Delivering on Data Management:

A Data Management Consultation for Berney Office Solutions

Clifton Gibbons

South University Montgomery

MIS6211: Data Management

Dr. Amber Epps

November 4, 2020

Introduction

Berney Office Solutions has been offering unique image duplication services to the Central Alabama region for over fifty (50) years. They have received numerous awards for their products, service and support. Having such an influential competitor in this region requires the acquisition of unique and talented personnel to meet and exceed their industry performance standards. Unfortunately, while the business is making great strides in providing powerfully unique and industry leading image duplication technology – a Xerox branded provider – to their clients, they have lacked in managing personnel matters in-house. This is due to the lack a simplistic, yet influential means of managing data for the organization: a personnel database. After discussion and administrative consideration, they developed a database that now meets their unique hiring needs. The Human Resource Personnel Tracker (HRPT) was designed to meet the logistical needs of the Human Resource Department (HR) through the development of logical database structures with industry standard normalization, implementation of inhouse data governance methodologies, physical design concepts for portability, development of security protocols and a plan for potential data warehousing efforts as the database grows.

Company Overview

Berney Office Solutions is a competitive image duplication product provider in the Southeastern United States. Their lineup of products includes copiers, fax and scanning devices and intuitive file management software. The combination of hardware and software solutions afforded by Berney have been so influential on the market that they have secured government contracts with defense agencies.

As a subsidiary of Xerox, Berney Office Solutions is bound by regulation to establish a secure database that, potentially, stores and protects highly sensitive data such as employee

Personally Identifiable Information (PII). For this reason, the ideal database will need to have safeguards in place that will protect and limit data access based upon a user's assigned position or their seniority within the position they currently hold.

Requirements and Use Case Relationships

According to Xerox standards, there are data requirements that must be met to implement into any database that is created to host vital, confidential and/or restricted. In order to meet the data sharing standards of "Classified" and "Confidential", Xerox specified certain criterial to be required database development. These criteria consist of the following:

- Manageability
- Change control demands
- Scalability
- Extensibility
- Security

While these are not an extensive list of elements to be included into the development of any database, they are essential to meet the needs and requirements as set forth by Xerox. The definition of these criteria has been defined and implemented based on Xerox standards and the Database Administrator (DBA) input.

Entity/Use Case Relationships

As part of the management of a database through various forms, or other elements of data retrieval, project managers will typically administer their projects with the use of relationships common to the Software Development Life Cycle (SDLC). In order to best create a relationship, there must, first, be a succinct development of the data tables that will be used within the project. An advanced project may span across many various tables, and across various applications

and/or forms. No matter the case, the importance of making appropriate associations (relationships) will make or break a well-developed database.

Berney Office Solutions was no exception these requirements. The beginning of the database occurs with the onboarding of new employees and staff. During this process, their personal information is collected and entered. This marks the beginning of the database tables. Below are some relevant examples of how such relationships are created and developed. See Figure 1.

