

# Design and Implementation of Public Data Warehouse

Nurul Farhanah Isha<sup>1</sup>, Nor Azuana Ramli<sup>1, a)</sup> and Qistina Batrisyia Azman Shah<sup>2, b)</sup>

<sup>1</sup>*Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah,  
Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang, Malaysia.*

<sup>2</sup>*Credence, 1 Jalan Damansara, Damansara Kim,  
60000 W.P. Kuala Lumpur, Malaysia.*

## Author Emails

<sup>a)</sup> Corresponding author: [azuana@umpsa.edu.my](mailto:azuana@umpsa.edu.my)

<sup>b)</sup> [qistina.azmanshah@credence.tech](mailto:qistina.azmanshah@credence.tech)

**Abstract.** There is a large amount of public data available that is spread across various sources. However, these public data is spread across different government bodies' websites and repository platforms. Data must be collected and stored in a warehouse for utilization. Hence, this research is conducted to design and implement an appropriate data warehouse that is able to merge all public data from different government bodies and public websites in one data warehouse. This research started by locating relevant sources from various open data platforms such as Public Info Banjir and the Official Portal iDengue for community version 3.0. Then, significant data were collected such as river water level data, rainfall data, dengue locality cases and weather data. A public data warehouse was constructed using the bottom-up approach of the Kimball methodology. Following this, a galaxy schema model was identified to design the fact and dimension tables, which is then converted into a logical model. Finally, the structured data warehouse was developed. The public data warehouse, containing data such as water level, rainfall, dengue locality cases and weather data will be of great benefit to both government bodies and the public. For future study, the data warehouse enables the development of advanced analytics and predictive models, which helps to enhance decision-making and optimize infrastructure management in relation to flood mitigation efforts.

## INTRODUCTION

There has been a dramatic surge in data in recent years. As [1] explains, data is a value assigned to an object, which can be seen or verified. When used properly, with a focus on data protection and privacy, data can help streamline service delivery, reduce human error, and influence significant operational improvements. Due to the challenges posed in managing this increase in data volume, most organizations and even governments have yet to fully utilize the richness of readily available data resources. Managing the ever-increasing amount of data has presented numerous challenges, leading many organizations and governments to remain unable to fully take advantage of the wealth of data available to them. A major issue is access, as the reliance on multiple databases related to specific functions rather than being linked together makes it difficult to make the most of the data [2]. Hence, it is critical for both public and private organizations to make better use of their data resources.

The term "public data" is used to describe information that is free for anyone to use, repurpose, and redistribute, provided that the sharing and attribution requirements are met [1]. Governments now provide public access to some data through open government platforms for citizens to retrieve publicly available government data. One example of a data supplier in the Malaysian Open Data Portal ([data.gov.my](http://data.gov.my)) is the Department of Statistics, Malaysia (DOSM). It is a single point of access where citizens may get publicly available government data online. Public data, also known as open government data, is a vast resource that has mostly gone unused up to this point [1]. Many people and organizations, including the government, can benefit from access to public data.

With the increasing demand for data, the issue of data management has become increasingly pressing. Data is often stored in separate datasets and managed by different government bodies for specific uses, leading to data being dispersed across multiple organizational silos. This raises the question of how data formats and standards can be kept