

Ethnostatistics of Single Integrated National Database: A Reflection on the National Program of Socioeconomic Registration for Social Protection Programs in Indonesia

Eko Wahyono¹, Pratiwi^{2*}, Nuzul Solekhah³, Dian Karinawati Imron⁴, Fatwa Nurul Hakim⁵, Nurbaety Setram⁶, Risyanto⁷, Atqo Mardiyanto⁸, Agus Pramono⁹, Agus Eko Nugroho¹⁰, M. Alie Humaedi¹¹

^{1,2,3,4,5,6,7,10,11} National Research and Innovation Agency, Indonesia

^{8,9} Statistics Indonesia, Indonesia

Corresponding Author: prat005@brin.go.id

Article Info

Article History;

Received:

2023-08-09

Revised:

2023-11-09

Accepted:

2024-06-19

Abstract: Lessons from participative de facto and de jure approaches to socioeconomic registration for integrated policymaking have been underexplored, especially in archipelagic states and middle-income countries. This study aims to analyze participants' perceptions of statistical data and assess how socioeconomic circumstances shape the Regsosek (socioeconomic registration) process in Indonesia. Using a qualitative method, the research incorporates ethnographic techniques and triangulation, with enumeration carried out by locals familiar with the socio-cultural characteristics of the communities. The data were validated through multiple cross-checking stages. The Regsosek initiative seeks to create a unified national database to improve policymaking, a key aspect of Indonesia's socioeconomic development that has received limited attention in terms of participative methods. While prior studies have focused on the technical aspects of socioeconomic data collection, this research highlights the influence of socioeconomic diversity on the process. Novel insights reveal that participants' perceptions of well-being vary based on socioeconomic backgrounds, geographic conditions, livelihood systems, gender, and exposure to previous development programs. Key challenges identified include community resistance and misconceptions linking data collection to social assistance, as well as the need for system integration and stability to establish a single national database. The study concludes that development and poverty alleviation programs must be mutually agreed upon to avoid disruptions, instability, and political influence, offering a fresh perspective on integrated policymaking in archipelagic and middle-income contexts.

Keywords: *Emic perspective; single integrated national data; social protections; REGSOSEK*

DOI: <https://doi.org/10.18196/jgpp.v11i3.19495>



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

This paper aims to explore aspects and perceptions affecting the collection of statistical data for policymaking, especially for social protection. Previous practices indicated that statistical data have been employed in decision-making in various contexts, such as academic organizations (Mills et al., 2006), financial sectors (Boje et al., 2006), tourism management (Valérian et al., 2022), corporations (Landrum & Boje, 2008), and human resource management (Stoycheva & Favero, 2020; Thomas et al., 2007). The research in the previous sentence was derived from all peer-

reviewed publications on the Scopus database. It depicts the employment of an ethnostatistics framework in public sectors, and policymaking has received the least empirical attention.

In policymaking, a historical perspective was discovered: The Roman Empire had collected data on the people to allocate taxes since the 27th century BC (Thorvaldsen, 2017). Until 1790, the United States public sector collected data to explore demographic and constituent data in general elections (Thorvaldsen, 2017). Census data collection could be affected by conflicts and community resistance (Thorvaldsen, 2017). Even though Thorvaldsen (2017) argued that register-based or de jure census eased state administrators' control over community resistance, in today's practices, some countries have adopted de facto census, including Indonesia, to obtain more comprehensive data for decision-making (BPS, 2020). Statistics Indonesia has prevented community resistance by conducting both participative de facto and de jure methods and public consultation forums by involving heads of the community actively (BPS, 2023).

Even though statistical data can serve as an instrument for precise and strategic policymaking (Oliver et al., 2014), creating statistics can be politically influenced by cultural and socioeconomic contexts and participants' perspectives. Olmsted et al. (1989) and Valérian et al. (2022) introduced the term ethnostatistics to analyze how *verstehen* perspectives can play in the making of positivist sociological data. Olmsted et al. (1989) emphasized that ethnostatistics presents the process, utilization, and interests in statistical data presentation. Statistical data are often used to serve various interests, for example, in providing incentives (Carlson et al., 2006), attracting investors (Smith et al., 2004), policy directions (Johar et al., 2019), and the needs of farmers (Liswanti et al., 2012).

Previous studies on ethnostatistics were limited to the context of private sectors and human resources management in high-income countries (Boje et al., 2006; Mills et al., 2006; Stoycheva & Favero, 2020; Valérian et al., 2022). Meanwhile, public sectors work in different natures, institutions, and motivations. This study fills in some evidence lacking in previous scientific literature by investigating the collection and utilization of statistical data in the Indonesian public sector to create a single integrated national database for socioeconomic policy.

Before the 1988 crisis, social protections in Indonesia were based on unofficial designs and initiated by non-state actors (Sumarto, 2018). After the crisis, Indonesia's designs for social protection and community empowerment were created based on data and evidence (Sumarto, 2018). Therefore, the validity of the business process of data collection is crucial. Investigating Indonesia contributes to drawing the pattern of bottom-up data collection and exploring how the process promotes data validity and reliability. Previous mechanisms of socioeconomic data in Indonesia were conducted through top-down processes without active engagements of the community (Wertheim & Nitisastro, 1971). In 2022, the Indonesian government has transformed its method of collecting socioeconomic data through more integrated and bottom-up methods in the hope of delivering more efficient social services (Simbolon et al., 2023). Additionally, this study responds to Kukutai and Thompson (2015), who suggested future studies to reveal the pattern of synergized multi-sector census data and technological utilization to increase data validity and reliability.

Indonesia's socioeconomic census began in Java during the Dutch East Indies era in 1775 (Wertheim & Nitisastro, 1971). During this period and in subsequent years, census data included population size, marital status, deaths, births, gender, ethnicity, and land ownership, serving as the basis for tax and income policies (Wertheim & Nitisastro, 1971). The census data during the colonial period were not collected directly by enumerators but by village leaders (Wertheim & Nitisastro, 1971). After independence, population censuses in Indonesia began in 1960 but were not implemented extensively in preparing socioeconomic policies (Wertheim & Nitisastro, 1971). Since 2022, the process of collecting citizens' socioeconomic data has engaged more actors from the grassroots and village-level officials (BPS, 2023). Therefore, the data collection has become more political. By using the ethnostatistics framework, this study explored perceptions and political processes during socioeconomic registration.

