

**28th Meeting of the Wiesbaden Group on Business Registers
The Hague, The Netherlands, 2 - 6 October 2023**

Jodi Jhouranda Siregar, Muhammad Yunus Hendrawan, Yovi Maulana Nugraha, Irma Damayanti
BPS-Statistics Indonesia

Session 6 : New data sources: Opportunity and challenges

Establishment “Face” Recognition Using Artificial Intelligence

Abstract

Keywords: SBR, Data Integration, Artificial Intelligence, Text Analytics, Record Linkage

BPS-Statistics Indonesia has developed the Statistical Business Register (SBR) since 2013. One of the roles of SBR in statistical activities is as a sample frame, so updating SBR data is essential. Data updating is carried out by integrating various data sources, for example, BPS survey, Data from the Directorate General of Taxes, Data from the Ministry of Industry, and Profiling from companies' websites. There are problems in the SBR update process with various data sources, such as differences in the data structure, fill validity, duplication, and data quality. Integrating SBR in BPS-Statistics Indonesia with various data sources is carried out through SQL Based Matching and Manual Assessment Matching. The SQL Based Matching process filters new input data from other sources with high probability attribute matching with SBR data.

Meanwhile, the Manual Assessment Matching process includes data with a low probability attribute matching. The Manual Assessment Matching process needs to detect businesses with a high probability attribute matching with SBR data but is misinterpreted due to differences in the structure and format of data sources. Manual Assessment Matching is done by Experts every time new data exists, so it takes a lot of time and resources. The manual matching process causes an inefficient SBR updating process. To solve these problems, we propose a new method by utilizing Artificial Intelligence to replace Experts in manual matching. Artificial Intelligence will learn the behavior of experts in carrying out the manual matching process. The matching result from experts will be used as input for the Siamese Network model so that all matching decisions taken by miscellaneous experts can be compiled in an Artificial Intelligence model. The system will implement the model to update the SBR Sample Frame efficiently. Every new data will go through the AI model, such as a “face” recognition mechanism, to determine the eligibility of the data as a new SBR Data record.