Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

_____

Signature of Student Date

An Evaluation of The Brain Health Center Data Mart: Integrating Clinical and Research Data

By
Noy S Hawkins
Degree to be awarded: MPH
Executive MPH

_____
Susan Hobson, MPH Date
IT Product Manager, Data Warehouse and BI, Committee Chair

_____
Jeff Weaver, MBA Date
Director, IT Data Solutions, Committee Member

_____
Shirley Mullins, MIT Date
Informatics Analyst III, Library IT Services, Committee Member

_____
Dr. Laurie Gaydos, PhD Date
Associate Chair for Academic Affairs, Executive MPH

Table of Contents

An Evaluation of The Brain Health Center Data Mart: Integrating Clinical and Research Data

By Noy Hawkins
The Rollins School of Public Health; Applied Informatics
Emory University

ABSTRACT

Neurocognitive disorders pertaining to Alzheimer's dementia are recognized as one of the most important medical and social problems in the aging population world-wide. In 2013 Alzheimer's disease was the sixth leading cause of death in the United States and the fifth leading cause of death in Americans age ≥ 65 years. It is projected that by 2050, the number of people living with Alzheimer's disease in the United States will grow to 13.8 million which equates to nearly 1 million new cases per year (1). The need for an innovative system that incorporates clinical and health research data is imperative.

The Emory Alzheimer's Disease Research Center (ADRC) has a mission to improve the lives of individuals affected by Alzheimer's and other related diseases through innovative research, education, and compassionate care (2). The ADRC, in partnership with Emory's Data Solutions Team, seeks to bring research advances into the clinic and community that will reduce the burden of Alzheimer's and related diseases through early detection and effective intervention. The implementation of a system aimed to make research more relevant and move research into practice is essential for

this population. The ADRC and Data Solutions team sought to implement such a system with the Brain Health Center Data Mart. This data mart was created to provide a consistent and efficient way of identifying patient health information (clinical and research), abstracting non-discrete information, and tracking specimens. With this information available, investigators can conduct research and integrate the best research evidence into the decision-making process for patient care. A process evaluation of the Brain Health Center Data Mart was conducted to assess implementation strategies.

Five key stakeholders participated in qualitative interviews about scope management, time management, risk management, stakeholder involvement, and other lessons learned. During the interview, additional conversations pertaining to usefulness, satisfaction, and ease of use of the reporting and visualization capabilities generated by the system were elicited. System architecture and interoperability, data quality, and stakeholder engagement emerged as dominant themes reflecting the challenges experienced in the implementation process. Participants attributed the misalignment of project objectives to limited stakeholder engagement. The planning, design, and implementation of this solution as a response to past data silos provided new insight into the relationship between the system architecture and interoperability, data quality, and stakeholder engagement. This evaluation of the Brain Health Center Data Mart assesses the gaps in implementation utilizing Zachman's Framework and The Open Group Architecture Framework to provide new insight into the alignment between IT and the business.

INTRODUCTION

In 2013 Alzheimer's disease was the sixth leading cause of death in the United States and the fifth leading cause of death in Americans age ≥ 65 years with an estimated 5.4 million diagnosed Americans and contributing to an estimated 700,000 deaths. It is projected that by 2050, the number of people living with Alzheimer's disease in the United States will to grow to 13.8 million which is equivalent to one new case of Alzheimer's every 33 seconds or nearly 1 million new cases per year. What makes Alzheimer's challenging is that the actual number of deaths from the disease is much larger than the number of deaths from Alzheimer's disease recorded on death certificates. This discrepancy is because many do not die because of Alzheimer's disease but die because of complications caused by the disease (1). According to the World Health Organization (WHO), between 2000 and 2012, Alzheimer's has increasingly impacted the global population in the following ways (3):

- The number of disability-adjusted life years (DALYs) rose by 65%,
- Years of life lost (YLL) doubled from 4.2m to 8.3m, and
- The number of years lived with a disability (YLDs) increased by 40%.

The effort to reduce the morbidity and maintain or enhance the quality of life of those suffering from Alzheimer's Disease aligns with the Healthy People 2020 (HP2020) Objectives for Dementias to (4):

- Increase the proportion of adults aged 65 years and older with diagnosed Alzheimer's disease and other dementias, or their caregiver, who are aware of the diagnosis.
- Reduce the proportion of preventable hospitalizations in adults aged 65 years and older with diagnosed Alzheimer's disease and other dementias.

The Brain Health Center Data Mart was created in support of the mission and vision of The Emory's Alzheimer's Disease Research Center (ADRC) (2).

> **Mission**: To improve the lives of individuals affected by Alzheimer's and related diseases through innovative research, education, and compassionate care.
>
> **Vision**: We will bring research advances into the clinic and community that will reduce the burden of Alzheimer's and related diseases through early detection and effective intervention.

