Online Population Census: The Indonesian Experience

Budi, Brilian Surya; Munaf, Alfatihah R. M. N. S. P.; Ashiddiiqi, Sulthoni (Statistics Indonesia)
brilian.surya@bps.go.id, alfa@bps.go.id, sulthoni@bps.go.id

Abstract

In the 2010 population census, Indonesia involved approximately 700 thousand data collectors. With the increasing population and the continued expansion of territories since 2010, the cost of conducting a census continues to increase.

Statistics Indonesia states the use of the combined method in the 2020 population census in 2018. New methods were introduced, which not only reduced costs but also increased public participation in census activities. One of the methods used is an online population census, using Computer-Assisted Web Interviewing (CAWI) technology.

The online population census website is designed using the Integrated Collection Survey (ICS) framework, which allows integration in conducting multimode data collection. ICS is a modernization solution in data collection as part of the technology transformation carried out by Statistics Indonesia.

During the online population census, the COVID-19 pandemic hit Indonesia. The online population census, which was initially only an alternative, has become a solution for implementing the census during the pandemic. Unexpectedly, for the first online census conducted by Indonesia, 51 million people participated in this online population census (approximately 19% of Indonesia's population as a whole).

The Baby Steps

The 2020 Population Census has been prepared since the end of 2017, with one big goal, collaboration with the Ministry of Home Affairs (Kemendagri) for One Indonesian Population Data.

In 2018, Statistics Indonesia (BPS) gained access to the Kemendagri database, with only five searches per day. This initial data is used for planning the integration of Kemendagri data into the 2020 Population Census. The variables that can be used are studied, which ones will be used to implement the Census. Which variables cannot be used due to differences in concepts and definitions. Matching is also carried out on area codes because BPS and Kemendagri have different regional coding procedures.

In the second semester of 2019, BPS was given access to all Indonesian population data incorporated in the Kemendagri. Through data processing, it is still found residents whose information is not known where the local administrative unit (Satuan Lingkungan Setempat - SLS) resides. For this population, pairing is done by involving BPS employees throughout Indonesia, with the aim that these residents can be included in the Population List later.

In early 2020, BPS again received data updates from Kemendagri. So, the implementation of the Online Population Census (Sensus Penduduk Online - SPO) has used consolidated data until December 2019.

The Forms

Sensus.bps.go.id is built using the Integrated Collection Survey (ICS) framework owned by BPS. The construction of the website has been carried out since 2018, with several changes to the login mechanism.

The first SPO application was used in the trial in January 2019. The login mechanism into the application is designed using two-way verification, in which respondents are asked to input their Population Identification Number (*Nomer Induk Kependudukan* - NIK) and name according to their ID card. If the NIK and Name match, the respondent will be asked for the birth mother's name for verification. After all, stages have been passed, and respondents will be guided to a prefilled census form. Otherwise, respondents will be guided to a census form with no data (blank). The feedback obtained from this trial is that people feel uncomfortable with questions related to birth mothers, which are considered sensitive and closely related to banking data.

During the April 2019 rehearsal, the login mechanism changed to a pair of NIK and Family Cards Numbers (Kartu Keluarga - KK). If successful, respondents will face a form with prefilled data (Data that, according to Kemendagri, should not change). From this rehearsal, problems were found, such as 1) some people cannot log in because they do not have a NIK or their NIK and KK pairs do not match, and 2) A person who has data on a pair of NIK and KK who fills in the census data of another person. The follow-up to this rehearsal is the use of tokens which are planned to be given via Short Message Service (SMS), to those who make requests to specific numbers, and tokens will be given after verification whether the pair of Phone Numbers and NIK numbers used for token requests is the same as the phone registration data. However, this idea also hit a dead end because we had to coordinate with all cellular operators in Indonesia, and each had different regulations.
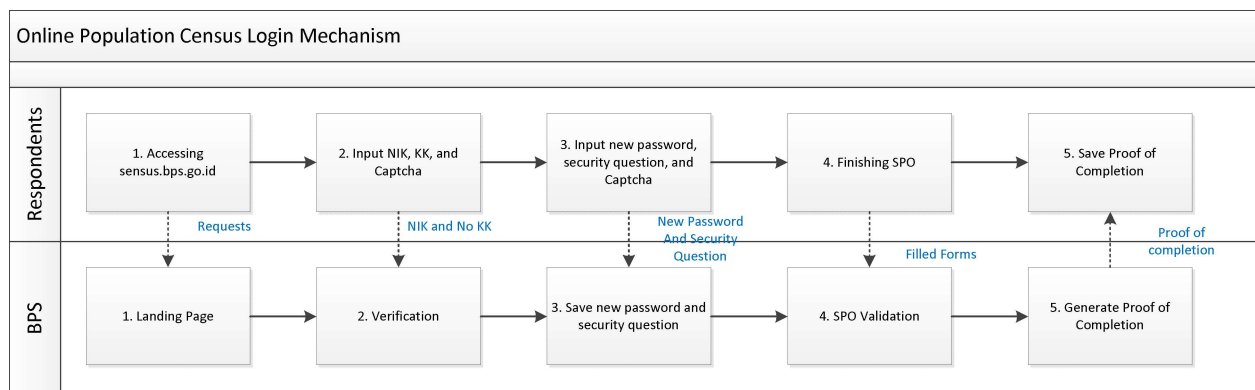

