CONFERENCE OF EUROPEAN STATISTICIANS
**Expert Meeting on Statistical Data Collection**
27-30 September 2021, Online

Session 1
27 September 2021

# Integrated Collection System (ICS): Modernization of Data Collection at Statistics Indonesia

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

***Abstract***
Data collection in Indonesia is a challenge in itself. Indonesia is a large country, with an area of 1,916,862.20 km$^2$, consisting of approximately sixteen thousand islands, and the fourth largest population in the world. Each data collection activity can involve hundreds of thousands of data collectors, the number is directly proportional to the time collected, and the data processing officers needed to process the data. Statistics Indonesia is currently working on technology transformation to support a modernized business process in data production, based on analysis as written in Statistical Business Framework Architecture (SBFA). One of the transformations carried out is to build an Integrated Collection System (ICS). ICS is a multimode data collection application enabling Statistics Indonesia to collect data in an integrated manner using several modes such as Computer-Assisted Personal Interviewing (CAPI), Computer-Assisted Web Interviewing (CAWI), Desktop Data Entry for Paper and Pencil Interviewing (PAPI), and External Data Acquisition. This system reforms how Statistics Indonesia collects data using various technologies and data transfers that were unintegrated. The ICS framework simplifies the questionnaire design process (one design for CAPI, CAWI and PAPI) and cuts down on the data collection process when using CAPI and CAWI modes. With CAPI and CAWI, data analysis can be carried out in near real-time, so the waiting time to generate statistics is much faster.

## Statistical Business Framework an Architecture (SBFA)

Statistics Indonesia (Badan Pusat Statistik / BPS) is committed to modernise its statistical production processes in order to respond to user expectations and the Bureaucratic Reform Programme. They will be based on better integration, coordination and standardisation of statistical production processes, the exploitation of modern technology and the use of a wider range of data sources, including administrative data.

The Statistical Business Framework and Architecture (SBFA) provides a blueprint as to how to structure BPS's processes and the ICT infrastructure and a vision for BPS until 2025. It provides the underlying structure for the modernisation programme, guidance to statistical design and redesign projects, and assurance that a consistent approach will be used now and in future.

The SBFA describes, based on the General Statistical Business Process Model (GSBPM), the integrated statistical production processes as value chains in which at each phase value is added to the products. It is a service oriented model in which separate organisational units like (meta) data management, Statistical Business Register, Large Business Unit and a Field Force Unit provide services and products to the value chains. A comprehensive set of

corporate statistical infrastructure (CSI) serve as the pillars of the new production processes including data bases and warehouses to support (meta-) data management, data collection, data processing and dissemination.

There are a number of forces driving the modernisation process, most notably the need to satisfy the users of BPS at national, regional and international levels; the need to improve response rates; the development of administrative data for use in official statistics; the Government's Open Data Initiatives; and Public Sector Reforms more generally. In combination these driving forces provide an excellent opportunity for BPS to modernise its statistical operations.

The implementation of the SBFA will provide a modernised BPS which will:

- Provide for the integration of statistical production and statistical outputs;
- Adopt international standards wherever appropriate;
- Adopt international best practice in its statistical activities;
- Adopt an integrated approach to the acquisition and use of technology, reducing the costs of acquisition and maintenance;
- Support innovative data collection and capture modes to improve efficiency;
- Use a common approach to data management, including a Dissemination Data warehouse, to ensure consistent and transparent outputs;
- Make greater use of administrative data which, together with the rationalisation of the survey programme, will reduce burden on respondents;
- Provide BPS staff with the competencies to work in a modernised statistical offices;
- Respect the specific needs of Provincial and District users, including improved services to regional users through better access to national data bases with a regional dimension;
- Actively lead the National Statistical System (NSS);
- Improve the relationship with the providers of data and the users of BPS.

BPS will rationalise its data collections. This will respond to the Bureaucratic reform agenda for improved quality and increased efficiency; ensure that resources are directed at producing high quality, timely statistics that best meet the priority needs of users and reduce respondent burden. The rationalisation will be based on a review of existing collections and their characteristics, assessing importance, the level of duplication and the scope for simplification and for using other sources. BPS will also put in place procedures for ensuring that new or modified collections, at national and regional level, meet corporate standards.

The implementation of the SBFA will ensure a cohesive IT environment based on an enterprise architecture and systems that are developed using the corporate toolbox and by using common application development standards.

The implementation of the SBFA will lead to changes in the way BPS has traditionally organised its activities. The Provincial and District Offices are now and in the new BPS an indispensable part of the BPS organisational arrangements with their own specific roles and responsibilities. This will be taken into account in the new organisational settings.

