02/11/2016

Requirements Specification (RS)

Disability Analysis with Machine Learning & Predictive analysis

Student Name: Declan Barnes

Student Number: x13114352

Course: BSCH-Data Analytics (4th Year)

College: The National College of Ireland

Requirements Specification (RS)

Document Control

Revision History

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **Version** | **Scope of Activity** | **Prepared** | **Reviewed** | **Approved** |
| 2/11/2016 | 1 | Create | AB | X | X |
| 18/11/2016 | 2 | Update | CD | X | X |

Distribution List

|  |  |  |
| --- | --- | --- |
| **Name** | **Title** | **Version** |
| Eamon Nolan | Lecturer |  |
| Paul Stynes | Supervisor |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Table of Contents**

[Requirements Specification (RS) 1](#_Toc467247575)

[Document Control 1](#_Toc467247576)

[Revision History 1](#_Toc467247577)

[Distribution List 1](#_Toc467247578)

[1 Introduction 3](#_Toc467247579)

[1.1 Purpose 3](#_Toc467247580)

[1.2 Project Scope 3](#_Toc467247581)

[1.3 Definitions, Acronyms, and Abbreviations 3](#_Toc467247582)

[2 Data Requirements 4](#_Toc467247583)

[3 User Requirements Definition 4](#_Toc467247584)

[4 Requirements Specification 5](#_Toc467247585)

[4.1 Functional requirements 5](#_Toc467247586)

[4.1.1 Output Requirements 5](#_Toc467247587)

[4.1.2 Use Case Diagram- Overall System 6](#_Toc467247588)

[4.1.3 Requirement 1 <Disability Dashboard > 6](#_Toc467247589)

[4.1.4 Requirement 2 <Identify Health Centre locations> 8](#_Toc467247590)

[4.1.5 Requirement 3 <Predictive Analysis> 10](#_Toc467247591)

[4.2 Non-Functional Requirements 12](#_Toc467247592)

[4.2.1 Performance/Response time requirement 12](#_Toc467247593)

[4.2.2 Availability requirement 13](#_Toc467247594)

[4.2.3 Recover requirement 13](#_Toc467247595)

[4.2.4 Robustness requirement 13](#_Toc467247596)

[4.2.5 Security requirement 13](#_Toc467247597)

[4.2.6 Reliability requirement 13](#_Toc467247598)

[4.2.7 Maintainability requirement 13](#_Toc467247599)

[4.2.8 Extendibility requirement 13](#_Toc467247600)

[4.2.9 Reusability requirement 13](#_Toc467247601)

[4.2.10 Resource utilization requirement 14](#_Toc467247602)

[5 Interface requirements 14](#_Toc467247603)

[6 System Architecture 14](#_Toc467247604)

[7 System Evolution 15](#_Toc467247605)

[8 References 15](#_Toc467247606)

# Introduction

## Purpose

The purpose of this document is to set out the requirements for the development of a final year project, this project will analyze disabilities in Ireland by showing key information that will be of use to anyone looking for a breakdown on disabilities in the country. Machine Learning will also be included in the project which will involve the prediction of where new health centres should be placed in the country in order to provide better coverage for those with disabilities. The final part of the project will involve predictive analysis and will show what equipment the health centres will require in order to provide the necessary services to the disabled.

The intended users for this project is anyone who wants more information about disabilities in Ireland.

## Project Scope

The scope of the project is to develop a dashboard that will show the breakdown of disabilities in Ireland, such as the amount of male’s and female’s with a disability, the amount of disabilities by each age group, and type of disability they have. The dashboard will also display the machine learning aspect of the project, which will show the prediction of where new health centres should be placed in Ireland. The Prediction Analysis part will involve predicting what equipment the health centres will need to in order to provide adequate services to the disabled.

## Definitions, Acronyms, and Abbreviations

* **Programming Application** – A programming Application will be used to program and conduct the necessary analysis required for this project.
* **Data Visualisation Software** – Will be used to visualise the data analysis that has been conducted.
* **Storage** – Is the database that will be used to store all the datasets.
* **Cloud Storage** – Will be used to save and back up the project throughout.
* **ML** – Machine Learning is used to automate analytical model building in data analytics.

