Census Data Capture in the 2010 Population and Housing Censuses: The Indonesian Case

Cahyono, Indra

Statistics - Indonesia

Jl. Dr. Sutomo 6-8

Jakarta 10710, Indonesia

E-mail: indrac@bps.go.id

Levin, Michael J

Harvard Center for Population and Development Studies

9 Bow Street

Boston, MA02115, United States

E-mail: jlevin00@yahoo.com

Introduction

Indonesia is the fourth most populous country in the world. In 2010, Statistics Indonesia (BPS) collected its sixth population census. The previous population censuses were conducted in 1961, 1971, 1980, 1990, and 2000, and those starting in 1971 used computer processing (See Table 1 below). In 1961, because of Indonesia conditions at that time and as well as limited resource of BPS manually calculated the population census. And only 10 percent of collected information could be processed. In 1971, BPS used two systems to conduct the census: a complete census with simple questionnaire, processed manually, and a sample census for 13,793 census blocks (3.18 percent of 361,843 census blocks) using Optical Mark Recognition (OMR) technology.

So, BPS began using data capture technology since 1971. BPS used a mainframe computer in 1980 to process the census centrally since processing using OMR technology would have been very costly, requiring high specifications and extra handling of documents. In addition, the innovation of data entry technology was more advanced. In 1990, population census data processing still used data entry to the mainframe.

In 2000, BPS used Optical Character Recognition (OCR). The use of this technology is based on the desire to be able to publish data until the lowest level of region (small area statistics), the number of questions in the questionnaire (15 questions) and the number of documents to process. Study and implementation of the processing was aided by the Japanese Government through Japan International Cooperation Agency (JICA). OCR software used was Nestor Reader and the scanner was Kodak DS3500. Data processing took 10 months for data capture, 8 months for editing and 20 months for publishing following census implementation.

Just as in 2000, the population census 2010 also used OCR technology. BPS conducted the population census for all individuals and housing units, including names and addresses. BPS made efforts to improve the data capture operation to get data to the public and private sector users more quickly, so requested bids from several companies. BPS selected Kofax, Inc. to provide data capture processing. Data capture occurred in 33 locations in each of Indonesia's provinces.

As this paper will show, we developed a program to transform the captured information into ASCII for CSPro processing. We applied adapted content editing programs from our pilot census to the scanned data. Hence, we were able to obtain a completely edited file for the whole country by November, 2010, and tables before the end of the calendar year. The whole process took 11 months from beginning of enumeration to printer-ready copy.

Year	Technology
1961	Manual
1971	Manual dan Optical Mark Reader (OMR)
1980	Data Entry – Mainframe
1990	Data Entry – Mainframe
2000	Optical Character Recognition (OCR)
2010	Optical Character Recognition (OCR)

Table1. Data Processing Technology Used

This paper discusses various aspects of our capture procedures, including check in from the field, slicing the forms for scanning, reading the forms, manually checking of the forms, and transfer of the final captured data to the central office for continued processing.

Planning for 2010 Population Census Data Processing

Indonesia consists of 33 provinces and BPS has branch offices in every province. The number of provinces to be a reference to the establishment of data processing centers. The estimated total population, household and census blocks for 2010 was 233,596,970 people, 63,645,055 households and 726,234 census blocks. Number of households reflected the number of documents to be processed. The information collected in the population census 2010 included individual questions, mortality, and housing information. The 42 questions provided details based on 21 individual questions, 8 mortality items and 12 housing questions.

Planning for the 2010 data collection started in 2003 when the Indonesian government assigned BPS to perform site data collection and data processing for the preparation of general election 2004 in Indonesia. The data collecting was similar to a population census of 15 individual questions. The data processing took 4 months with excellent results. The data capture technology applied was OCR using Kofax software and scanning machine Kodak DS3500. The study showed OCR obtains optimum system in terms of cost, resource and results.