## Computer Assisted Interviewing in the Context of the SBFA

For the realisation of the pointed modernized of BPS Computer Assisted Interviewing (CAI) is an essential mean. CAI is the collection of information with the use of an electronic questionnaire operating on a computer platform and linked to a local or remote data base. This could be shaped as computer assisted personal interviewing (CAPI), computer assisted self-interviewing (CASI) or computer aided web interviewing (CAWI).

The benefits of CAI are an enhanced quality of data by utilising the quality assurance mechanisms of electronic questionnaires and computer assisted interviewing technique. This will result in greatly reduced post-data collection editing and imputation processing. Next the timeliness of the data will improve because the CAI environment provides almost instantaneous access to preliminary raw data as they are collected in the field and processed in the central processing unit. A third advantage of CAI is the recording of a large amount of paradata about the survey process and related metadata. This information enables a profound analysis of the interviewing process.

The introduction of CAI will change the organisation of the surveys. To take full advantage of computer assisted interviewing there is the need for specialised questionnaire design specialists, who have good understanding of broader survey methodological. Centralisation of these skills within one team ensures higher standardisation of approaches to the measurements of the same concepts. Furthermore, centralisation facilitates the transition of knowledge and experience within the team of questionnaire designers and leads to higher quality and better efficiency of the team.

The use of electronic devices especially for personal interviewing (CAPI) will also change the work of the enumerators. Timely, almost instantaneous, access to the data provides live feedback information on the development of the fieldwork as well as possibility to escalate the levels of quality assurance by validating and verifying the collected information. The use of electronic questionnaire instead of paper questionnaire results in data entry and control during the interview which reduces this kind of work after the interview was held.

## Development Journey

Starting from the need to make surveys faster and the data collection process was getting shorter, in 2015, BPS began to pilot the use of CAPI. The applications used at that time were Survey Solution and Open Data Kit (ODK). From the trial results, it was concluded that doing surveys and collecting data using CAPI was faster than the traditional method (paper-based survey).

In early 2017, a CAPI study was conducted to strengthen readiness. The purpose of the study was to collect information on the functional requirements of CAPI following the BPS survey. To achieve this goal, the Directorate of Statistical Information Systems conducts discussions and interviews with several subject matters, which will use CAPI. Moreover, the CAPI study collects information about the types of CAPI applications available and their features. The study results conclude that the current CAPI application cannot fully meet the functional needs of the BPS survey, shown in the table bellows. Therefore, at the end of 2017, BPS decided to make its own CAPI application called the CAPI system framework.

Table 1. Comparison between CAPI applications

| No. | Kriteria | CAPI Application | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Open Data Kit (ODK) | Survey Solutions | CSPro for Android | Blaise | Collect | Survey System | SurveyBe | Survey ToGo |
| 1 | *Free* | √ | √ | √ | x | x | x | x | x |
| 2 | *Open Source* | √ | x | x | x | x | x | x | x |
| 3 | Does it resemble other programming languages? (VB / C# / Java / ...) | XML | C# | CSProLanguage | TBD | TBD | Web based | SQL dan HTML | TBD |
| 4 | Programming syntax using menus & command-line | √ | √ | √ | √ | Menu driven | TBD | √ | TBD |
| 5 | Support collaborative program development | TBD | √ | TBD | √ | √ | TBD | x | TBD |
| 6 | Extensible & the code is reusable (not disposable, can be upgraded to use | √ | Reusable but can't be developed | Reusable | Reusable | Reusable | Reusable | Reusable tetapi dalam potongan besar | TBD |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | other surveys) | | | | | | | | |
| 7 | Able to accommodate various types of questions (including roster) | √ | √ | √ | √ | √ | √ | √ | √ |
| 8 | Able to accommodate a large number of records | √ | √ | √ | √ | √ | √ | √ | TBD |
| 9 | Able to accommodate hierarchical questions | √ | √ | √ | √ | √ | √ | TBD | TBD |
| 10 | Have a basic navigation menu & complex skipping | √ | √ | √ | √ | √ | √ | √ | √ |
| 11 | Can pause, resume, save | √ | √ | √ | TBD | √ | TBD | Yes, but the resume feature does not automatically continue from the last question | TBD |
| 12 | Can easily add validation rules as needed | √ | √ | √ | √ | √ | TBD | √ | √ |
| 13 | Allows post-interview editing | √ | √ (*By Interviewer*) | √ | TBD | √ | √ | √ | TBD |
| 14 | Facilitates step-by-step corrections & checks | √ | √ | TBD | TBD | TBD | TBD | √ | √ |
| 15 | Able to send directly to the server | √ | √ | √ | √ | √ | √ | √ | √ |
| 16 | Provides back-up feature | √ | √ | √ | √ | √ | √ | √ | √ |
| 17 | Able to generate files in formats as needed (eg .csv, .xls) | √ (.csv, .xml, JSON,&.kml) | √ (.sav, .dta& .csv) | √ | √ (SQL lite, SQL server, oracle, mySQL, postgre) | √ | TBD | √ (.csv dan .do) | √ (SPSS, Excel, Word, Access,XML) |
| 18 | Can be added feature handbook | √ | x | √ | TBD | TBD | TBD | TBD | TBD |
| 19 | Can be integrated with picture / sound / video | √ | TBD | Picture | TBD | √ | √ | √ | √ |
| 20 | Support GPS | √ | √ | TBD | TBD | √ | TBD | √ | √ |

