NORTHWEST NAZARENE UNIVERSITY

Addressing Data Integrity for Individual Profiles and Building Data Flow Models at

NNU

THESIS

Submitted to the Department of Mathematics and Computer Science

in partial fulfillment of the requirements

for the degree of

BACHELOR OF ARTS

Zachary Smith

2023

THESIS

Submitted to the Department of Mathematics and Computer Science

in partial fulfillment of the requirements

for the degree of

BACHELOR OF ARTS

Zachary Smith

2023

Addressing Data Integrity for Individual Profiles and Building Data Flow Models at

NNU

Author: *Zachary Smith*

Zachary R Smith

Approved: *Barry Myers*

Barry L. Myers, Ph.D., Chair, Department of Mathematics and Computer
Science, Faculty Advisor

Approved: *Jill Nogales*

Jill Nogales, Writing Center
Second Reader

Approved: *Barry Myers*

Barry L. Myers, Ph.D., Chair,
Department of Mathematics & Computer Science

Abstract

Addressing Data Integrity for Individual Profiles and Building Data Flow Diagrams at NNU.

 SMITH, ZACHARY (Department of Mathematics and Computer Science), MYERS, DR. BARRY (Department of Mathematics and Computer Science).

Transitioning to a new enterprise resource planning system is a complex task that requires an extended period of time to complete. Two concerns during this process are related to the institution's data integrity and having an in-depth understanding of current business processes within the institution. In regards to data integrity, the current system at NNU was known to have duplicate profiles for some individuals because their information was split between at least two distinct sources, but there are cases where the profiles really are for two different people. Resolving these conflicts requires researching both the avenues for profile creation and any possible errors stemming from manually translating handwritten documents to digital profiles. After the research, a recommendation to either merge or keep the profiles separate would be given to the final decision maker. As for understanding business processes, it requires collecting all available facts from relevant sources and developing diagram models to capture how data flows and is used throughout the system. With accurate models and an in-depth understanding of the current system, higher level decision makers will be able to decide the best approach to transitioning to the new system without breaking any business processes during the ordeal.

Acknowledgements

Table of Contents

Table of Figures

Background and Overview

Transitioning from one tech system to another is an endeavor that requires extensive research and work during the process. In the case of one university, a new project, the Denali Project, began the transition from their current enterprise resource planning (ERP) system to a new system offered by a different company. This transition process requires an understanding on how data permeates throughout the system and addressing issues related to data integrity. Data that has lost its integrity would have lost its purpose and may leave the institution liable for legal trouble depending on the data lost. The ultimate goal for the transition process is to retain all business functionality once it is complete.

For this paper, use of names is redacted, so the different ERP systems will be referred to as the "Old ERP" or the "New ERP," and the university's name shall be "Uni." The reason for transition to the New ERP system was due to the Old ERP being roughly 20-years-old and becoming untenable. Causes for this include a high turnover rate for staff involved with the system in conjunction with a relatively high rate of customization on the Old ERP system. Additionally, the Old ERP was unintuitive for new staff members, so the cost of training them while suffering from the higher turnover rate meant much higher costs than typically expected.

The scope of this project will cover preliminary work dedicated to current state analysis of Uni's departments and addressing data integrity issues by resolving duplicate person records within Uni's own system known as CX/JX. In doing so, this project will have contributed to the preliminary work for the transition process from the Old ERP to

the New ERP by helping set up part of the new system's foundation. The work took place over about four months during the summer of 2022.

Background to Duplicate Records

Maintaining data integrity is important at all times, so the Denali Project lent its support to investigate suspected duplicate person IDs within CX/JX. Each person who interacts with Uni will have a unique ID generated and attached to them. Each person should only have one ID, but the system could generate another for the same person. A person who has their information split between more than one ID can experience issues because they may be able to use both to interact with the university and be left frustrated or confused when changes they have made are not reflected accurately. Resolving data related issues is important, especially for the students who represent the university's customer base.

Uni staff created a system for flagging potential duplicates that used a heuristic for scoring the likelihood of a duplicate case within CX/JX. The heuristic used data fields related to the IDs to score the chances of them being duplicates, with the more important fields being their name, date of birth (DoB), and social security number (SSN). The SSN is weighted highly because it should be unique for every person in Uni's system due to it also being a form of ID issued by the federal government and should be unique for everyone as well. When theSSN is paired with a DoB and the person's name, the likelihood of a duplicate case further increases. However, it is important to keep in mind that these fields are not absolute due to the potential of human error during any stage of the process. Errors can come from both Uni's side and the student's. Examples of human

errors include typos, misinterpreting handwritten documents, or filling in data from memory. A noteworthy example: a parent could fill out an application to Uni on their child's behalf, but they might mix up data between their kids like their SSN or DoB.

