DataMaster:
Software Requirements Specification
version 1.0

DnD
Computer Science Department
California Polytechnic State University
San Luis Obispo, CA USA

October 10, 2017

Contents

Credits

Name	Date	Role	Version
Griffin Aswegan	October 3, 2018	Author	1.0
Steven Bradley	October 3, 2018	Author	1.0
Christina Daley	October 3, 2018	Author	1.0
Larry Hu	October 3, 2018	Author	1.0
Shane Villalpando	October 3, 2018	Author	1.0
Dustyn Zierman-Felix	October 3, 2018	Author	1.0

Revision History

Name	**Date**	**Reason for Changes**	**Version**
Griffin Aswegan	October 8, 2018	Cleanup of Initial Revision	
Team DnD	October 8, 2018	Initial Revision	1.0
Christina Daley	October 11, 2018	Revise Intro	1.0

1 Introduction

1.1 Purpose

This document presents the requirements, restrictions, and limitations of the DataMaster. It will serve as a guideline for the functionality of the software, including major features, functional requirements, non-functional constraints, and other potential features that should be met.

1.2 Document Conventions

This document shall adhere to the following formatting conventions:

1. All documents referenced will be italicized as such:
   Software Requirements Specification
   Links to said documents can be found in the Appendix.
2. All Functional requirements will be labeled "FR-X", where X is based on the ordering provided in section 2.2.
3. All Nonfunctional requirements will be labeled "NFR-X", where X is based on the ordering provided in section 6.
4. All External Interface requirements will be labeled "IR-X", where X is based on the ordering provided in section 5.

1.3 Intended Audience and Reading Suggestions

1.3.1 MarkLogic

This document has been written with MarkLogic in mind, the primary benefactors of DataMasters. A suggested reading order for MarkLogic is as such:

1. Product Perspective
2. Functional Requirements
3. Non-functional Requirements
4. System Features

1.3.2 Professor Bruno daSilva

This document has been written with Professor Bruno daSilva in mind, the professor of the 405 Capstone class. A suggested reading order for daSilva is as such:

1. Product Perspective
2. Functional Requirements
3. Non-functional Requirements
4. User Personas
5. Use Cases
6. System Features

1.4 Project Scope

The main goal of this project is to provide a system that lets users classify and organize large pools of data sets within their databases. For further information, please refer to the team's *Vision and Scope* document in the Appendix below.

1.5 References

None, so far.

2 Overall Description

2.1 Product Perspective

Data has been exploding in technological popularity over the last decade. Data is used to predict advertising, manage personal data, track relevant user data, and manage important information. Modern data analysts skim through hundreds to thousands of data sets per day in order to find new information and relationships in that data, however searching through data sets can be a slow and laborious process. Our intent is to design a product that makes it easier to classify data sets and search through similar data sets to find new relationships and organize similar data sets through the data inside those sets.

2.2 Functional Requirements

FR-1a	The system shall classify data sets, provided by the user, based on data within the data set.
FR-1b	The system shall have already-existing common classifications for the system to use.
FR-2	The system shall allow users to search for data sets using keywords.
FR-3a	The system shall allow users to manually classify data sets.
FR-3b	The system shall allow users to manually edit classified data sets.
FR-3c	The system shall allow users to manually remove classified data sets.
FR-4	The system shall have to ability to find relationships between multiple data sets.
FR-5	The system shall observe relations between data sets.
FR-6	The system shall have the ability to output an organized graph of all user-specified connections.
FR-7	The system shall have the ability to display a table of customer information.
FR-8	The system shall have a feature to upload one or more data sets.
FR-9	The system shall provide the ability to provide a chart display all of each data set's information.

2.3 Operating Environment

This software has two parts - a server side management piece, and a client-side interface. The server side management piece is expected to run in any popular Linux distribution, such as Ubuntu or Debian. The client-side interface is expected to run in any popular modern web browser, such as Mozilla Firefox, and Google Chrome.

2.4 Design and Implementation Constraints

None, so far.