In 2018 - 2019 the CAPI system framework was used in several BPS surveys such as Migration Survey Pilot 2018, SP2020 Pilot, Sugar Cane Commodity Survey 2018, Potential Horticultural Survey 2018, Terrorism Risk Survey 2018, SPAK 2019, Podes 2019, and Tourism Survey 2019. However, there were still some obstacles faced, especially in the survey, which had many questions and a large sample.

In 2019, a re-evaluation of the functional requirements of CAPI was carried out. One of the main features to be added is the use of maps in CAPI. Besides that, the 2020 Population Census, for the first time, had been designed using CAWI. So it was necessary to create a collection system that supports CAWI. Therefore, a system that can be used for CAPI, CAWI, and PAPI (Paper and Pencil Interviewing) was designed, called the Integrated Collection System (ICS).

**Integrated Collection System Architecture**

Integrated Collection System (ICS) consists of several modules such as:

1. Survey Collection Management (SCM), this module manages survey creation and sample assignment to enumerators.
2. Questionnaire Designer (QD), this module functions to create digital questionnaires and apply validation rules. In QD, two sub-modules help the variable management process consist of Variable Metadata Repository and Validation Engine.
3. Field Force Management (FFM), this module functions to manage field officers and entry officers. Meanwhile, this module is connected with SCM for user management.
4. Collection Tools, this module functions to input data from the respondent answer into the database. The Collection tools consists of 4 modes such as Computer Assisted Personal Interviewing (CAPI), computer aided web interviewing (CAWI), Desktop Data Entry (DDE) for Paper and Pencil Interviewing (PAPI), and Acquisition External Data applications. The CAPI mode is used for mobile application enumeration, where field officers input data simultaneously as the interview process. The CAWI mode is used when the survey is conducted using self-enumeration, i.e., survey respondents fill out the questionnaire independently using a browser. The DDE for PAPI mode is used when the survey is conducted using a paper questionnaire and then inputted by the entry clerk either manually or by a scanner. Meanwhile, the Acquisition External Data line is used if the data is obtained from other Ministries/Agencies through the data transmission line.

The relationship between the ICS module and the system outside the ICS can be seen in the following figure:

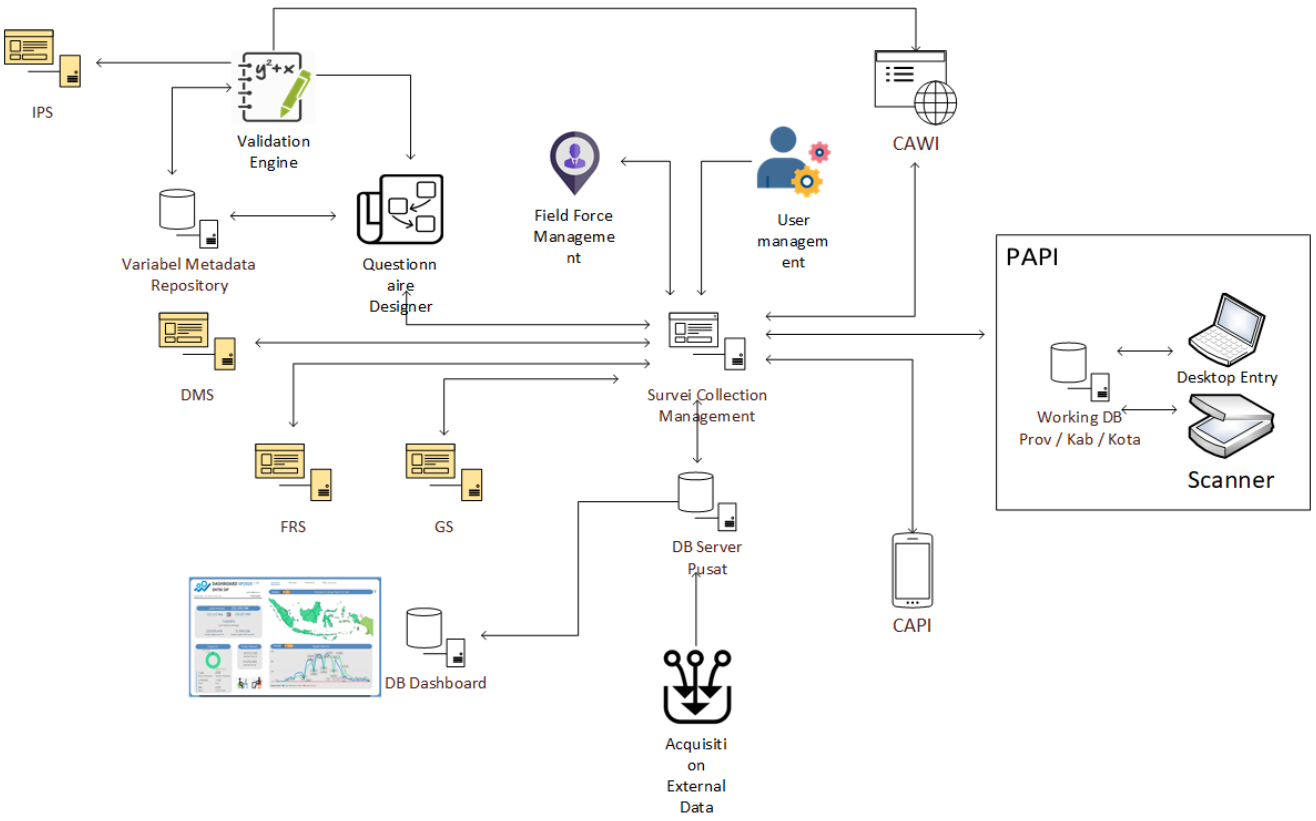


Figure 1. Relation between ICS modules and other systems

The implementation of ICS in a survey can be broadly divided into several stages, including:

1. *Designing a Questionnaire*

The first step in implementing ICS is to design a QD questionnaire based on a paper questionnaire design. The challenge is to adapt the paper design into the form of a CAPI/CAWI/PAPI questionnaire. This will undoubtedly be adapted for certain types of questions. In addition, at this stage, also enter a validation rule into each question, including the type of error and the error message.

