Optimizing Forms Processing

TAWPI

June 14, 2006

K. Bradley Paxton, Ph.D.

CEO, Advanced Document Imaging, LLC

brad.paxton@adillc.net

www.adillc.net

Exact Data accurate. fast. efficient.

ADI ADVANCED IMAGING DOCUMENT

Today's Main Topics

- Data Capture of Handprint Write-in Fields
- Reducing Data Capture Costs: $0.10 to $0.40 per form
- Improving Data Capture Quality
- Reducing Costs and Improving Quality at the same time!

Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

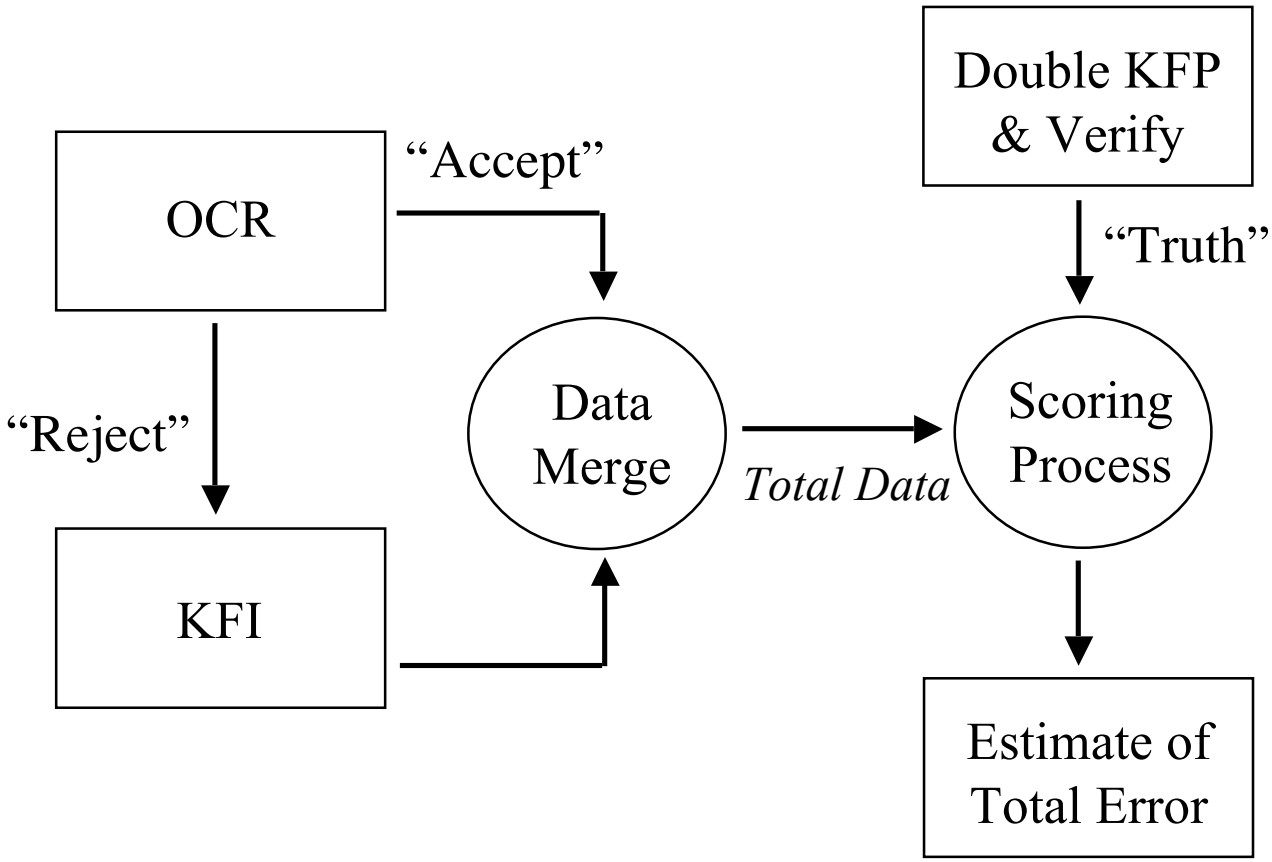
Estimating Data Capture Errors and Costs

- Error Sources
  - Key From Paper (KFP)
  - Key From Image (KFI)
  - Optical/Intelligent Character Recognition (OCR/ICR)
- Traditional Approaches
- Digital Test Decks™
- Data Capture Cost Model
  - What are potential “improvements” worth?

Traditional Testing Approaches

- Use only images (doesn't test front-end scanning or "real" throughput)
- Use hand-printed decks (never know "truth" for sure, only one "golden" deck)
- Use very small data sets (too much sampling error to measure accuracy very well)
- Rely on vendor's claims (don't test at all)