The Assumptions Used

Some of the assumptions used to calculate the planning for the processing of population census 2010 included: the maximum time to process was 6 months with 25 working days per month and 18 hours of work per day. Days and hours of work applied the minimum assumptions. The 42 questions required 8 pages for 1 questionnaire per household. The average number of members in a household was assumed to be 6 persons. BPS then added 5 percent as a reserve. So to calculate the number of pages to be processed for each province, the following formula was applied:

$$\sum \textit{total pages} = (\sum \textit{block census} + (\sum \textit{household} * 6))* 105\%$$

The calculation was carried out for each province to determine estimated for processing in each province. Table 2 shows the results.

No	Province	Est Total of Block Census	Est Total of Household	Est Total of Population	Est Total of Pages
(1)	(2)	(3)	(4)	(5)	(6)
1	Aceh	12.478	1.087.735	4.133.360	9.150.076
2	Sumatera Utara	35.770	3.084.056	13.261.393	25.943.629
3	Sumatera Barat	11.535	1.103.340	4.744.329	9.280.168
4	Riau	15.721	1.434.682	5.595.241	12.067.836
5	Jambi	10.018	740.369	2.887.432	6.229.619
6	Sumatera Selatan	20.729	1.883.701	7.346.410	15.844.854
7	Bengkulu	4.556	446.496	1.696.671	3.755.350
8	Lampung	24.962	1.960.280	7.645.068	16.492.562
9	Kep Bangka Belitung	2.123	284.735	1.138.931	2.394.003
10	Kepulauan Riau	5.056	403.013	1.571.742	3.390.618
11	DKI Jakarta	33.920	2.627.114	9.194.888	22.103.374
12	Jawa Barat	128.583	12.867.992	42.464.331	108.226.145
13	Jawa Tengah	99.360	8.819.554	32.632.277	74.188.582
14	DI Yogyakarta	11.194	1.099.800	3.519.352	9.250.074
15	Jawa Timur	138.499	10.623.282	37.181.437	89.380.993
16	Banten	29.085	2.778.679	10.281.100	23.371.443
17	Bali	8.101	975.096	3.607.833	8.199.312
18	NTB	15.107	1.371.422	4.525.677	11.535.807
19	NTT	10.948	999.734	4.598.732	8.409.261
20	Kalimantan Barat	10.309	1.072.826	4.398.559	9.022.563
21	Kalimantan Tengah	7.847	598.337	2.153.983	5.034.270
22	Kalimantan Selatan	12.240	989.144	3.560.898	8.321.662
23	Kalimantan Timur	9.368	878.409	3.250.086	7.388.472
24	Sulawesi Utara	8.997	667.337	2.268.925	5.615.078
25	Sulawesi Tengah	5.832	575.485	2.532.112	4.840.198
26	Sulawesi Selatan	22.291	1.936.470	7.939.483	16.289.754
27	Sulawesi Tenggara	5.844	539.573	2.212.221	4.538.549
28	Gorontalo	3.146	285.639	971.163	2.402.671
29	Sulawesi Barat	3.351	251.673	1.031.851	2.117.572
30	Maluku	4.194	270.660	1.353.288	2.277.948
31	Maluku Utara	2.319	178.920	984.048	1.505.363
32	Papua Barat	2.717	209.354	753.661	1.761.426
33	Papua	10.034	600.148	2.160.488	5.051.779
		726.234	**63.645.055**	**233.596.970**	**535.381.008**

Table 2. Estimated Total of Pages to be Process

Planning for Data Processing Using Data Entry Technology

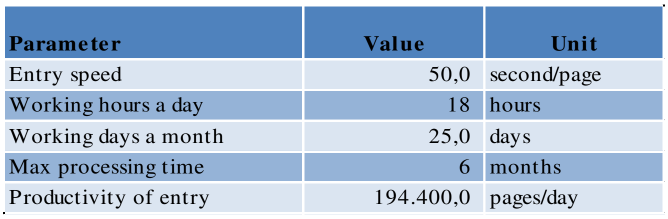
The use of technology in terms of making data entry application and implementation was easier. BPS owns application builder software, such as Clarion 6, Power Builder 10, Microsoft Visual Basic, Microsoft C#, etc. BPS also owns Relational Database Management System (RDBMS) like Sybase ASA, PostgreSQL, Microsoft SQL and MySQL. The data entry technology required only personal computers and a server to run the Client-Server-based applications. BPS already had System Analysts and Programmers to develop the system to perform the large amount of data processing, assisting in easier starting and cheaper procuring of software and hardware for data entry.

