THE PUBLIC LIBRARY of Cincinnati and Hamilton County

Catalog
Kidspace
Teenspace
MakerSpace
Fifty · Five +
Digital Library
My Librarian
The Library Foundation
The Friends
Anderson Township Library Association
Volunteers

2013 NATIONAL MEDAL FOR MUSEUM & LIBRARY SERVICE

LIBRARY JOURNAL FIVE-STAR LIBRARY

THE PUBLIC LIBRARY of Cincinnati & Hamilton County

Connecting people with the world of information and ideas.

Data Architecture Research Paper 2: Data Architecture
Prepared by Anne Sawyer
28 April 2019

Public Library of Cincinnati and Hamilton County Data Architecture Project Charter

Introduction

The importance of having an integrated, systemwide level of understanding for data and the systems that use it cannot be understated, particularly for an organization like the Library where providing high-quality information resources is at the heart of our Mission, Vision, and Values. The Library's recently updated Strategic Plan states that we will offer customizable service options, understand how and when customers use our services, introduce and utilize new technologies, and adapt to the changing technology needs of our customers. These tenets call for us to re-envision how the Library organizes, stores, and governs data to meet its obligations to customers and the community. Particularly, as we continue to shift resources from outdated legacy systems such as our old Microsoft Exchange servers to more modern Cloud-based enterprise solutions, it is more important than ever to establish a robust, flexible, extensible, and understandable enterprise data model. Information and data architecture specialist Andy Graham (2012) explains that, "an Enterprise Data Model promotes commonality, reduces risk, and improves the quality of data" (p. 15). Such a model will thus support the Library's Strategic plan while enabling us to better meet current and future customer needs.

PROJECT OUTPUTS

- Governance Council
- Business Glossary
- Conceptual Data Model
- Logical Data Model

KEY CONCEPTS

- Enterprise Architecture (EA) - "is explicitly describing an organization through a set of independent, non-redundant artifacts, defining how these artifacts interrelate with each other, and developing a set of prioritized, aligned initiatives and road maps to

understand the organization, communicate this understanding to stakeholders, and move the organization forward to its desired state" (EACOE, n.d.).

- Operating Model - "is the combination of roles, skills, structures, processes, assets and technologies that allow an organisation to deliver on its strategy" (Bevington Group, n.d.).
- Data Architecture - "is composed of models, policies, rules, or standards that govern which data is collected, and how it is stored, arranged, integrated, and put to use in data systems and in organizations" (Marcinko, 2019, p. 8).
- Enterprise Data Model - "differs from an individual project-based data model on two accounts - firstly, it is at a higher level of abstraction and secondly, it has a wider scope of coverage. It provides an integrated view of the data across the entire organization" (Graham, 2012, p. 14).

Goals & Objectives

PROJECT GOALS

1. Establish and promote shared understanding of organizational data between and among Library Services units and IT.
2. Improve data quality, integrity, and value for enhanced decision-making, particularly with regard to applications integration and updating.
3. Provide ongoing support for the people, processes, and technology designated for creating, retrieving, updating, and/or destroying data.

PROJECT OBJECTIVES

1. Within 1 month, form a data governance council to establish and enforce data quality standards.
2. Within 2 months, establish a Business Glossary and naming conventions for data entities.
3. Within 3 months, construct conceptual and logical data models that provide an integrated view of PLCH as an information organization.

OUTCOMES FOR THE LIBRARY

- Realize long-term cost savings by updating isolated legacy systems
- Increase efficiency of workflows and internal processes
- Respond more quickly to opportunities for improvement
- Improve our current business intelligence and analytics capabilities
- Reduce risk with more robust security and disaster recovery management

OUTCOMES FOR THE COMMUNITY

- Provide more in-demand and innovative services for customers
- Improve customer service with updated systems that can integrate data more easily
- Better stewardship of public funds
- Collections that are more usable and easily accessible
- Technology resources that are more sustainable, flexible, and customizable