Traditional Error Estimation



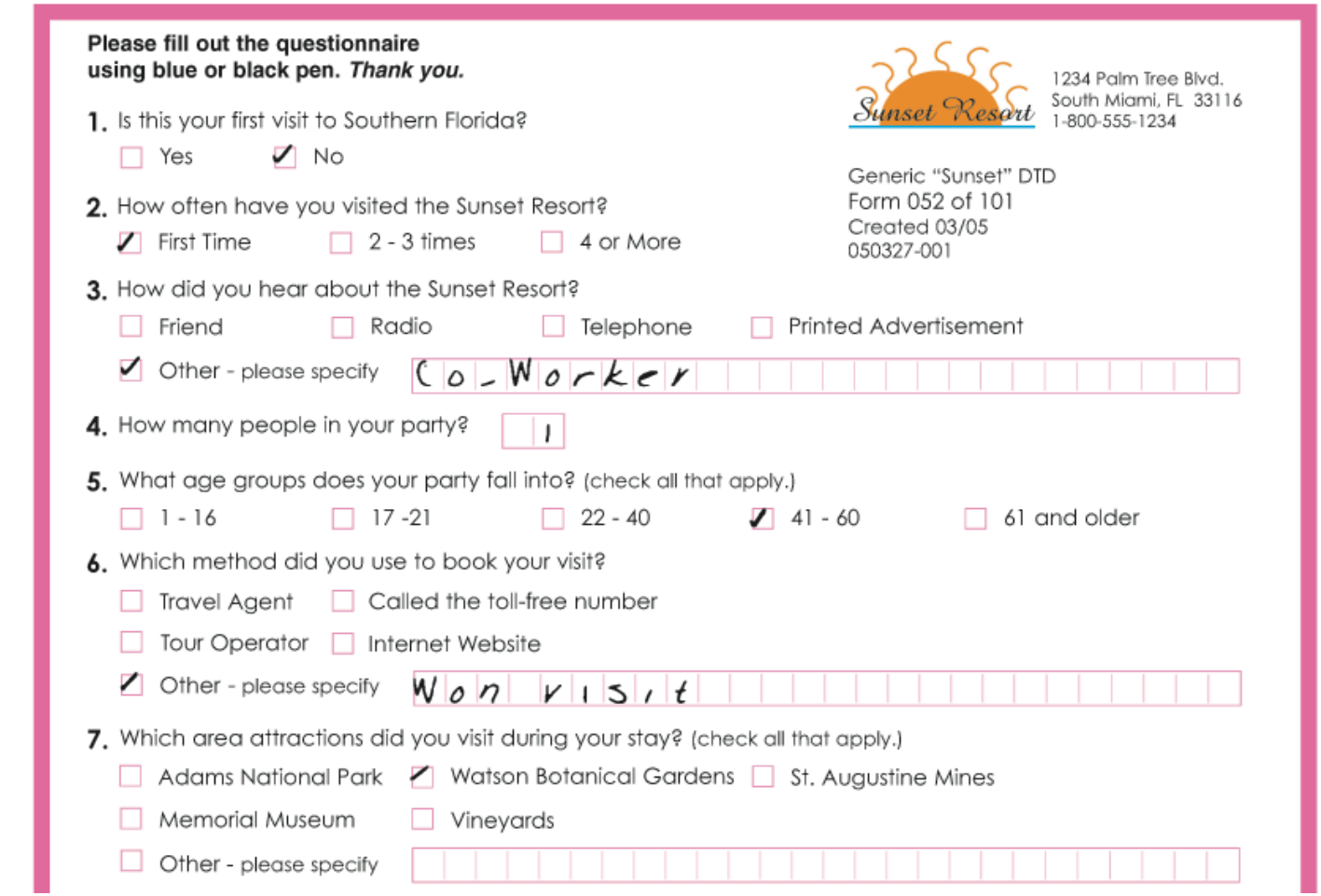
Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

Digital Test Deck™

- What a DTD™ is...
  - Hard Copy set of "Filled in " forms
    - ✓ That look real (to both humans and scanners)
  - Reproducible in Quantities
    - ✓ As required, for tests in parallel or over time
  - Perfectly known TRUTH
    - ✓ Usually “truth” files have errors
  - Generic or Custom DTD™
    - ✓ To meet particular testing and evaluation requirements
- Enables one to “benchmark” a data capture system’s performance, and really know “where you are”
- Gives a true “end-to-end” test of a data capture system

Generic Test Deck "A"



Please fill out the questionnaire using blue or black pen. *Thank you.*

Sunset Resort
1234 Palm Tree Blvd.
South Miami, FL 33116
1-800-555-1234

Generic "Sunset" DTD
Form 052 of 101
Created 03/05
050327-001

1. Is this your first visit to Southern Florida?
☐ Yes ☑ No

2. How often have you visited the Sunset Resort?
☑ First Time ☐ 2 - 3 times ☐ 4 or More

3. How did you hear about the Sunset Resort?
☐ Friend ☐ Radio ☐ Telephone ☐ Printed Advertisement
☑ Other - please specify Co-Worker

4. How many people in your party? 1

5. What age groups does your party fall into? (check all that apply.)
☐ 1 - 16 ☐ 17 -21 ☐ 22 - 40 ☑ 41 - 60 ☐ 61 and older