Figure 1. Final Online Population Census Login Mechanism

At the end of 2019, along with the completion of ICS development, SPO is targeted to be the first online survey built using the ICS framework. Improvements in terms of coding, performance, security continue to be made. In the process, BPS received evaluation and assistance from the Australian Bureau of Statistics (ABS), the National Cyber and Crypto Agency (BSSN), and professional institutions in data security. Regarding canceling the token distribution mechanism via SMS, based on discussions during mentoring with ABS, the login mechanism was returned to NIK and KK but without prefilled data. To ensure security, a captcha is added on the login page to avoid bot attacks. The final login mechanism is presented in Figure 1.

Figure 2. Insight from SPO Trial in January 2020

In early 2020, the first SPO trial with the ICS framework was carried out on a massive scale to all BPS employees throughout Indonesia (about 15,000 employees). The trial was carried out twice, January 13, 2020, and February 4, 2020. The main complaint found on the January 13, 2020 trial was not being able to log in. Some employees did not know that the NIK that he entered in the personnel employee database was wrong. As a result, the family data that we prepared did not match. Therefore, we also provide mini-apps, check NIK, so that respondents know whether the pair of NIK and KK match those in the primary data used by BPS.

From several trials, we also improve the UI/UX. We invite linguists to get the right words for each question so that respondents easily understand the sentence of the question. We do UI/UX testing by inviting various groups of people. In addition, we also invite high school students to participate in the SOP trial, assuming they will become assistants in filling out SOP forms in their households.

The Dashboard

The essential data used in the SP2020 dashboard comes from several sources. The databases used are also different, such as SQL Server, PostgreSQL, GeoJSON, Excel files to crawling results. Data providers are not only BPS internal stakeholders but also involve several stakeholders outside BPS. After gaining access to the data source, the next step is to process the data according to the needs of the dashboard. The tools used at this stage are Microsoft SQL Server Integration Services (SSIS) and Microsoft SQL Server Analysis Services (SSAS).

The following process is to create a dashboard interface. The tool used in this process is Microsoft Power BI Desktop. So that dashboard visitors can easily understand the data presented, the dashboard interface is made as attractive as possible. The dashboard's content is packaged in such a way that there is not too little information and not too much that it can confuse the reader. The form of data visualization chosen must also follow the data presented.

Figure 3. Dashboard Development for SPO

BPS has not only prepared a dashboard to monitor progress and content, but we have also prepared an infrastructure dashboard that describes activities related to the system. The information in this dashboard includes the area of origin of visitors to the sensus.bps.go.id site, the browser and operating system used by the visitor, the number of requests served per unit time, and the consumption of resources used by the server.
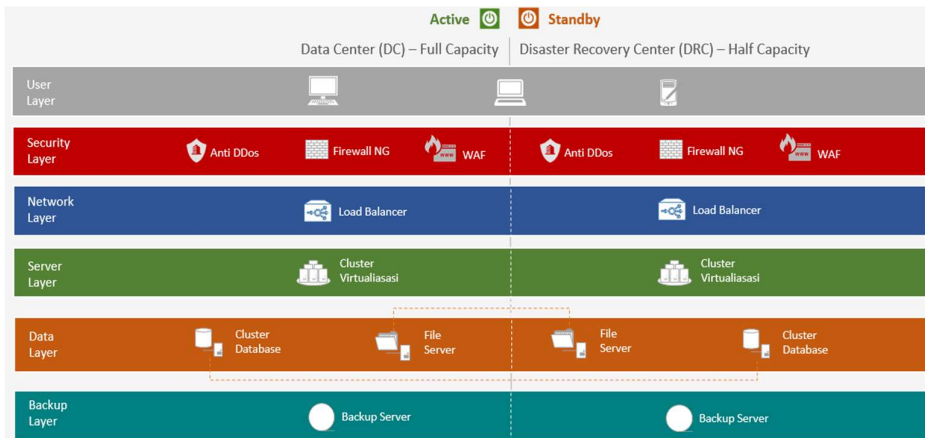
Infrastructure