Scope & Timeline

SCOPE

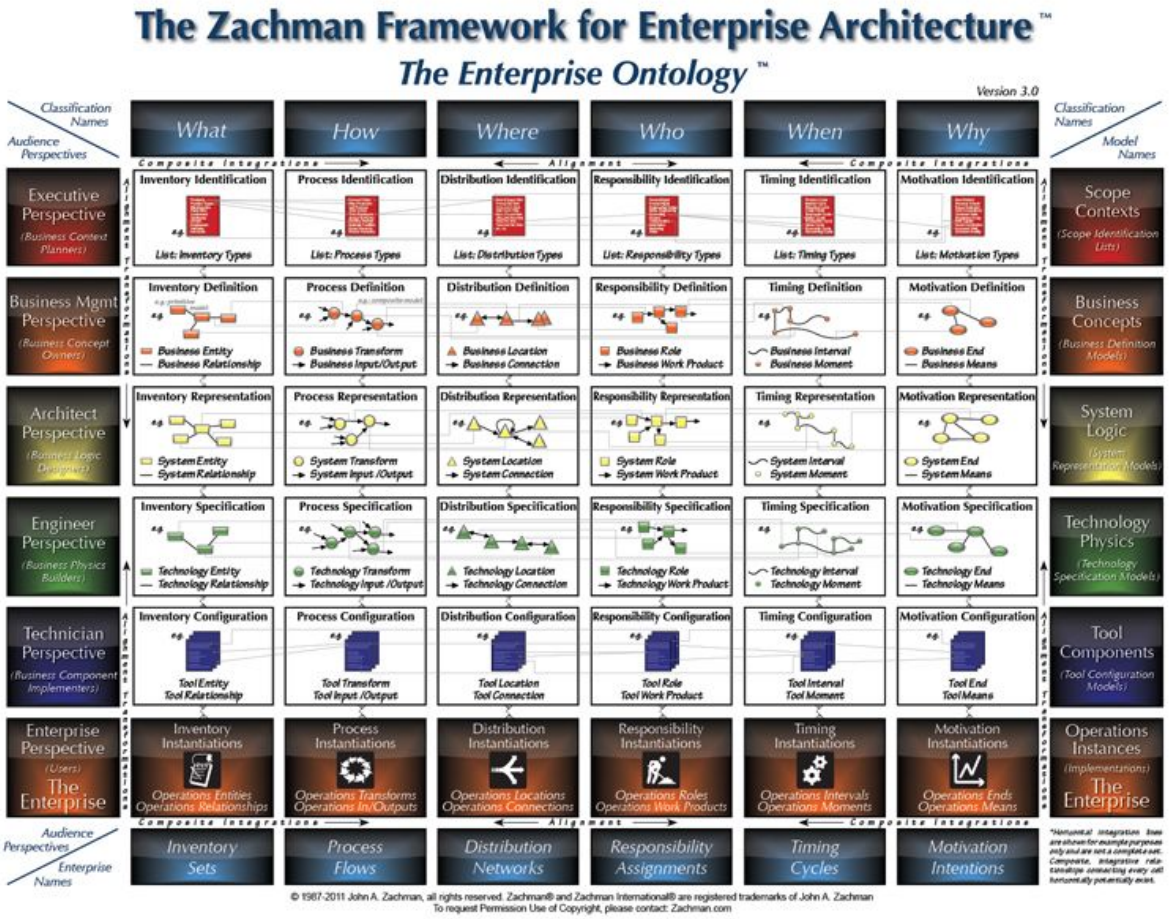
The PLCH Data Architecture Project will operate at a systemwide level with a focus on understanding the information needs of business-critical, value-added data for transactional and analytical processes. Due to constraints of time and budget, initial efforts will prioritize the information and processes that directly impact customers, such as collection development, facilities, programming, and library services, over internal or support services. Among the Project's many activities will be analyzing data needs and lifecycles, current data storage, disaster recovery, and metadata management to produce design outputs that improve efficiency, agility, cost reduction, and risk mitigation drivers. Because this project will rely in large part on extraction, documentation, and mapping of tacit knowledge from PLCH employees, a high level of interdepartmental collaboration and adoption from staff is critical for successful implementation of design deliverables. This includes the PLCH Data Governance Council, Business Glossary, and integrated Data Models. The Physical Data Model and Data Dictionary, which define information at the most fine-grained and technical level for database structures, are outside the scope of this project and will be implemented in a separate IT-led initiative.