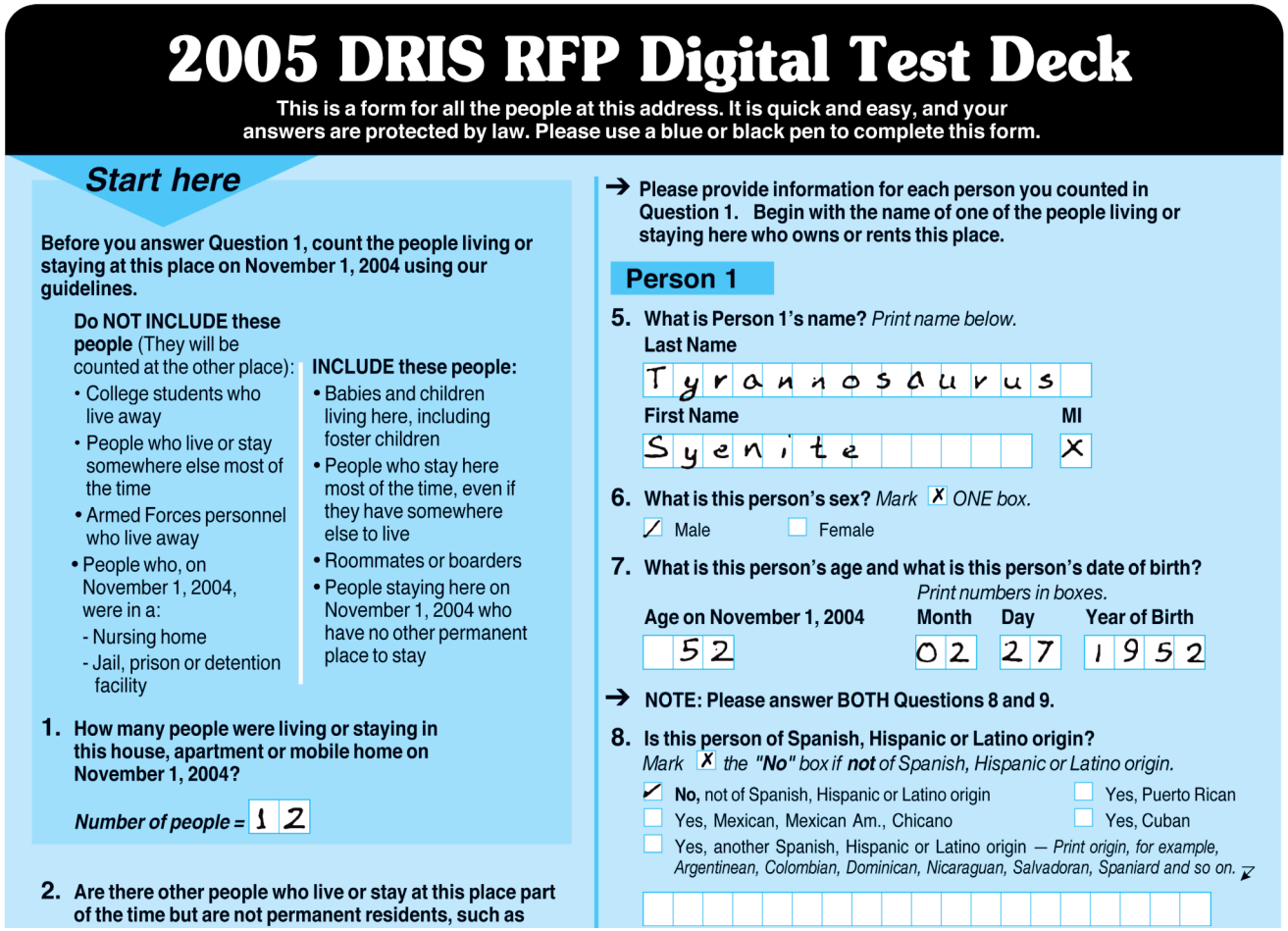
6. Which method did you use to book your visit?
☐ Travel Agent ☐ Called the toll-free number
☐ Tour Operator ☐ Internet Website
☑ Other - please specify Won visit

7. Which area attractions did you visit during your stay? (check all that apply.)
☐ Adams National Park ☑ Watson Botanical Gardens ☐ St. Augustine Mines
☐ Memorial Museum ☐ Vineyards
☐ Other - please specify

Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

Census RFP Test Deck



2005 DRIS RFP Digital Test Deck

This is a form for all the people at this address. It is quick and easy, and your answers are protected by law. Please use a blue or black pen to complete this form.

Start here

Before you answer Question 1, count the people living or staying at this place on November 1, 2004 using our guidelines.

Do NOT INCLUDE these people (They will be counted at the other place):

- College students who live away
- People who live or stay somewhere else most of the time
- Armed Forces personnel who live away
- People who, on November 1, 2004, were in a:
  - Nursing home
  - Jail, prison or detention facility

INCLUDE these people:

- Babies and children living here, including foster children
- People who stay here most of the time, even if they have somewhere else to live
- Roommates or boarders
- People staying here on November 1, 2004 who have no other permanent place to stay

1. How many people were living or staying in this house, apartment or mobile home on November 1, 2004?

Number of people = 12

2. Are there other people who live or stay at this place part of the time but are not permanent residents, such as

→ **Please provide information for each person you counted in Question 1. Begin with the name of one of the people living or staying here who owns or rents this place.**

Person 1

5. What is Person 1's name? *Print name below.*

Last Name

Tyrannosaurus

First Name | **MI**

Syenite | X

6. What is this person's sex? *Mark* X *ONE box.*

☑ Male ☐ Female

7. What is this person's age and what is this person's date of birth?

Print numbers in boxes.

Age on November 1, 2004	Month	Day	Year of Birth
52	02	27	1952

→ **NOTE: Please answer BOTH Questions 8 and 9.**

8. Is this person of Spanish, Hispanic or Latino origin?

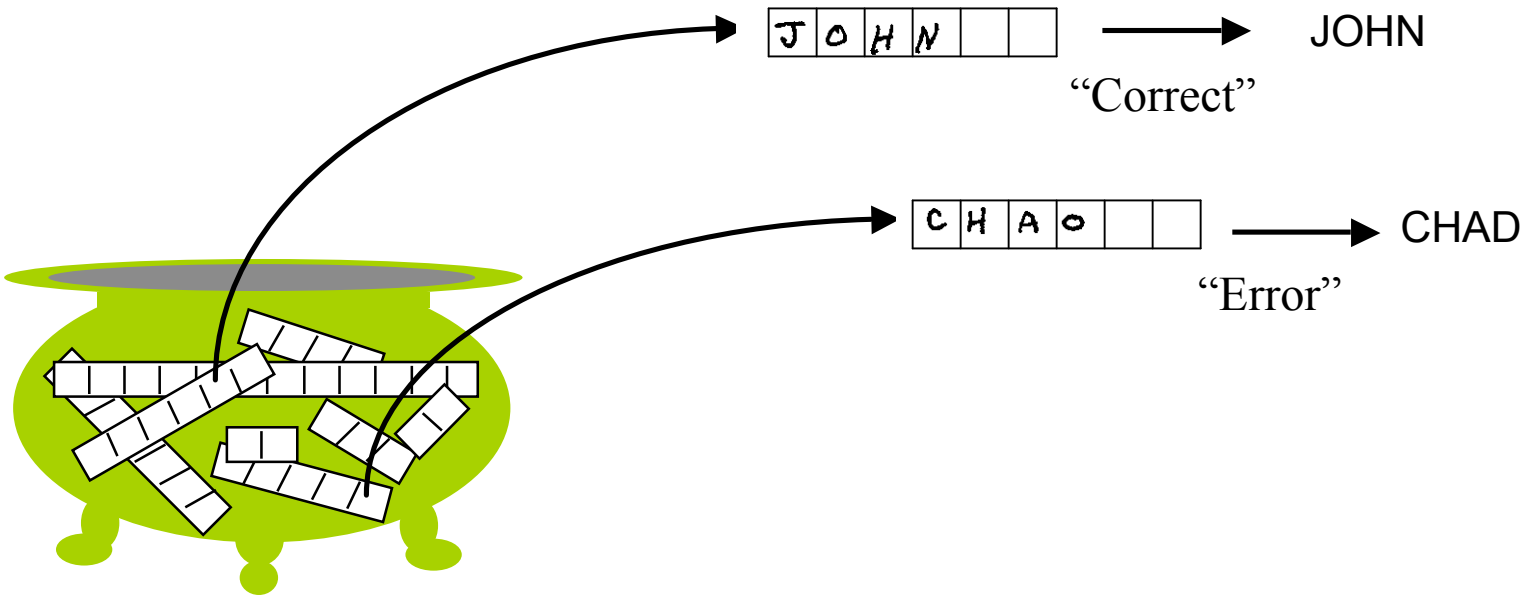
Mark X *the "No" box if* ***not*** *of Spanish, Hispanic or Latino origin.*