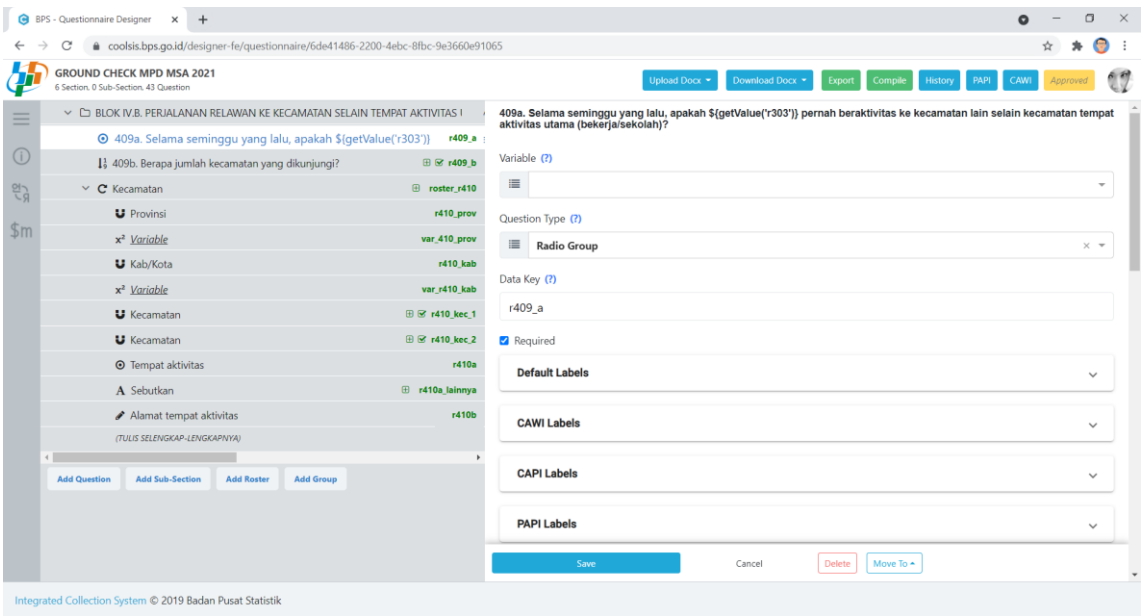


Figure 2. Questionnaire Designer

*2. Create and manage surveys*

After designing the questionnaire, the next step is to create and organize a survey on SCM. This stage is included in the survey design stage. In more detail, things that need to be regulated at this stage are giving the name of the survey, setting up users, adding samples, determining the survey period, determining the type of survey (whether enumeration or listing), determining user roles, determining the training schedule, and of course entering the questionnaire design that has been made. Previously. Samples are obtained from the Frame and Register System (FRS) or manually entered into the database via SCM.