Ethnostatistics was derived from a critical analysis of the objectivity and reality of numbers (Olmsted et al., 1989). Ethnostatistics studies humans' perspectives and the institutions that influence statistical data collection, presentation, and utilization (Mespoulet, 2015). There are three stages of ethnostatistics. The first stage produces statistical data using ethnography through qualitative observation to understand phenomena as variables into numbers and sizes. The second

stage categorizes and analyzes data using computer simulations for assessment or measurement. The third stage, statistical rhetoric, explores how statistics are adopted for purposes that influence and benefit society (Olmsted et al., 1989; Winiecki, 2008). The aspects of ethnostatistics lie in the process and the utilization of the data (Mespoulet, 2015).

An ethnostatistical study in the Netherlands examined the government's efforts to recognize and collect data on developing ethnicities as targets for policy distribution (De Zwart, 2012). Cultural and ethnic backgrounds became the concern of the government to encourage the integration of a multicultural society (De Zwart, 2012). The word 'ethnostatistics' was developed to describe and record ethnic classifications as part of handling social issues. De Zwart (2012) provided a comprehensive understanding that data on ethnicities can be applied to suppress more than to redistribute politically. However, this study was not based on systematic tests and transparent procedures in concluding except from picked previous literature.

Another study in the private sector in the United States unveiled that statistical data can be presented less objectively to attract investors and fix the company's poor financial performance (Boje et al., 2006). A study employed an ethnostatistic perspective on the company's reports and news report data, revealing knowledge disparity behind the intentionally questionable data presentation for the company's interest. However, Boje et al. (2006) did not explain detailed tricks in creating the company's reports. In another context, Carlon et al. (2006) argued that statistical data have become a fetish. The assumption that statistics is the only valid resource for a company's decision-making may not be true. For example, the use of data on a company's income as the only base to pay for managers' incentives and improve performance may not be contextual. Still, it can be deployed to serve the interests of the elites (Carlon et al., 2006). Other variables, such as customer satisfaction and product quality, may represent a more objective measurement of the company's performance. At that point, statistical data needed to be understood as representational instruments and can be biased (Olmsted et al., 1989). The actual meaning of statistical data cannot be understood except by understanding the data production, the participants involved in it, and when and how the data are produced. Ethnostatistics can be utilized to evaluate the reliability and validity of data.

In socioeconomic surveys, statistical data can benefit the community, such as farmers. CIFOR conducted a farmer survey aimed at accelerating facilitators' and farmers' understanding of local resource use systems and their management (Liswanti et al., 2012). The survey explored interactions in government decision-making with farmers and other participants. Therefore, opportunities for conflict resolution can be identified by taking the needs of the local community into account (Liswanti et al., 2012). To ensure this goal, the instrument should be made by considering some sampling techniques such as multilevel, multi-actor, and random sampling (Liswanti et al., 2012). To ensure statistical data are reliable and objective, ethnostatistics is required to reveal the efficiency, transparency, consistency, and representation of data collection and use (Connelly et al., 2016). The collection of statistical data also needs to be done collaboratively with the main participants and target group (UNICEF et al., 2021). The quantitative sociology perspective highlights that the question "how?" is insufficient to evaluate statistical data from the ethnostatistics perspective (Valérian et al., 2022). It should be complemented with other analyses of who compiled the statistical data and what their expertise is. Is it sufficiently representative? What forms of data are collected? When and at what moment are the data collected? In what context are the data collected? Is it national, regional or community and the location? As well as why the data are produced (Valérian et al., 2022). At the national level, statistical data standards are often compiled by the central government. Hence, they may not be contextual to regional needs; therefore, ethnostatistics became an evaluation instrument by paying attention to the role of each level of participants (Valérian et al., 2022).

In Indonesia, which is institutionally different from high-income countries with its decentralized system and different social protection programs, a study has evaluated a national social and economic survey (Susenas) using ethnostatistical instruments (Johar et al., 2019). The results of Susenas needed to be understood carefully because the expenditure variables in the survey were not representative of some other variables, such as informal household spending and subsidies (Johar et al., 2019). Interventions in health service reform can misunderstand health spending. Therefore, survey results need to be understood according to the context (Johar et al., 2019). Another factor influencing the presentation and utilization of statistical data is the historical

approach (Stoycheva & Favero, 2020). A historical approach can fill in the gaps in ethnostatistics by looking at the context of panel data (Stoycheva & Favero, 2020). Previous studies indicated that the conditions of society, socioeconomic processes, politics, and culture, as well as historical aspects, can influence the production of statistical data that are never neutral (Mespoulet, 2015; Valérian et al., 2022). However, ethnostatistics can be an evaluation method for the use of more reliable, valid, efficient, and transparent statistical data. The success of statistical data collection methods is related to geographical location, socioeconomic conditions of the community, and the active role of data collection officers recruited from the local area. Thus, officers have mastered the survey location, allowing them to easily communicate to acquire the desired data.

In the last four decades, several government organizations have comprehensively conducted censuses in various fields, including social, economic, agricultural, and health. One of the agencies entrusted with carrying out the census is Statistics Indonesia (BPS). Currently, the socioeconomic census in Indonesia is conducted differently from the previous ones. From October 15 to November 14, 2022, Statistics Indonesia officials conducted a socioeconomic registration data collection (Regsosek) in all provinces in Indonesia. According to an interview with an officer from Statistics Indonesia conducted on October 8, 2022, Regsosek is the collection of data on the entire population consisting of profiles, social conditions, economy, and level of welfare. Regsosek is the government's effort to build a single integrated population database by actively involving residents in door-to-door data collection and data verification. By using a single integrated data, the government hopes to carry out its various programs in an integrated, non-overlapping, and more efficient manner. The urgency of data for development is mandated in the Law of the Republic of Indonesia No. 25 of 2004 concerning the National Development Planning System Chapter VII Article 31 that development planning is based on accurate data or information. Data for development planning continues to be produced by the respective ministries or state agencies. With this new procedure, this study answers two research questions: (1) How do the socioeconomic conditions in several islands in Indonesia affect the collection of statistical data? (2) How is the utilization of statistical data according to the perceptions of participants?

RESEARCH METHOD

This study reflects on the activities of the Regsosek, which brings up urgent issues related to data collection (statistical processes) and data utilization (social aspects). To achieve these objectives, this research requires an appropriate framework and method to explore and analyze various social realities. This research was conducted using a constructivist paradigm with a qualitative approach. The constructivism paradigm promotes the extraction of information related to the reality of society and how people interpret their personal and social situations. It is relevant to the research objectives as this study tries to dig into the perceptions of participants in interpreting their experiences behind the data collection. There might be some explanation from the informants' perceptions that could give meaning beyond the quantitative data collection performed by Statistics Indonesia. Those perceptions might influence the participation of the informants during the social and economic registration data collection.