☑ **No,** not of Spanish, Hispanic or Latino origin
☐ Yes, Puerto Rican
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Cuban
☐ Yes, another Spanish, Hispanic or Latino origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard and so on.*

Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

Field Testing Simplified



Universe of Fields

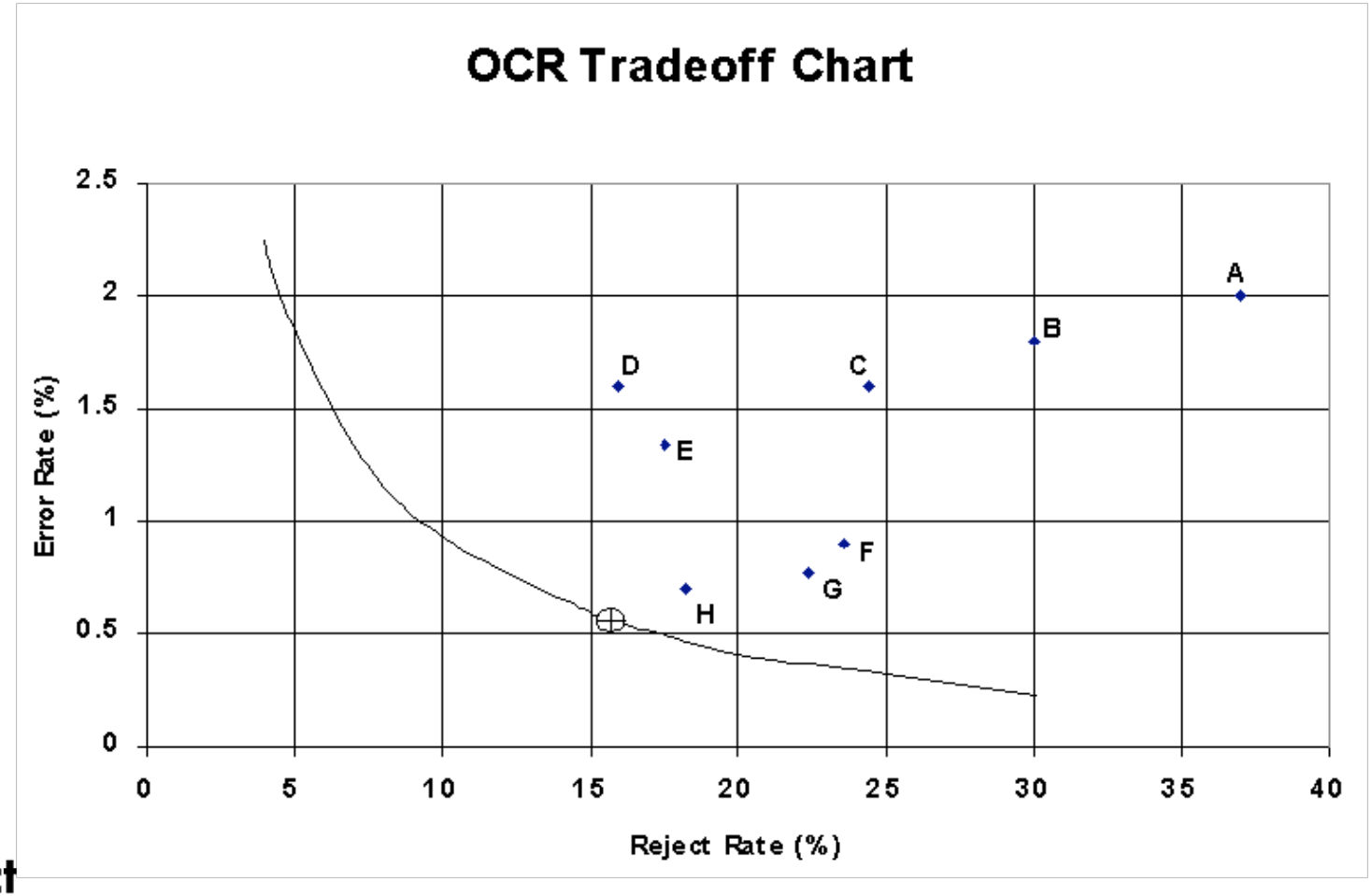
Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

Some Data Capture Improvement Opportunities

- Forms Design
  - Color, printing, layout
- Keying
  - Key from image vs key from paper
  - Field vs. Form keying
  - Organize workflow by field type
- Intelligent Character Recognition “Tuning”
  - Field vs. character level metrics and analysis
  - Reject Rate operating point selection
  - Multiple engine voting
  - Proper use of dictionaries
  - Context checking
- System-level QC on-line during production

Census 2000 “Journey”