TIMELINE

Week	Activity	Justification
1-3	Form data governance team: establish core members from IT and business strategy leaders. Recruit data stewards from among Library Services staff.	Governance council should be in place first to ensure Data Architecture project is implemented in a structured and stepwise manner. Recruitment via the Staff Blog will help educate and promote the Project to staff systemwide.
4-5	Requirements gathering: gather input about business-critical concepts, processes, and roles from staff at all levels via Staff Blog and location visits. Inventory current systems that use or store data (CRUD analysis).	Foments understanding of systemwide information needs and current data lifecycle. Generates top-down and bottom-up adoption of BG initiative.
5-6	Start building the Business Glossary: identify and define business-critical data entities, attributes, relationships, and metadata. Iterate as necessary.	Supports shared understanding of key concepts for front-facing and support areas of the Library.
7-8	Finalize Business Glossary and publish on Intranet. Updates will be made by data stewards on an ongoing basis.	Promotes the BG for knowledge sharing and consistent use of standard business language among staff.
9-10	Construct Conceptual Data Model: use the BG to identify and map major entities and processes and how they relate to each other.	Provides the basis for the Logical Data Model to follow.
11-13	Construct Logical Data Model: add notation, formalized relationships and attributes. Use business rules to decide granularity for dimensional modeling (OLAP functions) and normalized modeling for OLTP functions.	Provides the basis for the Physical Data Model to follow (future IT project).

Operating Model & Architectures

The Library is a complex, multi-tiered system comprised of many people, processes, and data objects that rely on interaction with a variety of business rules, tasks, and technologies to meet short and long-term business obligations and goals. Just as you need architectural blueprints to build a static physical structure like a new branch library, so too organizations like PLCH require a unified, strategic plan for leveraging data to move forward in operational maturit. “A data strategy ensures that all data initiatives follow a common method and structure that is repeatable. This uniformity enables efficient communication throughout the enterprise for rationalizing and defining all solution designs that leverage data in some manner” (Gurevich & Dey, 2018, p. 1). As a continuously-evolving, dynamic organization, PLCH must be able to align its business and IT resources to meet future needs, which is part of the value that a PLCH Data Architecture will bring to the table.



The Zachman classification scheme reduces complexity by providing context at a glance for architecture design activities and deliverables. Here, data architecture aligns with the Architect Perspective (Row 3).

The Library's business operating model is a hybrid Unified Coordination, with business functions that are highly integrated and relatively standardized, which will inform how the data architecture is planned and implemented. "Defining an operating model and a cadence of checkpoints for the business and IT to stay informed and engaged is a powerful governance approach to making a data strategy effective. Every major transformation — for example, modernizing a data warehouse — will need both a roadmap plan and an operating model before it can get started" (Gurevich & Dey, 2018, p. 8). Take for example one of our key functions, which is to loan materials to customers: Even though each branch library may vary in terms of specific items in their collections and associated customers, all locations follow the same circulation policies and rely on shared, centralized access to the ILS, which provides systemwide transactional data for cardholders and catalog items. This is the Unified part of our operating model. The Coordinated part refers to the level of autonomy at the business unit level: even though everyone follows the same broad policies, branch managers have the authority to direct specific procedures, services, and programs offered at their locations. The Library's enterprise data model must align with this model, providing enough dimensionality to support coordinated individual business needs while also reducing complexity to provide for unified processing needs at scale.