Work Process for Duplicates Records

The process for resolving duplicates requires fact-finding and identifying a possible source for the error. The issue can be resolved with greater confidence if there is strong evidence that points to either the IDs being duplicates that should be merged or that they are genuinely distinct people that should be kept as is. For the scope of this project's work, only recommendations with accompanying documentation were left for someone else to make the final decision after review.

Once the system has identified potential duplicates and added them to a queue, work can be done to investigate each case individually until a recommendation can be made to either merge the IDs into one or keep them distinct. The process for investigating duplicate cases involves reviewing data within CX/JX for the ID, checking documents, and checking *FastPeopleSearch* which is a website for finding publicly available information on people. If further information is required, then an audit of the IDs can be performed to review an abstract history of information related to the ID and how it has changed over time. If there is still not enough information, then the duplicate case would be added to another queue for specialist review.

The process starts by reviewing CX/JX because it is the source of the duplicates. Data associated with the IDs can be compared for similarities and differences between them and documented. The next step would be to review any available documents related

to both IDs to cross check them and look for errors caused by something like a typo. *FastPeopleSearch* should be used as a supplemental source of evidence, but it often does not have anything on a person if they are under the age of 20. If more information is needed, then an audit of the IDs would occur, but information is often sparse and unimpactful. Once all available sources of information have been examined, the development of a recommended course of action with accompanying documentation will be left for a superior to review and act on.
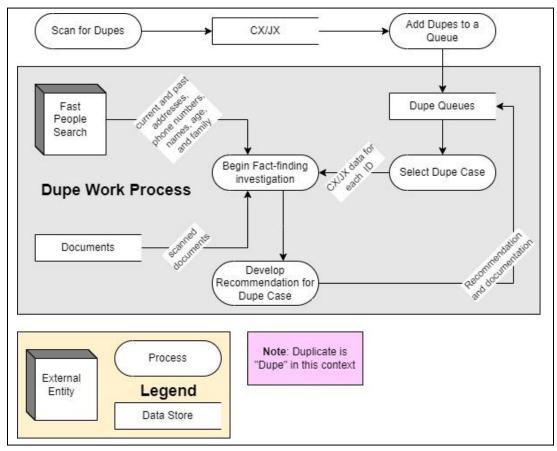

Figure 1. Duplicate Work Process

Results From Duplicate Work

After reviewing 200-300 duplicate cases as part of a small team, it was apparent that many cases had to be dropped due to the incredibly sparse level of information. These cases would be documented with as much information as possible before being left for a superior to move the case to a specialist queue.

Aside from the cases with sparse information, there were a few cases where duplicate IDs were more than two. These took additional work to document, but generally no more work than the other cases with one notable exception (see Figure 2). In this case, there were three IDs that were in conflict, but after reviewing all available material, running an audit, and creating a diagram for clarity, it became apparent that this was a father and son pair with the same name in conjunction with a deprecated ID that was still technically active. *FastPeopleSearch* did not have any records for the son in this case, but it did have many data points on the father. Because there were documents related to a young man's high school transcript by the same name, it was concluded that it was a father and son pair.
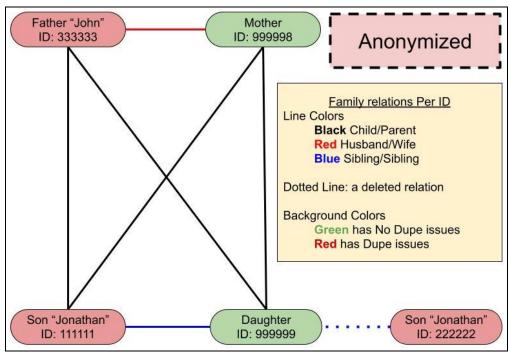
Figure 2. Example of a Difficult Duplicate Case

Finally, there were cases that were never properly picked up by the team due to those cases being out of scope for this project. Uni not only keeps people in the system by IDs, but also business entities that they work or have worked with. These cases where businesses are involved require more access to information over a much longer period of time than was available.

Background to Current State Analysis

In order for an institution to retain all of its business processes after transitioning to a new ERP system, the processes must be well documented to capture key functions. To accomplish this, current state analysis was performed on Uni departments or offices with work generally occurring in order of importance based on Uni's contract with the New ERP company. During this project, analysis was performed on Uni's Student Life

and University Advancement (UA) offices. The end goal for this analysis phase was for the development of data flow diagrams (DFDs).