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

Cost Model

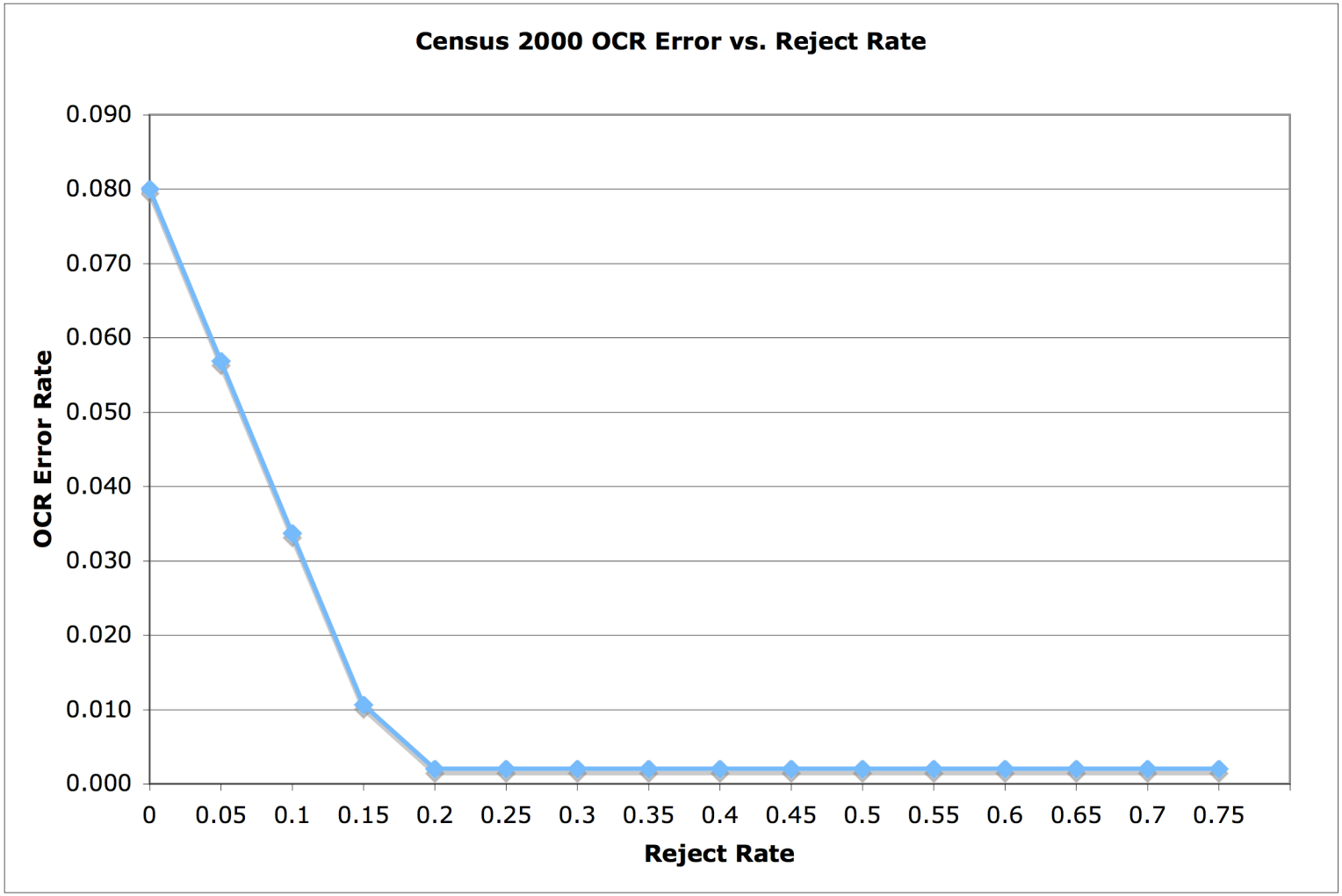
- Many factors enter into determining the true cost of data capture:
  - Number of forms processed in a year, say
  - Form complexity
  - System software and equipment
  - Keying personnel and support
  - OCR performance*
  - Keying performance*
  - Cost of an error downstream in your business process

- With good data capture system design, there is an optimal reject rate for a given system that minimizes data capture cost

---

*The performance of OCR and Keying subsystems are robustly measured using Digital Test Deck™ technology