One constraint emerged because the allocated time for processing duty was only 6 months. To complete the process in a timely manner, the verification process required many PCs and many operators. In calculating number of PCs and operators required for data entry, the following formulation was used as reference:

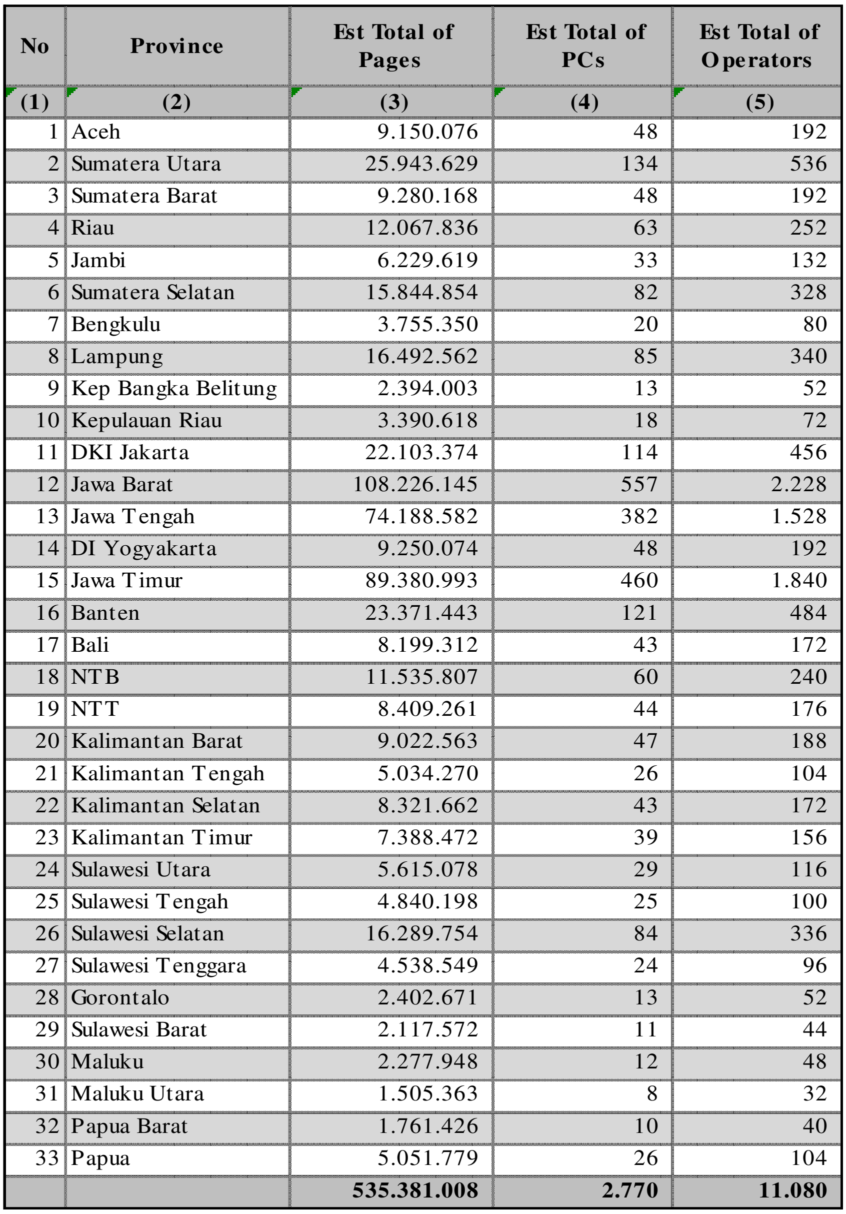
Productivity of entry = *max. processing time * working days a month * working hours a day * 60 minutes * 60 seconds / entry speed*