According to Berger and Luckmann (2016), the constructivism paradigm is a reality that is not a singular matter that can be generalized but is a multiple reality that humans can interpret differently. Humans act as subjects in constructing social reality. Reality, as understood by humans, is based on socially and culturally constructed knowledge (social relativity) (Berger & Luckmann, 2016). Additionally, Schutz asserted that the reality of everyday life is related to not only knowledge, reasoning, facts, and events but also representations of material things transformed into socio-cultural objects (Geniusas, 2020). This paradigm not only helps to understand the background of the community and its responses but also further explores how the community interprets the collection of data on Regsosek.

The ethnostatistical method combines ethnomethodology in the quantification and meaning through in-depth observations of social phenomena and social contexts into the cycle of statistics (Olmsted et al., 1989). Stoycheva and Favero (2020) argued that ethnostatistics is a method that emphasizes the premise that individuals understand authentic cultural systems and act on their social reality. Ethnostatistics is the study of the construction, interpretation, and display of statistics in quantitative social research to bridge the gap between quantitative and qualitative

research (Olmsted et al., 1989). The purpose of ethnostatistics is to describe, analyze, explain, and understand how statistics are constructed and function in a study.

Data collection was conducted through historical documents on socioeconomic data collections, in-depth interviews, focus group discussions (FGD), and observations. Data were taken using purposive sampling from nine provinces: Aceh, North Sumatra, Lampung, West Java, Central Java, East Java, West Kalimantan, Central Kalimantan, and East Kalimantan. Forty-five informants were interviewed and involved in discussions. Five informants in each province were involved, including a surveyor, a supervisor, a district coordinator, a survey respondent, and a village official. This research was conducted from October 15 to November 14, 2022. The research locations and units of analysis were selected by involving informants from Statistic Indonesia, field surveyors, village officials, community leaders, and citizens as the survey respondents. The collected data were analyzed using interactive dialogue data analysis techniques. The data were cross-checked with other data. Thus, the alignment and completeness of data were examined in stages according to research needs.

RESULTS AND DISCUSSION

The History of Socioeconomic Data Collection for Policymaking in Indonesia

The population census in Indonesia began in the colonial era. During this period, there were ten population censuses, with three periods: 1950, 1920 and 1930. The 1930s were considered the best time for the census, making it widely employed as a reference for census and population analysis in Indonesia (Tukiran, 2016). During the Japanese occupation (1942-1945), a local census was conducted. Meanwhile, after Indonesia's independence, seven population censuses were carried out in 1961, 1971, 1980, 1990, 2000, 2010 and 2020.

The 1961 census, the first census in Indonesia, was extremely simple; only two variables were considered: gender and population density. Data presentation and population density levels were also presented per level I area (province) and have not yet reached level II (regency) level. In 1971, more variables were added. Besides population, gender, and population density, the census also recorded age and other variables regarding citizenship (Biro Pusat Statistik, 1972). The 1980 census experienced a significant improvement related to the use of variables; 20 variables were utilized. In 1990, 27 provinces in Indonesia had completed data collection. The population census in 2000 was conducted in 30 provinces. A decade later, in 2010, the census experienced a notable change regarding the use of variables. Variables enumerated in 2010 included fertility, mortality, employment, migration, and housing facilities (Badan Pusat Statistik, 2010). The population census in 2020 has various breakthroughs. A single integrated Indonesian population data began to be designed using a combination method. The census in the previous six eras still utilized the traditional pattern of visiting door to door. In contrast, the 2022 census was a mixed one, using population administration data from the Directorate General of Civil Population Affairs (Ditjen Dukcapil) and the Ministry of Home Affairs (Central Bureau of Statistics, 2021).

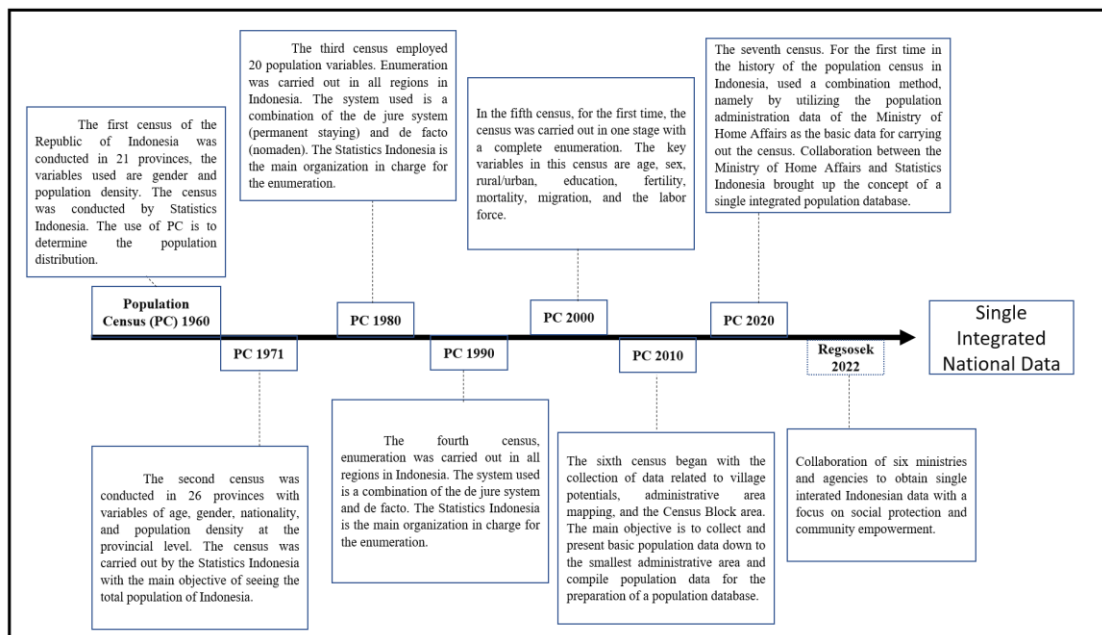


Figure 1. Population Census Trajectory in Indonesia
 Source: Central Bureau of Statistics (2021)

Interviews with officials from Statistics Indonesia unveiled that collaboration between BPS and the Ministry of Home Affairs in the 2020 census was a strategic step in conducting a census among participants across ministries. This collaboration generated data on the composition, number, distribution, and characteristics of the Indonesian population. The COVID-19 pandemic that struck during 2020-2022 has influenced the direction of the census and collaboration among participants. Regsosek was an effort to provide single integrated data. To achieve this goal, six ministries or agencies were involved in the population enumeration: the Ministry of Home Affairs, the Ministry of Communication and Information; the Ministry of Villages, Underdeveloped Regions and Transmigration; the Ministry of Finance; Statistics Indonesia; and the Ministry of National Development Planning (Bappenas).