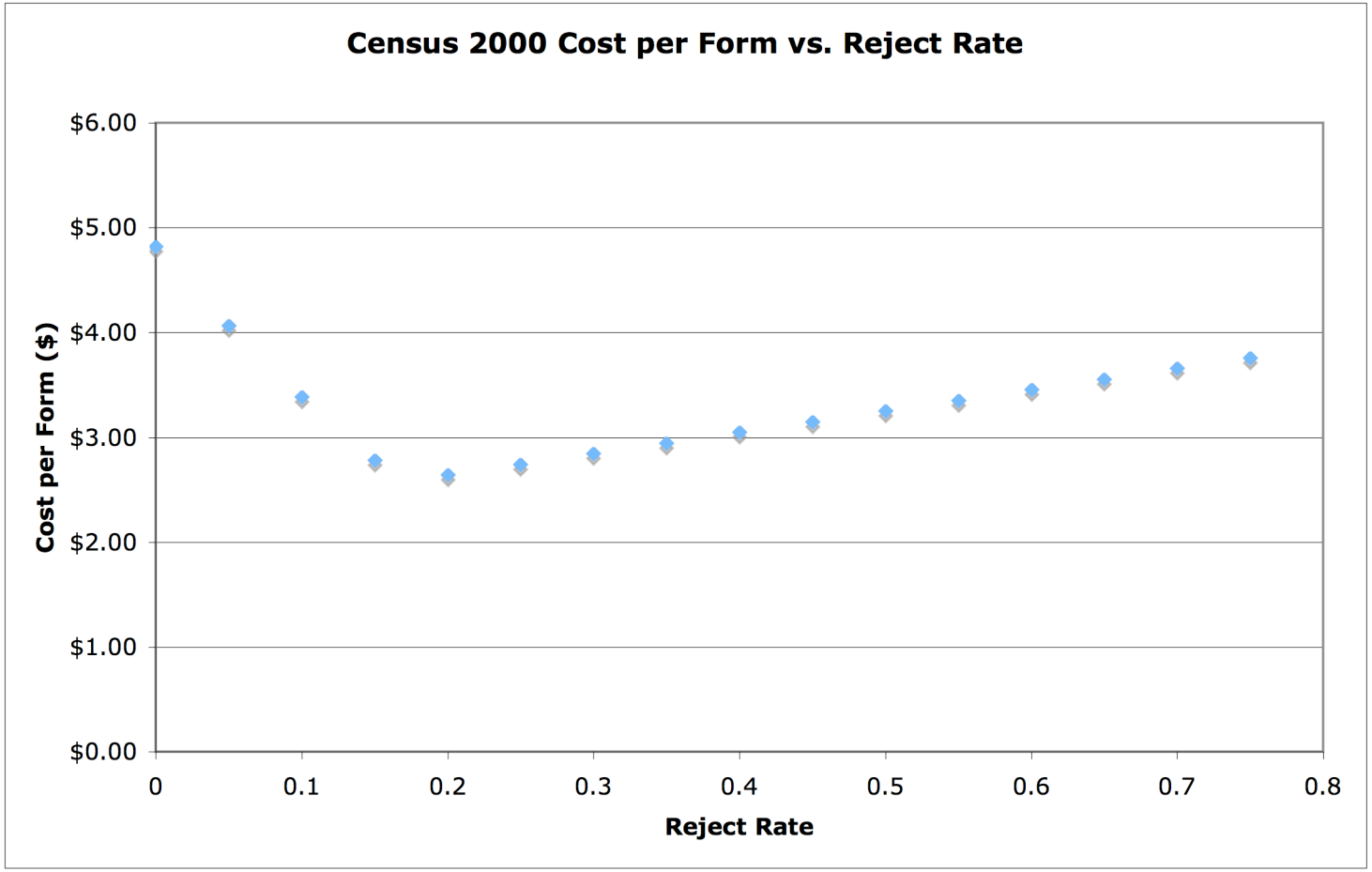
U.S. Census 2000 Example-OCR



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

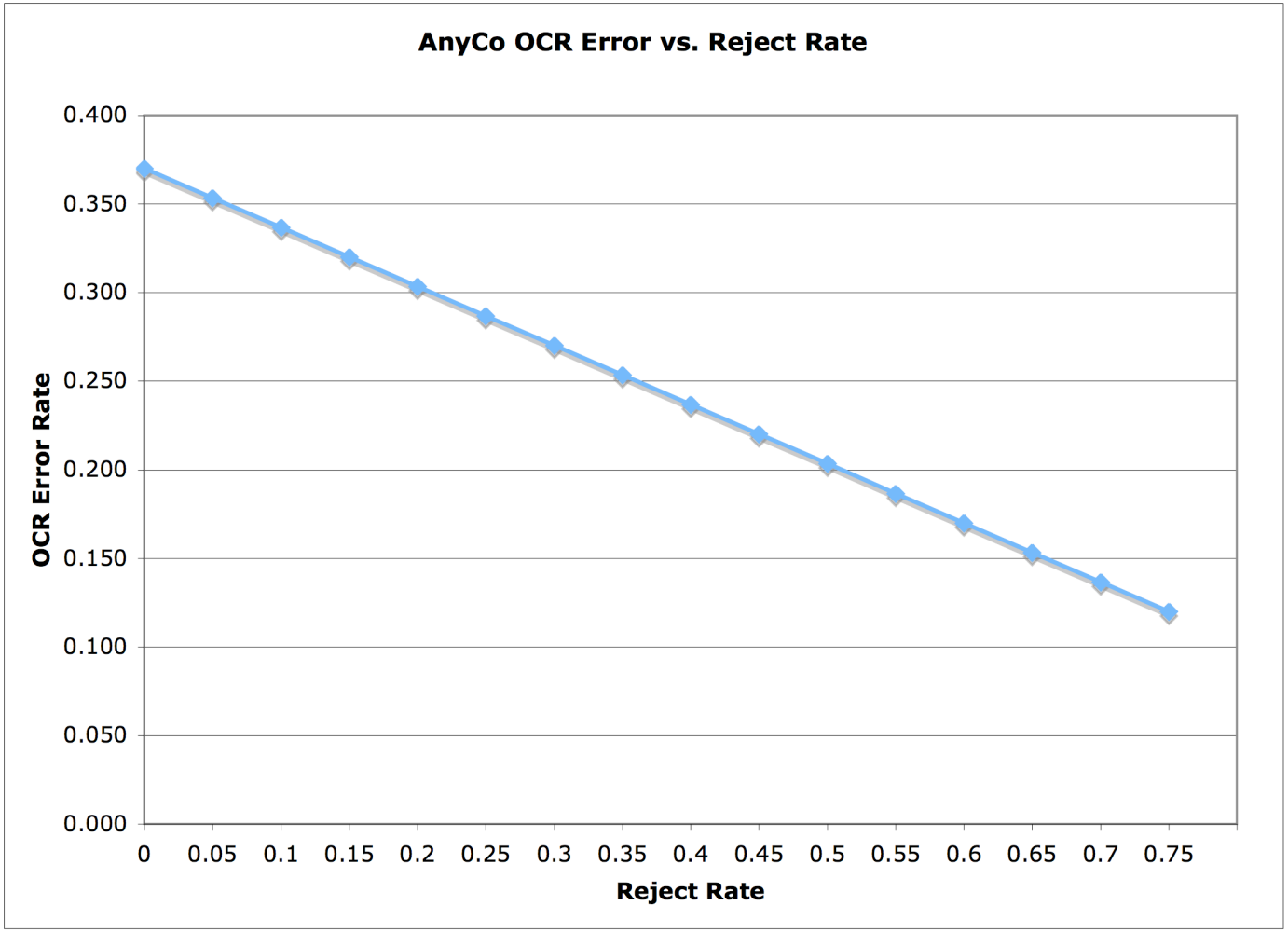
U.S. Census 2000 Example-Cost