2.5 User Classes and Characteristics

User Class	Description
Database Manager	A user that manages multiple data sets in a database.
Database User	A user that utilizes the capability of a database for their business.
Database Analyst	A user that analyzes data within databases to find unusual relationships or information.
DBMS Designer	A user that designed database management software.

2.6 User Personas

All user personas are described in more detail in the *User Personas* document.

2.6.1 George Willis

George Willis
Age: 35
Occupation: Database Analyst, Manager
Technological Level: Basic to Intermediate
George just received a list of clients that he needs to add in the Mark Logic DB. So first he needs to have the function to actually edit the DB but also His time is very pressed so he does not have much time to learn this new tool and would only use this tool for the project. So he believes the Mark Logic DB tool would have to be very simple to use with not too many functions. He simply wants to add the data in, and catogrize the insterted.

2.6.2 Janice Healy

Janice Healy
Age: 27
Occupation: Business Owner, Bakery
Technological Level: Basic to Intermediate

Janice Healy, owner of "Baker J's" has a lot of customers, and she manages all of the sales she makes in her database, as well as all of the customers she has. However, when setting up her database, there were multiple different fields that all had the same data, but with different names, and she struggled to keep everything coordinated and organized. She is looking for a way to search through her multiple data sets in order to figure out what data is stored where, since information for customers is broken apart in multiple data sets.

2.6.3 Patricia Martinez

Patricia Martinez
Age: 46
Occupation: Registered Nurse
Technological Level: Basic
Patricia Martinez is a 44 year old registered nurse. She works in Loomis, California for a small, endocrinologist clinic. Patricia is a mother to two kids and spends her time either at work or with her kids and family. Patricia went to school thirty years ago and has watched technology progress in her field and the world. She is not very technologically savvy but has been able to pick up any necessary tech for her job. Working in a specialized clinic Patricia receives and handles patients and their patient information from other clinics. Patricia handles files sent from different clinics and health care providers of patients and uploads this non-uniformed data into –Product Name–. She then uses –Product Name– to search for patients and update their information during and after consultations. –Product Name– is simple to use for Patricia. It's ability to classify differing kinds of patient information and organize it in a way Patricia can use it allows Patricia and the clinic to be efficient.

2.6.4 Jeffrey Miles

Jeffrey Miles
Age: 27
Occupation: Database Manager
Technological Level: Expert
Jeffrey Miles works as a database manager for "The Fictional Company" which handles a massive amount of data on a daily basis from multiple customers. With how the database is set up, each customer has their own database that keeps track of their own separate fields, and sometimes Jeffrey struggles keeping everything collected and organized, especially when one of the customers updates their databases with new data or fields. Jeffrey is looking for a way to keep a consistent naming convention between all of his data sets without modifying the data sets or the column names in the data sets.

2.6.5 Thomas Cadillac

Thomas Cadillac
Age: 33
Occupation: Database Manager
Technological Level: Expert
Thomas Cadillac is a thirty-three year old database manager who loves dissecting and integrating data. He often feels that the data he is working with gets classified into too many categories and he has trouble integrating all of his data sets in a maintainable fashion. He thinks there could be a better way to classify files in order to have more data in the same place, which would make his dissecting of data more effective.

2.6.6 Greg Barkley

Greg Barkely
Age: 41
Occupation: Zookeeper
Techonological Level: Intermediate
Greg Barkley is a 41 year old zookeeper. Greg had aspirations of becoming a veterinarian after finishing school, but when he was unable to pay for medical school he decided that he wanted to become a zookeeper instead. At the zoo, Greg is responsible for keeping track of the normal habits of the elephants. Everyday, he records information about the elephants and stores in a database. Greg uses —Product Name— to then classify the data he records. With the classified data, he can detect when an elephant is not behaving normally and can take the appropriate action to help the elephant.

2.6.7 Ryan Janzen

Ryan Janzen
Age: 55
Occupation: Business Owner
Technological Level: Intermediate
Ryan Janzen is a 55 year old business owner that works in Los Angeles, California. Because Ryan deals with private information, his clients expect their personal information to be safe. With —Product Name—, Ryan, nor his clients, will have to stress any private information being leaked to third parties.

