NATIONAL INSTITUTES OF HEALTH

enterpriseARCHITECTURE

Enterprise Data Architecture and Conceptual Data Model Overview

2/1/2006

NIH Enterprise Information Technology Architecture
Contact: enterprisearchitecture@mail.nih.gov

Objectives for Today

- Define key concepts in data architecture and explain how they apply at NIH.
- Describe NIH's approach to data architecture and the structure of NIH's data architecture framework.
- Introduce the NIH's Enterprise Conceptual Data Model NRFC and provide context needed to effectively use it.

Background

- An Enterprise Conceptual Data Model (E-CDM) has been developed for NIH.
  - The E-CDM will serve as the organizing taxonomy for data architecture deliverables.
  - The E-CDM was approved by the Architecture Review Board (ARB) on 1/24/2007 and will be published as NRFC0025.
  - The E-CDM will serve as the foundation for NIH's further data architecture efforts.

What is Data Architecture?

Enterprise Architecture

Enterprise architecture defines how information and technology will support the business operations and provide benefit for the business.
It illustrates the organization's core mission, each component critical to performing that mission, and how each of these components is interrelated.

Enterprise Data Architecture

Data architecture is the part of the architecture that focuses on the structure of information assets required to support the business, its management within systems, and the processes enabling effective use of information.

Why Does NIH Need Data Architecture?

- Improved information management in support of:
  - Organized approach to managing information
  - Data quality and integrity
  - Effective data sharing
  - Enabling reuse and SOA
  - Simplifying the environment
  - Improved decision-making
  - Improved alignment with the business
    - Support for multi-PI
    - Improved automation for people management
    - …

What Are the Components of the Enterprise Data Architecture?

Supports high-level IT planning and governance of data resources. Acts as a starting point for domain specific architecture.

Provides more detailed direction to support consistent data representation across implementation efforts and more detailed planning of systems within a particular segment. Align with Segment Application Architectures.

Provides explicit standards with respect to the representation of data and policies for data management that apply across the NIH enterprise.

Enterprise Data Architecture

- Enterprise Conceptual Data Model
- Data Architecture Current State
- Data Architecture Future State
- Enterprise Meta-Model

Segment Sub-Architectures

Grants Data Architecture
- Segment Conceptual Model
- Logical Data Models
- Detailed Service/System/ Process Mappings (CRUD)

People Data Architecture
- Segment Conceptual Model
- Logical Data Models
- Detailed Service/System/ Process Mappings (CRUD)

Research/ Clinical Data Architecture
- Segment Conceptual Model
- Logical Data Models
- Detailed Service/System/ Process Mappings (CRUD)

Additional Segments

NIH-wide Data Standards

Examples include - Person Name, Group Name, Address

Categories of Data Models

	Description	Purpose
Conceptual Data Model	A CDM represents the overall logical structure of a database, which is independent 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

Enterprise Conceptual Data Model Overview

- An Enterprise Conceptual Data Model (CDM) provides a high level model of the data within an enterprise.
- It is primarily a tool to support organization and planning of the enterprises information resources.
  - It is intended to identify the critical information required to support enterprise business process and provide a common nomenclature for describing data in “as is” and “to be” enterprise architecture artifacts.
  - It is not intended to drive the design and implementation of systems.
- Logical and Physical data models provide the details needed to ensure consistency of data structures within applications and to drive data standardization.

Why Does NIH Need an Enterprise CDM?

- An enterprise CDM provides a common vocabulary for the discussion of key enterprise data elements.
  - For example, everyone means the same thing when they say “grant” in the context of data architecture.
- Gives us a way to decompose NIH’s data needs and build more detailed sub-architectures on an incremental basis.
- Provides the basis for other enterprise architecture planning materials.
  - Allows us to map data to business processes:
    - Supports effective allocation of responsibility for managing data and ensuring data quality.
    - Provides a better understanding of data implications of evolving business processes.
  - Allows us to map data to systems providing a better understanding of required data flows between systems and identify opportunities for reuse.