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

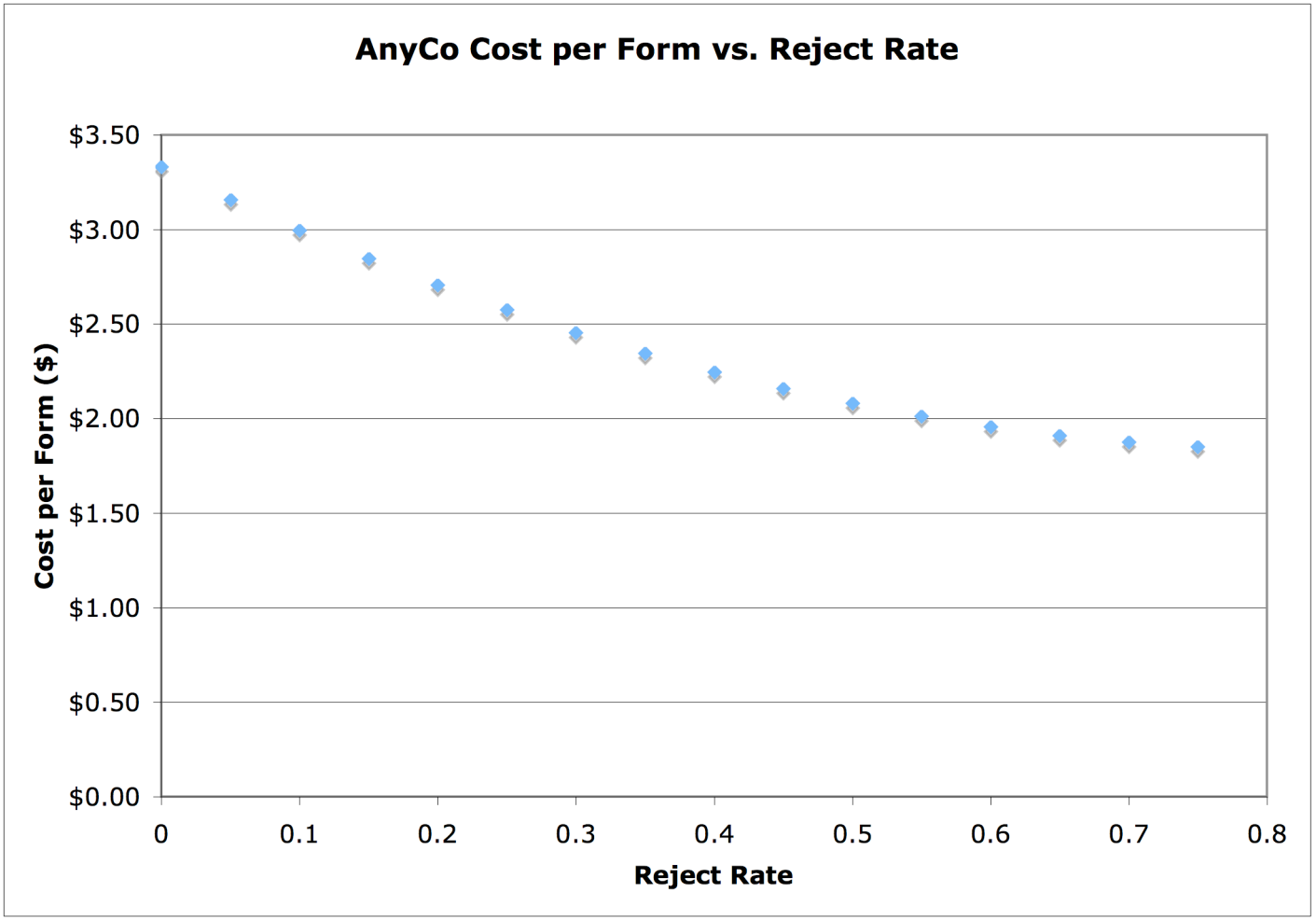
AnyCo Example-OCR



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

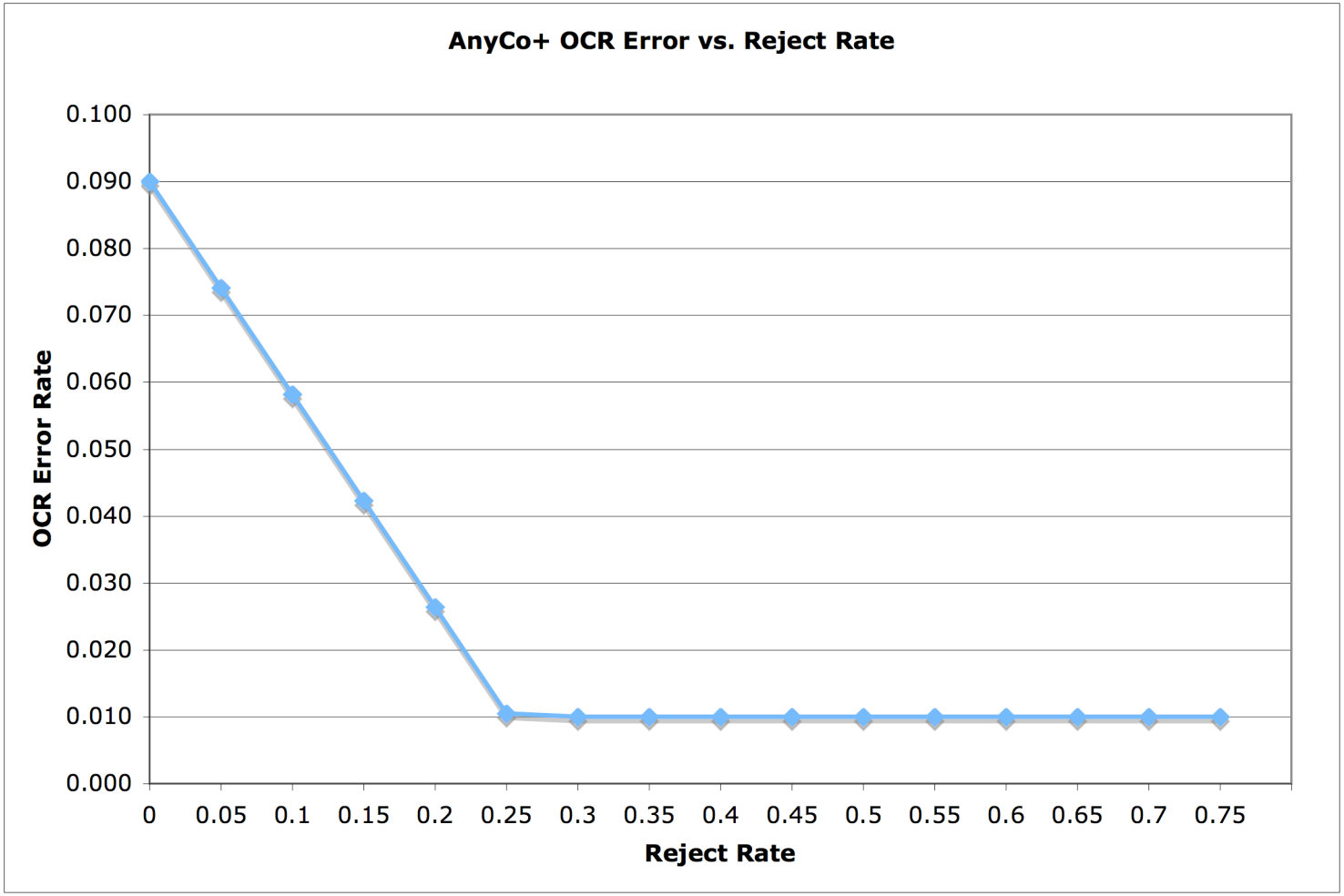