Business Glossary

Gartner, Inc., a long-time thought leader in EA, defines a Business Glossary as a business-led effort that forms "the foundation for all things 'semantic,' which concerns the different meanings of words" (2015). The PLCH Business Glossary will support consistent use of business language, clarify semantic associations between key concepts, and define all the metadata needed by various agencies to find organizational information. This is particularly important for expediency in customer service, as staff currently tend to rely more on experienced coworkers for important policy and procedural information rather than trying (and failing) to find it on the Intranet. "Even more mind-boggling is the amount of time that is wasted. The average information worker spends 3.5 hours each week searching for but not finding the information they were looking for. For a company of 1,000 employees, that is a staggering US$5,250,700 per year!" (IBM, 2008, p. 5).

Semantic associations mapped out by staff will form the backbone for the enterprise taxonomies and metadata architectures, master data management (MDM), and the Logical Data Model. As library employees, we're all aware of the importance of metadata for finding items in our collections, but we don't often think about how to apply this same semantic structure to other business items, such as customers, locations, resources, staff, and individual processes. "By recognizing the corporate investment in all of the disassociated metadata modeling activities and resolving to invest in and develop a core, corporate taxonomy, the complexities that threaten to overwhelm systems and threaten the effectiveness of business processes are managed and risk is reduced" (Roszkiewicz, 2010, p. 294). Developing an authoritative source of truth for Library Services definitions and associations will ensure that everyone is on the same page when delivering services, conducting programs, and communicating with each other and our customers.

Since it is primarily a business deliverable, the BG will be a coordinated effort led by Library Services leaders with support from IT. Once executive approval is in place, the next challenge for developing the BG will be getting staff on board and contributing to its success. This may be accomplished in a similar way to how we promoted and sustained interest in the Facilities Master Plan: regular blog posts that pose questions and encourage discussion in addition to top-down information dissemination at agency meetings and morale-boosting contests and low-cost swag such as buttons. Once we have enough information and ideas to move forward with constructing the BG, it will be published on the Intranet so that all staff can refer to it as needed and also suggest changes and additions over time. Data Stewards will be in charge of making these changes, once approved by the Governance Council.

Data Models

With a working BG in place, the Project Team can move onto analyzing and designing data structures. Data modeling involves identifying information entities (such as Cardholders), defining their attributes (Name, Address, Birthdate, etc.), and relating these to other entities (Cardholder → borrows→ Laptop). The ability of Entity-Relationship models to distill complex business functions into an elegant visual representation that can be understood at a glance helps business and IT users understand how data flows through various people, places, and

processes. This understanding will only become more important as the Library moves forward on strategic initiatives like the Facilities Master Plan, which will impact branch data needs as new technologies and services are adopted.

The Conceptual Model, which focuses on abstract, high-level mapping of key enterprise data, will be facilitated and scoped by the Data Architect, with input from Library Services team members. Since this model is more broadly focused on business information, IT time will not be best used here. Even though it does not describe specific technologies, the Logical Data Model will be developed more closely with contributions from IT, as finer details are added and cardinality and notation systems are applied. More technical considerations will need to be made at the logical level, such as dimensional modeling for analytical systems that require high query performance (e.g., using BI data from Savannah to create branch business plans). Other modeling needs may require normalization (for high volume of queries) and abstraction (less detail for broader application), particularly for the SQL-based relational databases necessary for daily ILS transactions. These models will not be a one-off effort, but rather will require re-tooling and iteration as new relationships are discovered and added, based on staff input.

Implementation & Governance

Governing PLCH organizational data will be an ongoing assurance effort to define and enforce rules for how PLCH data moves through its lifecycle. The DGI Data Governance Framework defines data governance as "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" (Thomas, n.d., p. 3). The PLCH Governance Council will define this system, along with its goals, roadmap, and KPIs for enabling Library leaders to improve decision-making processes and Library staff to use data more effectively and responsibly. "Given that it is a widely-held view that the amount of data in organisations will expand a hundredfold over the next five years, companies must increasingly depend on and develop a coherent and cost-effective data quality strategy" (Graham, 2012, p. 33). The Council should meet regularly (at least once per quarter) to set up the necessary governance structures and standards, refine

the rules and processes needed for effective data quality management, and produce reports that measure successes and areas for improvement.

