REKAYASA
Journal of Science and Technology
https://journal.trunojoyo.ac.id/rekayasa

Rekayasa, 2023; 16(2): 219-225
ISSN: 0216-9495 (Print)
ISSN: 2502-5325 (Online)

The Implementation of *Bantu Warga Apps* and Metabase for Spatial Data Processing In Surabaya

Alfandino Rasyid[1], Sritrusta Sukaridhoto[1*], Muhammad Agus Zainuddin[1],
Rizqi Putri Nourma Budiarti[2], Ubaidillah Zuhdi[2], Agus Imam Sonhaji[3],
Yohanes Yohanie Fridelin Panduman[4], Luqmanul Hakim Iksan[1]
[1]Politeknik Elektronika Negeri Surabaya, Jl. Raya ITS Kampus PENS Kota Surabaya 6011 Jawa Timur
[2]Universitas Nahdlatul Ulama Surabaya, Jl. Jemursari No. 51-57 Kota Surabaya 60237
[3]Badan Perencanaan Pembangunan Kota Surabaya, Jl. Pacar No. 8 Ketabang 60272 Kota Surabaya
[4] Okayama University, 1 Chome-1-1 Tsushimanaka Okayama Japan
[*]dhoto@pens.ac.id
DOI: https://doi.org/10.21107/rekayasa.v16i2.20509

Abstrak

Civil Registration and Vital Statistics (CRVS) dikenal sebagai sistem pendukung penting untuk pembuatan kebijakan dan penyampaian layanan yang efektif. Tujuan utamanya adalah menetapkan identitas hukum individu berserta semua peristiwa pentingnya untuk memastikan semua warga negara mendapatkan haknya dengan baik. Selanjutnya, CRVS tingkat lanjut dilengkapi dengan sistem prediksi untuk menyajikan data yang lebih bermanfaat. Penelitian ini membahas potensi manfaat dan tantangan penggunaan Metabase dan Bantu Warga sebagai sistem pendukung kependudukan. Sistem visualisasi data spasial menggunakan open-source Metabase berjalan dengan baik dan menghasilkan banyak visualisasi peta geospasial seperti kelahiran, perkawinan dan perceraian, pendidikan, dan pekerjaan. Konfigurasi model forecasting autoregression dan dataset sum kumulatif mencapai hasil error terbaik pada 1056618. Aspek lainnya, seperti sistem manajemen isolasi mandiri, sistem klasifikasi dokumen kependudukan, dan sistem OCR, juga memiliki hasil yang baik.

Kata Kunci : sistem pendukung, kependudukan, metabase, aplikasi Bantu warga

Abstract

Civil registration and vital statistics (CRVS) is known as an essential supporting system for effective policy-making and service delivery. Its primary purpose is establishing an individual's legal identity and vital events to ensure all citizens get their rights well. Furthermore, an advanced level of CRVS is equipped with a forecasting system to serve more useful data. This paper discusses the potential benefits and challenges of using Metabase and Bantu Warga as civil ministry support systems. The spatial data visualization system using open-source Metabase ran well and resulted in many visualization geospatial maps like birth, marriage and divorce, education, and employment. The forecasting autoregression model configuration and cumulative sum dataset reach the best error result in 1056618. The other aspects, like the self-isolation management system, civil documents classification system, and OCR system, have a good result.

Key words : *support system, demography, metabase, Bantu Warga appss*

INTRODUCTION

Civil registration and vital statistics (CRVS) systems are crucial in maintaining accurate records of births, marriages, and employment and are essential for effective policy-making and service delivery. Its primary purpose is establishing an individual's legal identity and vital events to ensure all citizens get their rights well. Furthermore, an advanced level of CRVS is equipped with a forecasting system to serve more powerful data. A comprehensive CRVS system can make population dynamics easier, affecting urban planning and resource allocation, especially in recovering from the COVID-19 pandemic (Cobos Muñoz *et al.*, 2020).