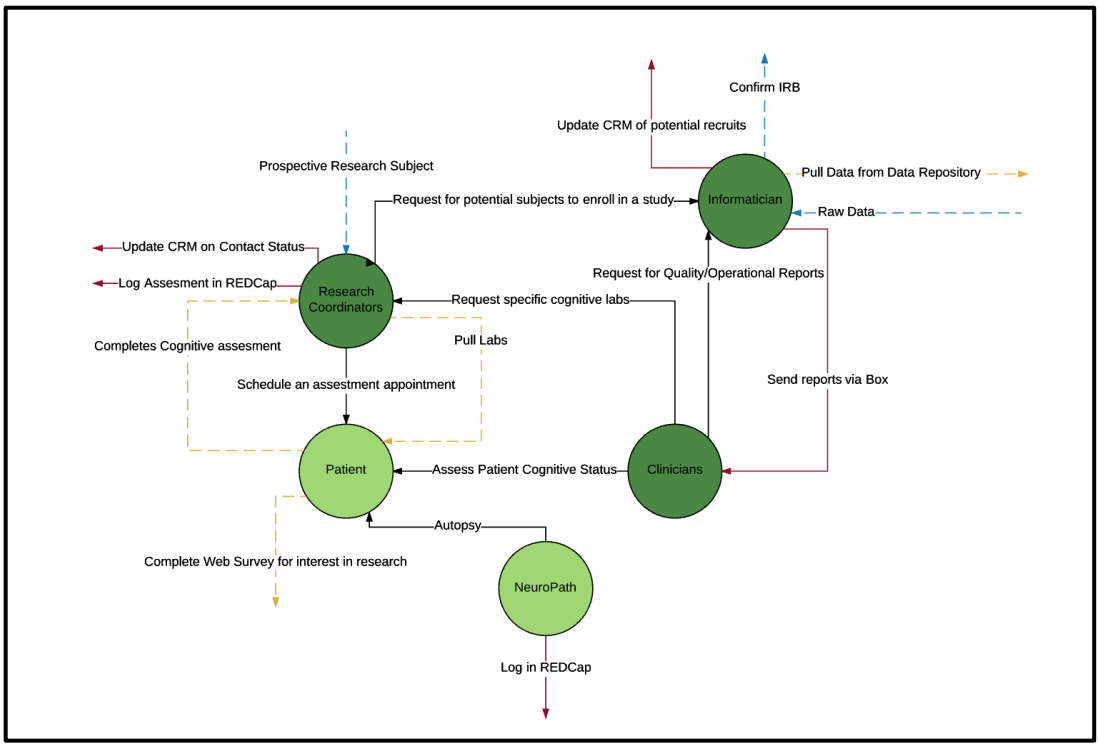
Through the implementation of the Brain Health Center Data Mart, the ADRC has the ability to:

- Increase efficiency of reports of disparate data sources,
- Provide an integrated view of the neurosciences of research and clinical population, and
- Provide broader access to the data.

The Brain Health Center Data Mart incorporates data from disparate Neurology data sources into one consolidated tool. The premise of this solution came about as the ADRC's outreach and recruitment efforts became more successful. This meant more data was being captured on the aging population resulting in more scientists becoming

interested in diverse and multifaceted opportunities to understand their population. Having disparate data sources started to present significant challenges, and the need to have a consistent and efficient way of identifying patient health information in both clinical and research data was becoming more evident. Scientists needed a way of abstracting non-discrete information and tracking specimens to help inform decision-making, enable greater efficiencies, and improve opportunities for research and quality outcomes.

Choosing a quick solution was imperative as the ADRC did not have time to wait for an implementation that could take several years which could have caused an undesirable implication on funding and research opportunities. It seemed the logical solution was a business-focused data mart that would be built with a prioritized set of business requirements. The ADRC built a consolidated data repository that embodies the current workflow which reduced the challenges experienced by having data silos. Figure one illustrates the current workflow of the ADRC that was considered during the implementation of the data mart. This provided a level of interoperability as it integrated data from the source systems to provide reporting, analysis, and decision support capability through the Oracle Business Intelligence tool, OBIEE.

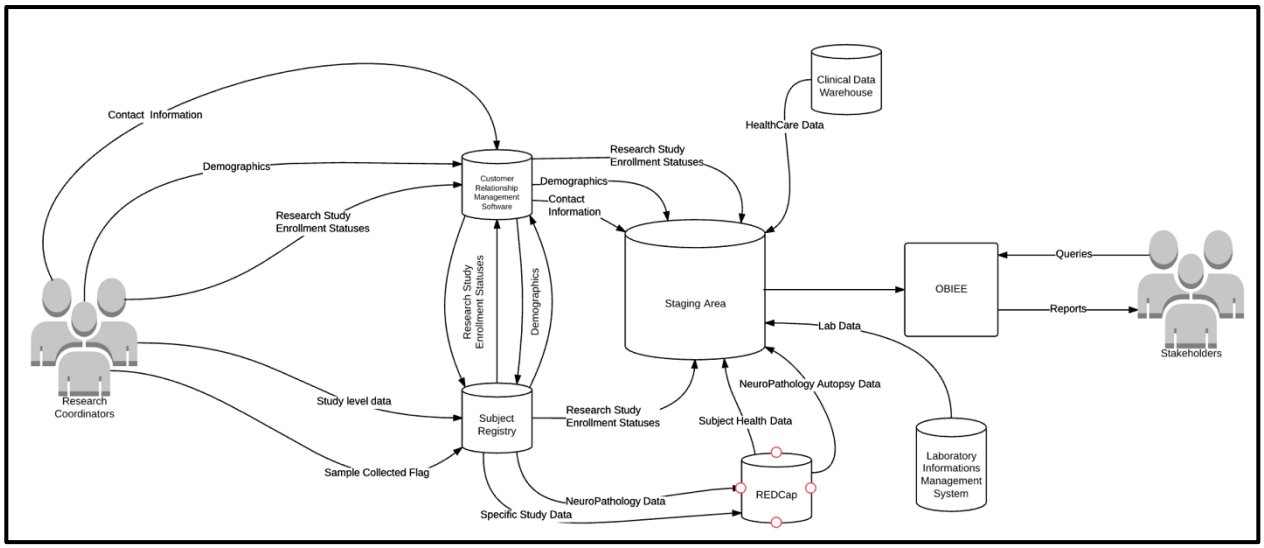


ADRC BUSINESS WORKFLOW
FIGURE 1

The business requirements for this system were essential to understand for the solution to encompass all aspects of the day-to-day workflow. Figure two illustrates how the data mart connects the workflows within each Neurology data source into a staging area where data is extracted, transformed, and loaded for analysis. Each of the following systems (shown in figure 2) were included in the implementation of the data mart solution:

- The Clinical Data Warehouse; Powernotes, Powerforms, Labs, Encounter information, and Medications
- An Emory proprietary research management system
- Laboratory Information Management System (LIMS)
- SalesForce

- REDCap



THE BRAIN HEALTH CENTER DATA MART INFORMATION FLOW
FIGURE 2

PURPOSE STATEMENT

This thesis will evaluate the Brain Health Center Data Mart to assess gaps in implementation and uncover potential opportunities for enhancements. This paper will use Zachman's Framework and The Open Group Architecture Framework to examine the system architecture, the systems interoperability, components of data quality, and stakeholder engagement. Furthermore, this paper will provide recommendations to enhance future iterations of development.

HYPOTHESIS

The architecture of the Brain Health Center Data Mart advances data mining capabilities by integrating the disparate data sources which will allow scientists the ability to access quality data efficiently and effortlessly.

- **Aim 1:** Understand the business rules of Neurology's clinical and research process
- **Aim 2:** Understand the data model and architecture of the Brain Health Center Data Mart
- **Aim 3:** Determine current capabilities and upcoming developments to recommend future improvements

SIGNIFICANT STATEMENT

The integration of clinical and research data is a relatively new concept to public health and healthcare informatics. As more medical specialties are requesting department solutions like data marts to meet the demands of their team, applying industry frameworks to better understand implementation strategies and reflect on lessons learned can provide a more effective implementation approach for future data marts.

LITERATURE REVIEW

A data mart is a subset of data generated for analytics and business intelligence users. Data marts are often created as a repository of pertinent information for a subgroup of workers or a particular use case. A data mart should not be confused with a data warehouse as a data warehouse is a central repository for all of an organization's data while a data mart is designed to meet the specific demands of a group of users within the organization. The concept of a data mart has evolved from being an innovative approach in a few progressive firms to being a necessity in many organizations. The “start small, think big” data mart approach typically begins with a specific business need for data, and is characteristically quicker, simpler, and less expensive to build providing the customer quicker access to their data (5).

According to Bill Inmon, who is recognized as the father of data warehousing, there are many advantages to the creation and operation of a data mart. Aside from the technology and architecture, the single most appealing advantage is the fact that departments within an organization can own their own data. When a department owns its own data mart, it can define specific key performance indicators (KPI) that help demonstrate how effectively they are achieving key business objectives. Owning the data mart implies that there is freedom to do whatever the end user department deems necessary and useful (6). There are other benefits to a data mart such as:

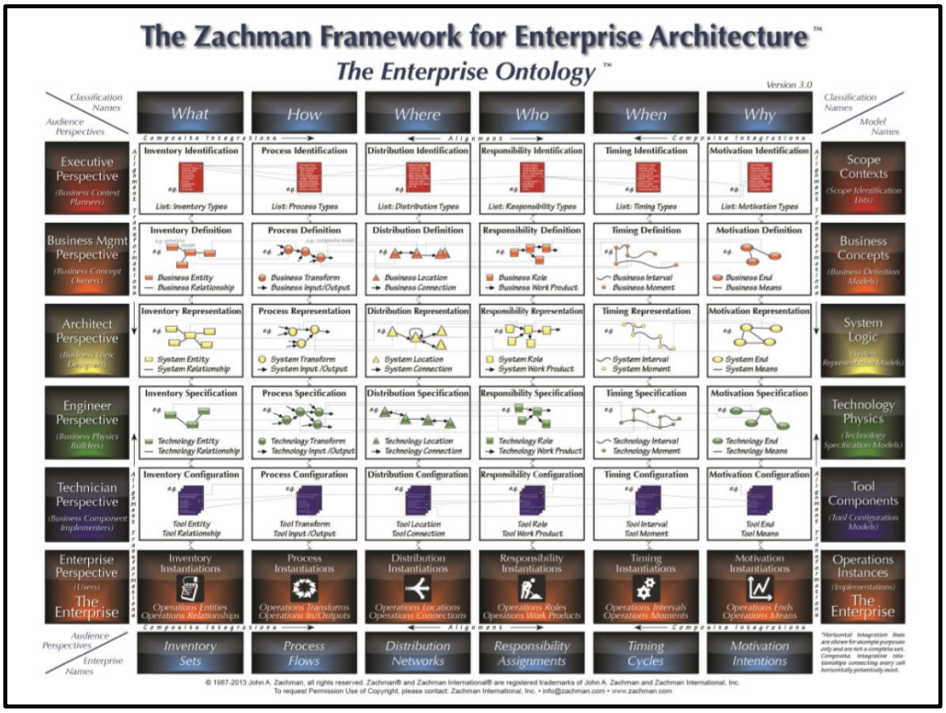
- The data can be tailored to the specific needs of the end user. Data in a data mart can be reshaped and restructured into a form that can specifically meet the needs of the end user (6)

- A data mart will most often times live in a separate processing flow which will greatly reduce the economics of analytical processing. The cycles of processing that occur on a large centralized processor are more expensive; therefore, by moving the data mart to a separate processor, the costs of processing can be significantly reduced as all processing does not have to occur on an expensive centralized processor (6)
- The data mart is the ideal place for the building and tracking of KPIs over a long period of time (6)

In order to create a successful data mart there are several frameworks available to help understand the differences between various approaches and assist in making the best possible choice. Each enterprise architecture framework encompasses the whole enterprise, more specifically the business processes, technologies, and information systems of the solution. This paper will focus on two frameworks used for understanding enterprise architectures. The perspective centric approach, also called Zachman's framework, focuses on the need to ensure that all of the different perspectives within the organization have been adequately considered. Another approach is the process centric method or The Open Group Architecture Framework (TOGAF) which emphasizes the need to focus on the process used to model the enterprise (7).