# Data Requirements

The following data requirements for this project are subject to change throughout the project. For the Disability dashboard, the data requirements are that the datasets acquired contain relevant information that will allow a decent depth of analysis to be performed, these would include: disability type, age-group, and sex. For the machine learning aspect to this project a dataset containing the latitude and longitude of the current health centres will be needed as well as further information to help with the prediction of new health centre such as: population. Lastly, for the predictive analysis, the data required will include the relevant types of equipment used by disabled people based on their disability type and needs. The following is an example of the variables needed in the datasets that will be used for the project:

**Dashboard**:

**Sex**: Male & female,

**Age** **Group**: 0-14,15-44, 25-44, 45-64, 65+,

**Types of disability:** Blindness or a serious vision impairment

Deafness or a serious hearing impairment

A condition that substantially limits one or more basic physical activities

An intellectual disability

Difficulty in learning, remembering or concentrating

Psychological or emotional condition

Other disability, including chronic illness

Difficulty in dressing, bathing or getting around inside the home

Difficulty in participating in other activities

**Machine Learning:**

**Locations**: Latitude/Longitude

**Population**: Amount by small areas

**Predictive Analysis:**

This dataset has still to be acquired.

# User Requirements Definition

The User requirements for this project is that the user should be able to look at all data associated with analysis of disabilities and the prediction of where new health centres should be placed and what disability equipment they should have in stock, that was undertaken for this project and can interact with it via a dashboard. The information shown should be easy to understand, while the dashboard itself will need to be clean and well presented for the user to understand the information being presented.

# Requirements Specification

## Functional requirements

### Output Requirements

#### Descriptive Analysis (Explorative)

Descriptive statistics, and the summarization of the data with visuals i.e. graphs: pie charts bar charts etc. Explore and analyze disabilities in Ireland using a number of variables such as: male/female, age groups, type of disability.

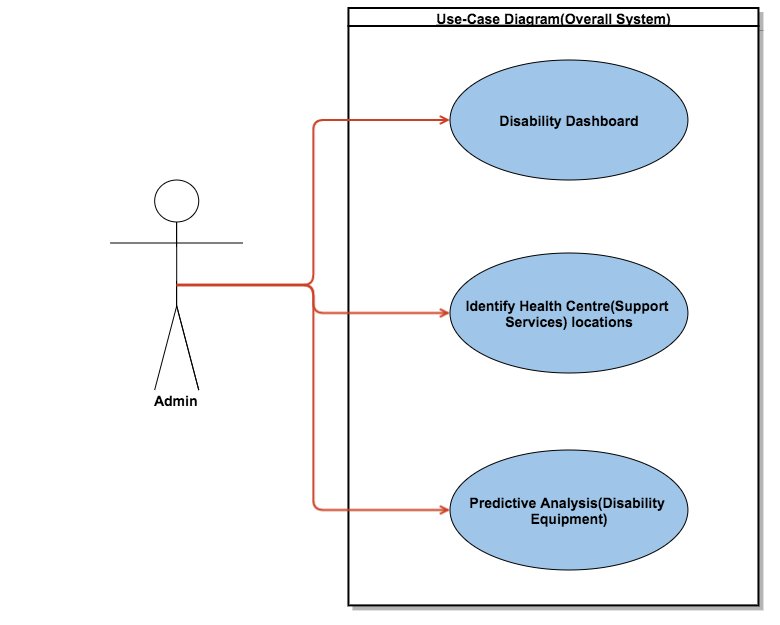
#### Machine Learning (Location Prediction)

As part of this project, new health centre locations will be predicted. By predicting where these centres should be placed there is hope that they will increase the quality of service to those with disabilities around Ireland by making more of these health centres readily available and reducing the amount of travel the disabled have to do to access these health centre services. This location prediction will be displayed on the dashboard using Tableau and it will be displayed as a spatial map.

#### Prediction Analysis (Equipment Prediction)

The prediction analysis portion of this project will involve having to predict what type of equipment the health centres should have such as, wheelchairs etc., so that the proper disability service and equipment can be provided to the disabled people attending the health centres.