In many countries, CRVS systems still need to be more cohesive and better integrated, resulting in data availability gaps (Adair *et al.,* 2020). Moreover, the data collected by CRVS systems are

Article History:
Received: June, 16th 2022; **Accepted**: July, 29th 2023
Cite this as :
Rasyid et al. 2023. The Implementation of Bantu Warga Apps and Metabase for Spatial Data Processing in Surabaya. Rekayasa. Vol 16(2). 219-225.

rarely linked with spatial data, which can provide valuable insights into population dynamics. The absence of a functional CRVS system can lead to gaps in the availability of accurate and timely data related to births, marriages, divorce, et cetera. This scheme can hinder the effective planning and delivery of services and evidence-based policy-making. However, making the CRVS system is also not an easy task, especially for a developing country like Indonesia (Cobos Muñoz *et al.,* 2020; Gizaw, 2020; Mills *et al.,* 2019; Muñoz *et al.,* 2018; Musadad & Kelly, 2023; Ukoji *et al.,* 2019; Yokobori *et al.,* 2021).

There is much research that has been done related to this study. Abou Zahr C. *et a*l. (2018) provides that civil registration and vital statistics systems provide a continuous, real-time, and locally representative flow of information. This system can easily make evidence-based government policies to plan and monitor health and social development progress (Abou Zahr *et al.,* 2019). While Ukoji *et al* (2019) examine the civil register and vital statistics in good governance in Nigeria and its challenges. The country has a poor quality of sociodemographic data, so it focused on how to make CRVS a robust system to make a high-quality sociodemographic data source. Therefore this research still needs more investment to scale up the CRVS system to make more good national policy-making for good governance (Ukoji *et al.,* 2019;).

On the artificial intelligence (AI) side there are Pastefanopoulos et al. (2020) discussed how to forecast the spread of COVID-16. This study aims to make significant preventive steps to strengthen healthcare facilities (Papastefanopoulos *et al.,* 2020). Bukuysahin *et al.* (2019) improve forecasting accuracy using ARIMA-ANN. This study aims to know the result of linear, non-linear, and a combination of both models. Several methods were tested, so ARIMA-ANN, which is the proposed method. This discussion is based on the advantage of combining the linear model, which is proper for stationary data, and the non-linear model, which is proper for non-stationary data (Büyükşahin & Ertekin, 2019). Audebert *et al.* (2019) made a system to classify a printed document. The dataset used is printed documents from tobacco industries by extracting the text from the documents and then classifying them (Audebert *et al.,* 2019). Ling *et al* (2020) also used Optical Character Recognition (OCR) technology to extract metadata from scientific images (Ling *et al.,* 2020).

For spatial data processing, there are Daniel Balla *et al* (2020) discussed performing several spatial data visualization tools. The tested tools are Keyhole Markup Language (KML) and Quantum GIS (QGIS), implemented in the cases of nitrate and phosphate contamination scope in east Hungary. The result is similar in all aspects except the data limitation, which is that KML has more defined limitations than QGIS (Balla *et al.,* 2020). The other research from Hayatpur *et al* (2020) discussed a visualization of spatial data in a virtual reality platform named DataHop. Spatial data visualization can be easier to explore with an unlimited canvas of data visualization on the virtual reality platform. It was beneficial, especially for multidimensional datasets. Priambodo *et al* (2020) built a digital system to manage a self-isolation program. With an IoT concept, he used an oximeter to detect the patient's oxygen saturation level. The oximeter's data is sent to the smartphone then the smartphone sends the oximeter data (along with the patient ID and patient to the server. Finally, the cloud can help servers visualize the data (Priambodo & Kadarina, 2020).