John Zachman's first outline of the challenges and visions of an enterprise architecture has guided the IT field since 1987. His value driven and agile framework was a response to the need for a holistic approach to systems architecture, considering all issues from all perspectives. The focus on a perspective centric approach to

enterprise architecture has been widely used as the strategy relies mainly on different perspectives within the enterprise, making sure that the different perspectives are incorporated in the project objectives. Figure 3 illustrates the Zachman Framework detailing the 36 cells that focus on a specific point between a stakeholder's perspective and a descriptive focus (7). The figure shows the significance of six stakeholders perspectives that range from the executive, business management, data architect, database engineer, technician, and enterprise roles to understand their perspectives answering the questions: what, how, where, who, when, and why of the enterprise architecture. Zachman often explains this framework using an analogy from the construction industry (7).

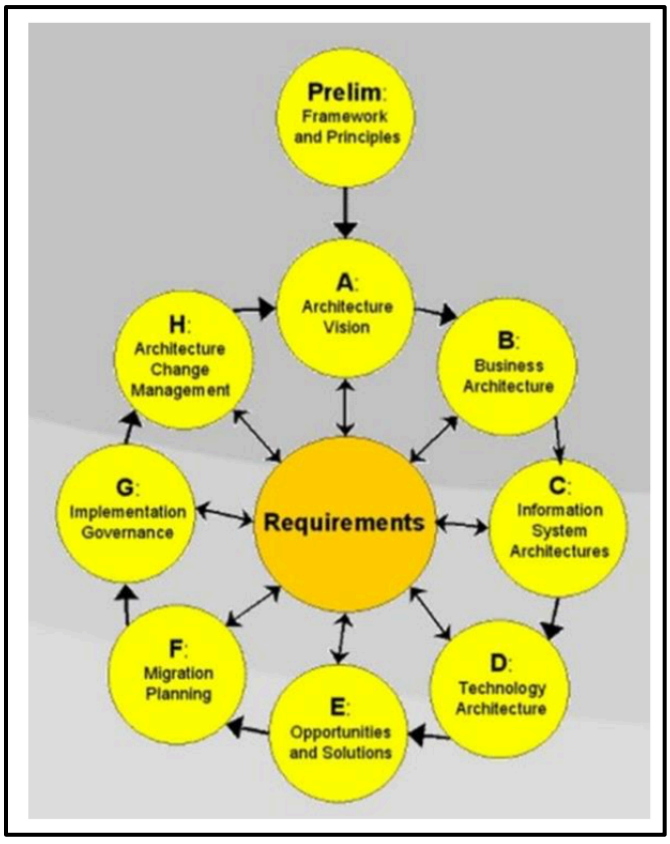


Zachman Framework
Figure 3 (7)

Typically, in a construction industry, architectural artifacts are organized using two-dimensions: stakeholders and business strategy. For the first dimension, in a physical building, stakeholders can be the owner, the builder, and perhaps a zoning board who ensures that construction follows building regulations. A builder prepares different artifacts for each stakeholder. Every stakeholder demands complete information, but what constitutes completeness will be different for each stakeholder. For example, the owner is interested in a complete description of the functionality and aesthetics of the building, and the builder is interested in a complete description of the materials and construction process. Essentially, the owner may not care about the placement of studs in the walls, and the builder may not care how the bedroom windows line up with the sunset. In the second dimension, Zachman details the descriptive focus addressing: the what, how, where, who, when, and why of the project. While this dimension is independent of the first dimension, both the builder and the owner need to know the answers to these elements to ensure they are meeting the complete needs and requirements of the project.

TOGAF takes a different approach to enterprise architecture. This process centric approach places emphasis on a good process which equates to a good enterprise architecture. TOGAF views enterprise architecture as a continuum of architectures, ranging from highly generic to highly specific. A typical TOGAF model is divided into four hierarchical categories: Business Architecture, Application Architecture, Data Architecture, and Technical Architecture. These categories all contain different perspectives of achieving a successful enterprise architecture. The business architecture describes the processes and strategies the business uses to meet its goals.

The application architecture focuses on the specificity of how the applications are designed and how they should interact with each other. The data architecture designates how the enterprise data stores are organized and accessed while the technical architecture details the hardware and software infrastructure that supports applications and their interactions. The most important part of TOGAF is the Architecture Development Method (ADM). It is important to know that the day-to-day experience of creating an enterprise architecture is driven by the ADM which is shown in Figure 4 (7).