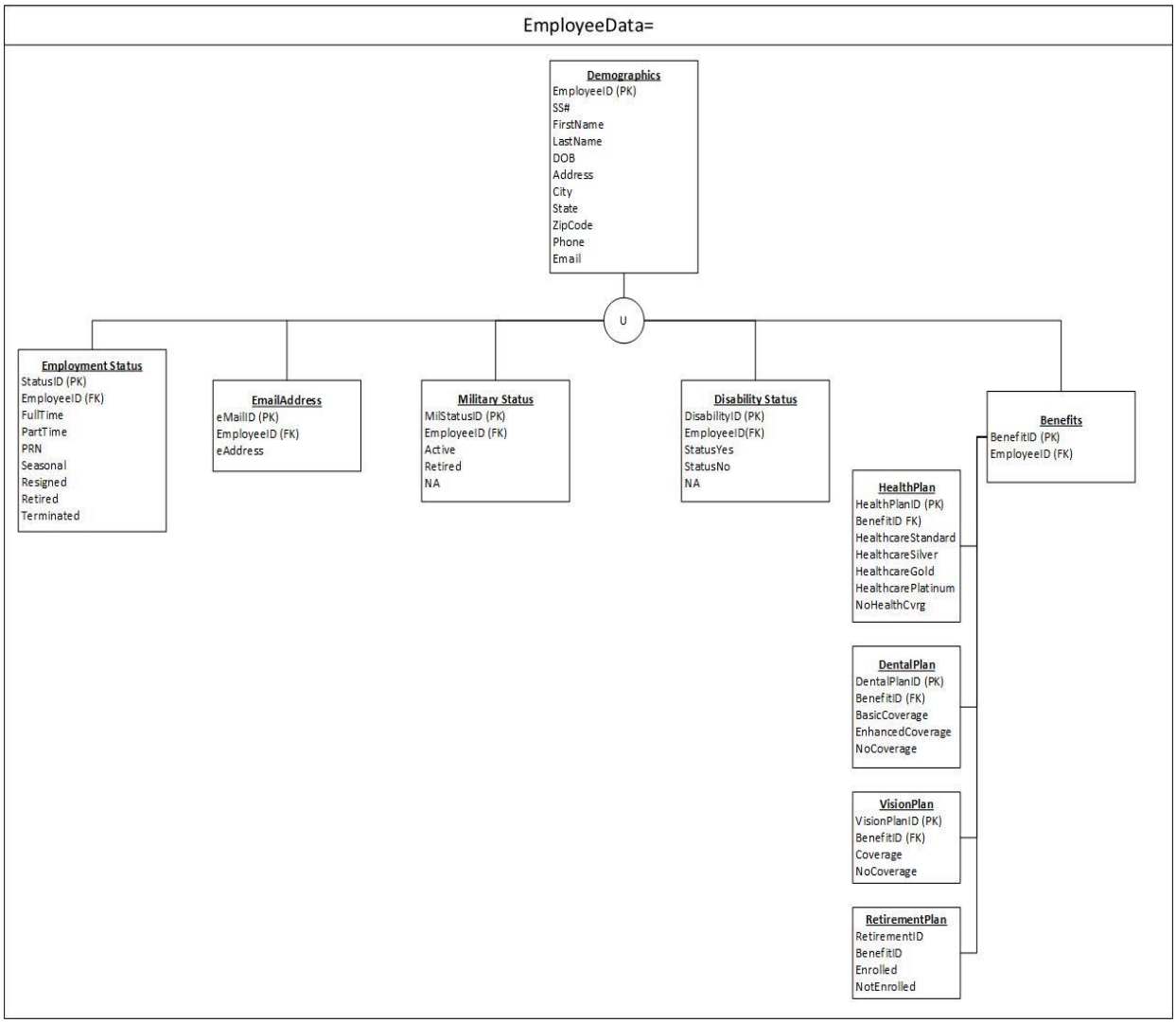

Figure 1: Creating and Entity Relationship Diagram will help developers to visually decipher how their database project tables will relate to each other. This diagram represents the base diagram of HRPT in its conception.

Database Logical Design, Normalization and Management

In the case of Berney's HRPT database, the elements of their Enhanced Entity Relationship Diagram are labeled according to key attributes that are specific to the employee. These include name, date of birth, address, etc. These elements create, essentially, the backbone of the database as they establish the most pertinent data needed for the project. In order to create a proper logical design, this entity/use case was made to encapsulate those pertinent bits of employee information. For this database, that item is called "Demographics". Within this entity resides the basic demographics that are needed for all employees. If the logical design for HRPT is not well-managed, then it would be easy to experience insertion anomalies. This occurs when you have two (2) employees that work in the same facility and share something unique in common such as an email address. To address potential anomalies, in the logical design, a new table was created for email addresses. This is what is considered industry standard normalization. When a DBA, or a data steward, manages this aspect of a database he/she must address, at least, three normalization forms that affect potential anomalies such as insertion, deletion and updates. When done properly, the Logical Database design is created which gives greater focus to the construct of the database.

Logical Design Model

In creating the HRPT database, Berney Office Solutions needed to create a working Logical Design model. The model helped the narrow the focus of the development of tables, relationships and designation of primary and surrogate keys along with potential foreign keys that are developed in entities/use cases. Below is a diagram of Berney's HRPT Logical Design diagram model. See Figure 2.

