



# Scanning

➤ Creates an Electronic bit map image of forms

# Scanned Image

RURAL (R) 5th ECONOMIC CENSUS - HOUSE LIST

Schedule No. 511 Page No. 01

01 DELHI 02 NORTH WEST 03 NARAZA 04 BAKHTAWAR PUR 05 0006

Line No.	House No.	House Category Code	Entrepreneurial Unit No.	Household No.	Name of the enterprise/head of the household	Whether any member of the household is engaged in entrepreneurial activity within the premises or outside without fixed capital (Part A, Para 2)		No. of employees (including the household unit)	No. of enterprises (excluding the household unit) having any structure		Total no. of enterprises (Part B, Para 2)	Type of the enterprise acquired by the household
						Without fixed location	With fixed location		Without fixed location	With fixed location		
001	1051	1		01	Davender		2				1	
002	1052	3	01		Doctor Prithvi clinic	01				01	1	
003	1052	1		01	Harvahu		2				1	
004	1054	1		01	Bah		2				1	
005	1055	1		01	Kinokhu		2				1	
006	1056	1		01	Safarsh Akhda		2				1	
007	1057	1		01	balbir		2				1	
008	1058	1		01	Om Prakash		2				1	
009	1059	1		01	Kiran Singh		2				1	
010	1060	1		01	Mukesh		2				1	
011	1061	1		01	Shyam Lal		2				1	
012	1062	1		01	Suresh		2				1	
013	1063	1		01	Gayatri chand		2				1	
014	1064	1		01	Mani Singh		2				1	
015	1065	1		01	Prem Singh		2				1	

0123456789

Signature of entrepreneur: Del Checked and signed correct: Del

Name of entrepreneur: Shamuel Kumar V Date: 6-10-05 Name of supervisor: Kuldeep Chand Mishra Date: 10-10-05

090605 100605

- Form-ID
- Manual Form-ID
- Recognition

## Content Recognition Modules:

### • Form-ID Module

- ✓ Form Identification
- ✓ Automatically identifies a form using Empty Form Images (EFI) which are defined during the design stage

### • Manual Form-ID Module

- ✓ When the Form-ID Module cannot identify the form due to poor scanning or other problems the user can manually match a scanned image with the proper Empty Form Image (EFI)

### • Processing/Recognition Module

- ✓ Character recognition using OCR /ICR/ OMR Recognition engines

## FORM-ID STATION

Integra - FormId [ECONOMICCENSUS] - 0207011000005

File Edit Search View Zoom Options Help

Run Pause Abort 2% Cur% 500% Zoom in Zoom out

Prepared EFI list cache  
Only page Id selected

	No. of enterprises engaged in entrepreneurial activity within the premises or outside without fixed structure (Yes:1, No:2)	No. of enterprises within the household/ unit	No. of enterprises outside the household/unit and NOT having any structure		Total no. of enterprises (Col. 8+9+10)	Type of the structure occupied by the household/unit
			Without fixed location	With fixed location		
Page 1 / 8 EFI: Man_Eng_51 (99%)	2					1
Page 2 / 8 EFI: Man_Eng_51 (99%)	2					1
Page 3 / 8 EFI: Man_Eng_53 (98%)	2					1
Page 4 / 8 EFI: Man_Eng_53 (99%)	2					1
Page 5 / 8 EFI: Man_Eng_55 (98%)	2					1
Page 6 / 8 EFI: Man_Eng_55 (99%)	2					1
Page 7 / 8 EFI: Man_Eng_57 (97%)	2					1
Page 8 / 8 EFI: Man_Eng_57 (99%)	2					1

- Form-ID
- Manual Form-ID
- Recognition

## Processing / Recognition

- Most important station
- Automatically Recognises data from image
- Automatic recognition is not 100% accurate
- Inaccurate data requires human intervention
- Positive effect on throughput as long as characters incorrectly identified are controlled

OCR/ICR/OMR engines  
used to recognize characters.



