

Key From Image Technical Experiences and Insights

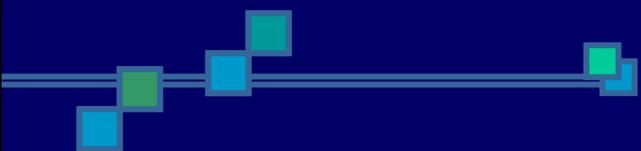
Philippine NSO Implementation



Core Statement

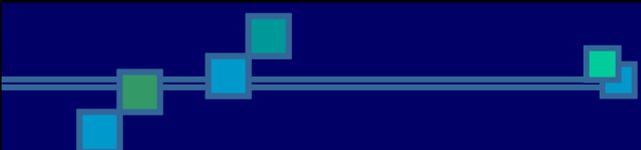
- In pursuit of the most faithful census data capturing methodology given the constraints on time and money, we dedicated ourselves to harnessing innovative technological strategies such as using document imaging systems.





Goals and Objectives

- Underline the pros and cons of using document imaging systems
- Showcase our hybrid data processing system
- Share insights and gain appreciation



Advantages in using Document Imaging Systems

- Much faster data capture
- More accurate (as claimed)
- Requires fewer resources
- Electronic copies of forms can be kept and are more conveniently archived



Disadvantages

- Expensive scanner hardware and software, expensive image processing software
- ICR accuracy is very low
- Difficulty in implementing data cleaning

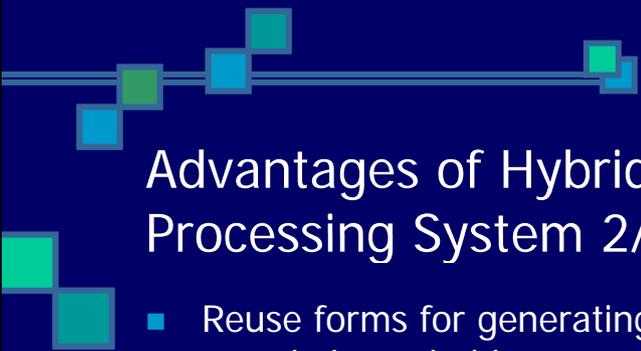
Hybrid Data Processing System

1. Scan the forms to produce image files
2. Extract OMR data from image files
3. Key-in, Key-verify and Clean handwritten data using Key-from-Image



Advantages of Hybrid Data Processing System 1/2

- Faster data entry of handwritten entries
- Lesser data items for encoding with OMR
- Verification of keyed-in and extracted data items
- Accuracy ranged from 99%-100% for sample key-verification in 2010 Census of Population and Housing



Advantages of Hybrid Data Processing System 2/2

- Reuse forms for generating list of sample households
- Corrections were reflected both in the data file and the scanned image of the forms



Key-from-Image Features 1/2

- CS Pro compatible for Data Dictionary and output data
- Drag and drop template preparation
- Imaging with/without interpreted data items
- Annotation of corrections in the image files

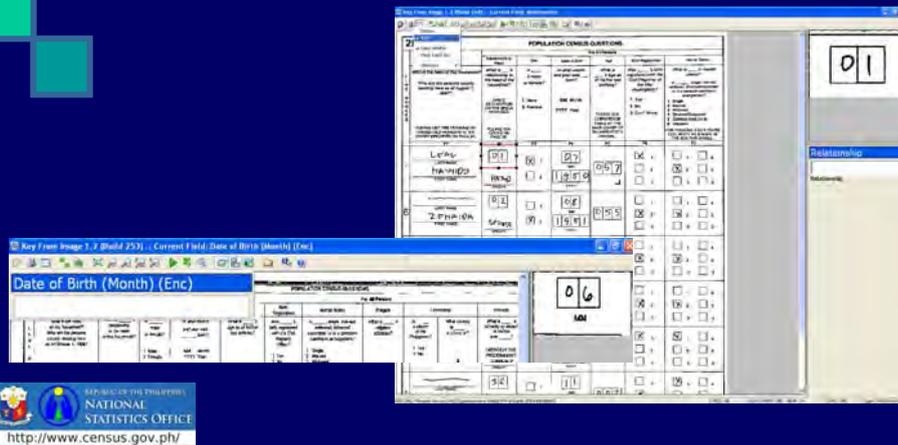


Key-from-Image Features 2/2

- Conventional data entry software
- Built-in electronic code book
- Supports custom skipping patterns
- Internationally aware (UTF support)



KFI Screenshots



Our Findings

- Accuracy of automatically extracted hand-written data from survey/census forms were poor
- Accuracy of extracted data from OMR fields ranges from 99% to 100%