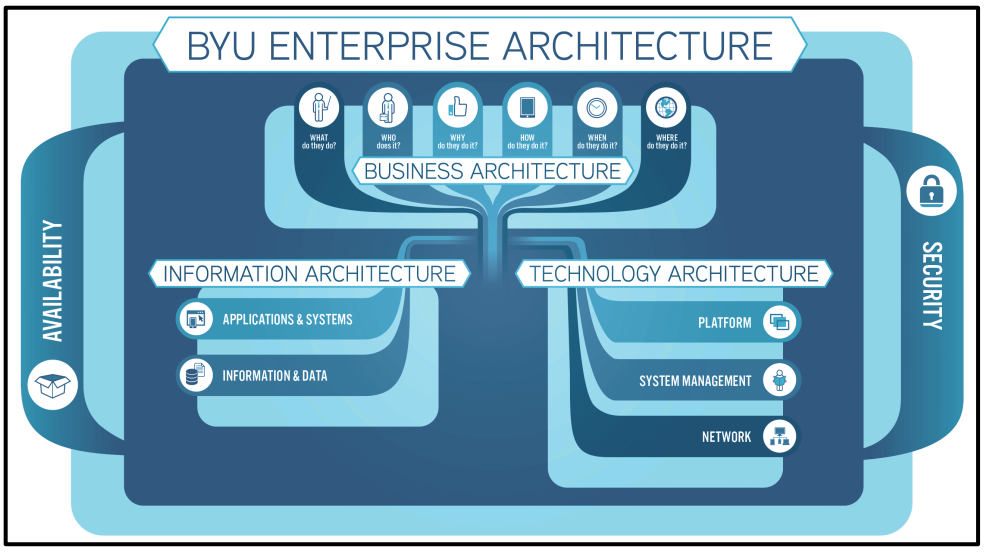


TOGAF Architecture Development Method (ADM)

Figure 4 (7).

TOGAF ADM consists of eight phases that are cycled iterations. To produce an "enterprise-specific" ADM successfully, the process and components should be reviewed for applicability and then tailored to suitably meet the needs of the individual solution (7).

While this framework seems to be different from Zachman's framework, when used in conjunction, TOGAF complements Zachman's framework. In figure 3, the illustration of Zachman's framework provides guidance on how to categorize elements that give guidance to the project while TOGAF provides a direction for creating each element. For example, once the what, how, where, who, when, and why are answered from various stakeholder perspectives it can be applied to the design of architectures such as the information architecture and the technology architecture represented in the TOGAF framework. Guidance from the Zachman framework can help shape the information architecture by transforming and cleaning the data to meet the needs of the customer. Then, this architecture can detail the technical architecture specifying the hardware and software needs. When creating an enterprise architecture, using a blend of these frameworks can be very beneficial to address the solution using both approaches. Brigham Young University (BYU) created a framework as shown in Figure 5 to blend both of TOGAF and Zachman's framework (8).



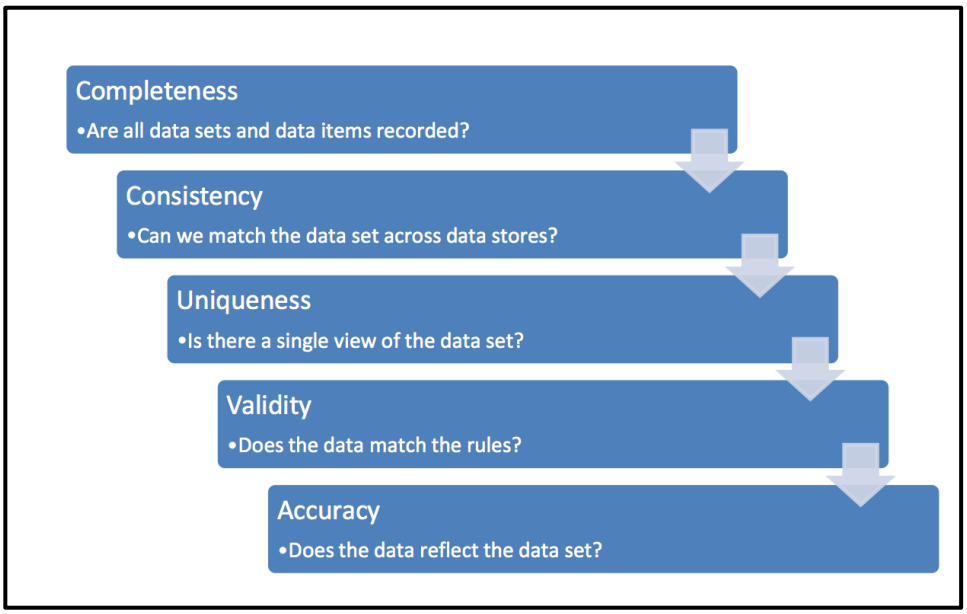
Brigham Young University (BYU) framework
Figure 5 (8).

BYU's Enterprise Architecture framework consists of three distinct architecture types: Business Architecture, Information Architecture, and Technology Architecture. In this blended framework, all three of the architectures lie within the context of information availability and security to emphasize that enterprise business priorities and risk are considered and assessed from the beginning of a project (8). Using a blended approach like the BYU framework can help organizations address both the perspective centric and process centric approach when trying to determine a solution.

In the BYU model, the Business Architecture addresses the documentation and development of models that reflect the organization's policies, processes, work activities, artifacts, and assets. Specifically, the Business Architecture answers the what, who, why, how, when, and where of the organization and processes. Understanding this architecture can provide guidance for a governance team that brings together key stakeholders to ensure that projects achieve both local and company-wide objectives (8). A study in 2006 conducted at Massachusetts Institute of Technology (MIT)

suggests that the most successful approaches to an effective governance team must address two fundamental goals: alignment between IT and the business as well as coordination across multiple organizational levels. IT governance is essential as this determines decision rights and accountability at the business-level. This provides the opportunity for stakeholders to define desirable needs in their IT solution. The study suggests that good engagement requires a good foundation built on clear, specific, and actionable objectives. The study also encourages providing motivation to stakeholders in the form of incentive plans such as bonus targets, annual reviews, and performance metrics which can help to focus and align activities. In addition, involving the right people to maintain a thorough and regular communication flow was shown as a successful engagement technique (9).

