

Metadata Updates – What if We Got it Wrong?

Dawn Michels

April 19, 2016

Agenda

- The NMDP (National Marrow Donor Program)
- Metadata as a focus
- Metadata within data modeling tools
- Techniques for gathering metadata early
- Process steps for making changes to existing metadata
- Costs and benefits for this work
- Tips for justifying to IT and business leadership -



Be The Match® is operated by the **National Marrow Donor Program®**, a nonprofit organization that matches patients with donors, educates health care professionals and conducts research so more lives can be saved.

Domestic and Global Presence

Be The Match® (NMDP®) is the hub of a global transplant network including partner centers throughout the U.S. as well as centers and registries in 41 countries



International Partners



World Leader in Marrow and Cord Blood Transplantation

Transplants

Improving access to transplant and remove barriers to treatment.

Patient Support

One-on-one support to patients and their families.

- Resources
- Guidance before, during and after transplant.

Research and Education

Center International Blood Marrow Transplant Research (CIBMTR), expand treatments and improve outcomes. Information, research and resources for health care professionals.

Engaging the Public

Support of thousands of individuals whose compassion and generosity help save lives.

We grow and manage the world's largest and most diverse donor registry

- 11 million donors and 193,000 cord blood units on the Be The Match Registry,
- 22.5 million potential donors and 601,000 cord blood units worldwide.

Be The Match volunteers also donated 104,000 hours—a value of \$2.5 million.