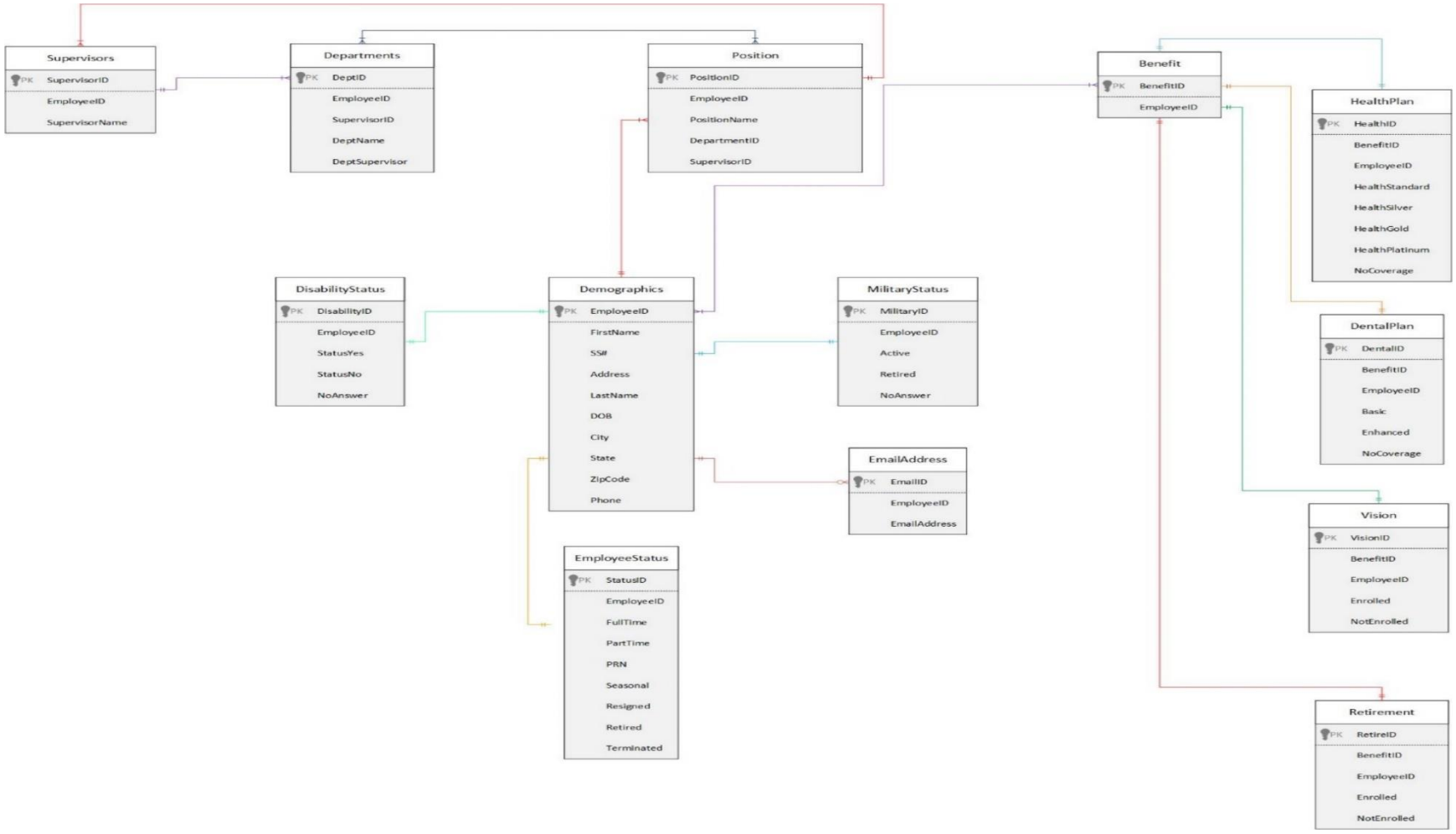

Figure 2: HRPT, after creation of the EER, showed to need various lengths of Normalization. The Logical Design Model, above, displays the application of Normalization to ensure better data integrity and avoidance of potential anomalies.

DBMS Selection and Physical Design

In selecting their DBMS, Berney Office Solutions had to conceptualize their design model of the databased and consider the data types and what management system would best serve the needs of the project. A database data type refers to the format of data storage that can hold a distinct type or range of values (Teach Computer Science, 2019). A data type can be considered much like a fingerprint to the type of program being used. Many data types may be the same, but are expressed within the program differently. A great example is the use of the data type "integer". This data type is represented in code for, nearly all, various DBMS's; however, when expressed in Microsoft Access, we see the term "number". They share the same data storage command/requirement, but they are expressed to the user in different ways.

Additionally, the use of data types can play a major role concerning the manner of development. Using Microsoft Access – this was the most logical choice for this project – affords the developer with a graphical user interface (GUI) that can allow for ease of development. In contrast, the use of something like MySQL may require more command prompt development. However, if used properly, the query features of this software are impeccable. Below are a few examples of how a developer may query HRPT.