Est. Total PCs = *est. Total Pages / Productivity of entry*

Est. Total Operators _= est. Total PC * 4_



Parameter	Value	Unit
Entry speed	50,0	second/page
Working hours a day	18	hours
Working days a month	25,0	days
Max processing time	6	months
Productivity of entry	194.400,0	pages/day



No	Province	Est Total of Pages	Est Total of PCs	Est Total of Operators
(1)	(2)	(3)	(4)	(5)
1	Aceh	9.150.076	48	192
2	Sumatera Utara	25.943.629	134	536
3	Sumatera Barat	9.280.168	48	192
4	Riau	12.067.836	63	252
5	Jambi	6.229.619	33	132
6	Sumatera Selatan	15.844.854	82	328
7	Bengkulu	3.755.350	20	80
8	Lampung	16.492.562	85	340
9	Kep Bangka Belitung	2.394.003	13	52
10	Kepulauan Riau	3.390.618	18	72
11	DKI Jakarta	22.103.374	114	456
12	Jawa Barat	108.226.145	557	2.228
13	Jawa Tengah	74.188.582	382	1.528
14	DI Yogyakarta	9.250.074	48	192
15	Jawa Timur	89.380.993	460	1.840
16	Banten	23.371.443	121	484
17	Bali	8.199.312	43	172
18	NTB	11.535.807	60	240
19	NTT	8.409.261	44	176
20	Kalimantan Barat	9.022.563	47	188
21	Kalimantan Tengah	5.034.270	26	104
22	Kalimantan Selatan	8.321.662	43	172
23	Kalimantan Timur	7.388.472	39	156
24	Sulawesi Utara	5.615.078	29	116
25	Sulawesi Tengah	4.840.198	25	100
26	Sulawesi Selatan	16.289.754	84	336
27	Sulawesi Tenggara	4.538.549	24	96
28	Gorontalo	2.402.671	13	52
29	Sulawesi Barat	2.117.572	11	44
30	Maluku	2.277.948	12	48
31	Maluku Utara	1.505.363	8	32
32	Papua Barat	1.761.426	10	40
33	Papua	5.051.779	26	104
		535.381.008	**2.770**	**11.080**

Table 3. Estimated Total of PCs and Data Entry Operators

Table 3 shows that big provinces needed as many as 400 personal computers. These computers would need 1,600 operators, based on 4 shifts per day. Most of the province had difficulty in finding enough computer operators to fill the shifts. This activity required large electric capacity and 24-hours operation to complete the task.

Planning for Data Processing Using OCR Technology

The calculation of data processing by applying OCR technology was more complicated. However, based on the calculation, the number of PCs needed was significantly smaller than would have been needed for straight data entry. For the OCR technology, the need of PCs was calculated based on each processing stage, from Scan, Recognition, Correction, Completion to Release. The operators were only required in the stages of Scan, Correction and Completion, while Recognition and Release did not require operators. Tables 4 and 5 show the estimates for the various processing stages.

Scan	Value	Unit
Scanner Speed (Potrait A4)	95	%
Daily Duty Cycle (mfd recommend)	100.000	pages/day
Working hours	18	hours
Working days	25	days
Max processing time	6	months