6,300 patients received a marrow transplant—**61,000** since 1987



\$3.2 million in patient assistance to **1,800 families**

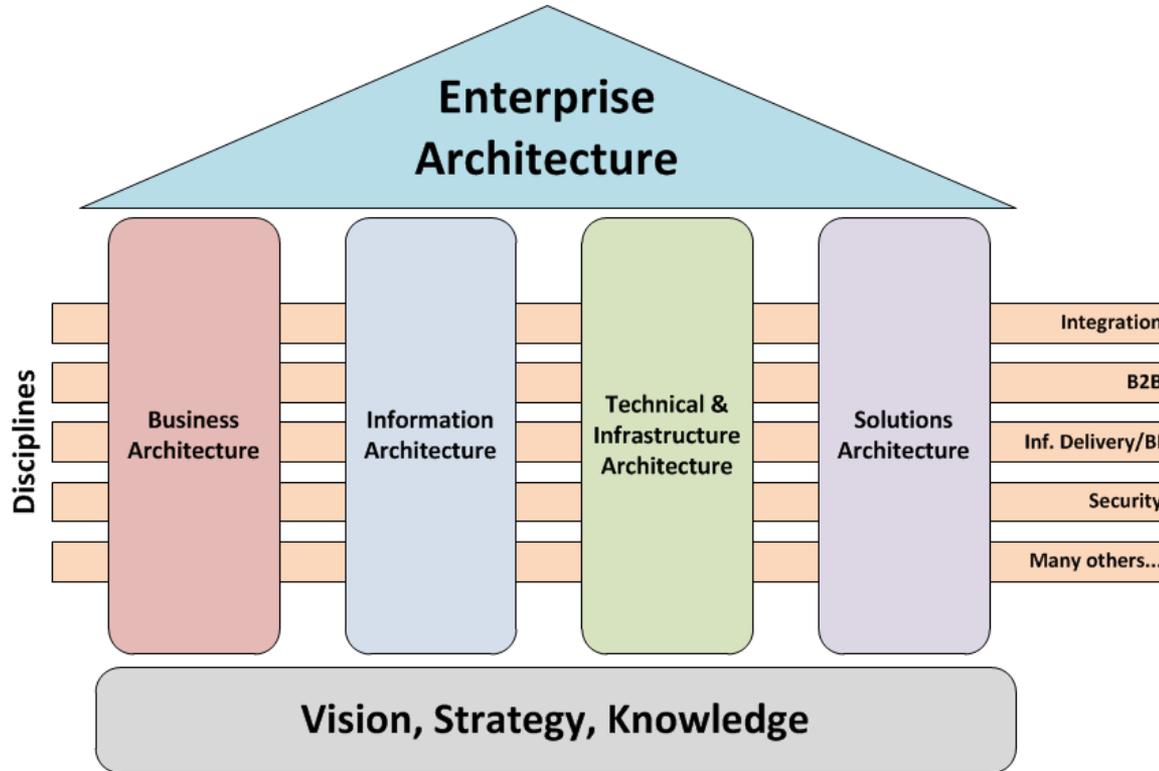


254 research studies under way

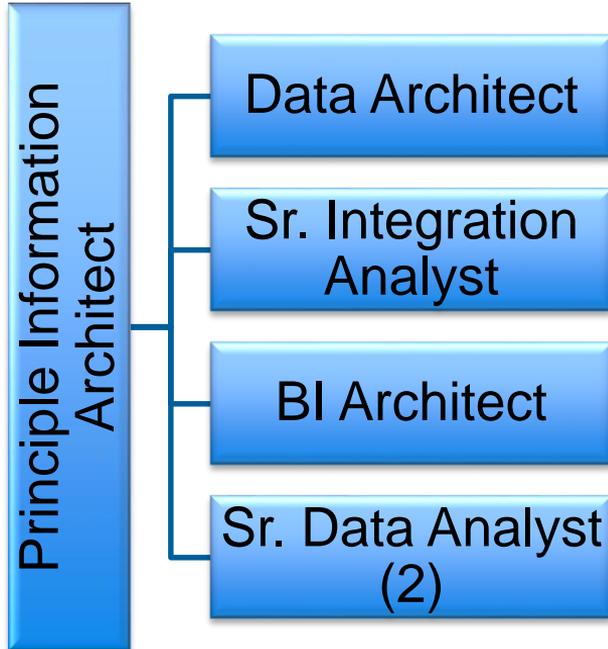
539,000 new potential donors added to Be The Match Registry®, including

44% with diverse ancestry

Enterprise Architecture at NMDP



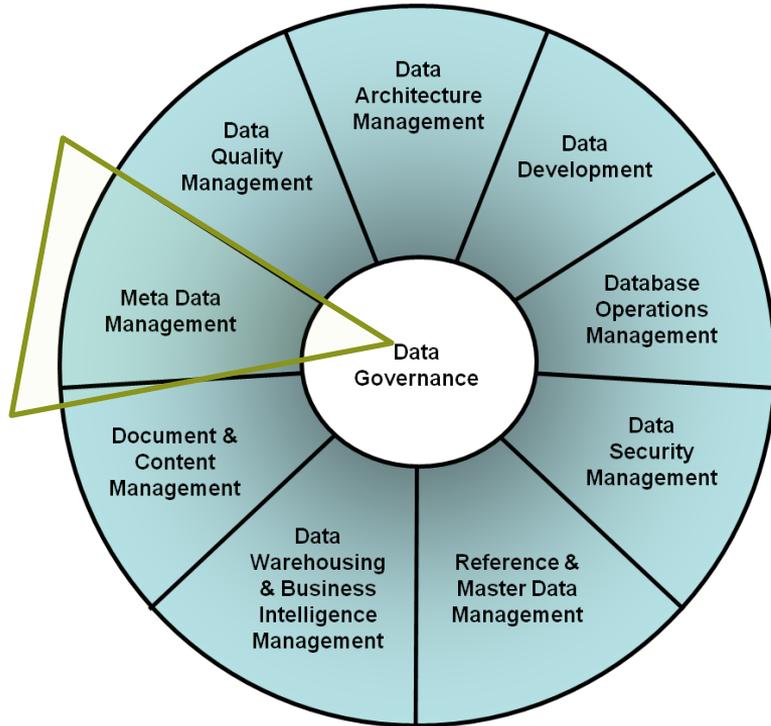
Data Architecture Team



Services

- Data modeling
- Metadata management
- Data analysis ~ 260 apps
- Naming standards
- Best practices
 - Normalization
 - Consistency
 - Minimal Redundancy
- Metadata maintenance for federated architecture

DMBOK© Functions of Data Architecture



- **Data Architecture** - Defining the enterprise data needs and a developing a blueprint or framework for data resources
- **Data Development** - Data modeling & design including- analysis, design, build, test, deployment and maintenance of data models – maps to an enterprise chosen SDLC
- **Database Operations** – Providing support from data acquisition to purging
- **Data Security**– Insuring privacy, confidentiality and appropriate access
- **Reference & Master Data** – identify and manage gold versions of contextual data values to share across the org
- **Data Warehousing and Business Intelligence** – Enabling reporting and analysis
- **Documents and Content Mgmt** – Managing data found outside of databases
- **Meta-data** - integrate, control and deliver meta-data
- **Data Quality**- define, monitor and improve data quality
- **Data Governance** – Planning , supervision and control over Data Management and use