Five Unique SQL Query Syntax Examples

- The Left Join Feature can be used to display all data from the left table and selections from the right tale
    - SELECT Departments.DepartmentName, Supervisors.SupervisorID
      FROM Departments
      LEFT JOIN Supervisors
      ON Departments.SupervisorID=Supervisors.SupervisorID
      ORDER BY Departments.DepartmentName

- While using SQL may have it downsides, it is a very powerful DBMS to use when performing queries. For HRPT, the where command can find all entries that may live in a particular city. It would be represented something like the following…
    - SELECT City
      FROM Employees
      WHERE City= 'Columbus'

- In order to find matching data across two (2) unique tables, the developer may use the INNER JOIN feature as such...
    - SELECT Supervisors.SupervisorID, Departments.SupervisorID
      FROM Supervisors
      INNER JOIN Departments
      ON Supervisors.SupervisorID=Departments.SupervisorID

- The developer could create a query that provides the names of all employees assign them by ascending/descending order using just the ORDER BY feature.
    - SELECT Employees.LastName, Employees.FirstName
      FROM Employees
      ORDER BY Employees.LastName [ASC]

- Another query element that will be useful to the developers is the use of the SELECT DISTINT. This could be used to determine how many distinct cities employees are coming from which could open doors to recruitment in other citys.
    - SELECT DISTINCT City
      FROM Employees

Technical Plans

As part of Data Governance, it is important to ensure that you have technical plans ready to address the various aspects of Operations Management and Security Management. These plans, while at this time, will not be exhaustive, they will, however, serve as a powerful launch post as the database grows over time. These plans are ever changing and adapting with technology, with management staff and with the core database needs.

Data Operations Management

To satisfy one of the first elements of data governance is to define the roles dealing with the operational management of data. Those who participate in the governance of operations management are tasked with elements of planning, controlling and supporting structured data

assets. Data assets refer to a system, application output file, document, database, or web page that companies use to generate revenues (Corporate Finance Institute, 2020). So, for a governance perspective, these assets would be databases, forms and applications. For Berney Office Solutions' project, this would reference HRPT along with support and technology management.

This support and technology management was created in the forms of asset management which determines baselines for developers to use to measure scalability of the database as well as develop necessary patches for vulnerabilities for future iterations. These assets – currently a set of portable hard drives – will significantly address issues of vulnerability management from a network perspective. From the perspective of physical management, these drives will be locked in a secure floor vault and must be checked out and back in upon use. Only the DBA/Data Steward will be able to check out both drives to perform data integrity functions.

Once asset management was addressed, then the development of Service Level Agreements (SLA's) was adopted. SLA's are important to establish both guidance expectations, support policies, reporting methods and prohibitions with the use of the designed product. This will help ensure that the database is used as prescribed by the terms of the SLA's and ascribes responsibility based on acquisition and use.

Data Operations Management – Technical Overview

In an overview of the information shared, the Director of Information Technology will serve as the chief point of contact on the management of operations for this project. He/she will be the developer of the SLA's to be provided to the users of the database as well as the technician who will implement data backup services. He will hold the storage backup the database on a single 2-terabyte storage drive that will be under lock-and-key for security.

Finally, he/she will conduct routine backups that will be every six (6) months until the storage level needs change. At this point in time, he/she will conduct a backup of data every three (3) months.

Data Security Management

Like all definitions, defining security for data is highly dependent on the company's vantage point. Many company's believe database security refers to the various measures organizations take to ensure their databases are protected from internal and external threats (Sumo Logic, 2019). As Berney is a Xerox Company, they must adopt the definitions assigned by Xerox in order to maintain security standards, access controls and vulnerability assessments.

Security Standards, set for by the Xerox Technology Governance Council (XTGC), are reflected in every aspect of Berney's daily technology usage. As a result of this, the Director of IT takes a huge role in the management of databases and services to ensure that Berney remains up to code on security measures and authorize user access.

The sole reason an internal data breach may occur in a database, can be attributed to weak access control policies. Data stewards or Database Administrators are typically charged with the development of access controls as outlined by XTGC; however, the DBA for HRPT will be given significant latitude to cultivate these controls due to the small staffing access. Since these drives are portable, the Director of IT elected to use a software access tool to manage the device access controls to users thus helping to reduce threats.

When it comes to the topic of threats, the tactics that can be used to minimize threats and create a safer environment is primarily derived from acknowledging risks and addressing them proactively. The implication of risk occurs when threat is exercised on vulnerabilities. Keeping in mind that definitions are unique to the environment they are ascribed to, typically

vulnerabilities are defined as weakness in your system that may allow a threat to materialize and cause harm (South University - Data Management, 2016). As a result, the best way to combat the materialization of threats is to understand the equation or risks. The idea of the equation is that risk is the culmination of threats factored by vulnerabilities (Risk = Threat x Vulnerability). Acknowledge and address your threats, then minimize your vulnerabilities as best as possible, then you are sure to reduce or eliminate your potential risk.

