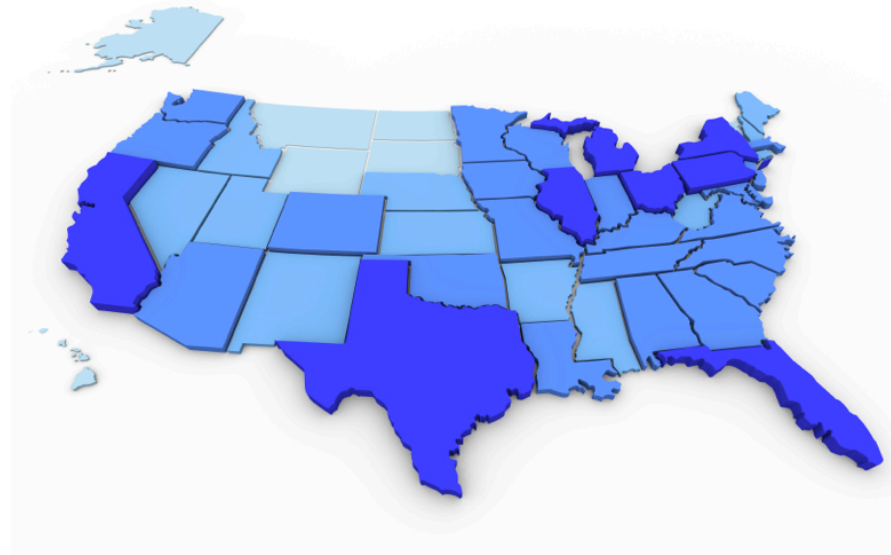
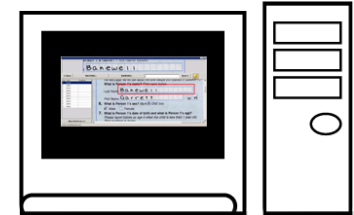