The collaboration among ministries or agencies enabled the single integrated national data to be a database for various programs and plans. To achieve data authenticity, the implementation of Regsosek covered the entire Indonesian population, utilized the same standard methodology across all areas in Indonesia, and involved regular and accessible updates for authorities. The focuses of the Regsosek data collection were information on geographic landscapes, socioeconomic conditions, housing conditions and clean water sanitation, asset ownership, conditions of the vulnerability of special population groups, geospatial information, level of welfare, and other socioeconomic information.

The Business Process of Regsosek: A Scenario for Data Updating

Regsosek is the government's effort to build a single integrated population data. It is a survey data collection program coordinated and implemented by Statistics Indonesia. Its scope covers the entire population of 514 regencies or cities based on single-family data. It provides a database for the entire population consisting of profiles, social and economic conditions, and levels of welfare, including databases for social protection and community empowerment. The data encompass demographic patterns, socioeconomic conditions, asset ownership, clean sanitation conditions, housing conditions, conditions of the vulnerability of special populations, geospatial information, level of welfare, and other socioeconomic information.

According to interview results with Statistics Indonesia officials, the socioeconomic database aimed to support healthcare programs, entrepreneurship programs, investment programs, and employment programs. The implementation of Regsosek went through several stages: coordination, preparation, instructor training, field data collection, and Regsosek nights, which comprised evening enumerations of homeless people and migrant workers. At the coordination

and preparation stage, socialization was carried out both with banners and through social media to call for the support of all participants. Before training, instructors first conducted the recruitment through the SOBAT application, which was utilized for the registration, selection, and assignment of officers. In contrast, the SIPMEN application was employed to monitor document flows. The data collection phase was conducted from October 14 to November 15, 2022, in all regions of Indonesia. The implementing component in the Regsosek data collection was through the recruitment of the District Census Coordinator (Koseka), Field Examination Officer (PML), and Field Data Collection Officer (PPL). Some of those officers worked in the village administration to carry out enumerations, and PML was in charge of monitoring and examining enumeration results and documents.

Statistics Indonesia is a vertical agency in every city or regency. Authorities and tasks are divided among the central, provincial and city or regency government levels related to social research and surveys. The role of the central level of Statistics Indonesia in Regsosek was to coordinate and support several ministries or agencies, for example, the Ministry of National Development Planning (Bappenas), the Ministry of Finance, the Ministry of Home Affairs, the Ministry of Village, Development of Disadvantaged Regions and Transmigration, the Ministry of Communication and Information, the Coordinating Ministry for Human Development and Culture, the Coordinating Ministry for the Economy and the National Team. Statistics Indonesia also played a role in the acceleration of poverty reduction by leading the initial data collection, standardized methodology, and updating Regsosek, as well as coaching statistics as a national instructor. Statistics Indonesia formulated the needs of officers throughout all areas in Indonesia by covering 514 regencies or cities, 7,274 districts, and 84,096 villages. From these initial data, it was determined that the average number of officers' workloads was one PPL collected data of 250 families, one PML supervised four PPLs, and one Koseka coordinated four PMLs.

The legality of the implementation of Regsosek was required, embodied in a formal decree. At the central level, a formal decree was issued for the Social Research Implementation Team, a circular letter of support related to the socialization of the implementation of Regsosek, and a Korwil team (regional coordinator) to strengthen communication among central, provincial, and regional Statistics Indonesia offices. Central Statistics Indonesia coordinated to solicit support from ministries or agencies regarding the preparation of logos, taglines, and other publicity materials, a series of webinars, appointment of Regsosek ambassadors, and socialization support in various print and electronic media. Central Statistics Indonesia was tasked with identifying the need for both printed and non-printed questionnaires. It also made a set of question-and-answer booths related to Regsosek.

Regarding the preparation and implementation of officers, Central Statistics Indonesia has prepared the SOBAT application for selection and the SIPMEN application for monitoring. Central Statistics Indonesia's training as a national instructor was tasked with compiling textbooks and conveying technical matters to regional instructors at the provincial level, in this case, 34 provinces, as a result of the technical information on the implementation of Regsosek. A program planning mechanism appeared in each provincial area, which would be further reduced to the city or district area. The technical information was utilized as a guideline for preparing the budget, stages, and stipulated implementation time. Central Statistics Indonesia also acted as a supervisor in the implementation of Regsosek. Supervision allowed for the identification of budgeting problems, other obstacles, and constraints. Thus, deficiencies could be resolved appropriately.

The role of Statistics Indonesia at the district level was as an executor. Before carrying out the Regsosek, they formed regional instructors who were locals and based in the district. The regional instructors understood the local conditions and technical barriers that may arise during the program implementation. District regional instructors were trained by regional instructors at the provincial level, and then regional instructors at the city or district level trained human resources who would go into the field, be it Koseka, PML, or PPL. The training covered technical matters in the Regsosek enumeration, for example, understanding related to instruments, how to do probing during interviews, and how to do geo-tagging of households whether classified as poor or not.

Regsosek was an attempt to transform data spread across several ministries and institutions into a single data unit. The data included the socioeconomic conditions of the community. With the single integrated data, both central and regional governments carried out evidence-based

development planning. Regsosek was a starting point for establishing the single integrated data. Each institution has a database; for example, the Ministry of Health has health profile data, the Ministry of Social Affairs has an Integrated Social Welfare Data (DTKS) mechanism, and the Ministry of Village, Underdeveloped Regions, and Transmigration has village information system data. The Ministry of Agriculture has data related to the landscape of its agricultural production areas. Therefore, Regsosek was designed and consisted of sectoral data such as the database of education (Dapodik), extreme poverty data, social welfare data, Social Security Administration for Healthcare (BPJS healthcare), employment, spatial data, law enforcement, sustainable development goal (SDG) data, village data, and lastly protocol data.

Regsosek was conducted from October 15 to November 14, 2022. PPL verified the initial data received from the central government through Statistics Indonesia. Verification was carried out through the lowest level of community associations or RT or RW to ensure whether the families recorded in the initial data were correct. Officers and managers understood the implementation guidelines for this Regsosek. Thus, the implementation of activities was based on these guidelines. Both PML and PPL officers underwent training before the data collection. After the training, they signed a contract and must complete the work within the allotted time. Officer recruitment was held in several stages: administrative selection, interview selection, and then the announcement of accepted candidates. The officers understood that Regsosek's data collection was not related to any social assistance. The understanding needs to be emphasized during the data collection.

Social-economic Conditions and Citizens' Openness Toward Regsosek

The findings of this study revealed that different socioeconomic conditions in three different islands, such as Sumatra, Java, and Kalimantan, affected the data collection. Participants who worked in informal sectors had more flexible working hours. Therefore, enumerators adjusted and utilized pick-up techniques to gather information from respondents with heterogeneous working hours. Some enumerators reported that some respondents with double income in informal sectors did not report their income completely as they were afraid of not getting social assistance.

