Literature Review

<< Project title>>

<< Team members>>

<<Submission date>>

Table of Contents

**Contents**

[**1.** **Introduction** 4](#_Toc33006866)

[**1.1.** **<<** **Core technical area 1>** 4](#_Toc33006867)

[**1.1.1.** **Sub technology 1 - core technical area 1** 4](#_Toc33006868)

[**1.1.2.** **Sub technology 2 - core technical area 1** 4](#_Toc33006869)

[**1.2.** **<<** **Core technical area 2>** 4](#_Toc33006870)

[**1.2.1.** **Sub technology 1 - core technical area 2** 4](#_Toc33006871)

[**1.2.2.** **Sub technology 2 - core technical area 2** 4](#_Toc33006872)

[**6.** **Conclusion** 4](#_Toc33006873)

Table of Figures

# **Introduction**

<< This section should outline the context of the work. What is the real world problem you are trying to address? What are the quantifiable issues which need to be overcome?>>

<<The section should state the major categories within which the literature review will be scoped>>

<<The section should provide high level reference into which the subsections interrelate>>

# **<<** **Core technical area 1>**

<< For core technical area 1 this section should outline the relevance of the technical area to the overall research agenda. Why is this technical area being discussed? How can this area be sub-categorised? How can these technologies address the overall challenge discussed in section 1?>>

# **Sub technology 1 - core technical area 1**

# **Sub technology 2 - core technical area 1**

# **Big Data**

Big Data is defined as “data that contains greater variety, arriving in increasing volumes and with more velocity” [1]. In the past data was stored mostly on paper records of local digital storage devices, if at all. As such the availability of data was much reduced when compared with today. In essence the dramatic increase in data collection then storage in a shareable way as potentially provides scientists a wealth of insights. With this technological leap scientists can enrich the data they collect themselves through observations and experiments with open-source data available collected from a multitude of organisations and institutions.

For epidemiologists, who undertake systematic, data-informed, investigations into the biology of pathogens and their complex interactions with their hosts to identify changes in behaviour that could contain an outbreak reduce or eliminate the spread of disease, Big Data can potentially be utilised to predict the development of an epidemic as well as the effectiveness of control measures.

Potentially valuable Data sources can be broadly placed in three categories.

Firstly, healthcare Big Data is generally collected from patients who have presented with disease. It will provide information related to age, gender and may be enriched with data relating to medical health, pre-existing conditions, vaccination status etc. This information is of course collected after an outbreak of disease has occurred, but it is specific to the disease as will contain information for individuals who have been proven to be infected. The data is normally collected by health care providers and government agencies. The parameters collected are often predefined and may be slow to be updated. The data collected can vary in content and completeness between organisations collecting it.

The second relevant Big Data category relates to the population and can be described as population demographics and population socio-economics. Demographic information will typically be collected by governments through a census. It describes the population geospatially, by ethnicity, religion, gender, and age. Socio-economics will be collected by government and other institutons. It is related to wealth of the individual and family. I have defined demographics and socio-economics in one category as they are often inter-related. For instance, within a given region of a given country individuals of a given ethnicity/ gender description may, on aggregate, fall have statistically significant differences in health than other individuals of other ethnicities/ genders. Of course, this data would not be considered specific to an outbreak of disease, but it can be relevant when epidemiologists identify risk factors associated with a disease.

The final relevant Big Data category is

<< For core technical area 2 this section should outline the relevance of the technical area to the overall research agenda. Why is this technical area being discussed? How can this area be sub-categorised? How can these technologies address the overall challenge discussed in section 1?>>

# **Healthcare Big Data analytics**

Ambigavathi and Sridharan [11] investigate Big Data Analytics in healthcare, studying its unique characteristics, different analytics phases, platforms, and tools used, challenges and their conclusions. Ending with a discussion of its further evolution. The unique characteristics, commonly referred to as V’s, describe the data

# **Demographic and Socio-economic Big Data**

Dsfdsfds

# **Social Media Big Data**

.

.

.

# **Conclusion**

<< This is a critical part of the literature review. Briefly state again the overall context of the work. Briefly state the literature review process including technical core areas and subcategories. Compare and contrast the relevance of the technologies under discussion. Compare and contrast the potential for existing technologies to address the overall research challenge. Identify limitations in existing approaches. Discuss future work which can enhance the existing state of the art in order to address the overall research challenge>>