Keeping this in mind, Berney is able to manage these variables of daily technology use by implementation of access controls for internal threats, routers with firewalls and DMZ's for external threats and a host of intrusion detection software to monitor potential vulnerabilities. Fortunately, for this project, the portable drives help address significant threats both internally and externally. Additionally, by assigning access rights to the drive(s), a log can be easily recorded to determine who has used and modified information on the device.

Data Security Management – Technical Overview

To ensure that there is adequate security, Berney will implement base security standards as ascribed under the Xerox Technology Governance Council. These standards will address elements of user authorization, user authentication practices, defining user access controls, assessing event logs periodically, ascertaining and documenting potential threats and vulnerabilities while implementing tactics designed to minimize or nearly eliminate risks through these assessments. As this database will be removable and designed with a user access list encoded with each authorized user's unique access policy, security is significantly enhanced so long as the users follow expectations as laid out within the Service Level Agreements established under the Data Operations Management governance practices.

OLTP to OLAP

As the nature of all databases, HRPT is destined to grow as the company grows. Berney Office Solutions must be ready to address this scale as it happens. This requires having a current plan and a future game plan. When faster query transactions are necessary, then the transfer from Online Transaction Processing (OLTP) must be converted to Online Analytical Processing (OLAP).

While many DBMS' have query capability, many times those features can be slow or time consuming to create (reference the SQL Queries). The reason for this is due to the fact that these management systems are meant for the transacting (input) of data. OLAP takes that data and transforms it into analytics that can be used to meet business needs. To ensure that this possible, the addition of hire, termination/retirement dates will provide a temporal function that meet search needs. This framework is, at this time, more beneficial to large databases such as used by Amazon. Fortunately, when that time comes for Berney, it is likely that a Data Mart will meet the needs of the company while keeping costs relatively low.

Management of Data Quality

The ability to become competitive and remain competitive, in any industry, can squarely fall on how much information you know and how accurate that information is. Data Quality focuses on that central theme to provide the competitive edge needed in business. This also applies to HRPT. In order for data to have value, there are some characteristics that have to be met.

Data quality is determined by the accuracy, completeness, conformance, consistency, currency, clarity, referential integrity, security, timeliness and uniqueness for which it possess. Those are quite a few elements to consider when you are managing data for quality; however,

there are unique tools that can be used to develop quality within data, but the question is: where do we start?

In the Data Quality Management area of governance, there are four (4) stages that must be met to help ensure the most accurate data quality possible. The first of these is planning. Berney's Director of IT planned this process well by determining the metrics for data quality at no less than 98% accuracy and establish audit periods. Once he had a plan of action for HRPT, he deployed a bottom-up software tool that will scan the database for anomalies, logical errors and duplication errors. His application of choice is the SaS Data Management software, which seems to works within the desired accuracy rate. Finally, he has developed guidelines for himself or other Data Stewards that details when they should take action on a given matter…as well as to what degree of action they should consider. As a result of all of these steps, he successfully created a management plan for quality that should meet the needs of Berney's HR.

Conclusion

Developing a data base is painstaking work filled with exceptional detail to task; however, the outcome of a well-designed database will carry over for years to come. From basic conception to development of EER's to development of governance guidance, we have found that database creations are no small task. Fortunately, Berney's Director of IT was up for the task. The Human Resource Personnel Tracking database will serve Berney for years to come due to its massive scalability, ease of use, portability, access controls and many more other features. The attention to detail along with the development of internal governance is a classic example of how exceptional Database Administrators are ***Delivering on Data Management***.

References

Corporate Finance Institute. (2020, June 3). *Data Assets - Overview, Types and How to Measure Returns*. https://corporatefinanceinstitute.com/resources/knowledge/other/data-assets/#:~:text=Data%20assets%20refer%20to%20a,dollars%20to%20manage%20such%20assets.

South University - Data Management. (2016). *Data and Information Security[Week 3 Lecture]*. https://myclasses.southuniversity.edu/content/enforced/70961-17091525/media/transcripts/Week3/SUO_MIS6211_W3_L2_Supplemental_1.pdf?_&d2lSessionVal=RyHHwHtDCDkuuH6Fjl8NyqKDo&ou=70961

Sumo Logic. (2019, June 6). *What is Database Security?* https://www.sumologic.com/blog/what-is-database-security/

Teach Computer Science. (2019, August 23). *Database Data Types: Computer Science*. https://teachcomputerscience.com/database-data-types/