All the past research also conducts the foundation of this research. The solution we build in this work scoop combines Bantu Warga and Metabase. Bantu Warga is a multi-service application to assist the civil ministry in Surabaya (Dispendukcapil Surabaya) with a self-isolation management system to help the civil officer to monitor the self-isolation patients, civil documents classification to assist civil officers in classifying the document of each citizen, and also Optical Character Recognition (OCR) for extracting the civil ID from the civil documents. Besides doing a helpful civil service, Bantu Warga was also designed to collect spatial data from the smallest administrative area. Also, spatial data processing can process make forecasting to make better-supporting data. At the same time, Metabase is an open-source platform data visualization system that can visualize spatial data into a region heatmap. By combining Bantu Warga and Metabase, it is possible to gain insights into population dynamics, such as live birth, marriage, divorce, education, and employment trends, which is the several aspects of CRVS standard (Mills *et al.,* 2019).

This paper discusses the potential benefits and challenges of using Metabase and Bantu Warga as civil ministry support systems for making a spatial data visualization system. We present case studies from Surabaya, Indonesia, to illustrate the benefits of this implementation and highlight best practices and lessons learned. Overall, the highlights are a system to support the self-isolation process, civil documents classification, OCR civil ID extractor, spatial data visualization, and population growth forecast system to become an evidence-based decision-making support system for the government, especially Dispendukcapil Surabaya.

METHODS

System Design

This research has two main features and three supporting features. The main features are spatial data visualization and the data forecasting system, and the supporting feature is a self-isolation management system, civil documents classification, and an OCR system. The whole feature is displayed in Picture 1.

Metabase

Metabase is a powerful platform for managing and visualizing data. It has many graph modes and a spatial region heatmap feature. With just an open-source version of Metabase, it was enough to make an excellent spatial visualization dashboard. Metabase use geoJSON is used in this system as a spatial data support element in the Metabase platform. Our geoJSON data contains detailed information for marking territory in Surabaya. It includes information on territory boundaries information until the district and sub-district level. The Metabase needs the boundary information with each area label to make the region heatmap visualization.

Civil Live Birth Forecasting

Spatial data forecasting in this research is used to estimate live birth data. This sector is called time series analytics and often uses just one column for the prediction target and the attributes. The methods used autoregression, this method is used several previous data to predict the targeted data. For example, if using 4 data attribute, to predict the X_t data needs X_{t-1} until X_{t-4} data to be considered. The autoregression tested with several sets up to be tested. The model is configured in several configurations and combined with several modes of the dataset, making a performance matrix to know the best configuration in this case. The dataset is divided into two parts: the train dataset to train the model and the test dataset to know the quality of each model.

Civil Data Base

Dinas Kependudukan dan Pencatatan Sipil Surabaya (Dispendukcapil Surabaya) is a organization on a government that. On a CRVS scheme, this organization naturally participates in the civil registration section. So they have a substantial civil database. The database is replicated for this research to the university's cloud storage. Combined with the Metabase, the spatial information within the database can be visualized in a geospatial interface. The data structure for the Indonesian citizen is displayed in Figure 2.