### Use Case Diagram- Overall System



### Requirement 1 <Disability Dashboard >

#### Description & Priority

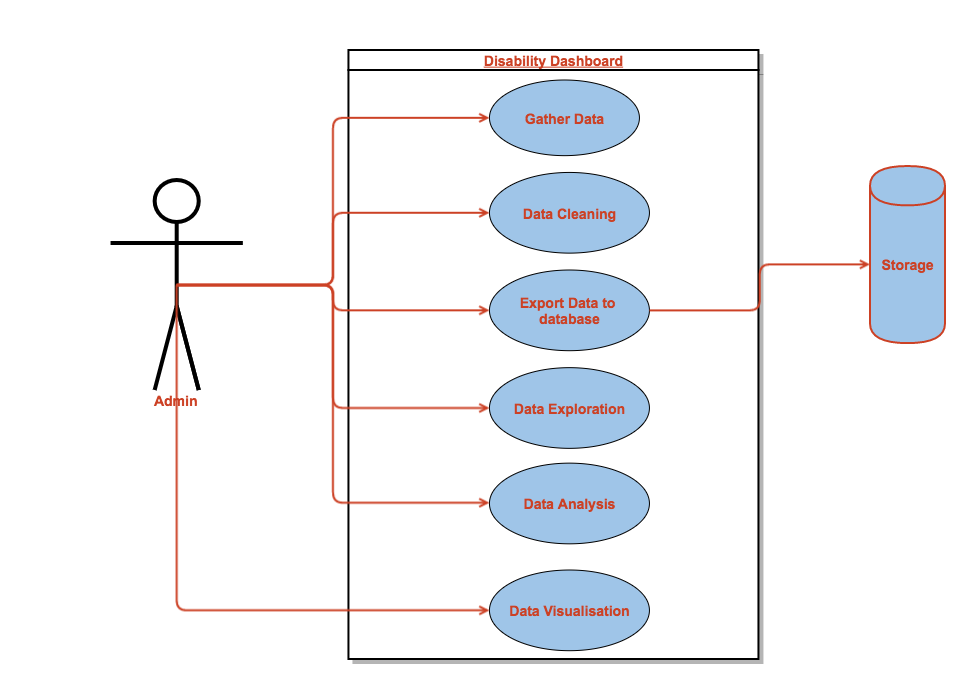
The <admin> researches and gathers the necessary datasets needed to start this project. Once the datasets have been collected the <admin> must then import them into RStudios where they will be cleaned. After the data cleaning the <admin> will import the datasets into a database for storage. Once all the datasets are in the database, the data exploration and data analysis of the data can be done. Finally the <admin> will link Tableau to the database and perform data visualisation of the data to show the findings.

#### Use Case

**Scope**

The scope of this use case is to conduct data analysis of disabilities in Ireland and visualise the analysis using a dashboard.

**Use Case Diagram**

****

**Flow Description**

**Precondition**

Research must be done by the <admin> to find out what data is required. Only then can they complete this requirement. If there is no data, then the <admin> cannot perform the data analysis required, and the dashboard cannot be created.

**Activation**

This use case starts when the <admin> completes research on what type of data is needed.

**Main flow**

1. The <admin> searches for datasets online and downloads them.
2. The <admin> performs data cleaning.
3. The <admin> exports data to the database.
4. The <admin> performs data exploration.
5. The <admin> performs data analysis.
6. The <admin> creates dashboard showing findings.

**Exceptional flow**

E1: <Corrupt data>

1. The dataset being downloaded is corrupted & unable to be used
2. The <Actor> searches & acquires a new dataset.
3. The use case continues at position 2 of the main flow

**Termination**

This use case will be terminated if there is a problem with the data that is being used. New data may need to be found.

**Post condition**

The <admin> will move onto the next stage of the project: identifying the locations of health centres that provide support services for disabled people.