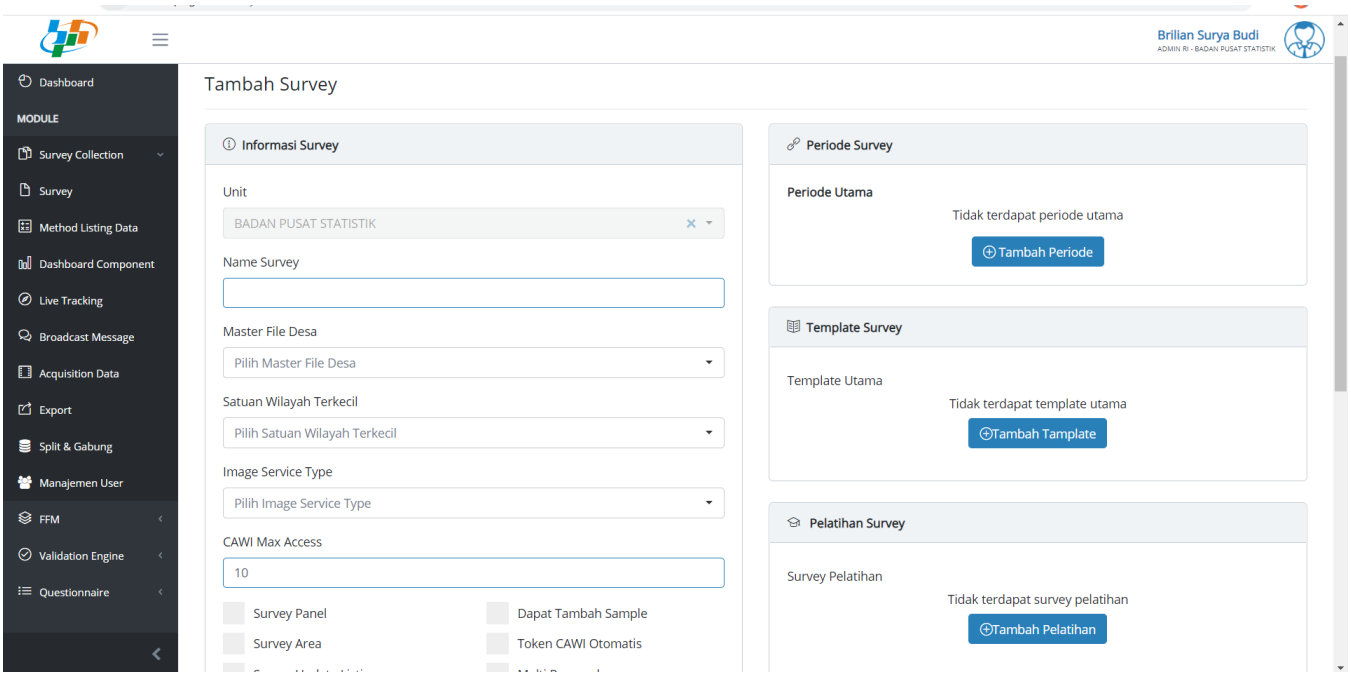


Figure 3. Create New Survey

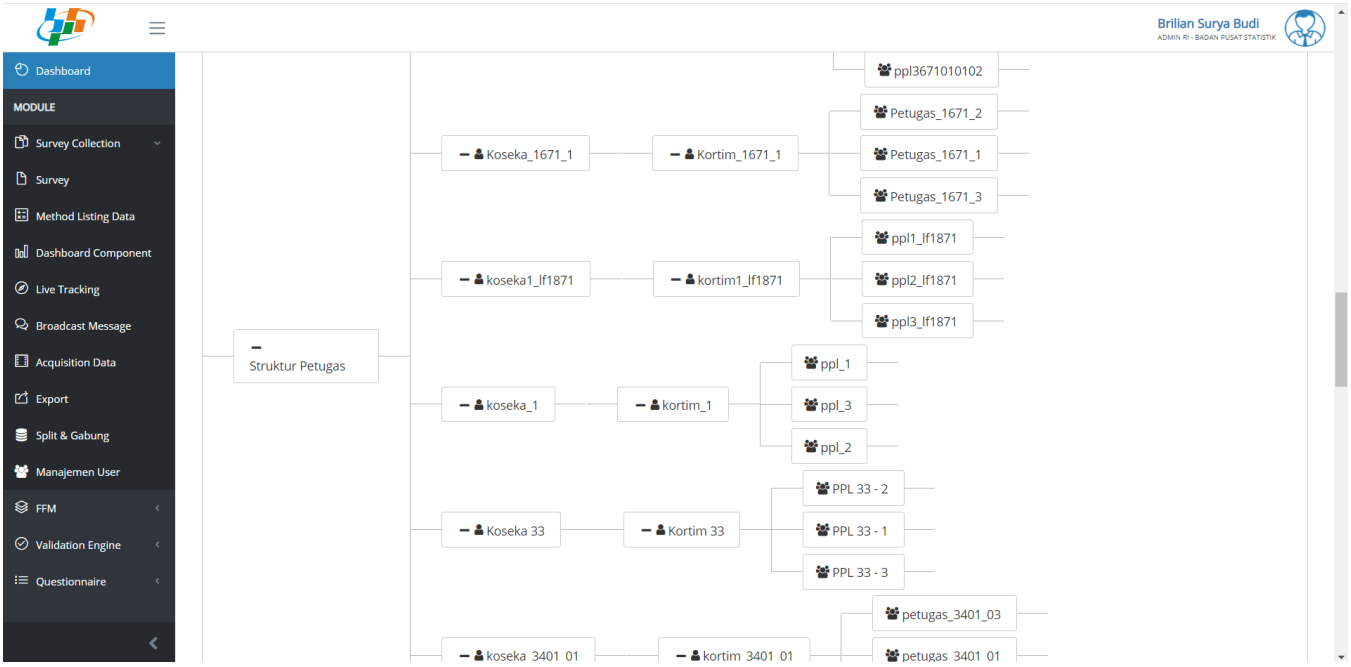


Figure 4. Field Force Structure

*3. CAPI / CAWI / PAPI Implementation*

The next stage is the data collection stage. At this stage, three modes of data collection can be used, namely CAPI, CAWI, and PAPI. The use of this mode depends on how the data is collected for each survey. CAPI is more suitable for surveys conducted by interviewing officers in areas with good internet connections. In comparison, CAWI is more suitable for self-enumeration surveys where respondents fill in their digital questionnaires. The last one is PAPI, which tends to suit surveys conducted by interviewing officers in areas with poor connections. Although there are three different modes, one survey can use all three modes at once. This is possible because there are surveys that have different business processes in each region. This condition does not mean that the questionnaire is designed as many as the modes used. In ICS, the questionnaire is designed only once and can be used in all modes in one survey period.

The role of the validation rule that has been installed on the questionnaire is to ensure that the inputted data is following the specified rule. This causes data that still violates the validation rules cannot be sent to the server. So that cross-checking the data can be done immediately.

During CAPI enumeration, there are times when a map is needed as an indication of the boundaries of the enumeration area. Map data is taken from the Geospatial System (GS). The map used by the enumerator is only the area that corresponds to the assignment area being worked on by the enumerator.