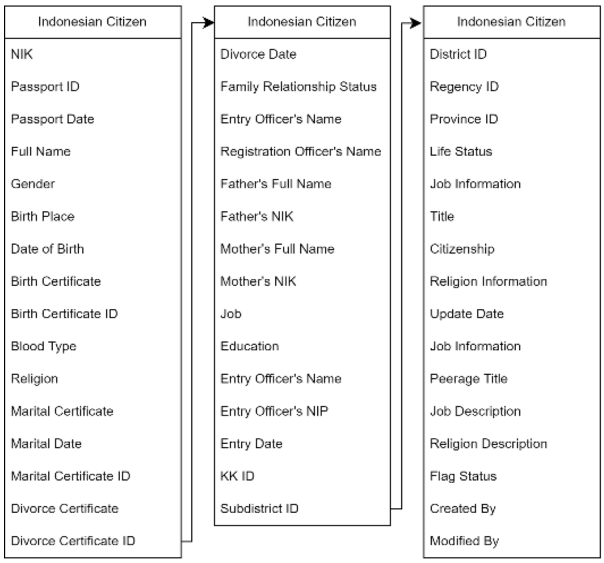


Figure 2. Dispendukcapil Surabaya's database structure

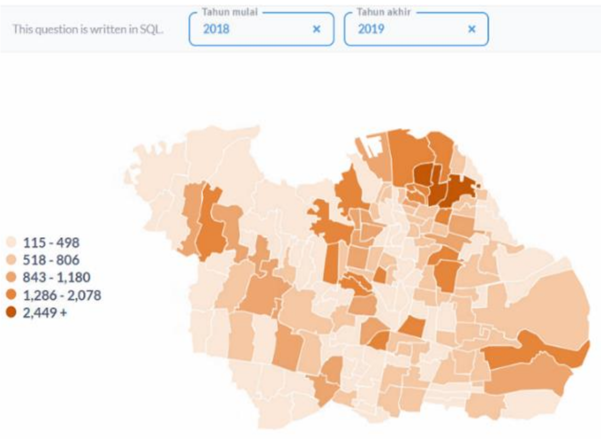
Elasticsearch

Real-time graph visualization is needed to show the oxygen saturation data communicatively. To make this real-time graph, a timeline-based database is needed. One of them is Elasticsearch. Elastic search is a comprehensive platform with the API to make the data more accessible from other systems. API streams the data to the flutter-based graph visualization.

RESULTS AND DISCUSSION

Spatial Data Visualization

Live birth rate and distribution is one of the essential aspects that the government must understand. By knowing the live birth rate, the government can read the warfare rates of the society. They can be processed to have higher and more distributed warfare. The regional heatmap visualization in Figure 3 shows birth rate data for each area. The data displayed is limited only from 2018 to 2019 through the data filter feature at the top of the visualization view. From the selected range of years of birth, it can be shown that the lowest birth rate is 115, and the highest birth rate is more than 2449.

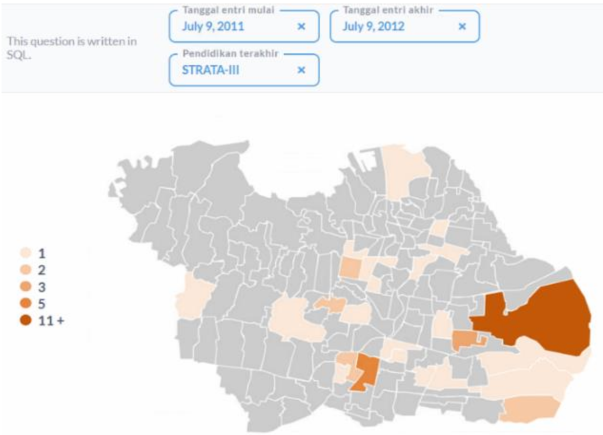



```
Algorithm 1: Live birth region heatmap visualization
  Data: Spatial Data
  Result: Live birth region heatmap visualization
  Maps preparation;
  Database connect;
  while input table do
      while table joining do
          joining the subdistrict code table;
          subdistrict.name ← code of subdistrict + district + city + province;
      end
      while filtering do
          birht.date.range ← userinput;
          filtering based on birht.date.range;
      end
  end
  changing table point of view based on subdistrict.name;
  while count aggregates do
      Aggregating each subdistrict.name;
      total ← subdistrict.name.aggregates;
  end
  Table with subdistrict.name & total;
  Maps data implementation;
```

Figure 3. Live Birth Region Heatmap Visualization

The process is carried out by preparing the required maps, after that connecting the visualization system with the spatial database. Then inputting tables from the database into the visualization system is carried out. The process is carried out by doing a join between the main table and the subdistrict code table, after that, filtering, aggregating, and finally visualizing it into the map that has been prepared.