Metadata as a Focus

- What is Metadata?
 - Per NISO (National Information Standards Organization) Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.
 - Three Types
 - Descriptive
 - Structural
 - Administrative
- “Metadata is the information and documentation which makes data understandable and shareable for users over time. Data remain useable, shareable, and understandable as long as the metadata remain accessible.”
(ISO/IEC 11179-1)
- Of high interest recently is the metadata of photography, images and unstructured content – this talk is focusing on the business metadata of structured content – not unstructured content

Metadata as an valuable Asset

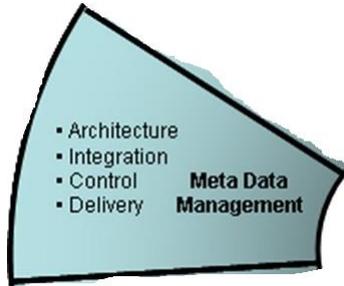
- Metadata provides a common business vocabulary with context for the data
- Metadata can give insight into the flow of the data through your organization
- Metadata can provide context for audit trails and for root cause analysis when problems arise.
- Metadata can help provide the producers and the consumer of the information context for root cause analysis.
- Metadata can aide in the assessment of the scope and impact of changes to systems and for understanding the environment before a change is made

Getty Institute Catalog of Metadata Types

3.3 Getty's definitions on types of metadata

Type	Definition	Examples
Administrative	Metadata used in managing and administering collections and information resources	<ul style="list-style-type: none"> • Acquisition information • Rights and reproduction tracking • Documentation of legal access requirements • Location information • Selection criteria for digitization
Descriptive	Metadata used to identify and describe collections and related information resources	<ul style="list-style-type: none"> • Cataloging records • Finding aids • Differentiations between versions • Specialized indexes • Curatorial information • Hyperlinked relationships between resources • Annotations by creators and users
Preservation	Metadata related to the preservation management of collections and information resources	<ul style="list-style-type: none"> • Documentation of physical condition of resources • Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration • Documentation of any changes occurring during digitization or preservation
Technical	Metadata related to how a system functions or metadata behaves	<ul style="list-style-type: none"> • Hardware and software documentation • Technical digitization information, e.g., formats, compression ratios, scaling routines • Tracking of system response times • Authentication and security data, e.g., encryption keys, passwords
Use	Metadata related to the level and type of use of collections and information resources	<ul style="list-style-type: none"> • Circulation records • Physical and digital exhibition records • Use and user tracking • Content reuse and multiversioning information • Search logs • Rights metadata

Meta Data Management our assessment



Our Maturity Rating

Current – 1-3

Future – 4

- Goals we wanted to include
 - Integrated from different systems
 - Tools metadata is used
- Success for us Looks Like
 - More integration
 - Better reporting capabilities
 - Defined process is in place to update metadata
 - Metadata definitions are stored and maintained in a controlled method

Meta Data - Governance



- Metadata helps catalog and identify enterprise data assets that require or would benefit from governance for the sake of compliance, security, privacy, system performance, and partner activities.
- Identifying accountability is fundamental – system of record – aka source of truth – aka authoritative source
- Identifies who gets to “approve” meta-data changes
- Clarifies decision making

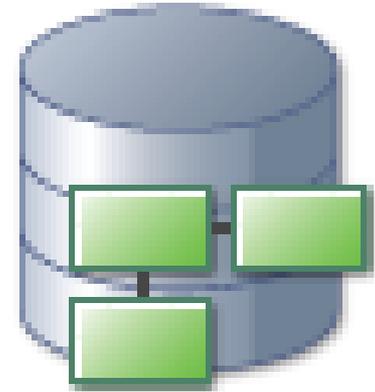
Techniques for gathering Meta Data

- Business Requirements – culling through requirements and trying to decipher data meaning
- User Interviews – ask the users to talk about their work and data and the way they understand the information
- Story Boarding – drawing out a day in the life of... business users
- Reviewing raw data – see what it really is



Metadata within Data Modeling Tools

- Our tool of choice is Power Designer
- What is your tool of choice?
 - Erwin?
 - Embarcadero?
 - Visio?
 - Excel?
- What about the structural (aka technical Metadata) found in DBMS or applications?



What do we do with the Meta Data?