Testing Production Data Capture Quality



A Presentation to the FCSM
Wednesday, 11 Jan 2012

The Production Data Quality (PDQ) Tool

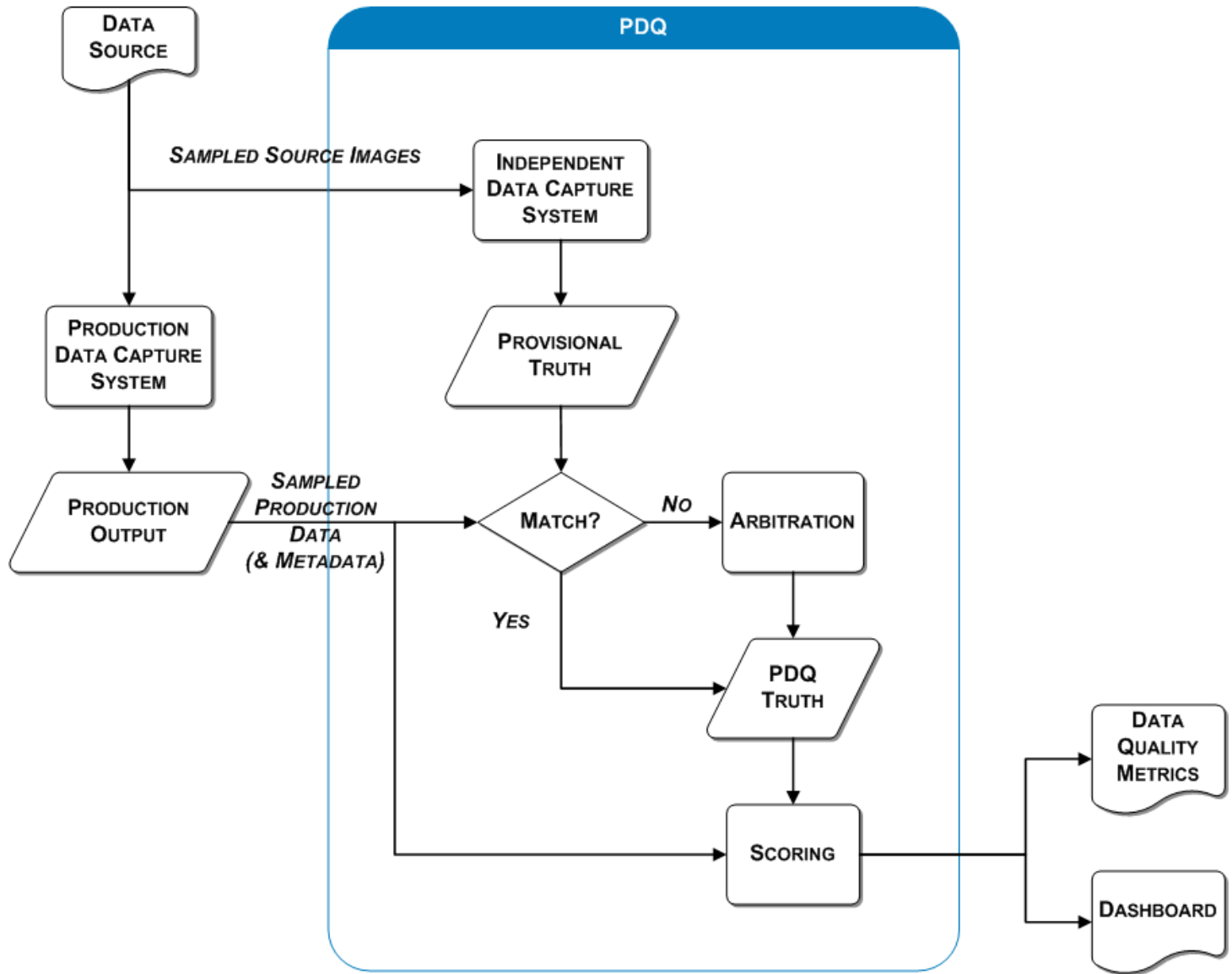


PDQ:

- Is a stand-alone data capture tool, used in the Census 2010 Decennial Response Integration System (DRIS) to measure data capture quality
- Uses different OMR & OCR software from the Census DRIS data capture system
- Independently evaluated over 800,000 sampled DRIS form images and production data in near-real time
- Reported on the mark recognition, character recognition, and keying data accuracy of the Census 2010 DRIS data capture system

PDQ Diagram

PDQ Workflow



PDQ Was Important to Census

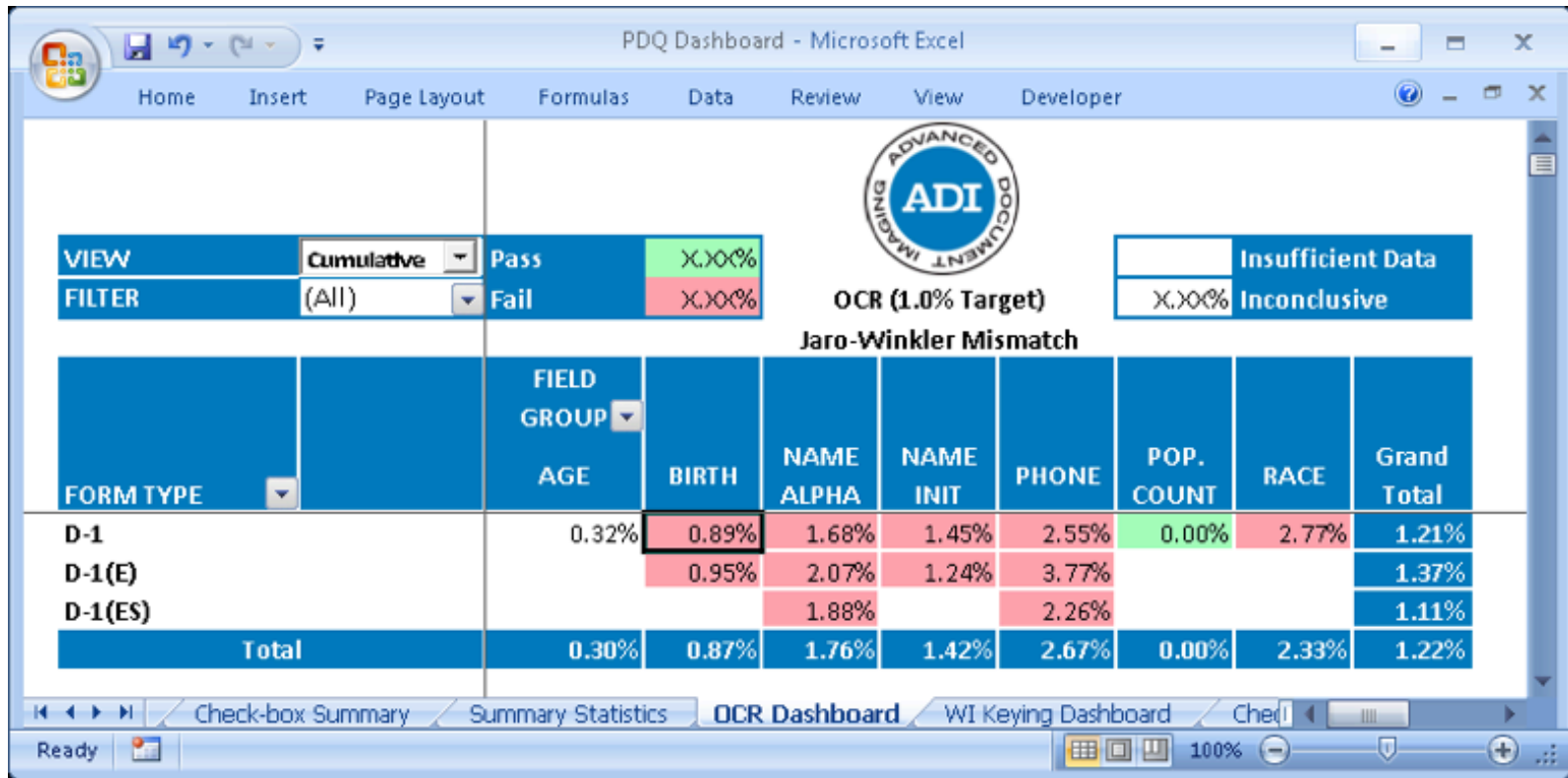
1. PDQ provided an independent measure of DRIS data quality in a timely manner
2. PDQ provided the capability to detect systemic anomalies in the DRIS production system, such as an accuracy shortfall due to a processing or template definition problem
3. PDQ provided precise evidence to Census that the Prime contractor (LM) met their Service Level Agreements (SLAs)