Another example is the education level and distribution map. The education level shown in Fugure 4 shows that the various education levels can also be aggregated based on their area. The filter feature consists of the start entry date and end entry date to select the time range of the data and the Last Education filter to select what education level data want to display.




```
Algorithm 2: Education level region heatmap visualization
  Data: Spatial Data
  Result: Education level region heatmap visualization
  Maps preparation;
  Database connect;
  while input table do
      while table joining do
          joining the subdistrict code table;
          subdistrict.name ← code of subdistrict + district + city + province;
          joining the education level code table;
          education.level ← code of education level;
      end
      while filtering do
          education.level ← userinput;
          entry.date.range ← userinput;
          filtering based on education.level & entry.date.range;
      end
  end
  changing table point of view based on subdistrict.name;
  while count aggregates do
      Aggregating each subdistrict.name;
      total ← subdistrict.name.aggregates;
  end
  Table with subdistrict.name & total;
  Maps data implementation;
```

Figure 4. Education Region Heatmap Visualization

Figure 5 Education Detail Bar Chart Visualization

For the different points of view in the education rate data, the metabase also can make a visualization in a neat bar chart. The data counted in a different aspect: the total of each education level in a specific district. Picture 6 shows that the parameter that can be chosen is not an education level anymore but a sub-district. From the selected range of entry dates and the sub-district, it can be shown that in the Keputih sub-district, the highest education level is no/yet school, and the second is senior high school.

Data Forecasting System

Forecasting is used to predict future data, especially for this study, which focuses on live births. This study uses the Autoregression method for doing forecasts. The autoregression predicts the new data based on several data before as an attribute, and of course, as a supervised learning model, the autoregression needs to be trained first. Different schemes of the dataset have also been tested, which are the normal scheme and extracted cumsum scheme.

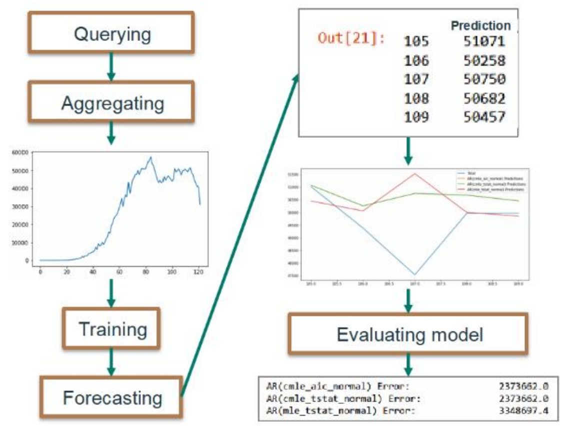


Figure 6. Normal Forecasting Scheme

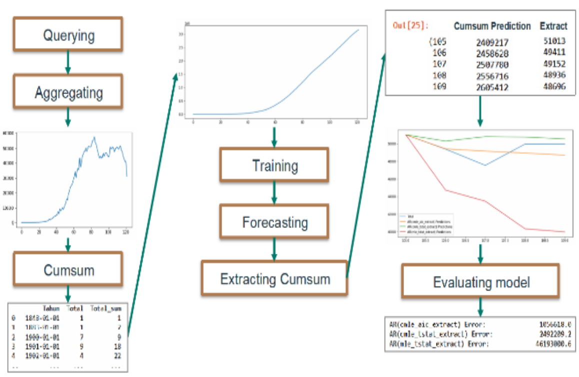


Figure 7. Extracted Cumulative Sum Scheme

Figure 6 shown the normal scheme. In a normal scheme, the dataset is the number of live births each year. After the data is forecasted, it will be compared to the original value of the testing dataset. Then, the system will take the error value for evaluation. Otherwise, Figure 7 shown the extracted cumsum scheme. Cumsum means cumulative sum, a scheme in which the original aggregated data is accumulated. After the training, it will be extracted to the normal form again. After the data is forecasted, it will be compared to the original value of the testing dataset. Then, the system will take the error value for evaluation.