- We include them in any excel reports or mapping documents and gain user buy-in during design
- As part of implementing our databases (Oracle especially – we place the definitions in the Comment field of DBMS)
- When possible we leverage those attributes in downstream reporting tools
- Our developers and QA testers use it to understand the data they are working with

Using the PowerDesigner Portal



PowerDesigner

Workspace

In-progress, review-ready and recent diagrams.

0

Diagrams



Repository

Published models, diagrams, and objects.



Dashboard

Charts visualizing model data.



Administration

Users, groups, settings, and extensions



Using the PowerDesigner Portal diagrams published and not

Folder

Modification Date: 12/27/2013 12:12 by lgittsov

Diagrams 35 Children Permissions

Diagrams

- All Tables - BI_CDC - v10
- All Tables - v108
- Dim - Business Party Snowfl...
- Star - Patient Case - v21
- Star - Order Management - v14
- Star - HLA Typing - v35
- Star - Product Infusion - v20
- Star - Sample Request - v1

Repository Root >

- ENTERPRISE INFORMATION SER... X
- BI Information Delivery > X
- Physical Data Models > X
- DataMarts

Models:

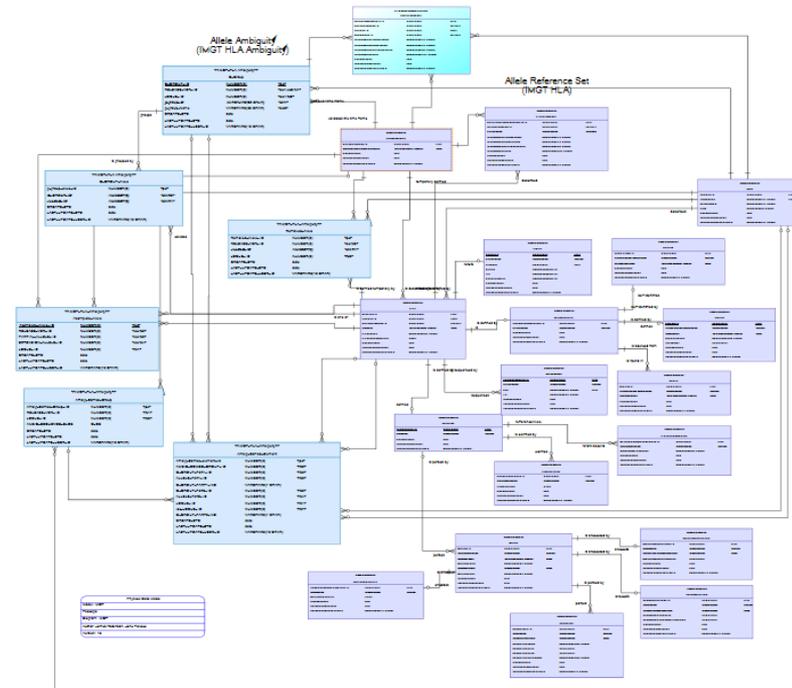
- BI_CDC - v22
- BI_EDW - v207
- BI_LANDING - v212
- BI_Staging - v46
- DAR_STG - v9
- ETL_TRACK - v9

Using the PowerDesigner Portal (cont)

PowerDesigner

(42) Facebook HML Product Plan Creatio... Enterprise Architecture - ... Application Catalog IT at NMDP - home Web Slice Gallery

Home ← IMGT (Read) ⏪ ⏩ Table



25%

Info Children Depends On Imp...

[Columns \(5\)](#) [Keys \(2\)](#)
[Permissions \(2\)](#)

Columns (5)

Name	Code
RELEASE_VE R_IID	RELEASE_V...
release_Ver_C urrent_Release	RELEASE_V...

Using the PowerDesigner Portal (cont)

The screenshot shows the SAP PowerDesigner Portal interface. The browser address bar displays the URL: `http://www.sap.com/PowerDesigner/portal/`. The page title is "IMGT (Read)". The main content area displays a complex database diagram with various tables and relationships. A tooltip is visible over one of the tables, providing details for the `release_Ver_Current_Release` table.

release_Ver_Current_Release
The particular version of the IMGT database for the given allele. The database is updated every 3 months.

Data...	VARCHAR2(10 CHAR)
Mand...	false
Modif...	10/10/2011 15:09 by jpollack

Below the tooltip, there are four icons representing different views: Info, Children, Depends On, and Impacts. The "Info" icon is selected, and the "No contents" message is displayed below it.