PDQ SLA Results

Requirement Name	Field Type	Capture Methods	DRIS Accuracy Rate Requirements	Weighted Accuracy Rate	Conclusion on DRIS Performance
OCR Accuracy	Write-in	OCR	≥99.0%	99.56%	Exceeded Min. Requirement
Write-in Keying Accuracy	Write-in	KFI, KFFI	≥97.0%	98.61%	Exceeded Min. Requirement
Check-box Accuracy	Check-box	OMR, KFI, KFFI	≥99.8%	99.98%	Exceeded Min. Requirement

PDQ Dashboard - Level 1



PDQ Dashboard – Level 2

PDQ Dashboard - Microsoft Excel

	B	C	D	E	G	H	J	K	L	M
1	FORM_TYPE	FIELD_GROUP	FIELD_NAME	TYPE	ERRORS	TOTAL	ERR_RATE	STD_ERR	CAPTURE	
2	D-1	BIRTH	EN::P01::DOB_DAY	OCR_NUMERIC	16	1465	1.09%	0.26%	OCR	
3	D-1	BIRTH	EN::P01::DOB_MONTH	OCR_NUMERIC	16	1477	1.08%	0.26%	OCR	
4	D-1	BIRTH	EN::P01::DOB_YEAR	OCR_NUMERIC	5	1364	0.37%	0.27%	OCR	
5	D-1	BIRTH	EN::P02::DOB_DAY	OCR_NUMERIC	11	931	1.18%	0.33%	OCR	
6	D-1	BIRTH	EN::P02::DOB_MONTH	OCR_NUMERIC	6	937	0.64%	0.33%	OCR	
7	D-1	BIRTH	EN::P02::DOB_YEAR	OCR_NUMERIC	4	893	0.45%	0.33%	OCR	
8	D-1	BIRTH	EN::P03::DOB_DAY	OCR_NUMERIC	9	520	1.73%	0.44%	OCR	
9	D-1	BIRTH	EN::P03::DOB_MONTH	OCR_NUMERIC	8	537	1.49%	0.43%	OCR	
10	D-1	BIRTH	EN::P03::DOB_YEAR	OCR_NUMERIC	2	510	0.39%	0.44%	OCR	
11	D-1	BIRTH	EN::P04::DOB_DAY	OCR_NUMERIC	3	256		0.62%	OCR	
12	D-1	BIRTH	EN::P04::DOB_MONTH	OCR_NUMERIC	1	267		0.61%	OCR	
13	D-1	BIRTH	EN::P04::DOB_YEAR	OCR_NUMERIC	3	263		0.61%	OCR	
14	D-1	BIRTH	EN::P05::DOB_DAY	OCR_NUMERIC	3	221		0.67%	OCR	
15	D-1	BIRTH	EN::P05::DOB_MONTH	OCR_NUMERIC	3	225		0.66%	OCR	