Observation results disclosed that the average respondents met by enumerators were elderly and female. It happened in Java, Sumatra, and Kalimantan. Hence, the data collected were mostly described from a women's perspective. In Java, mothers in some sample locations were mostly high school educated and dominantly played roles in parenting in the family, highlighting that stunting cases were affected by mothers' understanding of nutrition and participation in the community empowerment program for the well-being of the families (PKK). Families where the mothers participated in PKK, depicted no cases of stunting compared to those with mothers who never participated in PKK. Levels of education also influenced respondents' understanding of the questions asked during the Regsosek. Another finding unveiled that a household's livelihood was associated with who was at home and who was outside the home to work. Thus, the information conveyed to the data collector came from family members staying at home. One of the reasons why they were at home was that these women were developing a creative agency to manage their household finances. They tried to emerge in work schemes without having to leave the house and make money from home through the informal sector. The informal sector only became a concern after the 1950s and 1960s as part of the introduction of the third world and proved capable of contributing in difficult times during the crisis (Gërkhani & Van De Werfhorst, 2013).

What has become collective memory also influenced the way they responded to the data collection. Java emerged as the most dynamic province in terms of socioeconomics; every new trend could be spread quickly, one of which was due to internal migration. It influenced their perspective regarding data collection. There were quite a few respondents with identity cards from outside the city but lived in the city where the census was carried out. The age of the respondents in Java also affected their openness to data collection. Younger respondents were observed asking more questions about the objective of the Regsosek than elderly respondents. Participants who received previous social assistance programs from the government were also reported to be more open to Regsosek participation.

From the cultural perspective, one location in Central Java demonstrated a strict cultural hierarchy influenced by the sultanate kingdom, and ethnic exclusivity influenced the openness of these respondents. As a result, enumerators were guided by supervisors during data collection. Meanwhile, in Kalimantan, differences in social hierarchy and ethnicities did not seem to influence

the openness of the respondents. In this province, many respondents undertook plantations as their livelihood. Hence, data collection adjusted the working hours of the planters and farmers there. This kind of daily life influenced the way they responded to data collection, with some highlighting the economic problems caused by weather uncertainty.

On the other hand, development in Kalimantan has increasingly been carried out intensively. They worried that one day, they would be uprooted from the socioeconomic life they currently lived. This kind of concern also occurred among respondents in Sumatra. Things that have not been recorded in the census data were precious moments that required large expenditures in a one-year cycle, such as celebrations or crucial events such as weddings, circumcision ceremonies, and deaths. These things could not be answered using solely statistical indicators because they had cultural calculations determining moments. Therefore, solving this need in a long process should be a topic in the subsequent research. This study did not explore this much but discovered it as a finding that influenced data about their socioeconomic conditions.

Table 1. Social-economic Conditions and Respondents' Openness toward Regsosek

No	Island	Respondents' Openness
1	Sumatra	Their response was driven by their reaction toward the ongoing infrastructure development. The society feared that their livelihood and socioeconomic were affected by this development. Hence, it affected how they responded to the survey by adding their hope at every chance they answered the question.
2	Java	Respondents' answers, on average, were based on their response to assistance and comparisons with the success of government programs. Communities from areas with many development programs, such as Solo, tended to be more open. In contrast, communities in migrant worker enclaves, such as Indramayu, tended to answer with stories that emphasized the mechanisms by which they could be resistant to the socioeconomic conditions they faced.
3	Kalimantan	Their reaction toward the welfare indicators on housing and plantation programs drove their response. They worried that climate change and the change in government support for their livelihood would affect their income in the future. Thus, it also affected their openness to Regsosek.

Source: Authors' Analysis

Most of the respondents in Sumatra were micro-entrepreneurs. There was a phenomenon where a toll-road development program diminished one tourist route in Sumatra and reduced the income of those micro-entrepreneurs. The phenomenon influenced community openness toward Regsosek. Respondents who lived in another remote area in Sumatra hit badly by a Tsunami were reported to be more open to Regsosek as they hoped for social assistance from the government. In Java, Sumatra, and Kalimantan, the participation of enumerators from the local community has contributed to the openness of the respondents as they had strong social capital and were embedded in the community. Therefore, the recruitment of local enumerators has become the key to success in social-economic registration. The more experienced the local enumerators working with Statistics Indonesia, the better job quality that they reported, according to supervisors.

Citizens' Perceptions of Data Utilization for Well-being

The participants expressed various perceptions about the data utilization. Most participants associated the data utilization with social or financial assistance. The participants who previously received assistance expected to obtain the assistance continuously. In several locations, for instance, Sampang, Aceh, and West Kalimantan, the participants were open to answering the questions, although they were not the recipients of the assistance. Furthermore, the participants also thought the data utilization should lead to precise program planning and policy for poverty eradication and empowerment for long-term purposes. The participants also aspired to data utilization to data literacy for the community. It indicated that Regsosek had raised the literacy of the citizens to participate in policymaking and the importance of data collection. The data utilization needs to be linked to data utilization awareness, benefiting the community. The participants in Indramayu stated that the data would hopefully protect them from fraud offered by migrant workers. Participants in the Pulang Pisau Regency also expressed that data literacy was

an opportunity for farmer development. Most participants appreciated Regsosek. They mentioned that Regsosek's socioeconomic data collection had weaknesses because the data were gathered from different government boards, but the questions were similar. The participants perceived that there was a lack of coordination between government boards. The participants encouraged an effective system of data gathering and utilization.

Perspectives from Village Officials on Regsosek in Supporting More Coordinated Social Protection Policies

The high poverty rate in Indonesia requires a casuistic approach. Various social protection programs have been designed and implemented. However, according to some village officials, some of the implementations of social assistance and programs have resulted in multifaceted conflicts and chaos among communities. The multifaceted conflicts in social assistance policy implementations have occurred due to overlapping data among ministries or agencies. There was no single integrated data in various central and local agencies. The absence of single and integrated data was a problem for social policy because the programs and budgeting system were fragmented.

Furthermore, village officials highlighted two problems in using data in government. First, technical problems. In terms of technical flaws, each ministry or institution had various applications and different metadata. Each ministry and institution also had a different methodology for collecting field data. Second, non-technical issues. The silo mentality displayed in fragmented economic and social policies among ministries was a prevalent phenomenon. Regulations that were not yet supportive required a legal shelter and a single integrated database that participants with various understandings and insights could understand. The perceptions of village officials indicated an awareness from village officials toward the urgency of a single-integrated database for social policy.