### Requirement 2 <Identify Health Centre locations>

#### Description & Priority

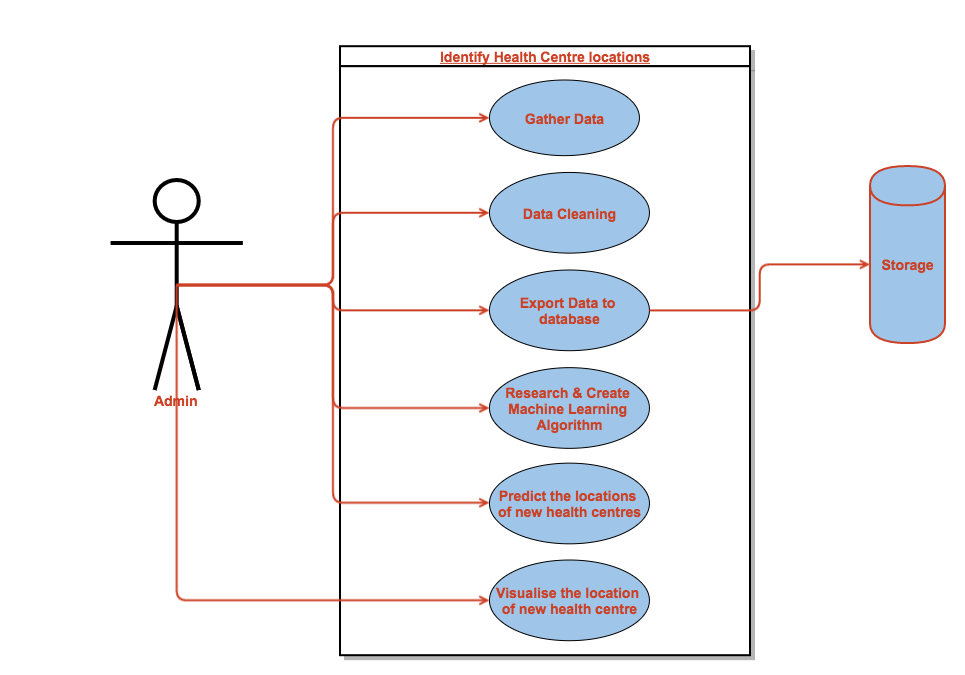
For this requirement datasets will be gathered and cleaned by the <admin>, the datasets will then be imported into the database for storage. Once the data is stored the <admin> will research two machine learning algorithms that can be combined and modified in order to create the <admins> own algorithm that will be used to predict the location of health centres that provide support services to disabled people. Once that is achieved this requirement will then be visualised with Tableau and attached to the dashboard.

#### Use Case

**Scope**

The scope of this use case is identifying the locations of health centres in Ireland that provide support services to the disabled

**Use Case Diagram**



**Flow Description**

**Precondition**

None. Requirement 1 does not need to be completed for this requirement to begin.

**Activation**

This use case starts when the <admin> has successfully researched and found the relevant datasets needed. These datasets will show the locations of current health centres, the population in each county etc.

**Main flow**

1. The <admin> searches for datasets online and downloads them.
2. The <admin> performs data cleaning.
3. The <admin> exports data to the database.
4. The <admin> performs research into machine learning algorithms and creates a machine learning algorithm for this requirement.
5. The <admin> identifies the locations of the new health centres using the created ML algorithm.
6. The <admin> visualises the health centre locations with Tableau.

**Exceptional flow**

E1: < Machine Learning Algorithm Error >

1. The system cannot identify health centre location using the algorithm.
2. The <Actor> troubleshoots & modifies the algorithm to fix the error.
3. The use case continues at position 5 of the main flow

**Termination**

This use case will be aborted if there is no data to be found, and if there is a problem with the database.

**Post condition**

The <admin> now has everything required to move onto the next stage of the project: Predictive Analysis.