PDQ Dashboard – Level 3

Table Tools PDQ Dashboard - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Developer Design

Field History THIS SCREEN CONTAINS VERIFIABLY FICTITIOUS INFORMATION Image Viewer

FORM_TYPE	CTR	SCANNER	DRIS_BATCH_ID	FORM_ID	DRIS_TIME	FIELD_NAME	DRIS_VALUE	PDQ_TRUTH	CAPTURE	QAS
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003950191_001	10910020329139111230	11-Jul-09 5:54:58 PM	EN::P03::DOB_DAY	3	9	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003950191_001	10910020347133111269	11-Jul-09 5:55:01 PM	EN::P03::DOB_DAY	1	7	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003980175_002	10930040364143111133	11-Jul-09 5:47:09 PM	EN::P03::DOB_DAY	1	6	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003990105_001	10950059958115111330	11-Jul-09 5:43:23 PM	EN::P03::DOB_DAY	10	18	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003930137_002	10950060112120111376	11-Jul-09 6:03:21 PM	EN::P03::DOB_DAY	5	3	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003940164_001	10950060470130111341	11-Jul-09 5:56:17 PM	EN::P03::DOB_DAY	5	3	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003910180_002	11910020472183111288	11-Jul-09 5:57:19 PM	EN::P03::DOB_DAY	8	3	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003920110_002	11930040676126111135	11-Jul-09 6:07:48 PM	EN::P03::DOB_DAY	10	18	OCR	0
D-1	MST	MSTPZ-SDB01	PPDQMSTG00003950191_003	11950060567141111348	11-Jul-09 5:55:19 PM	EN::P03::DOB_DAY	17	13	OCR	0

WI Keying Dashboard Check-box Dashboard Details (by field name) Details (by form id) Field History Form History Manual Entry Oracle Query Settings Scratchpad 1 Scratchpad 2

PDQ Dashboard - Level 4

FORM ID	DRIS	E	PDQ
109100 2032913911...	3		9

A Major Finding

Overstatement of Multiple Mark Check-box Responses

What is Person 1's race? Mark one or more boxes.

White
 Black, African Am., or Negro
 American Indian or Alaska Native — Print name of enrolled or principal tribe. ↴

Asian Indian Japanese Native Hawaiian
 Chinese Korean Guamanian or Chamorro
 Filipino Vietnamese Samoan
 Other Asian — Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on. ↴ Other Pacific Islander — Print race, for example, Fijian, Tongan, and so on. ↴

Some other race — Print race. ↴



Alternative Response Patterns

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? Mark all that apply.

Children, such as newborn babies or foster children
 Relatives, such as adult children, cousins, or in-laws
 Nonrelatives, such as roommates or live-in baby sitters
 People staying here temporarily
 No additional people

3. Is this house, apartment, or mobile home — Mark ONE box.

Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
 Owned by you or someone in this household free and clear (without a mortgage or loan)?
 Rented?
 Occupied without payment of rent?

4. ~~What is your telephone number? We may call if we don't understand an answer.~~

Area Code + Number

- -

Conclusions

- PDQ was used very successfully for 2010 Census data capture to measure Production Data Quality
- It could be applied to any production forms processing system to do a data quality audit or to guide continuous improvement
- This technology is extensible to other types of classifier systems, such as Record Linkage

Any Questions?

K. Bradley Paxton
ADI, LLC
200 Canal View Boulevard
Rochester, NY 14623
(585) 239-6057
brad.paxton@adillc.net
www.adillc.net

General Reference: Paxton, *Handprint Data Capture in Forms Processing – A Systems Approach*, 2011 (Amazon.com)