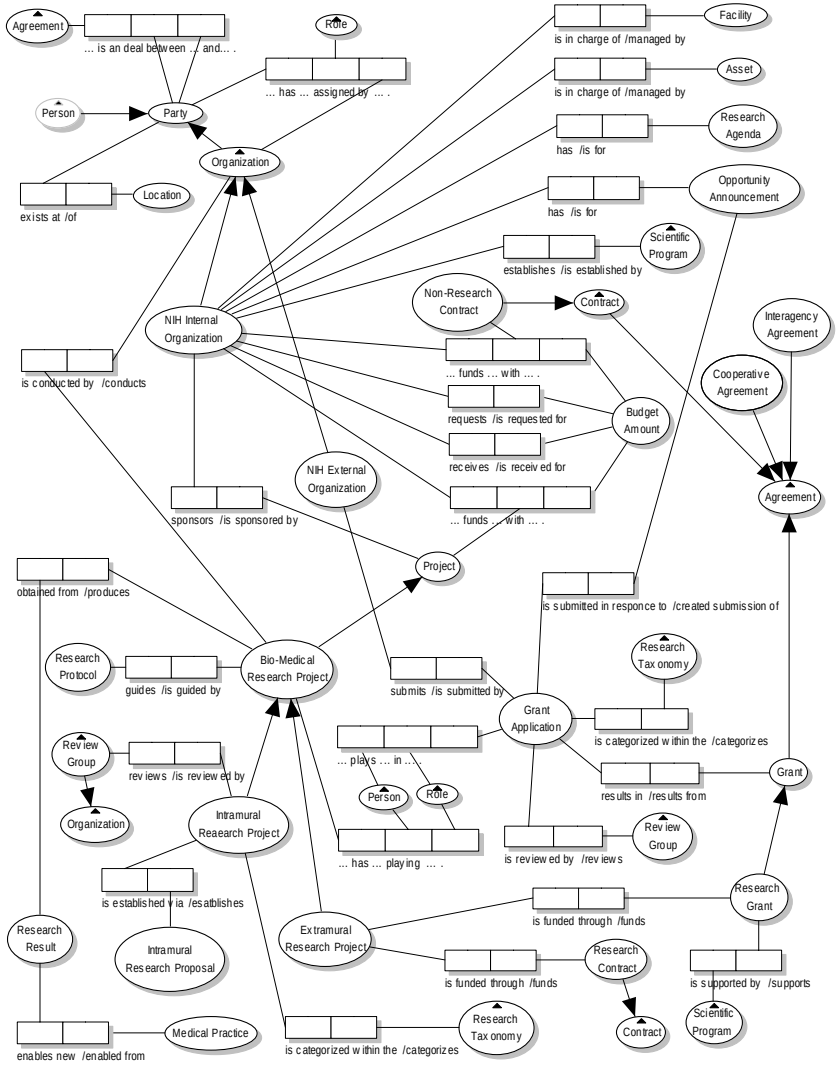
NIH E-CDM Purpose (From the NRFC)

- The purpose of the Enterprise Conceptual Data Model is to provide a conceptual view the key data entities and their relationships that support NIH's mission.
- NIH requires such a model in order to provide an organizing framework for further enterprise data architecture efforts. The Enterprise CDM acts as a high-level taxonomy organizing all of NIH's data assets into groups and rules.
- This high-level presentation allows NIH management and stakeholders to effectively understand the current state of the data architecture and to plan for a future state data architecture that will enhance NIH's ability to share information across the enterprise and build more integrated, flexible systems.

E-CDM Overview: CDM Components

Component	Description
Subject Areas	A topic of interest shared within a community. The full list of subject areas of interest to a community form the context for that community. A super type is a subject area that spans multiple Communities of Interest (COIs); (DRM usage). In NIH's CDM, subject areas are used to group entities and align them with business processes.
Entities	An abstraction for a person, place, object, event, or concept described (or characterized) by common Attributes. For example, "Person" and "Agency" are Entities. An *instance* of an Entity represents one particular occurrence of the Entity, such as a specific person or a specific agency.
Relationships	Describes the association between two Entities. Relationships may also be described as business rules that specify the nature of the interaction between two entities.
Attributes	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 a "Person" Entity may be "Social Security Number (SSN)".