The next type of architecture BYU emphasizes is the Information Architecture. In this architecture, the team creates documentation and models the fundamental information assets. This includes things like identifying an application that will best enable business processes as well as defining how applications and information will work together to support the enterprise. The information architecture also identifies which parts of the business process are supported by each application and where each type of data is stored and managed. This level requires extensive data management competencies as data quality is at stake. According to the DAMA UK Working Group, a premier international organization for data management professionals, there are six key dimensions recommended to be used when assessing or describing data quality as described in figure 6.



Data Quality Dimensions (10)

Figure 6

The ability to isolate data flaws at this level allows analysts and developers to apply data management techniques to address the completeness, consistency, uniqueness, validity, and accuracy of the data. This will improve the processes that will create and manipulate the information to maintain data quality (10).

The final architecture BYU describes is the Technology Architecture. This level represents the technical infrastructure as it relates to hardware and software technologies that support the information systems. This architecture must support interoperability (10). According to the Healthcare Information and Management Systems Society (HIMSS), interoperability is defined as "the extent to which systems and devices can exchange data and interpret that shared data. For two systems to be interoperable, they must be able to exchange data and subsequently present that data such that it can be understood by a user (11)." This is the level that will enable users to abstract data

from various disparate sources to help inform decision-making, enable greater efficiencies, and improve opportunities for research and quality outcomes.

METHODOLOGY

To gain a greater understanding of how the team implemented the Brain Health Center Data Mart, an inductive qualitative design was selected to understand each team member's perspectives and opinions of the process. In depth discussions allowed for fluid conversations to ascertain details of each person's experience. The interviews involved open conversations pertaining to scope management, time management, risk management, stakeholder involvement, and other lessons learned. This gave participants the opportunity for each person to talk in some depth, choosing their own words, to help develop a real sense of the person's roles, tasks, challenges, and successes.

Participants

Participants were recruited from the Brain Health Center Data Mart implementation team based on the major responsibilities: Data Model expert, Data Architect, Business Analyst, and the Technical Lead. All participants are employees of the Emory LITS Data Solutions team, who specialize in supporting researchers and the academic health sciences through the design and adoption of information technology solutions. The products and services rendered by this group include but are not limited to assisting researchers in data collection, management, and resource discovery, managing research laboratory and specimen data, and analyzing significant volumes of data through high performance computing (12).

A second set of participants were selected from the Department of Neurology to explore the implementation from the perspective of the customer with knowledge of the business needs and expectations of the data mart. Interviewing participants from both

the IT team and business owner provided the opportunity for a comprehensive view of the implementation process. The sample sizes were fixed prior to interviewing due to the availability of the participants.

Data collection

After obtaining oral consent, each individual participated in a series of 1-2 unstructured interviews lasting approximately 30-60 minutes. Participants engaged in an unstructured qualitative interview from February 2017 to March 2017 consisting of open ended questions pertaining to scope management, time management, risk management, stakeholder involvement, and other lessons learned in order to elicit participants' experiences in their own words. Follow-up questions relating to the system architecture, stakeholder engagement, data quality, and interoperability were asked based on the experiences shared by participants. Interviews were documented using Microsoft Word by the author during the session and after each interview the transcripts were analyzed into common themes.

Data Analysis

A general review of each interview session was conducted first to gain familiarity with the content and tone of the discussion. The next step of analysis required organizing discussion content into underlying themes. These themes were condensed and classified into categories: system architecture, stakeholder engagement, data quality, and interoperability. Interviews allowed for better comparison and contrast of the current state of the Brain Health Center Data Mart vs Zachman's EAIS framework and the TOGAF framework.

Ethical Considerations

The study was considered minimal risk and did not meet the definition of "research" with human subjects or "clinical investigation" as set forth in Emory policies or federal rules. Therefore, approval from Emory's Institutional Review Board or written informed consent was not necessary. Participants received verbal instruction that participation was voluntary and could be withdrawn at any point without consequences. The study was approved by the Rollins School of Public Health, Atlanta, Georgia.

RESULTS

Four of the six members of the implementation team participated in this study. Participants identified themselves as having a specific integral role in the design and implementation of the Brain Health Center Data Mart and offered technical influences within their specific role. Having defined roles and tasks was necessary to take into account each team members' needs and perspectives of the solution. Study participants described primarily positive outcomes from the execution of the Brain Health Center Data Mart. Participants discussed four components: stakeholder engagement, data quality, interoperability, and system architecture which presented as directly related to the enterprise architecture.

Stakeholder Engagement

The implementation team was able to address the main questions of the ADRC: the who, what, why, and where. This was described to be the smoothest process to date as the LITS Data Solutions team has worked extensively with Emory's ADRC team for several years. A former embedded analyst was able to act as the technical advocate during the ADRC business process evaluation. The former embedded analyst was aware that the ADRC was capturing data in various disparate data sources and the only way to get access to patient data was through the support of the current embedded analyst writing SQL queries to try and connect the sources. During each interview, it was expressed that because the project had a devoted analyst that knew the business process and data very well, there was a heavy dependency on the analyst for this knowledge. Some participants voiced concern regarding a lack of a governance team. It was explained that the stakeholder engagement from the customer in this process were

limited to two persons. This was another shared frustration among the implementation team as it restricted the project and impacted scope, timelines, and overall business knowledge.