2.7 User Documentation

Developers will be keeping track of all progress on the application using JIRA, and will produce more documentation as the need arises. Preliminary documentation that will be produced include this document (the *Software Requirements Specification*, a *Vision and*

Scope document, multiple UML diagrams pertaining to the architecture of the product, and user interface mock-ups for review and testing.

3 Use Cases

3.1 Use Case 1: Upload and Update Data

Use Case ID:	1
Use Case Name:	Upload and Update Data
Created By:	Christina Daley
Last Updated By:	Christina Daley
Date Created:	October 9, 2018
Date Last Updated:	October 9, 2018
Actors:	Patricia
Description:	Patricia is a registered nurse who handles patient information from different providers and clinics. She is not technologically savvy. Part of her job is to compile the data and update it.
Preconditions:	1. Patricia is logged into the system. 2. Patricia is authorized to see and edit the patient's data.
Postconditions:	None
Normal Flow:	1.0 Upload and Update Data 1. Patricia opens the home page of the web portal. 2. Patricia clicks add patient. 3. Patricia uploads different excel files to the new patient. 4. The patient's information updates with new categories. 5. Patricia checks in the patient. 6. Patricia adds to and updates patient information. 7. Patricia displays information for doctor.
Alternative Flow:	None
Exceptions:	None
Includes:	FR-2, FR-3a, FR-3b, FR-3c, NFR-F
Priority:	High
Frequency of Use:	Medium to often, depending on the user
Business Rules:	None
Special Requirements:	1. Patricia shall be able to return to the home page at any time. 2. Patricia shall be able to search for and update any patient.

Assumptions:	Patricia has security clearance for all patient information.
Notes and Issues:	None

3.2 Use Case 2: Search For Data Sets

Use Case ID:	2
Use Case Name:	Search For Data Sets
Created By:	Griffin Asweagn
Last Updated By:	Griffin Aswegan
Date Created:	October 9, 2018
Date Last Updated:	October 9, 2018
Actors:	Users
Description:	A User uses the search bar at the top to highlight certain data sets that contain key words in their title, classifications, or column names.
Preconditions:	1. User is logged into the system. 2. User is authorized to see the data sets they are searching for.
Postconditions:	None
Normal Flow:	1.0 Searching for Data Sets 1. A User opens the home page of the web portal. 2. A User types key words into the search bar at the top of the screen and either hits "Enter" or clicks the Search icon to the right of the search bar. 3. The system will scour through all of its available data sets and filter out data sets that do not contain the keywords in their title, classifications, or column names. 4. The system will display the results of the search, ordered by data set name alphabetically. 5. Once the user has finished, they can click "Back to Home" or the browser's "Back" button to return to the home page.
Alternative Flows:	1.1 Viewing Data Set Details (after step 4)

	1. User clicks on a data set. 2. The system will display details about the data set as well as provide the option to edit or delete the data set. (a) If the user chooses Edit, the system will begin Use Case X ¡¡replace with the ID of the use case of editing data sets¿¿ (b) If the user chooses Delete, the system will begin Use Case Y ¡¡replace with the ID of the use case of removing data sets¿¿ (c) If the user chooses View, the system will display more detailed information about the selected data set, such as classifications, column names, import date, and type of data set. 1.2 Viewing Multiple Data Set Details (after step 4) 1. User highlights multiple data sets. 2. The system will display common details between the data sets, as well as edit only common details about that data set. 3. The system will also display a "Delete" option that will allow users to delete multiple data sets at once. 1.3 No Results Found (after step 3) 1. The system displays "No items were found". 2. The system returns to step 2. 3. other stuff
Exceptions:	None
Includes:	FR-2, FR-3a, FR-3b, FR-3c, NFR-F
Priority:	High
Frequency of Use:	Medium to often, depending on the user
Business Rules:	None
Special Requirements:	1. Users shall be able to return to the home page at any time. 2. Users shall see their current search keywords at any time.
Assumptions:	None
Notes and Issues:	None