No	Province	Est. Total Of Pages	Daily Duty Cycle	Scanner	Productivity
1	Aceh	9.150.076	61.001	1	61%
2	Sumatera Utara	25.943.629	172.958	2	86%
3	Sumatera Barat	9.280.168	61.868	1	62%
4	Riau	12.067.836	80.452	1	80%
5	Jambi	6.229.619	41.531	1	42%
6	Sumatera Selatan	15.844.854	105.632	2	53%
7	Bengkulu	3.755.350	25.036	1	25%
8	Lampung	16.492.562	109.950	2	55%
9	Kep Bangka Belitung	2.394.003	15.960	1	16%
10	Kepulauan Riau	3.390.618	22.604	1	23%
11	DKI Jakarta	22.103.374	147.356	2	74%
12	Jawa Barat	108.226.145	721.508	8	90%
13	Jawa Tengah	74.188.582	494.591	5	99%
14	DI Yogyakarta	9.250.074	61.667	1	62%
15	Jawa Timur	89.380.993	595.873	6	99%
16	Banten	23.371.443	155.810	2	78%
17	Bali	8.199.312	54.662	1	55%
18	NTB	11.535.807	76.905	1	77%
19	NTT	8.409.261	56.062	1	56%
20	Kalimantan Barat	9.022.563	60.150	1	60%
21	Kalimantan Tengah	5.034.270	33.562	1	34%
22	Kalimantan Selatan	8.321.662	55.478	1	55%
23	Kalimantan Timur	7.388.472	49.256	1	49%
24	Sulawesi Utara	5.615.078	37.434	1	37%
25	Sulawesi Tengah	4.840.198	32.268	1	32%
26	Sulawesi Selatan	16.289.754	108.598	2	54%
27	Sulawesi Tenggara	4.538.549	30.257	1	30%
28	Gorontalo	2.402.671	16.018	1	16%
29	Sulawesi Barat	2.117.572	14.117	1	14%
30	Maluku	2.277.948	15.186	1	15%
31	Maluku Utara	1.505.363	10.036	1	10%
32	Papua Barat	1.761.426	11.743	1	12%
33	Papua	5.051.779	33.679	1	34%
		535.381.008	**3.569.207**	**55**	

Table 4. Estimated Scanner and Productivity

Recognition			Correction			Completion			Release		
Parameter	Value	Unit	Parameter	Value	Unit	Parameter	Value	Unit	Parameter	Value	Unit
Speed	0,9	second/page	Speed	6.500,0	char/hour	Speed	15,0	second/page	Speed	0,62	second/page
Working hours	18	hours	Working hours	18	hours	Working hours	18	hours	Working hours	18	hours
Working days	25,0	days	Working days	25,0	days	Working days	25,0	days	Working days	25,0	days
Max proc. time	6	months	Max proc. time	6	months	Max proc. time	6	months	Max proc. time	6	months
Productivity	10.800.000,0	pages/station	Productivity 1	17.550.000,0	char/station	Productivity	648.000,0	pages/station	Productivity	15.677.419,0	pages/station
			Num. of Char.	125	/page	Compl. Rate	40	%			
			Accuracy	90,0	%						
			Productivity 2	1.404.000	pages/station						