CLEQS

PARASCRIPT

NESTOR

JUSTICR

CGK

Clear Image

TISICR

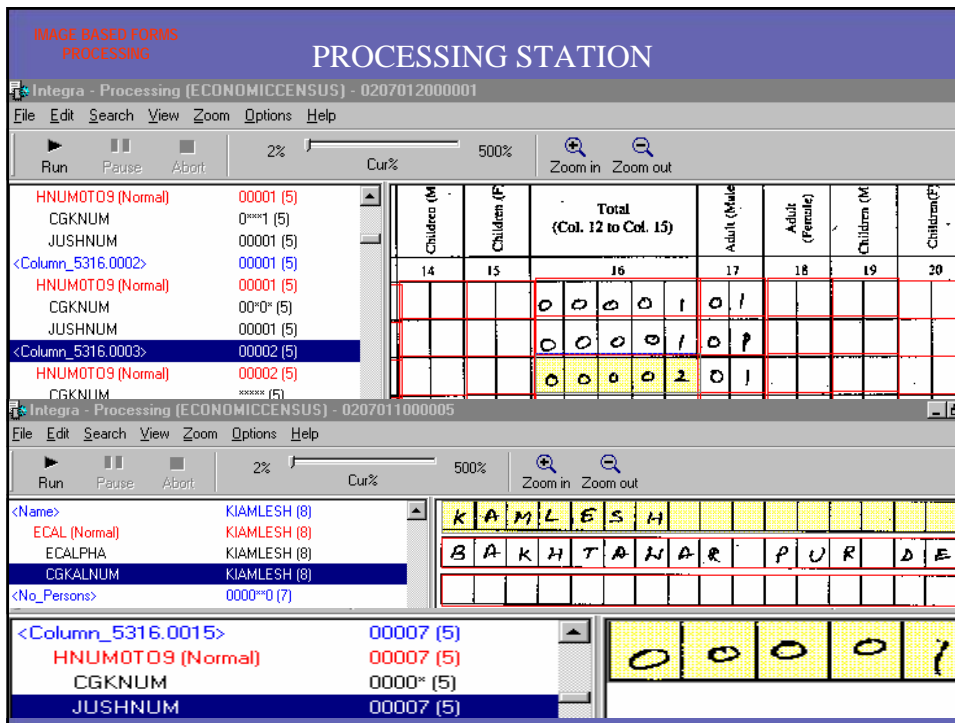
KADMOS

LIGATURE

EXPERVISION

RecoStar

AEG



### Key from Image

Certain types of data do not lend themselves to automated extraction via a recognition engine of one kind or another.

Cursive handwriting in a form, is an example.

Such data can be extracted efficiently from a form using key from image techniques.

Software displays zoomed predefined areas of an image--or zones--to a human operator for keying.

As soon as the contents of the area is keyed, the software automatically displays the next zoomed area, and so on.

## Content Validation

- Content Validation includes 3 modules:
  - Tile
  - Completion
  - Exception

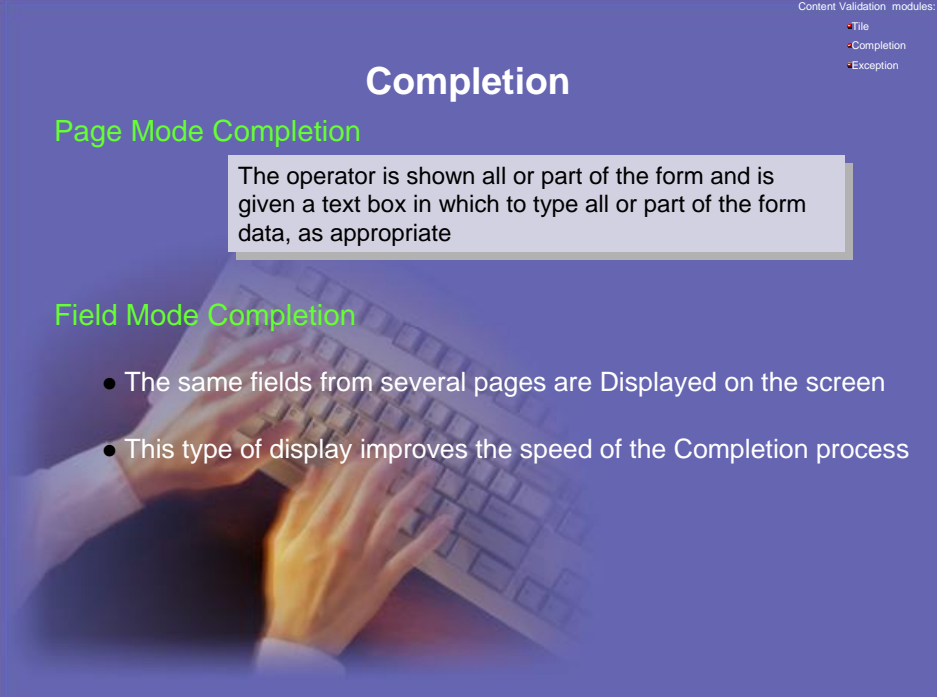
## Tile

- The unique TILE module optimize data accuracy with a systemized display of characters grouped together to allow easy identification
- Possible to identify which characters are correct and which are not and allows to mark as reject.
- Makes the completion more accurate



## Completion

- The Completion module
  - ✓ Receives data from the Tile Station
  - ✓ Enables the operator to correct unrecognized characters
- The data is completed manually with the help of
  - ✓ Data dictionaries
  - ✓ Rules
  - ✓ Validation functions etc.
- Modes of display - Page and Field mode



Content Validation modules:

- Tile
- Completion
- Exception

## Completion

### Page Mode Completion

The operator is shown all or part of the form and is given a text box in which to type all or part of the form data, as appropriate

### Field Mode Completion

- The same fields from several pages are Displayed on the screen
- This type of display improves the speed of the Completion process

