ORIGINAL: ENGLISH

SECRETARIAT OF THE PACIFIC COMMUNITY
STATISTICS FOR DEVELOPMENT DIVISION (SDD)

2020 WORLD ROUND OF POPULATION AND HOUSING CENSUSES –
PACIFIC ISLAND COUNTRIES' CENSUS PLANNING MEETING:
INTERNATIONAL RECOMMENDATIONS/STANDARDS, CONTEMPORARY TECHNOLOGIES AND REGIONAL COOPERATION

Noumea, New Caledonia, 27 July – 31 July 2015

PART 3b: SCANNING
(Paper presented by Phil Bright, GIS Specialist, Secretariat of the Pacific Community)

1. BACKGROUND

There is a comprehensive set of recommendations on how best to prepare for and utilise scanning in *UNFPA-SPC/Census Review 2012/Working paper 5* which will be referred to in this paper, though not duplicated.

Scanning was used in several PICTs in the 2010 census round with varying success. Being the first time it was undertaken in the region, there were teething issues and hence lessons we can learn from for the 2020 census round.

2. THE KEYS TO USING SCANNING IN A CENSUS

Pre-scan **planning** and **testing** is particularly important along with a **well-managed scanning production flow**.

An incorrectly designed questionnaire without correct dropout colours can be so hard to interpret that the manual corrections required result in more work than if manual entry was used in the first place.

During processing, forms need to be checked and correctly ordered within their households (if there are multiple forms per household) and within EAs. The scan process then needs to be well supervised so there is no double scanning of forms and errors such as forms getting stuck together are quickly resolved. Assuming the interpretation does a fairly good job and the verification controls are well structured, the small amount of cleaning should be fairly streamlined. Changes which are made in the interpretation checks are quite detrimental and require the whole dataset to be reinterpreted and verified, whereas small changes to verification rules can be changed during processing.

3. THE MAIN CATCH WITH VERIFICATION

One of the patterns discovered during verification in the 2010 census round is that verifiers think they can simply push buttons without really thinking. It is easy to get into a routine when there are no or minimal errors and then verifiers easily miss corrections and start making mistakes. A verifier who does twice as much work as anyone else is generally a warning sign.

4. EXCELLENT ARCHIVAL AND RETRIVAL

Unique file names are created for scanned questionnaires which are also referenced in the corresponding datafile. If GPS is also used and the GPS codes correctly recorded then it is possible to create a hot-linkable map layer which allows households to be interrogated within a GIS to perform quality checks and identify areas which were potentially missed.

5. RECOMMENDED SCAN-OPERATOR AND VERIFIER TRAINING:

As with all other aspects of a census, training is crucial, though corners are often cut. A little bit too much training is far better than not enough.

1. Produce a sample set of questionnaires which clearly have "TRAINING ONLY" written on them so they can't be confused at a later date with the real questionnaires.
2. Carry out the same detailed explanations for each question as with manual entry. Verifiers need to fill out many forms to properly understand the flow and structure.
3. All verifiers need to have a go of scanning and all scan operators a go at verifying so they better understand the whole process.
4. Run many tests of sample datasets through the scanner, deleting the sets and restarting over.
5. Verify the test datasets many times – they can be reused without having to rescan every time.
6. Once the real scanning starts, process a few EA's then stop and get verifiers to check each others' work in Excel. If there are problems work through them in detail with the verifiers, deleting whole EAs if necessary to redo them.

6. CONCLUSION

If scanning is correctly prepared for, and well managed, it can significantly speed up data processing and improve data quality by automating part of the process and avoiding manual data entry errors. There is also the added advantage of being able to effectively reference the digital questionnaires for archival and retrieval. There are many technical and infrastructure requirements though which are very important to properly assess before using this method of data capture.

However, looking back over the last 10-15 years, the logical next step to improve enumeration speed and quality would have been to simply find a way to incorporate all the controls of a data entry system (like CSPRO), in the field. The technology gap was too significant at the time, and costs too high to allow enumerators to walk around with laptops. With great advances in smartphones and tablets, coupled with reductions in cost, and improvements in mobile networks, the scanning step can realistically be skipped for a full digital enumeration.