No	Province	Est Total Of Pages	Est. PC Recognition	Province	Est Total Of Pages	Est. PC Correction	Province	Est Total Of Pages	Est. PC Completion	Province	Est Total Of Pages	Est. PC Release
1	Aceh	9.150.076	1	Aceh	9.150.076	7	Aceh	9.150.076	6	Aceh	9.150.076	1
2	Sumut	25.943.629	3	Sumut	25.943.629	19	Sumut	25.943.629	17	Sumut	25.943.629	2
3	Sumbar	9.280.168	1	Sumbar	9.280.168	7	Sumbar	9.280.168	6	Sumbar	9.280.168	1
4	Riau	12.067.836	2	Riau	12.067.836	9	Riau	12.067.836	8	Riau	12.067.836	1
5	Jambi	6.229.619	1	Jambi	6.229.619	5	Jambi	6.229.619	4	Jambi	6.229.619	1
6	Sumsel	15.844.854	2	Sumsel	15.844.854	12	Sumsel	15.844.854	10	Sumsel	15.844.854	2
7	Bengkulu	3.755.350	1	Bengkulu	3.755.350	3	Bengkulu	3.755.350	3	Bengkulu	3.755.350	1
8	Lampung	16.492.562	2	Lampung	16.492.562	12	Lampung	16.492.562	11	Lampung	16.492.562	2
9	Babel	2.394.003	1	Babel	2.394.003	2	Babel	2.394.003	2	Babel	2.394.003	1
10	Kepri	3.390.618	1	Kepri	3.390.618	3	Kepri	3.390.618	3	Kepri	3.390.618	1
11	DKI Jakarta	22.103.374	3	DKI Jakarta	22.103.374	16	DKI Jakarta	22.103.374	14	DKI Jakarta	22.103.374	2
12	Jawa Barat	108.226.145	11	Jawa Barat	108.226.145	78	Jawa Barat	108.226.145	67	Jawa Barat	108.226.145	7
13	Jawa Tengah	74.188.582	7	Jawa Tengah	74.188.582	53	Jawa Tengah	74.188.582	46	Jawa Tengah	74.188.582	5
14	DI Yogyakarta	9.250.074	1	DI Yogyakarta	9.250.074	7	DI Yogyakarta	9.250.074	6	DI Yogyakarta	9.250.074	1
15	Jawa Timur	89.380.993	9	Jawa Timur	89.380.993	64	Jawa Timur	89.380.993	56	Jawa Timur	89.380.993	6
16	Banten	23.371.443	3	Banten	23.371.443	17	Banten	23.371.443	15	Banten	23.371.443	2
17	Bali	8.199.312	1	Bali	8.199.312	6	Bali	8.199.312	6	Bali	8.199.312	1
18	NTB	11.535.807	2	NTB	11.535.807	9	NTB	11.535.807	8	NTB	11.535.807	1
19	NTT	8.409.261	1	NTT	8.409.261	6	NTT	8.409.261	6	NTT	8.409.261	1
20	Kalbar	9.022.563	1	Kalbar	9.022.563	7	Kalbar	9.022.563	6	Kalbar	9.022.563	1
21	Kalteng	5.034.270	1	Kalteng	5.034.270	4	Kalteng	5.034.270	4	Kalteng	5.034.270	1
22	Kalsel	8.321.662	1	Kalsel	8.321.662	6	Kalsel	8.321.662	6	Kalsel	8.321.662	1
23	Kaltim	7.388.472	1	Kaltim	7.388.472	6	Kaltim	7.388.472	5	Kaltim	7.388.472	1
24	Sulut	5.615.078	1	Sulut	5.615.078	4	Sulut	5.615.078	4	Sulut	5.615.078	1
25	Sulteng	4.840.198	1	Sulteng	4.840.198	4	Sulteng	4.840.198	3	Sulteng	4.840.198	1
26	Sulsel	16.289.754	2	Sulsel	16.289.754	12	Sulsel	16.289.754	11	Sulsel	16.289.754	2
27	Sultra	4.538.549	1	Sultra	4.538.549	4	Sultra	4.538.549	3	Sultra	4.538.549	1
28	Gorontalo	2.402.671	1	Gorontalo	2.402.671	2	Gorontalo	2.402.671	2	Gorontalo	2.402.671	1
29	Sulbar	2.117.572	1	Sulbar	2.117.572	2	Sulbar	2.117.572	2	Sulbar	2.117.572	1
30	Maluku	2.277.948	1	Maluku	2.277.948	2	Maluku	2.277.948	2	Maluku	2.277.948	1
31	Maluku Utara	1.505.363	1	Maluku Utara	1.505.363	2	Maluku Utara	1.505.363	1	Maluku Utara	1.505.363	1
32	Papua Barat	1.761.426	1	Papua Barat	1.761.426	2	Papua Barat	1.761.426	2	Papua Barat	1.761.426	1
33	Papua	5.051.779	1	Papua	5.051.779	4	Papua	5.051.779	4	Papua	5.051.779	1
		535.381.008	**68**		**535.381.008**	**396**		**535.381.008**	**349**		**535.381.008**	**54**

Table 5. Estimated PCs for Recognition, Correction, Completion and Release

Scan

Scan process of total scanner was calculated by adopting a parameter that one unit of scanner recommended by manufacturer was able to process 100,000 pages per day (daily duty cycle manufactured recommend). About 100,000 pages per day could be achieved if the rate of scanner applied was 95%. It is important to know first the total pages per day to process (daily duty cycle) for every province.

Daily duty cycle *= est. total of pages / working days a month * max. processing time*
Est. scanner *= roundup(daily duty cycle / daily duty cycle (mfd recommend)*

"Total scanner" is equivalent with total PC needed for scanning process. Based on the calculation result, scanner productivity in each province may be disclosed using the following formula:

Productivity *= daily duty cycle / ∑ scanner / daily duty cycle (mfd recommend) * 100%*

Based on table 4, the estimated total scanner and PC scan showed the scanner productivity imbalance. That is, 6 provinces showed scanner productivity below 20 percent and 5 provinces showed productivity above 80 percent. So, a decision was made that we would implement a process whereby machines would move to some big provinces after the small provinces completed the scanning process for all their documents. The productivity in small provinces had to be increased by additional working hours and working days so as to immediately divert the scanners to big provinces.