Figure 4. Core Architecture Infrastructure for SPO

The 2020 SPO is the first online Census held by BPS. The calculation of infrastructure needs is carried out carefully so that SPOs can run smoothly. In addition to using the infrastructure available at the BPS head office, DRC is also used, as described in Figure 4. Performance tests and security tests are carried out to ensure the system can run properly and is safe from possible security attacks.

Technically, performance testing is done by adding the number of users continuously simulated through the Apache JMeter application. In this test, the maximum number of users who can enter is obtained. Various test scenarios were carried out starting from the login mechanism, loading, forms, saving and sending entries, to logging out, as for the other results of the last application test, obtained 8,000 artificial connections that can enter simultaneously through the SPO website. In addition, several causes became obstacles in filling out the SPO. Some of the causes were finally used for suggestions for improvement in application development.

Security testing is carried out using the Vulnerability Assessment and Penetration Testing (VAPT) method. Vulnerability assessment is an effort to discover vulnerabilities in applications and infrastructure owned by

BPS, while Pentest is an effort to penetrate defense systems or activities to see if Intruders/hackers can enter the system at a specified time. The follow-up of this security test is the addition of a captcha on the password page to prevent attacks from entering through this page. The risk assessment conducted by BSSN categorizes the web sensus.bps.go.id in Low Risk (Low – Med – High range). Moreover, for SSL implementation, this web rank is at A.
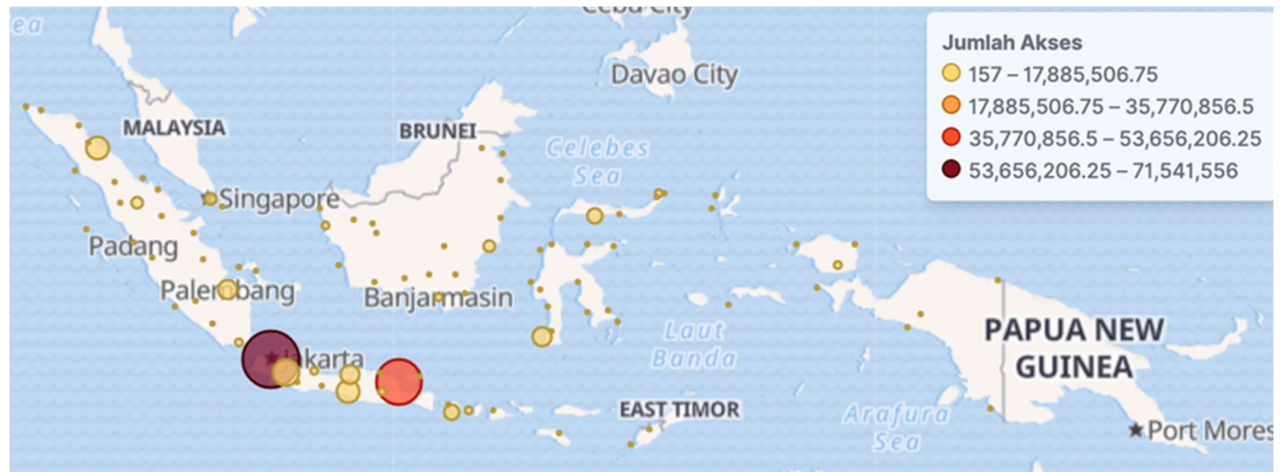
Census War

