

2008年人口センサスの調査票入力にOCRを導入しない理由(メモ)

1. 会議の出席者

Ms. Hang Lina, Deputy DG  
Mr. Yem Sourng, Director for Social Statistics Dept.  
Mr. Sin Serey Vuth, Chief for Statistical Information  
Mr. Meng Kimhor, Officer for Data Processing  
Mr. F. Nishi, JICA Study Team Member

2. 会議の日時

2004年10月19日

3. OCR(文字読み取り装置)を導入しない理由について

Ms. Lina から 2008 年人口センサスのデータ入力は、OCR を利用したいとの要望があった。これに対して、個人的な意見と前置きした上で、調査団員から以下のとおり回答した。

データ入力の方法には、PC を利用した方法とスキャナーを利用した方法(OCR/OMR システム)の2通りがあるが、2008 年人口センサスにはPC を利用した方法を薦めたい。その理由について、インドネシアの例と比較しながら以下のとおり説明した。

(1) 人口の違い

インドネシアの人口は2億人を超えており、PC を利用した方法を採用すると、仮に1年以内の結果公表を目指すとするれば、約2,000台のPCが必要となる。このような大量の台数になると、設置場所やデータ入力要員の確保が困難である。したがって、スキャナーを利用した方法に移行する必要がある。

一方、カンボジアの人口は、約1,140万人であり、必要とされているPCの台数が60台~90台と見込まれていることから、設置場所やデータ入力要員の確保は比較的容易である。

(2) センサス実施計画の違い

インドネシアでは、西暦年末尾が0の年に人口センサスを実施する、3の年に農業センサス、6の年に経済センサスとすることが規則に定められているので、OCRを継続的に有効利用する見込みがある。

一方、カンボジアでは、10年に1度、人口センサスを実施することが統計法(案)に定められているのみで、人口センサス後に、その他のセンサスを実施する予定はない。したがって、OCRを継続的に有効利用する見込みが立っていない。

(3) 他の業務への転用可能性の違い

PCを利用した方法の場合には、人口センサスの後は、PCをNIS内に分配して、経常業務に利用するなどの有効利用の方法がある。一方、スキャナーの場合には、経常業務に利用する機会は、PCと比較すると、かなり少なくなる。

( 4 ) 調査票にかかるコストの違い

PCを利用した方法の場合には、通常の調査票で差し支えないが、一方、スキャナーを利用した方法の場合には、調査票を厚くしてより丈夫にする必要があり、また、読み取る範囲を示すマークの印刷、読み取るべき文字と質問文を区別するためにカラー印刷をする必要があるなど通常の調査票よりも、調査票にかかるコストがかなり高くなる。

( 5 ) 読み取り精度の問題

PCを利用した方法の場合には、人間の目で見えて入力するので、スキャナーを利用した方法の場合よりも、データの精度は高くなる。

( 6 ) メンテナンスの問題

カンボジアにはスキャナーやOCR ( 又はICR ) ソフトウェアのメンテナンスを依頼できる代理店が未だない。このため、トラブル発生時の対応に時間がかかることが予想される。現状では、トラブル発生時には、バンコクからメンテナンスの業者を招聘することになる。

以上の回答について、1998年人口センサスのProject ManagerであったMr. Yem Sourngから賛同のコメントがあったが、Ms. Hang Linaは、NIS局長のH.E. San Sy Thanがスキャナーを利用した方法を希望しているとコメント。

その直後、NIS局長のH.E. San Sy Thanに対して、全体会議の前のインフォーマルな会話の中で、上記のことを説明した。この会話には、UNFPAのコンサルタントMr. Raoも同席しており、彼からも、南アメリカでのスキャナーを利用した方法の問題例を挙げた上で、PCを利用した方法に賛同するコメントがあった。