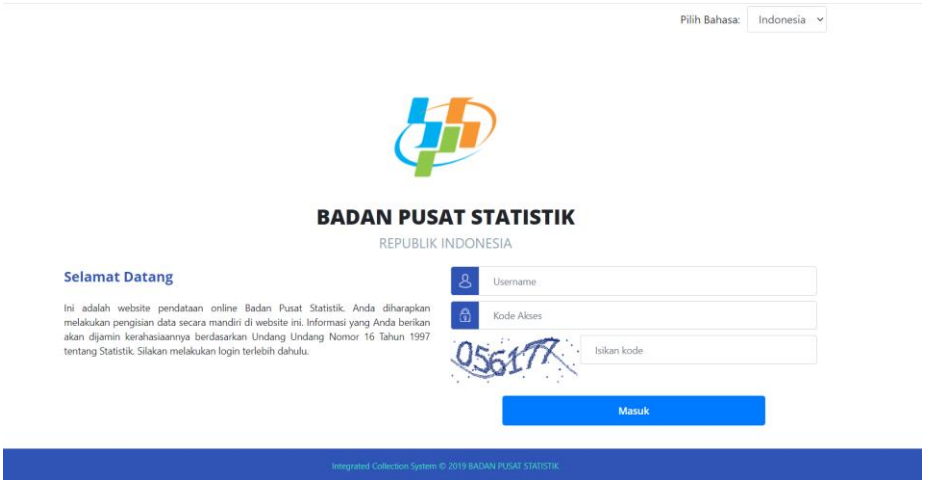


Figure 5. Landing page of CAWI

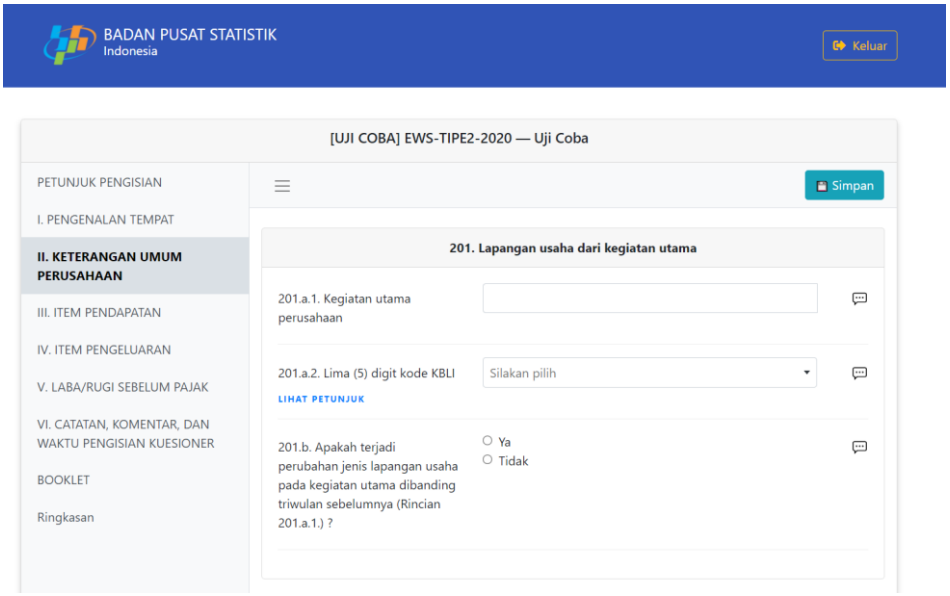


Figure 6. CAWI e-form

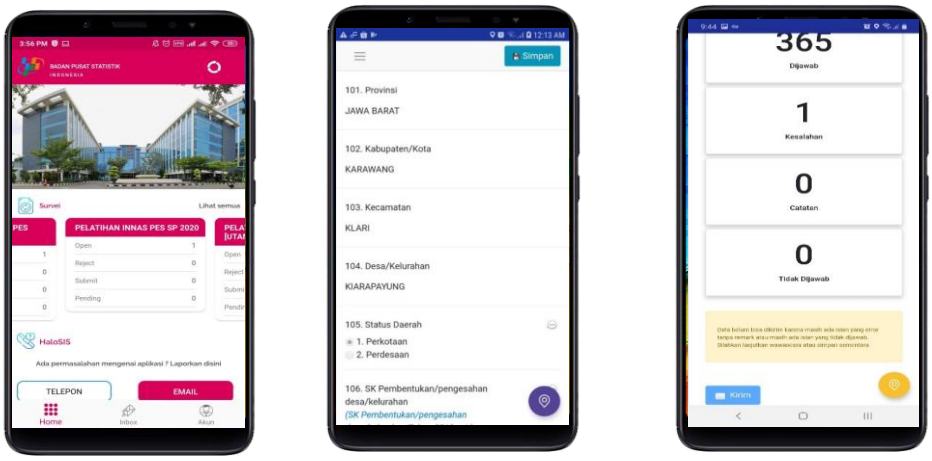


Figure 7. CAPI Mobile Application

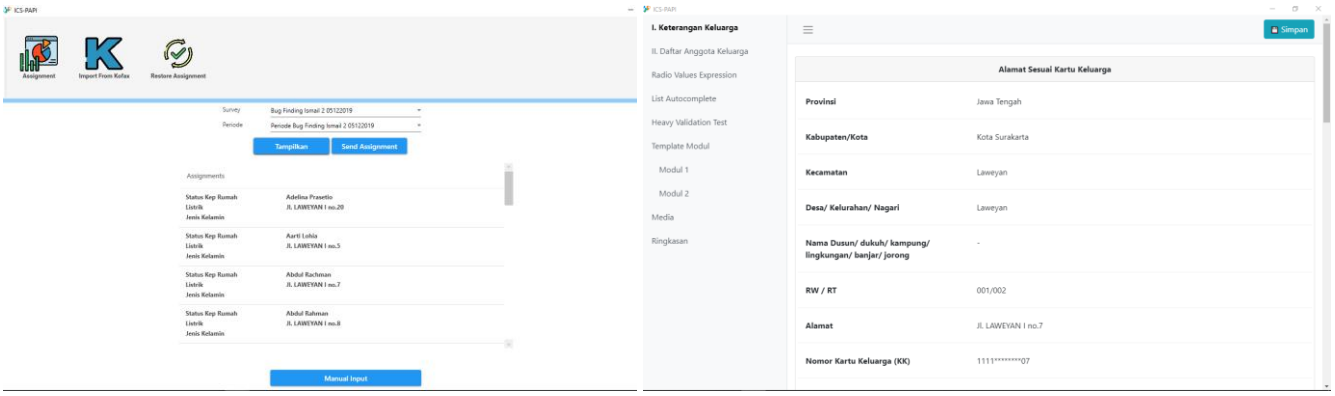


Figure 8. Desktop Data Entry for PAPI

*4. Inspection*

Examination of the results of data collection is the next stage. Examinations are carried out by officers who have the role of inspectors. More checks on the fairness of the data. Data that is successfully sent to the server indicates that no validation rule has been violated. However, it is still possible for there to be unnatural stuffing. Examples of coding business classifications that do not match the description provided, the size of the entries compared to the number of family members, the number of hours worked compared to the number of workers, and others.

The examination results are marked with a note on each question that needs to be reconfirmed by the survey officer. If there is such a condition, the examiner can notify the office by rejecting it. With the reject status, the officer is obliged to confirm every question that has a record. However, if all the entries are considered reasonable, the examiner can approve them.

*5. ETL enumeration results to other systems.*

All data that has been entered into the ICS server, the data is processed through the ETL (Extract-Transform-Load) process to be displayed in the dashboard, further processed by the Integrated Processing System (IPS) or directly saved to the Data Management System (DMS).

## Impacts and Changes with ICS

Currently, only 2 ICS Collection Tools are used, namely CAPI and CAWI. Therefore, this section will only describe the impact and changes of using ICS on CAPI and CAWI.

For BPS, ICS (especially CAPI and CAWI) impact the speed of the data collection process and improved data quality and data availability. The increase in the data collection process at CAPI and CAWI occurs because many business processes can be trimmed, such as the absence of procuring paper questionnaires, editing coding, and entry in the processing center. Meanwhile, the increase in data quality can be seen when filling out the CAPI and CAWI,

