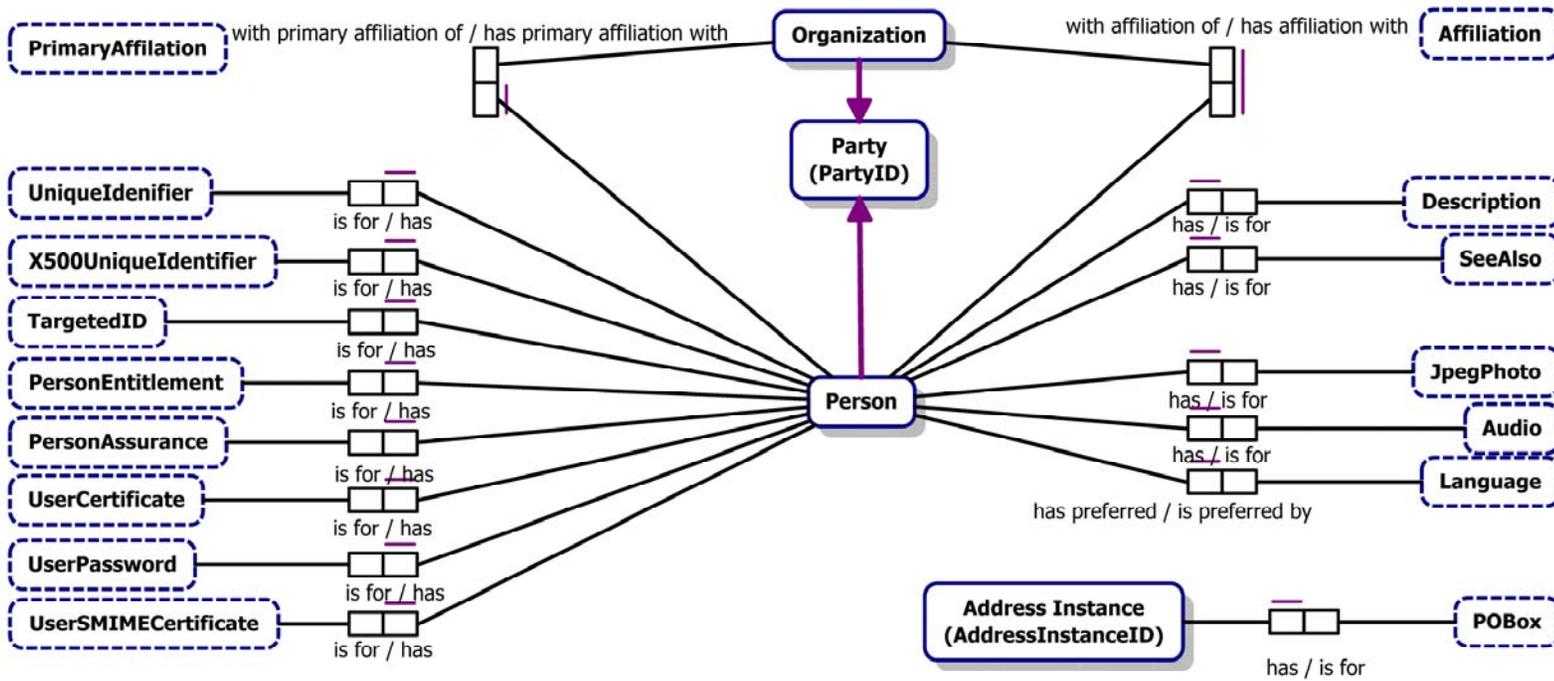


Table 12—EDU Person CDM Relationships and Rules

#	Relationship/Rule	Example	Constraint	Assumption	Comment
1	Person with party id <PartyID> has Primary Affiliation with the Organization identified by party id <PartyID>.	Person with party id 7890000001 has Primary Affiliation with the Organization identified by party id 1230000009 .	M:1		
2	Person with party id <PartyID> has an Affiliation with the Organization identified by party id <PartyID>.	Person with party id 7890000001 has an Affiliation with the Organization identified by party id 1230000009 .	M:M		
3	Person with party id <PartyID> has unique identifier <UniqueIdentifier>.	Person with party id 7890000001 has unique identifier 19271932 .	M:1		
4	Person with party id <PartyID> has X500 Unique Identifier <X500UniqueIdentifier>.	Person with party id 7890000001 has X500 Unique Identifier xxx .	M:1		
5	Person with party id <PartyID> has Targeted ID <TargetedID>.	Person with party id 7890000001 has Targeted ID WGVEAC .	M:1		
6	Person with party id <PartyID> has Person Entitlement <Person Entitlement>.	Person with party id 7890000001 has Person Entitlement xxx .	M:1		
7	Person with party id <PartyID> has Person Assurance <Person Assurance>.	Person with party id 7890000001 has Person Assurance xxx .	M:1		
8	Person with party id <PartyID> has User Certificate <User Certificate>.	Person with party id 7890000001 has User Certificate xxx .	M:1		

#	Relationship/Rule	Example	Constraint	Assumption	Comment
9	Person with party id <PartyID> has password <UserPassword>.	Person with party id 7890000001 has password *****.	M:1		
10	Person with party id <PartyID> has User SMIME Certificate <UserSMIMECertificate>.	Person with party id 7890000001 has User SMIME Certificate xxx .	M:1		
11	Person with party id <PartyID> has description <Description>.	Person with party id 7890000001 has description This person...	M:1		
12	Person with party id <PartyID> has “See Also” reference to Person with uid <SeeAlso>.	Person with party id 7890000001 has “See Also” reference to Person with uid 001-9876-543 .	M:M		
13	Person with party id <PartyID> has JPEG photo <JpegPhoto>.	Person with party id 7890000001 has JPEG photo Photo 23.jpg .	M:1		
14	Person with party id <PartyID> has audio clip <Audio>.	Person with party id 7890000001 has audio clip Voice Sample 23.mp3 .	M:1		
15	Person with party id <PartyID> has preferred language <Language> for official communication.	Person with party id 7890000001 has preferred language German for official communication.	M:1		
16	Address Instance with id <AddressInstanceID> has P.O. Box of <POBox>.	Address Instance with id 93726341 has P.O. Box of 8807 .	M:1		