The Council will be an interdepartmental body comprised of various leadership roles, including:

- Data Architect - will lead the production of design deliverables, helping to understand data structures and define data models, using business requirements as input to ensure accuracy of PLCH data.
- CTLO - will be the primary liaison between the Council and Library Administration and the Board of Trustees; will guide the direction of architecture efforts, ensuring that they align with the Library's Strategic Plan.
- Technology strategist - will be the primary liaison between business and IT for data integration activities, and take an active role in developing and maintaining the Logical Data Model.
- Network Administrator - will help analyze data lifecycles, contribute to the Logical Data Model, and enforce data security, privacy, and compliance standards.
- Communications strategist - will lead formal communications efforts to ensure transparency and facilitate broader staff understanding and adoption of enterprise architecure initiatives.
- ILS Manager - will help develop metadata standards and related architectures, define operational process standards, and negotiate with ILS vendors to make changes that support transactional data requirements.
- Data Stewards - will enforce data quality standards, maintain the BG, and serve as liaisons between Library Services staff and the Council, ensuring that the data architecture continues to balance the needs of daily business operations and broader strategic objectives.

References

Bevington Group. (2018). *Operating model design*. Retrieved from: https://www.bevingtongroup.com/services/operating-model-design/

Enterprise Architecture Center of Excellence. (n.d.). *Enterprise architecture defined from an information & technology perspective*. Retrieved from: https://www.eacoe.org/ea-definition

Graham, A. (2012). *The enterprise data model: A framework for enterprise data architecture*. (2nd ed.). Koios Associates Ltd.

Gurevich, A., & Dey, S. (2018). Defining a data strategy. *DXC Technology*. [PDF document]. Retrieved from: https://assets1.dxc.technology/analytics/downloads/MD_7257a-18_Data_Strategy_White_Paper.pdf

IBM. (2008). *The business value of a business glossary*. [PDF document]. Retrieved from Blackboard Learn: https://learn.kent.edu/bbcswebdav/pid-9531442-dt-content-rid-127724241_1/courses/12723.201910/Week%2002/RequiredReading/The-business-value-of-a-business-glossary.pdf

Marcinko, M. (2019). *IDERA slides: data architecture – the foundation for enterprise architecture and governance*. [Lecture slides]. Retrieved from: https://www.dataversity.net/idera-slides-data-architecture-the-foundation-for-enterprise-architecture-and-governance/#

NASCIO. (2003). *Enterprise architecture maturity model*. [PDF document]. Retrieved from: https://www.nascio.org/Portals/0/Publications/Documents/2003/NASCIO-EAMM.pdf

Roszkiewicz, R. (2010). Enterprise metadata management: how consolidation simplifies control. *Journal of Digital Assets Management*. [Vol. 6, 5, pp 291-297]. Retrieved from Blackboard Learn: https://learn.kent.edu/bbcswebdav/pid-9585573-dt-content-rid-128700492_1/courses/12723.201910/Week%2009/Required%20Readings/EnterpriseMetadataManagement.pdf

Thomas, G. (n.d.). The DGI data governance framework. *The Data Governance Institute*. [PDF document]. Retrieved from Blackboard Learn: https://learn.kent.edu/bbcswebdav/pid-9585661-dt-content-rid-128701382_1/courses/12723.201910/Week%2010/Required%20Readings/dgi_framework.pdf

Zachman, J. (1996). The framework for enterprise architecture: background, description and utility. *Zachman International*. Retrieved from: https://www.zachman.com/resources/ea-articles-reference/327-the-framework-for-enterprise-architecture-background-description-and-utility-by-john-a-zachman

Zachman, J. (2011). *The Zachman Framework for Enterprise Architecture*. [JPEG image]. Retrieved from: https://www.zachman.com/images/ZI_PIcs/ZF3.0.jpg