3.3 Use Case 3: Find New Relationships in Data Sets

Use Case ID:	3
Use Case Name:	Find New Relationships Between Data Sets
Created By:	Steven Bradley
Last Updated By:	Steven Bradley
Date Created:	October 9, 2018
Date Last Updated:	October 9, 2018
Actors:	Users
Description:	A user uses the 'discover' feature to find new relationships between multiple data sets.
Preconditions:	1. User is logged into the system. 2. Data sets that the user is interested in are uploaded to the database.
Postconditions:	None
Normal Flow:	1.0 Find New Relationships Between Data Sets 1. The User opens the home page of the web portal. 2. The User clicks on the discover tab of the web portal. 3. The system searches for data sets that are available to the user and displays them on the screen. 4. The user selects the data sets they are interested in by clicking a checkbox next to each data set. 5. Once the user has selected their data sets, the 'discover' button becomes clickable. 6. The user clicks on the 'discover' button. 7. The system runs the data sets through the data classifier and displays the results.

Alternative Flows:	1.1 Viewing Data Set Details (after step 3) 1. User clicks on a data set. 2. The system will display details about the data set as well as provide the option to edit or delete the data set. (a) If the user chooses Edit, the system will begin Use Case X <<replace with the ID of the use case of editing data sets>> (b) If the user chooses Delete, the system will begin Use Case Y <<replace with the ID of the use case of removing data sets>> (c) If the user chooses View, the system will display more detailed information about the selected data set, such as classifications, column names, import date, and type of data set. 1.2 User has no data sets (after step 3) 1. The system displays "No data sets found". 2. The system displays an 'add data set' button. 3. If button is clicked, then the system proceeds to Use Case Z <<Replace with ID of the use case for adding a data set>>
Exceptions:	None
Includes:	FR-4, FR-3a, FR-3b, FR-3c, NFR-F
Priority:	High
Frequency of Use:	Often
Business Rules:	None
Special Requirements:	Users shall be able to return to the home page at any time.
Assumptions:	None
Notes and Issues:	None

3.4 Use Case 4: Add Classifications to Data Sets

Use Case ID:	4
Use Case Name:	Add Classifications to Datasets
Created By:	Dustyn Zierman-Felix
Last Updated By:	Dustyn Zierman-Felix
Date Created:	October 9, 2018

Date Last Updated:	October 9, 2018
Actors:	Thomas Cadillac
Description:	Thomas Cadillac is a database manager who wishes to integrate a lot of data sets into one. He is happy with the way that —Product Name— classifies data, but would like to add classifications and rename some that the system came up with.
Preconditions:	1. Thomas is logged into the system. 2. Thomas has data sets that he wishes to integrate.
Postconditions:	None
Normal Flow:	1.0 Add Classifications to Data Sets 1. Thomas runs —Product Name— on three different data sets. 2. Thomas observes the way that —Product Name— has classified the data. 3. Thomas thinks of another classification he could add to the data. 4. Thomas adds a classification to further clarify his data.
Alternative Flow:	None
Exceptions:	None
Includes:	FR-1a, FR-3a, FR-3b, FR-3c, FR-5
Priority:	Medium
Frequency of Use:	Depends on the user.
Business Rules:	None
Special Requirements:	1. Thomas shall be able to return to the home page at any time.
Assumptions:	None
Notes and Issues:	None

3.5 Use Case 5: Display Graph of Data Sets

Use Case ID:	5
Use Case Name:	Display Graph of Data Sets
Created By:	Shane Villalpando
Last Updated By:	Shane Villalpando
Date Created:	October 10, 2018
Date Last Updated:	October 10, 2018

Actors:	Ryan
Description:	Ryan is a business owner and wants to see a graph by connecting products, manufacturers, shipments, parts, and customers all in a single graph. As a business owner, he has to analyze this data but is struggling since the data is scattered.
Preconditions:	1. Ryan is logged into the system. 2. Ryan is authorized to select the connections he wants to make and display a graph.
Postconditions:	None
Normal Flow:	1.0 Display Graph of Data Sets 1. Ryan opens the home page of the web portal. 2. Ryan clicks display graph. 3. Ryan chooses the connections he wants to make. 4. Ryan clicks display, and the graph is presented neatly.
Alternative Flow:	None
Exceptions:	None
Includes:	FR-1b, FR-2, FR-4, FR-6
Priority:	Medium
Frequency of Use:	Depends on the user
Business Rules:	None
Special Requirements:	1. Ryan shall be able to return to the home page at any time.
Assumptions:	None
Notes and Issues:	None