Most village officials stated that Regsosek was brought in to address various issues of overlapping national data and unravel these problems. Because one of Regsosek's concerns is social protection, the data are focused on profiles, socioeconomic conditions, and the welfare of the population. The involvement of various ministries and agencies as active participants from the beginning of the Regsosek process has become a mediating tool for a more synergic step. Most village officials hoped this step would strengthen development planning, budgeting, monitoring, and evaluation based on strong, accurate, and regularly updated data. Most village officials expressed more awareness that Regsosek has a strategic position for integrating population data.

Crucial Aspects for the Regsosek

Observations during the fieldwork of this study reflected crucial aspects of the Regsosek. This section elaborates on the epistemological and axiological aspects of Regsosek data for various community social protection needs. To begin with, as the spearhead of socioeconomic data enumeration, PPL had a key role. They directly dealt with residents with various characteristics. The strategy was to determine PPL based on their domiciles. By implementing this strategy, the efficiency of the implementation of activities was carried out quicker than previous socioeconomic data collections. Citizens who were being enumerated were quick to trust and freely provide data as PPL had social proximity with the citizens. Once the data were gathered, they were cross-checked by PML.

In addition, after PML checked the data, Koseka checked the validity, depicting a double cross-check. Epistemologically, the validity of the Regsosek data collection could be ensured. The data obtained by PPL, responsible for around 250 families, were checked by PML. After PML validated the data, Koseka checked it again. These Koseka were generally Statistics Indonesia's employees, although, in the field, some were not. Checking data in stages by each of these officers allowed the results to be accountable, verified and valid. Authentic data could be collected through various methods, tools, models, and simulations (Kjelvik & Schultheis, 2019). Regsosek data were collected using questionnaires, geo-tagging, and recorded photos of residential buildings. The use of geo-tagging provided accurate geographic information regarding longitude and latitude coordinates. By combining various techniques and data collection methods, the field team continued to strive for validity and authenticity. Along with the complex development of society and the times, various data collection techniques were also developed.

In several conditions in the field, there were input errors. However, verification and data monitoring allowed for immediate correction. This situation was also supported by observations focused on the character of the local environment in terms of social, cultural, and regional topography. The system has represented Regsosek as a holistic and integral form of data collection. Moreover, the process was supported by emic narratives from researchers from the National Research and Innovation Agency (BRIN).

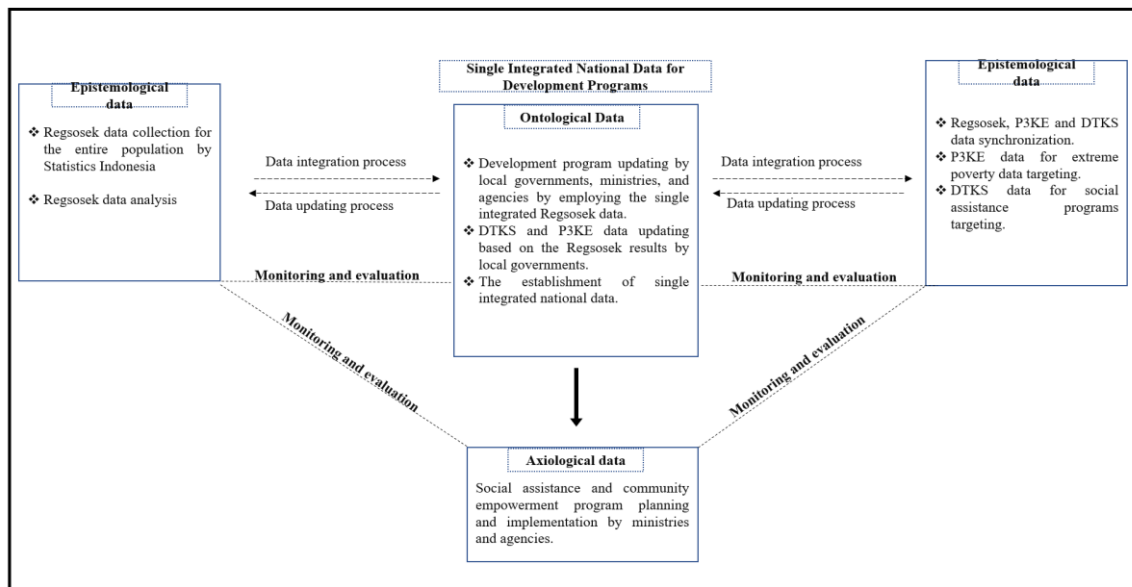


Figure 2. The Mechanism of the Single Integrated National Data of Regsosek
 Source: Authors' analysis

Throughout 2022, the preparation and execution of data were conducted by the Statistics Indonesia team. Data collection was carried out simultaneously in all regions of Indonesia. Various data variables were collected, and one of the data becoming the focus of this Regsosek was related to the acceleration of extreme poverty elimination (P3KE). These data would be beneficial in complementing the integrated social welfare data (DTKS), serving as the basis for targeting extremely low-income families not receiving financial assistance from the government. According to some officials in Statistics Indonesia, data on the poverty rate were the subject of an issue among ministries. Through the integration of data between P3KE and DTKS, it is expected that the integration overcomes the overlapping data and policies in alleviating extreme poverty as a form of social protection.

This study discovered three phenomena regarding socioeconomic perceptions and the data collection for Regsosek. To begin with, socioeconomic differences in different islands were not much associated with collecting statistical data. Livelihood only affected the procedure and timing of data collection; for example, household heads working in agriculture and palm oil plantations required permission from the head of plantations to be interviewed. Instead, previous experiences in participating in government social assistance and other development policies tended to affect participants' attitudes toward Regsosek. Therefore, the finding corroborated with Mespoulet (2015) that previous government social policies could politically influence statistical data production. Other than that, this study unveiled a slightly different suggestion to Stoycheva and Favero (2020), who stated that historical events affected data utilization in policymaking. This study discovered that previous conflicts in social assistance data motivated the attempt to initiate reform in socioeconomic data collection for more single-integrated data. Appreciation from citizens and village officials has illustrated bottom-up participation in Regsosek data collection, thereby minimizing conflicts, as suggested by N. et al. (2012) and UNICEF et al. (2021).

Another new finding of this study was that participants with different socioeconomic and geographical locations demonstrated different perceptions of well-being and poverty. In a regency in Java where the majority of the population worked as migrant workers abroad, they perceived

that the quality of housing determined a family's well-being as housing was easily observable by others. The perception could hamper the stunting eradication program because the community did not value balanced nutrition for children.