### Requirement 3 <Predictive Analysis>

#### Description & Priority

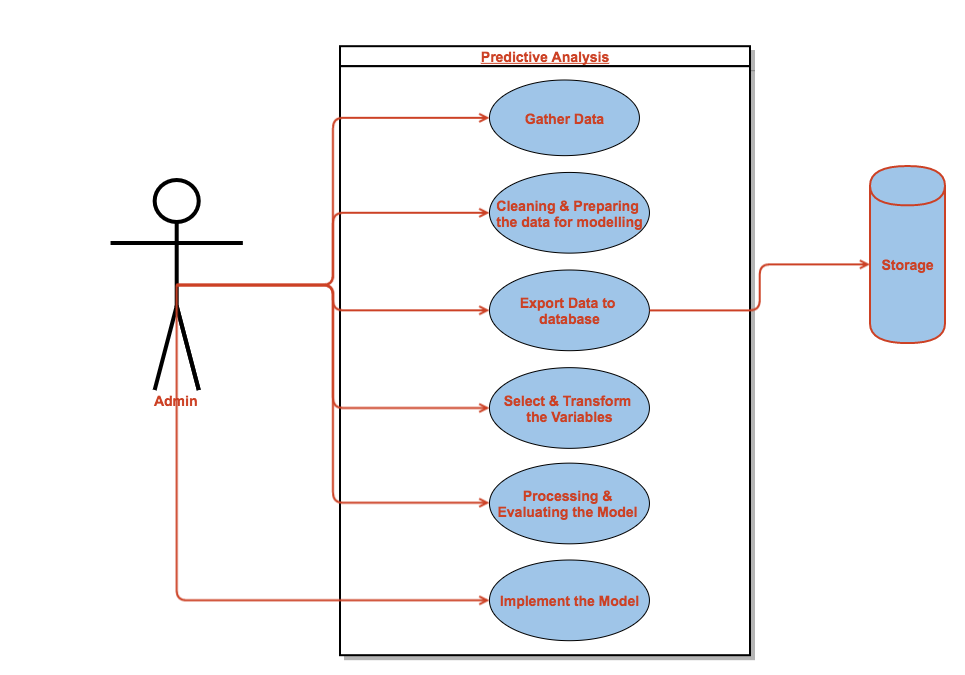
The purpose of this requirement is to use predictive analysis to predict the equipment that is required at each health centre location from requirement 2. This is to ensure the health centres in question will have the necessary equipment to support disabled people. The requirement will be completed as follows: The <admin> research and gather the datasets needed to start this requirement, once these datasets have been acquired the <admin> will clean and prepare the data for modelling. After this has been done the <admin> will then import the data into the database. Once the data is stored it will be accessed by RStudios and the <admin> will then select and transform the data’s variables. After which the model will be processed and evaluated before being implemented.

#### Use Case

**Scope**

The scope of this use case is to predict the equipment require at each health centre that provides support services to the disabled.

**Use Case Diagram**



**Flow Description**

**Precondition**

Requirements 2 will need to be completed before this requirement can be started. To conduct predictive analysis on the health centres they need to be located first and foremost.

**Activation**

This use case starts when the <admin> has successfully found datasets that will be of use for this requirement.

**Main flow**

1. The <admin> searches for datasets online & downloads them.
2. The <admin> cleans & prepares the data for modelling.
3. The <admin> exports data to the database.
4. The data’s variables are selected & transformed by the <admin>
5. The <admin> processes & evaluates the model
6. The model is implemented by the <admin>.

**Exceptional flow**

E1: <Model Implementation Problem>

1. The Predictive Model is incorrect and can’t be implemented by the <admin>.
2. The <Actor> must re-evaluate the model and fix the problem that has occurred.
3. The use case continues at position 4 of the main flow

**Termination**

This use case will be aborted if the <admin> has found no dataset that can be used for this requirement.

**Post condition**

N/A

## Non-Functional Requirements

### Performance/Response time requirement

Even though performance and response time is an important aspect of many systems, it does not apply to my project as the volume of data being used as part of the data analysis does not need to be analysed at a fast speed or with any real urgency. Therefore, the speed of performance and response time is not an important aspect to this project.

### Availability requirement

The data will remain available to the system throughout.

### Recover requirement

In the event of a hardware or software failure the data and database being used for this system will need to be backed up on a regular basis. The data folders containing the datasets and database schema’s will be backed up on GitHub using version control. The data folder will also be copied and saved to multiple locations such as google drive and Dropbox for extra safety.

### Robustness requirement

The robustness requirement does not apply to this system.

### Security requirement

The data has been acquired from several websites such as the Central Statistics Office which is a public site, the datasets that I have downloaded from the CSO site are all publically available and require no special permissions to use them. The system itself can only be accessed on my personal laptop including the MySQL database which is protected with a username and password. The database is also hosted using gearhost.com which is also being protected by a username and password.

### Reliability requirement

The data that is being used for this system came from the Central Statistics Office which compiles this data on a yearly basis. Therefore, the system is dependent on the accuracy and availability of data produced by the CSO.

### Maintainability requirement

This system does not necessarily need to be maintained once it is created. It may need to be updated with new data in the future, but for this final year project the data that will be used has already been acquired.