IMAGE BASED FORMS PROCESSING

## COMPLETION STATION [Field mode ]

Integra - EC51\_52 (EconomicCensus) - 030701100001#5

Integra - EC53\_54

Integra - EC55\_56 (EconomicCensus) - 030701200001#5

File Tools View Options Help

Get Save Discard Exceptions Zoom Skip Pin Image

**HouseList**

1	D16
2	017
3	*0*
4	019
5	020
6	021
7	022
8	023
9	024
10	02*
11	026
12	027
13	028
14	029
15	030

**Enterprise Schedule**

C5301
1 167
2 169
3 171
4 171
5 172
6 173
7 174
8 175
9 176
10 177
11 178
12 180
13 181
14 182
15 183

**Enterprise Abstract**

No. of persons usually working in Enterprise										No. of one-time (part-time, casual) working in Enterprise			No. of change having 15 more workers than last year	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	5	0	0	0	4	7	0	0	2	7	0	0	7	4
1	4	0	0	0	2	4	0	0	0	2	4	1	4	0
1	4	0	0	0	2	2	0	0	0	2	9	1	4	0
0	7	0	0	0	1	0	0	0	0	1	0	0	7	0

Integra - EC55\_56 (EconomicCensus) - 020701200001#5

File Tools View Options Help

Get Save Discard Exceptions Zoom Skip Pin Image

**Column\_55**

1	014
2	---
3	012
4	
5	
6	
7	

Extracted data

IMAGE BASED FORMS PROCESSING

## COMPLETION STATION [Page mode ]

Integra - EC57 (EconomicCensus) - 0307011000004#7

File Tools View Options Help

Get Save Discard Exceptions Zoom Skip Pin Image

Code: 0000013110

Investigation unit (17) No.:

CVS (Basic) table IV set:

IV: Recording unit no. of enterprises within a page: 08

V: Total Number of persons working: 0000040

VI: Year of start of operation: 1982

NIC: NIC Code: (to be filled up in separate form)

MOOL: BAKHITANARPUR DELHI - 26

ES\_No: 53 ES\_PageNo: 01 SNo: 08 No\_Persons: 000040 Non\_Hired: 000 Year\_Start: 1982 NIC\_Code:

Name: KAMLIEISH

Address\_1: BAKHITANARPURDELHI

Address\_2:

II

PinCode: 1100\*\*

STDCode: TelephoneNo: STDCode: FaxNo:

Email: J

Address Slip

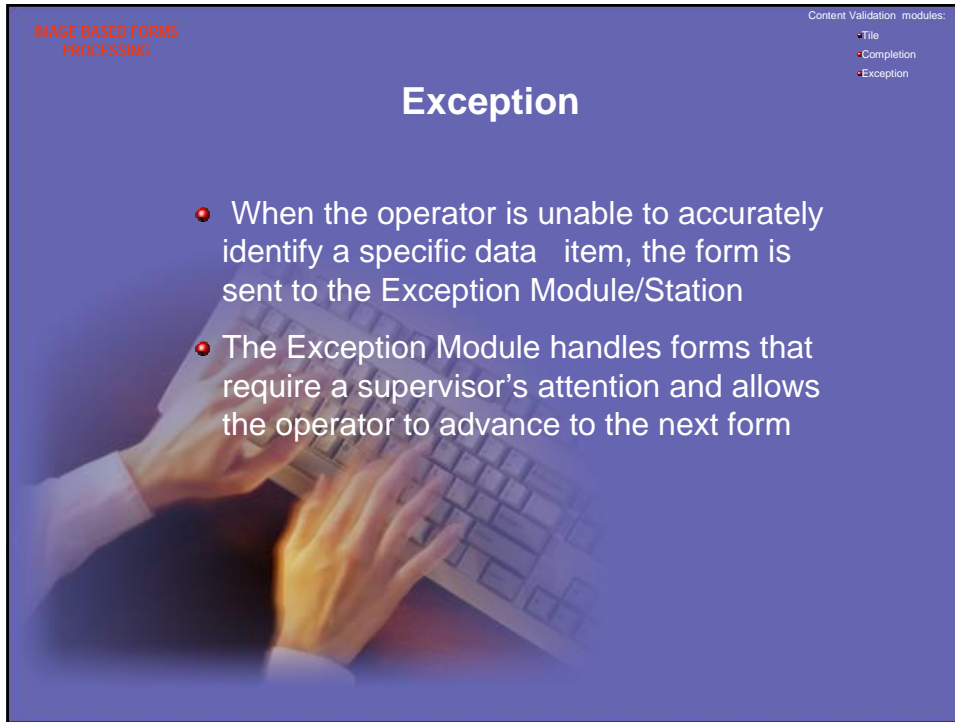
Form: FP0007 Page: 1 / 1 REJECT Insert AM 10:42:24



- Title
- Completion
- Exception

## Exception

- When the operator is unable to accurately identify a specific data item, the form is sent to the Exception Module/Station
- The Exception Module handles forms that require a supervisor's attention and allows the operator to advance to the next form



## EXCEPTION STATION

Form Image

Exception cases

Order	Color	Batch	Form #	Field	Exception	Description
X 1	1	03070110...	1	Column_5316\$0...	Exception	Exception
X 2						

## CONTENT DELIVERY

- EXPORT the final data in ASCII or Database file format.