Recognition

The recognition process was calculated using such parameter that one page required 0.9 seconds. So, the productivity and total PC for recognition needed were as follows:

Productivity *= Max. processing time * working days a month * working hours a day * 60 minutes * 60 seconds / recognition speed*
Est. PC recognition *= Roundup(daily duty cycle / productivity)*

Correction

The correction process was calculated using parameter of one hour for 6,500 characters. Thus, the productivity of characters per PC was:

Productivity1 *= max. processing time * working days a month * working hours a day * correction speed (char./hour)*

Next, upon the parameter of 125 characters per page (number of characters) and 90 percent accuracy, then the productivity of number of pages per PC and the number of PCs for Correction needed were:

Productivity2 *= productivity1 / (100 – accuracy)/100* number of characters*
Est. PC correction *= roundup(est. total pages / productivity2)*

Completion

The completion process was calculated using the parameter of completion speed of 15 seconds per page. Then, completion rates or estimated documents which should be passed completion process were 40 percent. The productivity and number of PCs for Completion needed were:

Productivity *= max. processing time * working days a month * working hours a day * 60 minutes * 60 seconds / completion speed*

Est. PC completion *= roundup(est. total of pages * completion rate / productivity)*

Release

The Release process was calculated using a parameter of release speed of 0.62 seconds per page. Then, the productivity and number of PCs for Release needed were:

Productivity *= max. processing time * working days a month * working hours a day * 60 minutes * 60 seconds / release speed*

Est. PC release *= roundup(est. total of pages / productivity)*

Based on the aforementioned calculations, the number of PCs needed to process using data entry when compared with that using OCR technology, on average, showed that OCR technology only required approximately 40 percent as much as PCs for data entry. The big provinces with huge numbers of documents significantly benefited from OCR technology as number of PCs needed was only around 30 percent, if data entry is applied. On the other hand, the small provinces with smaller numbers of documents only required approximately 70 percent. For all of Indonesia, we would have needed 2,770 units for data entry technology, while only 919 units were required by using data capture technology.

The number of PCs by using OCR technology, saved on cost of thePCs themselves, but also saved the cost of leasing processing room and operator. The number of operators for data capture technology needed was, on average, only 30 percent of what the data entry technology would have been. We needed only 3,188 operators for the data capture technology throughout Indonesia, rather than the 11,080 persons we would have needed when using the data entry.

Implementation of Population Census 2010 Data Processing With OCR Technology

The original plan was to have 55 scanners in the 33 provinces for the implementation of population census 2010 data processing. BPS conducted an open bidding for the OCR implementation. In order to obtain optimum results from OCR technology, standardized procedures had to comply, including type of pencil, paper, document printer, document arrangement, etc. Standard incompatibility would impede processing activities which would eventually require additional processing or make the process does not run at all. Six companies made bids, but only three included OCR software: ABBY, IBM Filenet, and Kofax. All three had excellent image capture. BPS selected Kofax.

The Population Census 2010 questionnaire was printed in the form of household booklets. Documents were grouped by census block. Staff started by using a paper cutter to prepare for capture by cutting each batch of documents. Batches were named, and recorded at the start of phase scan using the identity of the regions and provinces, districts, sub districts, villages and census blocks. Once completely scanned, the batch would automatically go through the stages of recognition. Stages of recognition performed document classification and image interpretation. Correction then followed. This stage showed the field when recognition had a confidence threshold level below 80 percent, but for the name and address, the confidence level was set at 100 percent due to the special needs of BPS. Operator had to correct the fields by looking at the image.