### Extendibility requirement

At the time of this requirement specification, there is no plan to further the study at this, but the methods and processes that are being used could be used again in the future.

### Reusability requirement

This system currently has no requirement for reusability.

### Resource utilization requirement

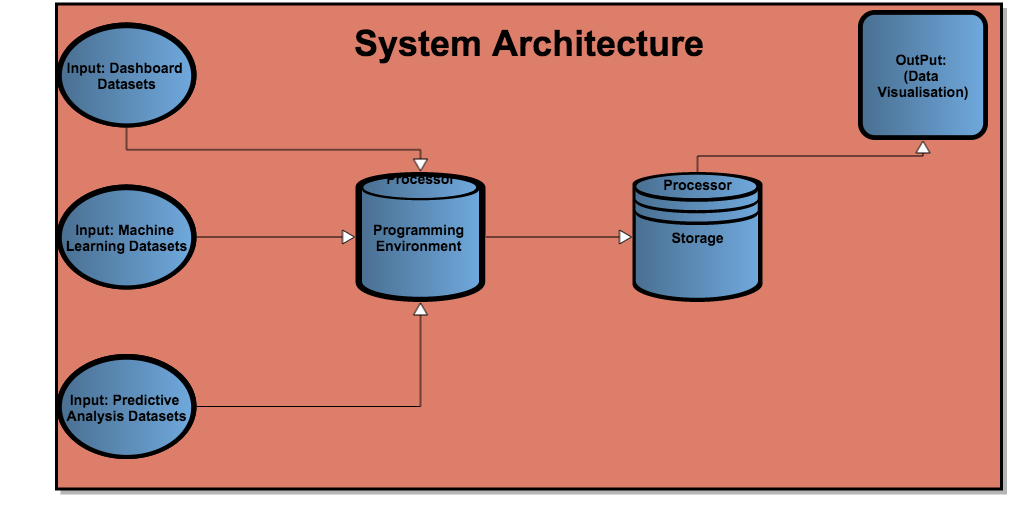
There is a need for a laptop, and cloud storage for backups. Some of the software that will be used for this system still needs to be installed, they are: RStudio, Tableau, MySQL Workbench, Gearhost, Google Drive and Dropbox.

# Interface requirements

This requirement does not apply to this system now, but there is a scope to add an end user interface for this system at a future date if the user would like to few specific statistics pertaining to disabilities and health centres in Ireland.

# System Architecture

This section describes the system architecture and shows the components that make up the project and show how they interact with each other and how the extraction of the data works. The components that make up this project include an input source which is the datasets that will be acquired from various data sources such as websites etc. There is a MySQL database component which will be used to store the datasets as tables and RStudios will be used to connect to the database to access the data and conduct all the analysis required for the project. After all the analysis is complete the datasets will be re-stored to the database and visualized using Tableau. All the components listed above make up the key aspects of the system architecture and they are subject to change as the project progresses.



# System Evolution

This system could evolve in the future to include further analysis if more datasets are acquired. There could also be further machine learning added, for example a prediction of learning disabilities in school children. Another way this system could evolve is by expanding the geographic focus past Ireland and have it conduct analysis of disabilities in the whole of Europe, and North America. By expanding the focus, it will allow the comparing of disability statistics from several different countries or regions to analyze the difference if there is one in the amount of disabilities affecting men, women and children of various types disabilities.

# References

"An Exploratory Study into The Association Between School Expenditure Levels Per Student and School Performance in High Stake Examinations - TRAP@NCI". *Trap.ncirl.ie*. N.p., 2016. Web. 18 Nov. 2016.

"Data.Gov.Ie — Ireland's Open Data Portal". *Data.gov.ie*. N.p., 2016. Web. 18 Nov. 2016.

"National Disability Survey 2006 - First Results - CSO - Central Statistics Office". *Cso.ie*. N.p., 2016. Web. 18 Nov. 2016.

Cran. (2016). *sqldf-Package*. [online] Available at: https://cran.r-project.org/web/packages/sqldf/sqldf.pdf [Accessed 30 Nov. 2016].

En.wikipedia.org. (2016). *RStudio*. [online] Available at: https://en.wikipedia.org/wiki/RStudio [Accessed 30 Nov. 2016].