where the data can be directly validated so that the level of data errors can be minimized and directly confirmed to the respondents. Because many post-data collection business processes are trimmed, and data entered in CAPI and CAWI can be sent directly to the server, data availability in the data center can be accessed and viewed in real-time so that the data can be analyzed without having to wait for the survey activity to end.

However, because CAPI and CAWI are primarily viewed through smartphones, with minor screen displays, changes must be made to the flow of filling out the questionnaire and the layout of the questionnaire display.

For the Provincial / Regency / City BPS in several surveys, there is a shift in roles and responsibilities from the Provincial BPS to the Regency / City BPS. This is because the district/city BPS mainly carries out CAPI field management (both officers and electronic documents) compared to the provincial BPS. Therefore, the role and function of the Provincial BPS can be transferred to the analysis of survey data.

For Enumerators, additional requirements are needed to become a CAPI enumerator: familiar with using gadgets to make it easier for the enumerator to understand the CAPI ICS application. However, they will still receive instructions on the operation of the CAPI ICS application.

For programmers, it is faster to meet the needs of providing data collection applications. Changes to validation rules can also be done more quickly on the ICS QD module. Before ICS, programmers in building applications had to code from scratch. When changes occurred, both in design and content, it took a long time to adjust to these changes.

On the other hand, before ICS, when one application was custom-built for one survey, the application development load increased, and the knowledge maintained varied. With ICS, programmers can focus more on developing ICS applications. Collaboration for data collection applications becomes easier because it starts with the same knowledge and technology.