In another Java area close to the coast, the community also prioritized good housing to prevent the damage caused by abrasion. Therefore, the spending of the family was focused on renovating the house more than children's nutrition. The finding in Java highlighted that the perception of household spending, livelihood, cultural aspects, and geographical conditions has influenced children's nutrition. In Sumatra, the government's infrastructure programs and business permits seemed to affect the participants' perception of well-being and poverty. The development of toll roads and the opening of coal mining have decreased the income of the local community, particularly in local souvenir businesses and agriculture. Government programs have changed socioeconomic spaces. It is essential to prepare and assess the economic impact of infrastructure development on the local community. In Kalimantan, local natural resources seemed to affect the disagreement of the community in housing standards for prosperous families defined by Statistics Indonesia. As most of them lived in peatlands, they utilized ironwood to build their house and not concrete, as suggested in the household well-being indicators. The local community perceived ironwood as the most suitable for housing than other materials.

The second finding of this study complied with one of Valérian et al. (2022) suggestions that data context, such as socioeconomic conditions, genders, previous government assistance, and timing of the data collection affect the openness of the respondents. On the other hand, the findings of this study offer a different perspective to bridge central government goals and local needs, a concept questioned by Valérian et al. (2022). It is possible by employing local enumerators embedded in the local Statistics Indonesia office and local community. The third finding unveiled a similarity of local participants' views toward the outcome of Regsosek. The interviews with the local community as the primary participant of Regsosek implied that all of them believed the Regsosek data collection result would affect their chances for social protection programs such as cash transfers. The perception was built particularly because of the timing of the social protection program. The government's program implementation timing and data collection timing might influence the participants' perceptions. In the future, data collection and social assistance programs need to be carefully timed and designed. For example, a census or survey collecting employment data should be conducted in a month other than August, September, and October as during the time many students graduate and have yet to get a job.

This study contributes to filling the gap of previous studies in different contexts. Previous studies have highlighted that ethnostatistics reveal the misuse of statistics in the private sector (Boje et al., 2006). In contrast, this study demonstrated that variables and data integration could be a tool for reducing flaws in designing development programs. Regsosek is one of the initiatives to establish a single integrative national data. The implementation of this registration is expected to be the improvement of data flaws for national program planning. Some basic information regarding socioeconomic conditions could be adopted as the main reference in determining the government's priority programs. The data enumeration encountered various technical obstacles, such as limited access to internet signals for geospatial tagging and running the Wilkerstat platform. In general, technical problems were properly handled by enumerators and supervisors.

The primary problem was not the epistemology of Regsosek data but the ontology and axiology of data integration. Ontologically, processing and integrating data to establish agreements among participants (ministries and agencies) are crucial. The intended agreement is that all ministries and agencies acknowledge the authenticity, validity, and reliability of the Regsosek data. Recognition of Regsosek data as an integrated, comprehensive, and holistic entity could help plan a program. Sectoral data synchronization with Regsosek is significant. It was performed as a basis for updating development programs by regional governments, ministries, and agencies. Ontologically, data system stability is crucial in realizing the formation of a national data center. The targeting of national programs and development, especially for social protection, needs further attention. Axiologically, programs planned and implemented for the community must always be based on a single integrated national data. They should not be based on political interests but based on evidence. In other words, in this case, data lead policies. With the change in political power, the single integrated national data and program planning could change following

the new regime. Thus, there needs to be a short-term and long-term development plan regarding the use of data and a stable national program during the political dynamics.

CONCLUSION

Socioeconomic data collection in Indonesia has evolved since the 1960s with more variables. Lately, regarding Regsosek, the data collection was conducted door-to-door, integrated among sectors, and bottom-up. It indicated that there was a willingness among sectoral ministries to address fragmented policies, budgeting, and data for social policies. Village officials also appreciated Regsosek's ease of social assistance delivery at the village level. This study has highlighted that in social policies, statistical data collection and presentations were not affected by firms' interests but livelihood of the citizens, citizens' previous involvement in receiving social assistance, cultural hierarchy, supportive government development programs for citizens' income, disasters experienced by the citizens, and the level of citizens' formal education. The involvement of local people as enumerators or PPLs and supervisors contributed to minimizing the data biases given by the citizens. Training for enumerators and supervisors mitigated risks and enabled enumerators to collect data and explain to the citizens. PPLs had social proximity with citizens, therefore, citizens could easily trust and understand Regsosek and gave more valid answers. The leveled data verification also helped to raise socioeconomic data validity during Regsosek.

There are some policy implications of this study: (1) Timing to collect socioeconomic data influenced the result of the number of employment. The data collection was conducted during university graduation, therefore, it resulted in an increased number of employment. Statistics Indonesia needs to consider how timing could influence data results. (2) Beyond the technical side, some things need to be anticipated—political stability, considering the presidential election in Indonesia in 2024. It affected program implementation and data utilization. The crucial thing is that all parties must agree to and complete the necessary tasks regardless of the political circumstances. This study suggests to not change the direction of the data and the program in the future. Short-term and long-term development need to be agreed upon despite changing political conditions. Joint agreements among participants, especially the ministries and agencies, have become crucial in executing the national program. The agreements should contain the point that in planning the program, the database is the single integrated Regsosek data. This agreement enables all participants to commit to integrative decision-making.

The limitation of this research is that the fieldwork was conducted during the implementation of the Regsosek data collection. This research has not been able to see the planning and implementation of Regsosek at the top-level policymakers in a deeper manner. Upcoming investigations may examine deeper from key informants regarding the Regsosek planning for formulating policies for social security and protection. This research suggests future studies to analyze the integration of Regsosek data with other data. The national program planning and implementation regarding the use of Regsosek data also needs to be monitored and analyzed by academics and researchers to produce more results on statistical data utilization in policymaking.

ACKNOWLEDGEMENT

We are grateful for the assistance from Statistics Indonesia across Sumatra, Java, and Kalimantan, particularly during the data collection and for providing the required resources and infrastructure.