Figure 8 shows the differences between the tested configuration. The forecasted data is the birth rate from 2003 to 2007. As a result, almost all the data need to meet the test dataset as a reference. The forecasted data are also the birth rates from 2003 to 2007. The experiment shows that almost all configurations meet the tested dataset in the first year. Then there is one configuration that meets the tested dataset in two years, which was the cmle + aic configuration.

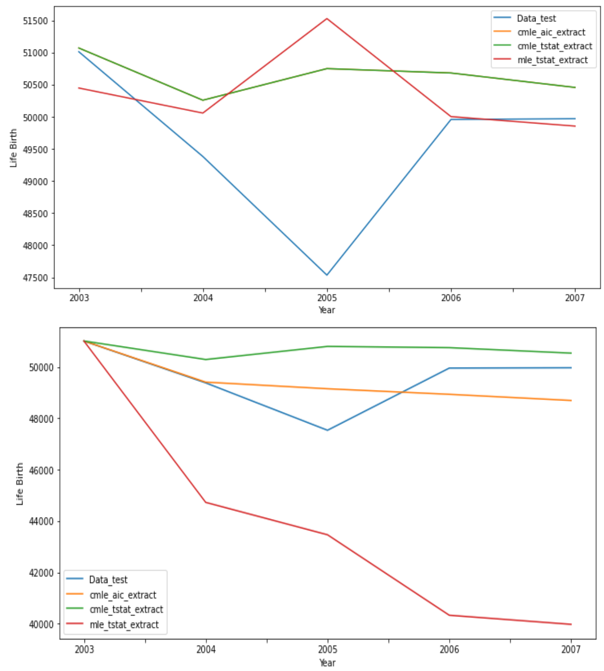


Figure 8 Differences Between The Tested Configuration

Some testing was held to know the best model configuration for this case. Several configurations are as cmle + aic, cmle + t-stat, and mle + t-stat, also In the data section, there are two kinds of pre-processed data: the normal data and the extracted cumsum data. The specific data is an original birth dataset. It produces the original birth forecasting data. Then the other is extracted cumsum, a forecasting data extracted from the cumulative sum dataset. The extracted cumsum schema is conducted from an insight that the data trends of cumsum data were relatively stable so that this dataset can produce more accurate forecasting data. The result shows that the extracted cumsum schema has the best prediction value when processed with cmle + t-stat configuration. It is marked with the lowest error.

CONCLUSION

The design and implementation of the Bantu Warga Application combined with the open-source metabase are suited well to the needs of Dispendukcapil Surabaya. The spatial data visualization system using open-source Metabase ran well and resulted from many visualization geospatial maps like birth, marriage, divorce, education, also employment. The forecasting autoregression model configuration and cumulative sum dataset reach the best error result in 1056618. The self-isolation management system works well with the nurses and the COVID patients/ex-patients. The civil documents classification system reached a 0.68 accuracy level using DenseNet 201 architecture with 300 epochs, 0.0001 learning rate, and 1 batch size for the experiment set in some datasets is just 390 training data separated into 26 classes. Based on The confusion matrix, the model has 11 classes that reach above 70% correct classification and 15 below 70% correct classification. The OCR system works well inside to assist the civil documents classification system. All these features were an excellent element in the CRVS scheme to make more robust evidence-based policy in every specific district.