The analysis for Uni's Student Life office was narrowed down to just the housing process while the UA office underwent a more broad analysis. The UA office handles a wide range of responsibilities related to Uni's external relations and communications.

Special attention was given to discovering where and how data is stored. This priority was set by the Denali Project because discovering all sources of data with relevant fields is needed for handling the transition to the New ERP system. The staff at Uni use many third-party applications and keep spreadsheets of data related to their work.

Work Process for Current State Analysis

To capture the current state of business processes within an office, work must be done to research the past and present. Reviewing past documentation already available on an office should offer a framework with which to start, but interviews with Uni staff who perform the business functions must also take place. This work is important, but it was also crucial to understand that the staff also had their own work to do and that any contact with them is essentially more work for them. A balance between respecting their time while still gleaning relevant information from them was essential.

After past documentation has been reviewed, preparations for interacting with the targeted office should be started. Communication with the office should clearly state the purpose and provide some priming questions for thought before personal interviews or meetings occur. During any interactions, a professional standard of communication and note taking was required. After each interaction, time must be taken to compile and

review the notes to form an understanding of what the targeted office does. A task inventory should be generated which is just a spreadsheet of what processes have been discovered so far. With a basic understanding, DFDs can be drafted, and any errors or oversights can be corrected with future interactions with the targeted office to clear up any misunderstandings until an accurate draft is completed. This final draft should capture the current state of business processes for the targeted office while explicitly showing where data lives within the current system.

Results of Current State Analysis

In regards to the Student Life office, eight DFD drafts were created, and the final version broke the housing process down into specific cases which were represented with smaller, more focused DFDs. Great effort was spent on making the DFDs not appear tangled with arrows crossing over each other. This is because these DFDs could serve as a reference point for future work by individuals who had been uninvolved in the current state analysis process. By the analyst spending extra effort to make the DFDs easy to understand, it reduces the wasted time of future analysts who will continue the work.

The Student Life DFD started as a whiteboard drawing from a staff member. After eight iterations not including the original whiteboard draft, the housing process was passed and stored by the Denali Project. It was a relatively small piece of the overall picture, but it was the first DFD by the Denali Team which meant it took more time to address formatting standards for DFD.

As for the UA office, it was much more complicated with many points of overlap with other offices or departments. The UA office handled donations, some marketing,

external relations, and some relations to alumni of Uni. This was confusing because Uni has a dedicated Alumni office, and external relations can mean someone sponsoring scholarships which go through the Financial Aid office. Several DFD drafts of over a dozen business processes were completed, but not all processes were captured with the degree of accuracy considered desirable. However, the heavier emphasis of this current state analysis was on the discovery of data stores and external entities, which was successfully completed. Future work to iron out the processes related to the UA office can be done if the need arises.

Conclusion and Future Work

The Denali Project is a two year long project, so future work is quite extensive. As of the writing of this thesis paper, data migrations and system integrations are still aspects that must be finished.

In the end, the work performed over the summer was done to progress the Denali Project. In terms of contributing to the project's success, my work was successful. That said, the work I initially planned and prepared to do over the summer was very different from what actually happened. Instead of integrations of Uni systems for the new ERP system, preliminary work for data migration was done. It is probably accurate to say that no one on the Denali team had their plans meet their expectations, so we all had to adapt to the dynamic situation.
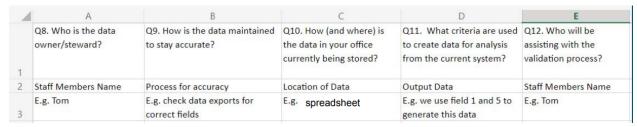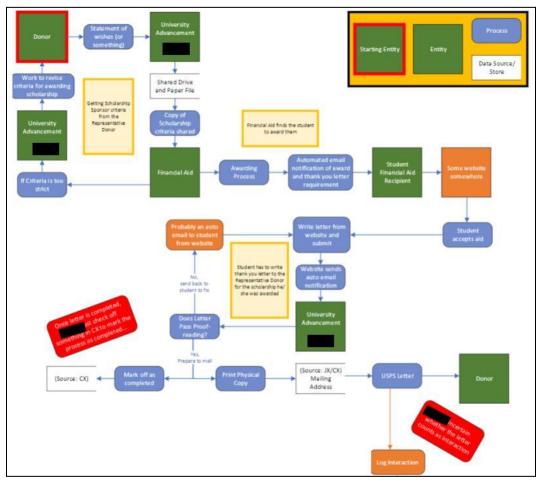
Appendix