Data Quality

The Data Solutions implementation team was most familiar with the data captured in the various sources as this team was the team that implemented the current informatics solutions for the ADRC. All data collected in the ADRC research databases are targeted to those who have volunteered for research. Integrating this research population with their clinical population was the biggest challenge expressed by the participants. Participants shared that this integration made apparent many process gaps that made data management challenging. Upon analyzing each data source individually, the implementation team gained insight into the data cleaning needs as there were many fields that did not have values and there were many examples of inconsistent data or duplicate records as well as other random errors. The need for extensive and tedious data cleaning to prepare for integration was necessary. For example, some sources did not collect any person identifiers that could be used to link to other subjects and some users did not follow the process that was in place, leading to incomplete data. The implementation team came to a consensus that the best way to match subjects within the disparate data sources were to use the following patterns:

1. Persons with matching Medical Record Numbers and Date of Birth
2. Persons with matching Medical Record Numbers, First and Last names
3. Persons with matching First and Last names and Date of Birth

4. Persons with matching Enterprise Master Patient Index Number (EMPI) and Date of Birth
5. Persons with matching EMPI, first name and last name
6. Persons with matching Social Security Number if available

The implementation team felt this was the best technique to identify persons in the data silos. While this was able to identify the majority of the necessary population, the team will face the potential of missing some patients who did not match the six patterns.

System Architecture and Interoperability

The decision to use OBIEE came about as the implementation team assessed the technical specifications. In order to create a solution that would best meet the needs of the ADRC and align with HIPAA requirements, the implementation team assessed the following before choosing OBIEE as the final technology.

Service Access and Delivery:

- Access Channels - users will need to have access using a web browser
- Delivery Channels - a Virtual Private Network must be established as users will access the site using the internet
- Service Requirements - the system must be HIPAA Compliant and access will be restricted to certain personnel which will require authentication by individual user

Service Platform and Infrastructure:

- Software Engineering - the solution will require design, development, maintenance, testing, and evaluation
- Database/Storage - a new schema will be created to store the data flowing in the data mart

Component Framework:

- Security - logic will be set to only allow users to see particular data as their IRBs allow
- Business Logic - solution will require heavy logic to account for various visits, labs, medications, etc., pertaining to each person for longitudinal studies

Service Interface and Integration:

- Integration - the solution will need to merge disparate data sources to allow data flow to OBIEE
- Interoperability - the solution will need to connect patients that brings about new meaning, context and clinical insights through the combination of diverse sources of data.

The participants expressed that the capabilities of this tool and the budget available for this project made OBIEE a superior choice. Participants shared that OBIEE's access control capabilities were advantageous because the data mart includes data related to various research studies which require users to have specific IRB approval to access and view the data. Participants also shared that the availability of the Repository Database (RPD) to update the tool as needed was very valuable. Users expressed they were happy with the tool, but less happy with the reporting interface as they felt it looked antiquated.

A major challenge that impacted the project was the change in all data sources mid-project. Approximately one year into the work of the Brain Health Center Data Mart project, the ADRC decided to halt all work and to overhaul all data source systems. This was a common frustration presented by all participants, both stakeholders and the

implementation team. This overhaul not only impacted previous project work, but participants lamented it impacted the scope and cost of the project.

During each interview all participants expressed that they felt the Brain Health Center Data Mart was a success. Some deemed it a success that heterogeneous data was made homogeneous while others deemed it a success that the Brain Health Center Data Mart now provided the ADRC with the ability to identify both clinical and research patient health information comprehensively.

DISCUSSION

The primary goal of this project was to create a data repository used for reporting, analysis, and decision support. According to the stakeholders interviewed, the Data Solutions team was able to successfully make this happen. The implementation team was able to take data from the various information technology systems and software applications within the ADRC and make them interoperable by creating a data mart that helped the customer exchange data accurately, effectively, and consistently. This was a major victory for the team considering all the challenges during the project such as limited stakeholder engagement, tedious data preparation work, and overhauling of the systems midway through the project.

Three points relevant to Zachman's framework stand out in assessing the Brain Health Center Data Mart implementation. The first point is that the Data Solutions team effectively isolated each element in the framework. Zachman states that every architectural element should live in one and only one cell. If it is not clear in which cell a particular artifact lives, then there is most likely a problem with the artifact. The Data Solutions team was successful in defining the roles and responsibilities of each team member and their tasks. Because the elements that pertain to the Data Solutions team will generally not be of interest to the customer, their assigned specific roles and responsibilities allowed them to isolate their perspectives and address the majority of the components in the framework. The second point is that each cell should contain enough information to completely define the system for one specific stakeholder. This can only happen once every cell is populated with appropriate information to fully describe the system from the various perspectives of both the implementation team and

the stakeholder. This comprehensive discussion to define the system for multiple stakeholder needs was not achieved in this project. The third point is that cells from the Zachman grid should be related to each other. Consider, for example, the inventory column in Figure 3. From the ADRC perspective, inventory is information about the business; however, from the data mart design perspective, inventory refers to rows and columns in the database. Each person defines inventory differently, so there should be some relationship between these perspectives. Applying the Zachman grid can provide guidance to the implementation team to ensure that appropriate discussions are occurring between all of the important stakeholders to understand their perspectives and incorporate business needs.

Zachman's framework provides a groundwork for documenting important elements of a project that can guide team members through the database design and ensure that it is being driven by the customer's business requirements. Zachman emphasizes that if the customer has requirements that are not traceable down to the database design, then the implementation team should question if the business needs will be met by the current architecture. Conversely, if there are database design elements that do not trace back to business requirements, then the implementation team might question if they have included unnecessary design at the database level. The Zachman framework can benefit the Data Solutions team in their future data mart projects in the following ways:

- Ensure that every stakeholder's perspective has been considered for every descriptive focal point.

- Understand the ADRC perspectives by sharpening each of their focus points to one particular concern for one particular audience.
- Ensure that all of the ADRC business requirements can be traced down to some technical implementation.

However, the Zachman framework alone is not a complete solution for a successful data mart implementation. Considering the BYU blended approach, the business architecture defines six elements of the business: the what, how, where, who, when, and why. This can be integral to data governance as receiving an agreement on the answers of the six elements may allow for more stakeholder buy in and overall investment.