CDM Overview: Graphical View



- See handouts

E-CDM Overview: Subject Areas

Subject Area	Description
Global	Entities that are shared across process areas.
Party	Information about people, organizations, and other actors in NIH processes, and their roles.
Sustaining Services	Information related to NIH supporting processes such as acquisition of assets, property management, telecommunications, information technology, budget formulation, funds tracking, and funds allocation.
Research Administration	Information about the management of grants, research contracts, cooperative agreements and other aspects of NIH funded research. Information about the management and administration of intramural research.
Clinical and Research	Clinical information about the treatment of patients at the clinical center and data resulting from research, both extramural and intramural.

E-CDM Overview: Entity Examples

#	Entity Name	Definition	Subject Area	Super Type	Examples
1	Agreement	Statement of an exchange of promises.	Sustaining Services		
2	Asset	Property, funds, receivables, and other items that positively impact an organization's balance sheet.	Sustaining Services		
3	Biomedical Research Project	A basic and clinical investigation in biology or medicine.	Research Administration		See subtypes – Extramural and Intramural Research Project.
4	Budget Amount	The amount of funds available to an organization and the allocation of those funds across various accounts, programs, offices, initiatives, operations, etc. A budget may be composed of several budgets allocated to sub-organizations.	Sustaining Services		$250,000.00
5	Contract	A binding agreement between NIH and an external party with specified performance requirements.	Sustaining Services	Agreement	
9	Grant	Financial assistance from Public Health Service agency for approved activities. Performance responsibility rests primarily with a grantee with little or no government involvement in the research; term covers grants and cooperative agreements	Research Administration	Agreement	Grant Number: 1R41EY017781-01 Project Title: Recombinant T-cell Receptor Ligands for Treatment of Uveitis
10	Grant Application	Application for financial assistance from a Public Health Service agency.	Research Administration		Application with title "sar-mediated regulation in Staphylococcus aureus" from Univ. of Arkansas for Medical Science

E-CDM Overview: Entity Examples (cont'd)

#	Entity Name	Definition	Subject Area	Super Type	Examples
11	Interagency Agreement	An agreement between two government entities that defines rules for interaction.	Sustaining Services	Agreement	
12	Intramural Research Project	A biomedical research project performed within NIH.	Research Administration	Biomedical Research Project	Intramural Project Number: 1 Z01 EY000396-04 DIR, Project Title: Evaluation And Treatment Protocol For Patients With Retinal Disease
15	Medical Practice	A formal position or recommendation for clinical practice, disease treatment, or disease prevention that results from a scientific body that has concluded a review of an amalgamation of available scientific information.	Research Administration		
16	NIH External Organization	An organization created and managed outside of NIH.	Party	Organization	
17	NIH Internal Organization	An organization created and managed by the NIH regardless of whether it is composed of NIH employees, external people, or a combination.	Party	Organization	NIAID
18	Non-Research Contract	A contract that does not involve the conduct biomedical research.	Sustaining Services		Contract: DHHS-NIH-OD-2004-IW2-1037-T00 Title: DEAS Workflow Management and Monitoring System Contractor: National Micrographics Systems, Inc. Amount: $493,044.00