Figure 5. Access to sensus.bps.go.id by location

During the Online Population Census period on February 15 – May 29, 2020, there were 252 million accesses to the web sensus.bps.go.id. Based on the distribution, it can be seen that the most considerable access is still from the island of Java, followed by the island of Sumatra, as shown in Figure 5. Most are accessed via mobile phones with the Android operating system and via the mobile browser Google Chrome.
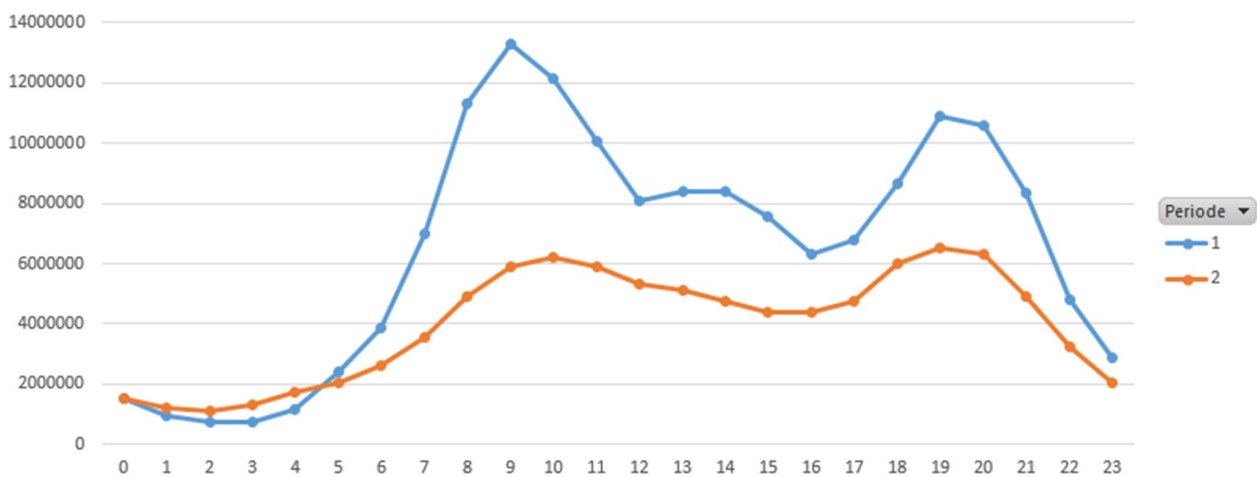

Figure 6. The difference in access patterns in the two periods

There are differences in access patterns in the two periods, the initial SPO period - 1 (15 February-31 March 2020) and the extended SPO - 2 (1 April-29 May 2020). In Figure 6, the access pattern shows that the data has two peaks, between 9–10 am and 7–8 pm. However, if you divide the time into different periods, Regular (February 15 - March 15, 2020), Work From Home - WFH (March 16 - April 23, 2020), Fasting + Lebaran

(April 24 - May 24, 2020), and at the end of the SPO – WFH 2 (May 25 – May 29, 2020), it was seen that there was a difference in the pattern between the regular SPO period, the peak of filling occurred in the morning when the "ngibar" event (filling in the form together in public facilities) could still be carried out. While in the WFH period, the peak of charging occurs at night.
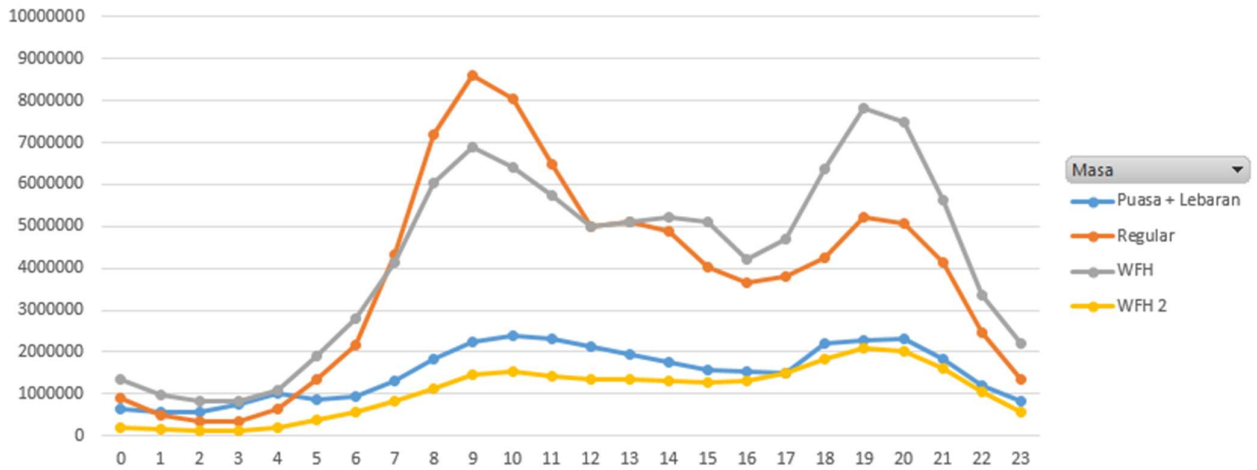