REFERENCES

- Badan Pusat Statistik. (2010). Penduduk Indonesia : Hasil Sensus Penduduk 2010. *Badan Statistik Indonesia*, 706. <https://www.bps.go.id/>
- Berger, P., & Luckmann, T. (2016). The social construction of reality. *Social Theory Re-Wired: New Connections to Classical and Contemporary Perspectives: Second Edition*, 110–122. <https://doi.org/10.4324/9781315775357>

- Biro Pusat Statistik. (1972). *Sensus Penduduk 1971 Penduduk Diperinci Menurut Provinsi dan Kabupaten/Kotamadya*. 1.
<https://www.bps.go.id/publication/1972/05/31/a6b0fd245504c17bc991e652/sensus-penduduk-1971-diperinci-menurut-provinsi-dan-kabupaten-kota--angka-sementara.html>
- Boje, D. M., Gardner, C. L., & Smith, W. L. (2006). (Mis)Using numbers in the Enron story. *Organizational Research Methods*, 9(4). <https://doi.org/10.1177/1094428106290785>
- BPS. (2020). *Pelaksanaan Sensus Penduduk*.
<https://www.bps.go.id/subject/12/kependudukan.html#subjekViewTab2>
- BPS. (2023). *Audiensi FKP Regsosek dan ST2023*.
<https://semarangkab.bps.go.id/news/2023/03/31/262/audiensi-fkp-regsosek-dan-st2023.html>
- Carlson, D. M., Downs, A. A., & Wert-Gray, S. (2006). Statistics as fetishes: The case of financial performance measures and executive compensation. *Organizational Research Methods*, 9(4). <https://doi.org/10.1177/1094428106291056>
- Central Bureau of Statistics. (2021). Berita Resmi Statistik. *Bps.Go.Id*, 7, 1–52.
<https://papua.bps.go.id/pressrelease/2018/05/07/336/indeks-pembangunan-manusia-provinsi-papua-tahun-2017.html>
- Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59.
<https://doi.org/10.1016/j.ssresearch.2016.04.015>
- De Zwart, F. (2012). Pitfalls of top-down identity designation: Ethno-statistics in the Netherlands. *Comparative European Politics*, 10(3), 301–318. <https://doi.org/10.1057/cep.2012.8>
- Geniusas, S. (2020). “Multiple Realities” Revisited: James and Schutz. *Human Studies*, 43(4), 545–565. <https://doi.org/10.1007/s10746-020-09548-1>
- Gërkhani, K., & Van De Werfhorst, H. G. (2013). The effect of education on informal sector participation in a post-communist country. *European Sociological Review*, 29(3).
<https://doi.org/10.1093/esr/jcr087>
- Johar, M., Soewondo, P., Pujisubekti, R., Satrio, H. K., & Adji, A. (2019). In Data We Trust? An Analysis of Indonesian Socioeconomic Survey Data. *Bulletin of Indonesian Economic Studies*, 55(1). <https://doi.org/10.1080/00074918.2018.1515474>
- Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education*, 18(2), 1–8. <https://doi.org/10.1187/cbe.18-02-0023>
- Kukutai, T., & Thompson, V. R. (2015). Censuses, Population: Comparative International Aspects. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*.
<https://doi.org/10.1016/B978-0-08-097086-8.41004-4>
- Landrum, N. E., & Boje, D. M. (2008). Using an ethnostatistical analysis to interpret data: the Nike case. *Organization Management Journal*, 5(3). <https://doi.org/10.1057/omj.2008.12>
- Liswanti, N., Shantiko, B., Fripp, E., Mwangi, E., & Laumonier, Y. (2012). Panduan praktis untuk

survei mata pencaharian sosial-ekonomi dan hak dan kepemilikan lahan untuk digunakan dalam perencanaan penggunaan lahan kolaboratif yang berbasis ekosistem. In *Panduan praktis untuk survei mata pencaharian sosial-ekonomi dan hak dan kepemilikan lahan untuk digunakan dalam perencanaan penggunaan lahan kolaboratif yang berbasis ekosistem*. CIFOR. <https://doi.org/10.17528/cifor/004031>

- Mespoulet, M. (2015). Prouver et gouverner. Une analyse politique des statistiques publiques, A. Desrosières. *Sociologie Du Travail*, 57(3). <https://doi.org/10.4000/sdt.1455>
- Mills, J. H., Weatherbee, T. G., & Colwell, S. R. (2006). Ethnostatistics and sensemaking: Making sense of university and business school accreditation and rankings. *Organizational Research Methods*, 9(4). <https://doi.org/10.1177/1094428106290786>
- Oliver, K., Lorenc, T., & Innvær, S. (2014). New directions in evidence-based policy research: A critical analysis of the literature. In *Health Research Policy and Systems* (Vol. 12, Issue 1). <https://doi.org/10.1186/1478-4505-12-34>
- Olmsted, A. D., Gephart, R. P., Gubrium, J., Noblit, G. W., & Hare, R. D. (1989). Ethnostatistics: Qualitative Foundations for Quantitative Research. *Contemporary Sociology*, 18(6). <https://doi.org/10.2307/2074243>
- Simbolon, I., Manurung, N., Andini, S., & Suadaa, L. (2023). A Sentiment Analysis and Topic Modelling of The Socioeconomic Registration 2022. *Proceedings of The International Conference on Data Science and Official Statistics*, 78-83. <https://doi.org/10.34123/icdsos.v2023i1.301>
- Smith, W. L., Boje, D. M., & Gardner, C. (2004). Using the ethnostatistics methodology to reconcile rhetoric and reality: an examination of the management release of Enron's year-end 2000 results. *Qualitative Research in Accounting & Management*, 1(2). <https://doi.org/10.1108/11766090410813337>
- Stoycheva, S., & Favero, G. (2020). Research methodology for ethnostatistics in organization studies: towards a historical ethnostatistics. *Journal of Organizational Ethnography*, 9(3). <https://doi.org/10.1108/JOE-03-2019-0016>
- Sumarto, M. (2018). *Perlindungan Sosial dan Klientelisme: Makna Politik Bantuan Tunai dalam Pemilihan Umum*. Gadjah Mada University Press.
- Thomas, M., Bloor, M., & Frankland, J. (2007). The process of sample recruitment: An ethnostatistical perspective. *Qualitative Research*, 7(4). <https://doi.org/10.1177/1468794107082300>
- Thorvaldsen, G. (2017). Censuses and Census Takers: A Global History. In *Censuses and Census Takers: A Global History*. <https://doi.org/10.4324/9781315148502>
- Tukiran, T. (2016). Sensus Penduduk Di Indonesia. *Populasi*, 11(1), 17-34. <https://doi.org/10.22146/jp.12328>
- UNICEF, UNDP, Prospera, & SMERU. (2021). Analysis of the Social and Economic Impacts of COVID-19 on Households and Strategic Policy Recommendations for Indonesia. In *UNICEF*.
- Valérian, G., Victor, A. R., Davide, C., Fabrice, B., Roser, C., Giovanni, F., & Mathis, S. (2022). *Analysis of statistical datasets and actors of the quantification of tourism. Methodological propositions*.

Wertheim, W. F., & Nitisastro, W. (1971). Population Trends in Indonesia. *Pacific Affairs*, 44(1).
<https://doi.org/10.2307/2755870>

Winiiecki, D. J. (2008). An ethnostatistical analysis of performance measurement. *Performance Improvement Quarterly*, 20(3-4). <https://doi.org/10.1002/piq.20010>