3.6 Use Case 6: Click to Access and Display Data

Use Case ID:	6
Use Case Name:	Click to Access and Display Data
Created By:	Larry Hu
Last Updated By:	Larry Hu
Date Created:	October 9, 2018
Date Last Updated:	October 9, 2018
Actors:	George

Description:	George is a database manager who handles patient information from different providers and clinics. he is more of a manager than an engineer and only wants to access the data to record info. Part of his job is to compile the data and update it.
Preconditions:	1. George is logged into the system. 2. George is authorized to see and edit the patient's data.
Postconditions:	None
Normal Flow:	1.0 Upload and Update Data 1. George opens the home page of the web portal. 2. George clicks on patient. 3. A drop down of the customer is listed in a table. 4. The patient's information with categories are displayed. 5. George marks the recorded customer information to save.
Alternative Flow:	None
Exceptions:	None
Includes:	FR-2, FR-3a, FR-3b, FR-3c, NFR-X
Priority:	High
Frequency of Use:	Medium to often, depending on the user
Business Rules:	None
Special Requirements:	1. George shall be able to return to the home page at any time. 2. George shall be able to search access and save any patient. 3. George can access any other patient with a click.
Assumptions:	George has security clearance for all patient information.
Notes and Issues:	None

4 System Features

4.1 System Feature 1: Searching For Data Sets

4.1.1 Description

The system will allow users to filter out unwanted data sets by providing a list of keywords to search for in a search bar.

4.1.2 Stimulus/Response Sequences

Stimulus	Response
User presses enter while focused on the search bar	The system displays all data sets
User enters keywords into the search bar and presses enter	The system displays all data sets with that keyword contained in the title, classifications, or column names
User clicks on a found data set	The system displays options and information for the data set

4.2 System Feature 2: Finding Relationships in Data Sets

4.2.1 Description and Priority

The system will allow users to select multiple data sets and find relationships between them. This feature is extremely high priority.

4.2.2 Stimulus/Response Sequences

Stimulus	Response
User clicks on 'discovery' tab.	The system displays all data sets available to user.
User clicks on check-boxes for desired data sets	The system makes the 'discover' button clickable.
User clicks on the 'discover' button	The system runs the selected data sets through the data classifier and displays the results.

4.3 System Feature 3: Display Data

4.3.1 Description

The system will allow users to display a table when accessing a customers table. (Larry)

4.3.2 Stimulus/Response Sequences

Stimulus	Response
User scrolls through the database and clicks on a customer.	The system displays a table of customer into
User clicks away	The System hides away the display table
User clicks on a found data	The system displays options and information for the data set

5 External Interface Requirements

5.1 User Interfaces

None, so far.

5.2 Hardware Interfaces

None, so far.

5.3 Software Interfaces

None, so far.

5.4 Communications Interfaces

None, so far.

6 Other Nonfunctional Requirements

6.1 Performance Requirements

NFR-A	The system shall respond to user interaction in under 5 seconds.
NFR-B	The system shall find search results in less than 5 seconds.
NFR-C	The system shall navigate through edit features with less than 5 clicks.

6.2 Safety Requirements

NFR-D	The system shall not allow third parties to see a users private information.
NFR-E	The system shall be able to process up to 100 data sets at a time.

6.3 Security Requirements

None, so far.

6.4 Software Quality Attributes

NFR-F	The system shall use common design conventions for the User Interface.
NFR-G	The system shall adhere to a similar color convention that MarkLogic's systems use.

7 Other Requirements

None, so far.

A Documents

Vision and Scope:
https://bit.ly/2IQRI9X

User Personas:
To be written later...

B Glossary

None, so far.

C Issues List

None, so far.