A major component to also consider is data quality. Data quality issues are inevitable when integrating heterogeneous data sources, and poor data quality can alter the accuracy of insights or could lead to incorrect insights. Even though it is time consuming and the least enjoyable task of the process, data cleaning is of utmost importance. It was apparent to the team that the clinical and research data sources have not been designed to work well together as there were differences in formatting, data types, unrelated identifiers, processes, and different levels of data collection. It has been an ongoing effort for the LITS team to evaluate the data, validate the data, and conduct the appropriate data preparation process which includes data cleaning, data integration, and data transformation.

Analyzing each data source individually provided insight into the data cleaning needs as there were many fields that did not have values and there were many examples of inconsistent data or duplicate records as well as other random errors. How

the implementation team would handle inconsistent data was quite challenging as there were many different processes and varying needs for each data source. The work of the former embedded analyst was vital to the process as she was the sole person to find connections with each data source and relay that information back to the technical team and the customer to find a solution. The team effectively addressed the six key dimensions: completeness, consistency, uniqueness, validity, and accuracy to fully understand the data quality issues they were up against; however, limited stakeholder engagement made addressing those dimensions more challenging.

Data cleaning was an integral part for successful integration of systems as well as resolving data conflicts that were discovered in the data cleaning process. The transformation process of merging sources into a standard form, applying aggregations, incorporating relevant filters, translating coded values, and creating necessary pivots of the data was now easier to accomplish.

Using enterprise architecture frameworks to guide projects like the Brain Health Center Data Mart can create a unified IT strategy for implementation. More specifically, these frameworks can provide guidance to align business needs, create standardization, reuse existing IT assets, and lead to sharing of common methods for project management and software development. In addition, enterprise architecture frameworks can help businesses assess current workflows and implement necessary process changes to align with business goals. As the mission, goals, budget, technology, and people of the enterprise change, every enterprise architecture created must be revisited periodically. Therefore, using the Zachman and TOGAF framework to

problem solve each time can be beneficial to understand the big problem and break it into smaller pieces.

LIMITATIONS/ RECOMMENDATIONS

Zachman's framework does not provide a step-by-step process for creating a new enterprise architecture, nor does it provide much help when deciding if the architecture being created is the best architecture for your business need. However, this framework provides guidance to dive as deep as the user would like to create a well-supported architecture. A general recommendation is to keep a well-documented enterprise architecture. While it is difficult to create quickly, it is likely that by the time the enterprise architecture project is completed a new set of technologies will be available and new business processes will be used that need support. This documentation can easily turn into an endless cycle. It is recommended that the enterprise architecture team focus on documenting the current state of the organization and keep the focus primarily on the future state. To better ensure success, it is also recommended that the enterprise architecture team work closely with other parts of the organization both to inform thought leaders across the enterprise and to stay informed of the changes taking place across the organization.

The main limitation within the project was the lack of involvement and collaboration of stakeholders. It is imperative to get a balanced perspective even when one person is more engaged in order not to create bias in the solution. The implementation team must consider all end users, which can be quite challenging and require consensus building. It is important to use human factors by scheduling regular focus groups to receive feedback on the solution being created. This will provide the option to fail early, often, and cheaply. Allowing the business to drive the

implementation will result in delivering more value as the implementation team markets the small scaled pilot and show the potential of the solution.

This thesis presented a few limitations as well. It was later learned that unstructured interviewing was particularly useful for exploring this topic broadly, but each interview was unique with no predetermined set of questions asked of all respondents. This inconsistent structure made analyzing more difficult, especially when synthesizing across respondents. Also, out of the second set of participants from the Department of Neurology, only one was willing to participate, limiting the feedback from the stakeholders' perspective.

REFERENCES

1. 2016 Alzheimer's disease facts and figures. (2016). Retrieved October 8, 2017, from https://www.ncbi.nlm.nih.gov/pubmed/27570871.

2. Emory University Alzheimer's Disease Research Center. (n.d.). Retrieved October 08, 2017, from http://alzheimers.emory.edu/

3. Disability and life lost. (n.d.). Retrieved October 08, 2017, from https://www.dementiastatistics.org/statistics/disability-and-life-lost/

4. 2020 Topics and Objectives – Objectives A–Z. (n.d.). Retrieved March 09, 2018, from https://www.healthypeople.gov/2020/topics-objectives

5. Rose, R. (1998). Data marts: A key to reviving the enterprise data warehouse. HP Chronicle, 15(11), 1-1, 18. Retrieved from https://login.proxy.library.emory.edu/login?url=http://search.proquest.com/docview/223226346?accountid=10747

6. The Good Data Mart by Bill Inmon - BeyeNETWORK. (n.d.). Retrieved March 09, 2018, from http://www.b-eye-network.com/view/15040

7. Sessions, R., & DeVadoss, J. (n.d.). A Comparison of the Top Four Enterprise Architecture Approaches in 2014.

8. Enterprise Architecture Framework. (n.d.). Retrieved October 08, 2017, from http://ocio.byu.edu/ea-framework.html

9. Fonstad, N. O., & Robertson, D. (n.d.). TRANSFORMING A COMPANY, PROJECT BY PROJECT: THE IT ENGAGEMENT MODEL. Retrieved March 9, 2018.

10. THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT. (2013, October). Retrieved March 9, 2018, from https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf

11. What is Interoperability? (2017, October 13). Retrieved March 09, 2018, from http://www.himss.org/library/interoperability-standards/what-is-interoperability

12. University IT Services and Research Solutions. (n.d.). Retrieved March 09, 2018, from http://it.emory.edu/about_us/rsch_health_sci/index.html