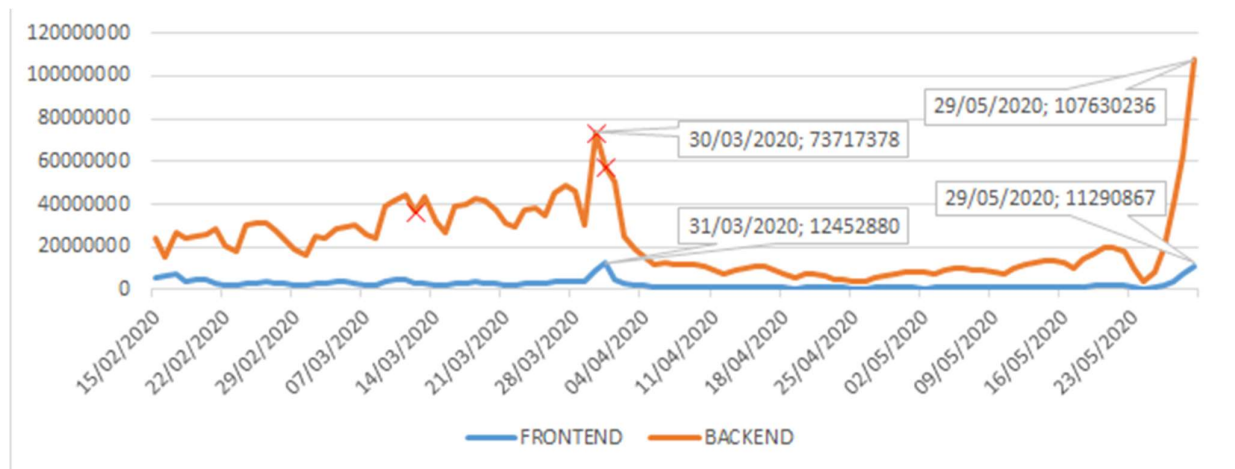*Figure 7.* The difference in access patterns in the four periods

Figure 8. Sorry Page

During the implementation of the SPO, it was noted that the system posted a sorry page 3 times when the system encountered problems that could not be resolved quickly: March 12, 2020 (2 to 6 pm), March 30, 2020 (8 to 10:35 pm), and March 31, 2020 (9 am to 7 pm). This happened because there was a significant increase in traffic, and there were several processes experienced a slowdown, resulting in long queues. Improvements include increasing access capacity by using other ISPs available at BPS, optimizing some queries, and adjusting the threshold for the number of requests that can be handled in a queue. On the last day of closing the SPO, our hard work paid off when we got 51,360,072 residents participating in the SPO.

The Pandemic

The next stage after the SPO is preparation for the September 2020 Population Census (SP2020S). In SP2020S, it is planned that surveyors bring the Population List (which contains population data based on Kemendagri data after being updated based on the results of the SPO) to the SLS visited.

March 2020 is the period in which Indonesia begins to deal with Covid-19. The first mitigation from implementing the population census is to extend the implementation of the SPO until May 2020. The pandemic has made adjustments in several business processes. A significant change occurred in the training of surveyors, which was carried out entirely online, in collaboration with Television of the Republic of Indonesia (TVRI) for distance learning. The Population Census training has also become the highest number of e-learning participants since BPS first used e-learning.

The processing of this census data is also unusual because it happened during a pandemic. Legal instruments are prepared to maintain data confidentiality. For areas where it is impossible to set up a processing center and the operator plans to do data entry from home, the operator must sign a Non-Disclosure Agreement (NDA) attached to the contract. The focus of this NDA is the obligation to maintain the confidentiality of data, especially regarding personal data. For areas that have processing centers, it is not without problems. Space that cannot be fully used due to social distancing policy results in longer processing times. In addition, several areas must be locked down because there are operators who have been confirmed positive for Covid-19.

The End of 2020

In each stage of the Population Census implementation, the dependence on work between teams is very high. Intense communication and synchronization are carried out continuously, collaborating to achieve a big goal: One Indonesian Population Data. Moreover, Statistics Indonesia and the Ministry of Home Affairs jointly announced on January 21, 2021 the total population of Indonesia in 2020 was 270,203,917.