The next stage was Completion, which was used to ensure that empty fields are filled by looking at the image and checking the two types of answers – markings and handwritings for different contents. The last stage was the Release, with release storing the data to the RDMS and the image to a folder in the Server. The name of the batch became the folder name to facilitate searches, if necessary. Two stages controlled the batch. First was the Document Viewer which checked if any unclassified documents or document did not order properly. The second was Quality Control when a batch error was due to network interruption, etc.

Geographic constraints caused problems in getting the documents to four of provinces from the field to the processing centers. In these provinces, BPS processed the non-scanner documents by data entry. The scanners assigned to these provinces could be moved to other provinces. The 55 scanners were reduced to 52 at some point. Two major provinces reduced the number of scanners: Jawa Barat was reduced by 2 units and Jawa Timur was reduced by 1 unit. Later, the scanner needs to Jawa Barat and Jawa Timur were met by borrowing from other provinces. Small provinces were able to optimize the hours and days of work to promptly complete the scan, so the scanners then went to Jawa Barat and Jawa Timur

Table 6 shows processing time for each stage of data capture, using data obtained from Banten Province as example. The average time for one page to process was 5 seconds. As it turned out, a bottle neck occurred, as the speed of Scan and Recognition was slower than the Correction and Completion stages. This problem was solved by reducing working hours for operators during Correction and Completion, while running Scan and Recognition 24 hours nonstop.

Process	Time Proc. In seconds	Total Pages	Pages/Second	Seconds/Page
Scan	4.832.542	20.405.806	4,22	0,24
Recognition	36.422.040	20.405.806	0,56	1,78
Correction	52.989.460	20.405.806	0,39	2,60
Completion	14.181.101	20.405.806	1,44	0,69
Release	2.050.310	20.405.806	9,95	0,10

Table 6. Results Data Processing for Each Stage from Banten Province

Conclusions

Because Indonesia is so large, keying the 2010 Census data would have been logistically difficult, would have been more costly that scanning, and would have taken much longer. This paper has discussed the methods used to obtain and implement the scanning for the 2010 Census, based on previous work for the 2000 Census and improvements in the hardware and software technologies as well as cooperation with the BPS offices. By almost any measure, the scanning of the 2010 Indonesia Population Census must be considered to be successful.

Table Titles

REFERENCES

Suharto, Sam, and Abdulmadjid, M. 1973. *Progress Report on 1971 Population Cansus of Indonesia.* http://www.disc.wisc.edu/INDO/indo_report.html

Suwito, Sugito. 1998. *Exploiting New Information Technology in Indonesia Census Operations.* http://www.ancsdaap.org/cencon98/papers/indones/indones.pdf.

UNESCAP. 2001. *Guidelines on the Application of New Technology to Population Data Collection and Capture*. http://www.unescap.org/stat/pop-it/pop-guide/index.asp#guide-capture.

ABSTRACT

Indonesia is the fourth most populous country in the world and started using computers in the 1960 census. Statistics Indonesia (BPS) first employed scanning for the 2000 Census data using scanning equipment provided by Japan International Cooperation Agency (JICA) with BPS developing the software to capture and process the census data. The capture period was 12 months, the editing took 8 months, and the results were produced in April, 2002, 23 months after the enumeration.

For the 2010 Census, BPS made efforts to improve the data capture operation to get data to the public and private sector users more quickly, so requested bids from several companies. BPS selected Kofax, Inc. to provide data capture processing. Data capture occurred in 33 locations in each of Indonesia's provinces over a period of 11 months after enumeration.

Clearly, in a population the size of Indonesia, keying is not a real possibility if we want to see results in a relatively short period. Also, scanning provides more efficient use of both human resources and data processing facilities. Scanning reduces the numbers of needed the data entry operators and the human errors resulting from fatigue.

We developed a program to transform the captured information into ASCII for CSPro processing. We applied adapted content editing programs from our pilot census to the scanned data. Hence, we were able to obtain a completely edited file for the whole country by November, 2010, and tables before the end of the calendar year. The whole process took 11 months from beginning of enumeration to printer-ready copy. This paper discusses various aspects of our capture procedures, including check in from the field, slicing the forms for scanning, reading the forms, manually checking of the forms, and transfer of the final captured data to the central office for continued processing.