	A	B	C	D	E
1	Q8. Who is the data owner/steward?	Q9. How is the data maintained to stay accurate?	Q10. How (and where) is the data in your office currently being stored?	Q11. What criteria are used to create data for analysis from the current system?	Q12. Who will be assisting with the validation process?
2	Staff Members Name	Process for accuracy	Location of Data	Output Data	Staff Members Name
3	E.g. Tom	E.g. check data exports for correct fields	E.g. spreadsheet	E.g. we use field 1 and 5 to generate this data	E.g. Tom

Figure 3. Task Inventory

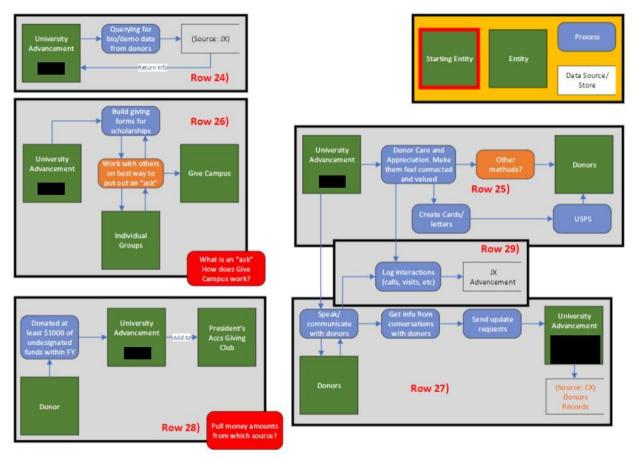
Figure 4. Early Draft for Part of UA

10

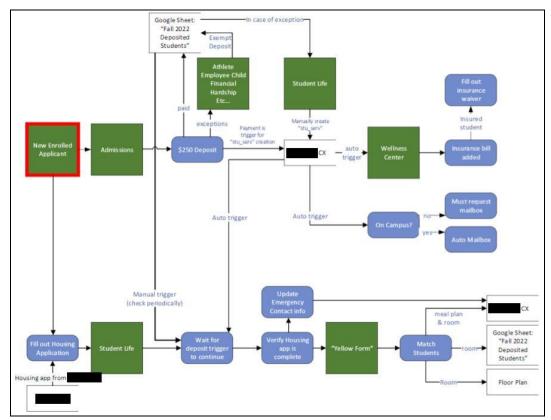**Figure 5. More Task Items from UA**
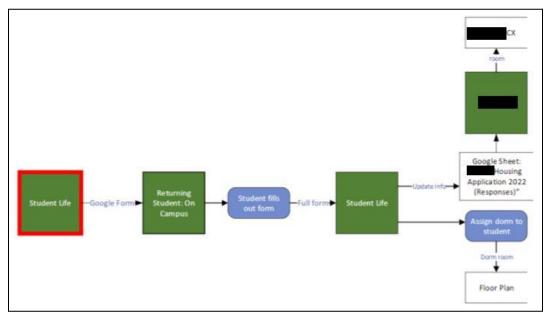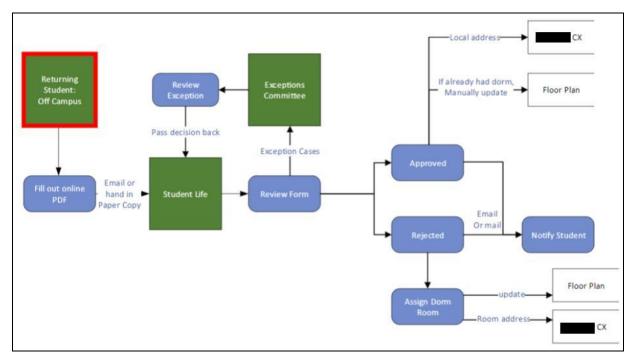
Figure 6. Housing One
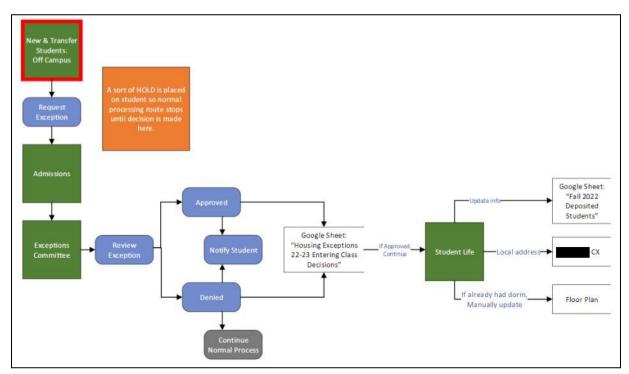

Figure 7. Housing Two

Figure 8. Housing Three

Figure 9. Housing Four