Using the PowerDesigner Portal (cont)

PowerDesigner

(42) Facebook HML Product Plan Creatio... Enterprise Architecture - ... Application Catalog IT at NMDP - home Web Slice Gallery

Home ← IMG T (Read) << >> ← Column

25%

Comments(0) X

Add a comment...

Filtered by release_Ver_Current_Release

No comments

Info Children 0 Depends On 0 Impacts

No contents

The image shows a screenshot of the PowerDesigner Portal interface. The main area displays a UML diagram with various entities and relationships. A comment box is overlaid on the right side of the diagram, showing a search filter 'Filtered by release_Ver_Current_Release' and the text 'No comments'. The interface includes navigation icons, a search bar, and a sidebar with icons for 'Info', 'Children', 'Depends On', and 'Impacts'. A green circle highlights the comment box and the sidebar icons.

Our planned process steps for getting metadata changes implemented * (Happy Path)

- Either notice ourselves, or receive feedback from users of data models
- Identify who the business owner or key stakeholder is of the data (entity or diagram level)
- Review proposed change with business owner,
- Discuss implementation timing
- Update data model (document in model notes – changes made)
- Check model in (accept changes – will render new version of diagram)
- Produce DDL (data definition language – ie script for change)
- Submit change requests – first in lowest environment – will require stepping up the stack
- Based on timing decision above decide when to actually execute changes
- **Communicate broadly**

Costs & Benefits of making these changes

Costs

- Time to change in data models (data architect time)
- Resources to change in physical databases (DBAs)
- Resources to verify (QA test changes to database)
- Update any isolated data documentation (spreadsheets, reports etc) BI staff and Users

Benefits

- Information is more accurate
- Building consensus with business
- If classification is in place it can help users handle data correctly
- People new to the business can find current correct information

Tips for justifying to Leadership

- Metadata enables faster response to change
- Data sharing across the organization
- Improves reporting capabilities – carrying it forward in reporting tools or viewers – the metadata is accurate
- Demonstrates collaboration with business and commitment to their concerns
- Improved demonstration of compliance and reduced risk (think data classification)

Tips for justifying to Leadership (cont)

- $((h_1 - h_2)d)ps$ = Money saved (in payroll dollars) via metadata management
- **Where:**
 - h1** = Hours spent by technical staff collecting and documenting source metadata
 - h2** = Hours spent extracting the same information from a metadata management solution (An assumption is that h2 will be a small fraction of h1)
 - d** = Average number of developers on a project
 - p** = Average hourly pay of developers, in dollars or other currency
 - s** = Number of applications or data management solutions that draw from the metadata repository

TDWI surveys have shown that three full-time workers per data management project is fairly common. Assuming that each worker is paid \$100,000 per year and works 40 hours a week for 52 weeks a year, the average hourly pay per worker is approximately \$48. Finally, let's say that the repository is used and reused in 10 projects per year. All totaled, the savings in developer time (expressed in payroll dollars) amounts to \$103,680 per year

Key Take Aways

- Metadata is an important asset of any organization
- Accuracy is an important but often overlooked aspect
- Managing metadata will save you time, money and confusion
- Collaboration with business makes the information more holistic and likely to stand up to scrutiny and audit
- Developing a process can take time, especially if it is an “other duties as assigned” proposition
- Be confident explaining what the changes are and why they benefit the company.

Resources

- http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html
- <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- http://www.damachicago.org/wp-content/uploads/2012/08/TDWI_ChecklistReport_CostJustificationforMetadataManagement.pdf
- Formula is based on work by Stephen Putnam, senior consultant at Baseline Consulting and a recognized expert in metadata management. Read Steve's blog posts at www.basewlineconsulting.com.
- Dublin core: [dc_presentation_corp_circle_rogers.ppt](#)
- https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/white-paper/metadata-management-data-governance_white-paper_2163.pdf

About Dawn Michels

Connect or Questions

- dmichels@nmdp.org
- michels.dawn@gmail.com
- Cell: 651-246-6913

Qualifications

- CDMP – 2013
- President DAMA MN – 12 years intermittent
- Adjunct Faculty, University of St. Thomas - 2008-2009
- Adjunct Faculty, University of St. Catherine - 1990-2010
- VP Chapter Services DAMA-I - 2000-2002
- Current Pres NMDP Toastmasters
- ACM Bronze