REFERENCES

Abou Zahr, C., Mathenge, G., Brøndsted Sejersen, T., & Macfarlane, S. B. (2019). Civil Registration and Vital Statistics: A Unique Source of Data for Policy. In *The Palgrave Handbook of Global Health Data Methods for Policy and Practice* (pp. 125–144). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-54984-6_7

Adair, T., Richards, N., Streatfield, A., Rajasekhar, M., McLaughlin, D., & Lopez, A. D. (2020). Addressing critical knowledge and capacity gaps to sustain CRVS system development. *BMC Medicine*, *18*(1), 1–6. https://doi.org/10.1186/s12916-020-01523-y

Audebert, N., Herold, C., Slimani, K., & Vidal, C. (2019). *Multimodal deep networks for text and image-based document classification*. http://arxiv.org/abs/1907.06370

Balla, D., Zichar, M., Tóth, R., Kiss, E., Karancsi, G., & Mester, T. (2020). Geovisualization techniques of spatial environmental data using different visualization tools. *Applied Sciences (Switzerland)*, *10*(19). https://doi.org/10.3390/APP10196701

Büyükşahin, Ü. Ç., & Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing*, *361*, 151–163. https://doi.org/10.1016/j.neucom.2019.05.099

Cobos Muñoz, D., De Savigny, D., Sorchik, R., Bo, K. S., Hart, J., Kwa, V., Ngomituje, X., Richards, N., & Lopez, A. D. (2020). Better data for better outcomes: The importance of process mapping and management in CRVS systems. *BMC Medicine*, *18*(1), 1–10. https://doi.org/10.1186/s12916-020-01522-z

Gizaw, M. E. (2020). The status, challenges and opportunities of civil registration and vital statistics in Ethiopia: a systematic review. *International Journal of Scientific Reports*, *6*(5), 200. https://doi.org/10.18203/issn.2454-2156.intjscirep20201604

Hayatpur, D., Xia, H., & Wigdor, D. (2020). DataHop: Spatial data exploration in virtual reality. *UIST 2020 - Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 818–828. https://doi.org/10.1145/3379337.3415878

Ling, X., Gao, M., & Wang, D. (2020). Intelligent document processing based on RPA and machine learning. *Proceedings - 2020 Chinese Automation Congress, CAC 2020*, 1349–1353. https://doi.org/10.1109/CAC51589.2020.9326579

Melis, G., Kočiský, T., & Blunsom, P. (2019). *Mogrifier LSTM*. http://arxiv.org/abs/1909.01792

Mills, S., Lee, J. K., & Rassekh, B. M. (2019). An introduction to the civil registration and vital statistics systems with applications in low- and middle-income countries. *Journal of Health,* *Population and Nutrition*, *38*(S1), 23. https://doi.org/10.1186/s41043-019-0177-1

Muñoz, D. C., Abouzahr, C., & De Savigny, D. (2018). The "Ten CRVS Milestones" framework for understanding Civil Registration and Vital Statistics systems. *BMJ Global Health*, *3*(2), 2017–2019. https://doi.org/10.1136/bmjgh-2017-000673

Musadad, D. A., & Kelly, M. J. (2023). *Implementation research for mortality statistics development: Findings from Indonesia*. 1–22.

Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (2020). COVID-19: A comparison of time series methods to forecast percentage of active cases per population. *Applied Sciences (Switzerland)*, *10*(11), 1–15. https://doi.org/10.3390/app10113880

Priambodo, R., & Kadarina, T. M. (2020). Monitoring Self-isolation Patient of COVID-19 with Internet of Things. *2020 IEEE International Conference on Communication, Networks and Satellite, Comnetsat 2020 - Proceedings*, 87–91. https://doi.org/10.1109/Comnetsat50391.2020.9328953

Ukoji, U. V., Okoronkwo, E., Imo, C. K., & Mbah, C. S. (2019). Civil Registration and Vital Statistics as Sources of Socio-Demographic Data for Good Governance in Nigeria. *The Nigerian Journal of Sociology and Anthropology*, *17*(1), 102–120. https://doi.org/10.36108/NJSA/9102/71(0170)

Yokobori, Y., Obara, H., Sugiura, Y., & Kitamura, T. (2021). Gaps in the civil registration and vital statistics systems of low- and middle-income countries and the health sector's role in improving the situation. *Global Health & Medicine*, *3*(4), 243–245. https://doi.org/10.35772/ghm.2020.01103