E-CDM Overview: Relationship Examples

#	Relationship/Rule	Example
1	<Agreement> is a deal between <Party 1> and <Party 2>.	Contract 14998 is a deal between CIT and System Inc.
2	<Biomedical Research Project> has <Person> playing <Role>.	XXX project has Jill Smith playing the role of principle investigator.
3	<Biomedical Research Project> is a type of <Project>.	
4	<Biomedical Research Project> is categorized within the <Research Taxonomy>.	Project xxx is categorized within the lung cancer; disease taxonomy.
5	<Biomedical Research Project> is conducted by <Organization>.	Project xxx is conducted by NCI/xx.
6	<Biomedical Research Project> is guided by <Research Protocol>.	Project xxx is guided by research protocol yyy.
7	<Biomedical Research Project> produces <Research Result>.	Project xxx produces research result of new treatment for urinary tract infections.
8	<Contract> is a type of <Agreement>.	
9	<Cooperative Agreement> is a type of <Agreement>.	
10	<Extramural Research Project> is a type of <Biomedical Research Project>.	
11	<Grant Application> is categorized within the <Research Taxonomy>.	Grant application 'Non-asbestos caused mesothelioma' is categorized within the lung cancer; disease taxonomy.
12	<Grant Application> is reviewed by <Review Group>.	Grant application 'Effects of Cranberry Juice on Urinary Tract Infections' is reviewed by 'NCCAM Initial Review Group.'

Using the Enterprise CDM

- Starting a New Project
  - New projects can use the E-CDM to help define the high-level structure of key data elements. For example, the E-CDM shows a party as the supertype for people and organizations. How might this effect the design of a system.
  - Using the E-CDM as a base, the new project can build a solution-specific model that extends the E-CDM by adding new entities
- EA Alignment and Compliance for Ongoing Projects
  - The E-CDM provides a consistent way to specify data architecture across NIH organizations and solutions.
  - Projects can map their specific data needs to the E-CDM providing a means to ensure that alignment exists with the enterprise model, identify any gaps, and identify opportunities for sharing or reuse.
  - This mapping to the E-CDM provides a means to collect consistent metadata across systems for inclusion in NIH's Enterprise Architecture Repository.

Example – Starting a New Project

- NIH organizations may need to expand on the Enterprise CDM to address the specific needs of a solution or an organization.
- The approach results in a documented mapping between Enterprise CDM entities and the data model for the specific organization or solution. This mapping provides:
  - A management tool to determine dependencies on other systems or projects and sources for required data.
  - A source of input for subsequent iterations of the Enterprise CDM and as a means to promote reuse across NIH solutions.

Example – Starting a New Project

	Activity	Comments
1	Identify relevant Enterprise CDM entities and subject area focus.	Provides a means to understand the solution's requirements in the broader context of NIH's Enterprise Architecture.
2	Identify additional attributes for existing Enterprise CDM entities.	Additional attributes should be identified as part of this process. The primary key, or unique identifier for any entity should align with that specified in the Enterprise CDM. If no unique identifier is specified in the CDM, one may be established within the scope of the solution/organization-specific model.
3	Identify additional entities required.	In almost all cases, specific solutions will require additional entities that are not in the E-CDM. These should be included in the solution-specific conceptual model.
4	Identify additional relationships required.	As with entities, additional relationships are likely to be needed in support of the specific solution implementations.
5	Identify existing database mappings to the E-CDM and additional entities identified in the solution specific model.	Mapping existing databases to a higher-level conceptual model will allow dependencies between databases to be understood and will allow the solution to be placed in the broader context of the Enterprise Data Architecture.
7	Document key data flows between systems in terms of the E-CDM and additional entities identified in the solution-specific model.	Describing the flow of data between systems using a common taxonomy of data entities will allow NIH to analyze dependencies and more effectively develop strategies for growing and extending the system over time.

What's Next

- Additional Data Standards
  - Person Name approved by the ARB on January 24th, 2007
  - Budget Entities approved by the ARB on January 24th, 2007
- Segment Sub-Architectures
  - Progress on "People" data in support of NED II.
  - "Grants" conceptual model is under development based business process modeling efforts.
- Data Governance
  - Change Management processes for enterprise-shared data.
  - Data ownership and stewardship roles and responsibilities.

Questions

Contact Information

- Web site

  http://EnterpriseArchitecture.nih.gov

  Subscribe to updates -
  http://enterprisearchitecture.nih.gov/Tools/EmailUpdates/

- The NIH Enterprise Architecture Community in the NIH Portal

- Email

  EnterpriseArchitecture@mail.nih.gov

- Subscribe to the EA LISTSERV:

  http://list.nih.gov/archives/enterprise_architecture.html