AnyCo Example-Cost



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

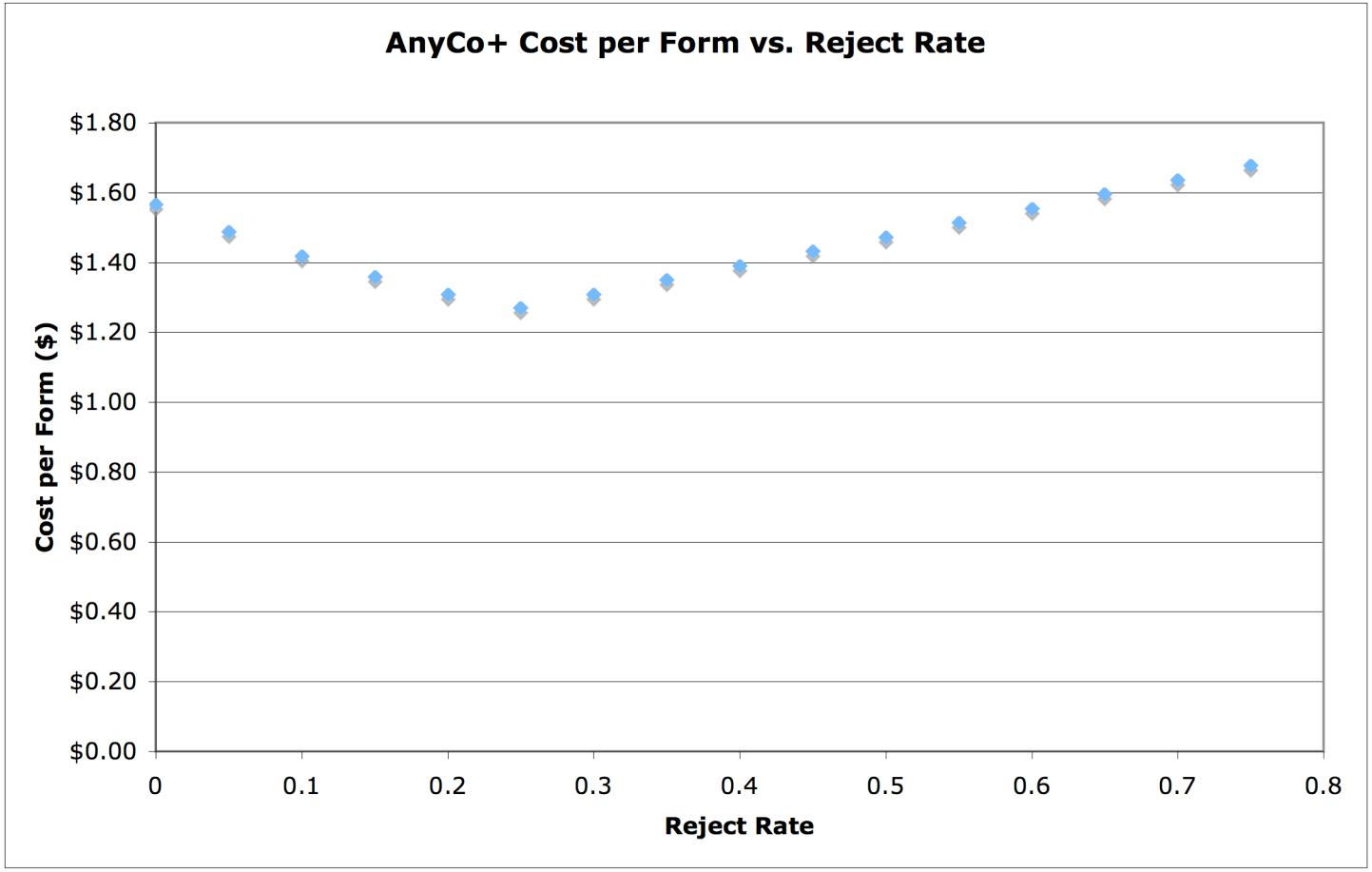
AnyCo+ Example-OCR



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

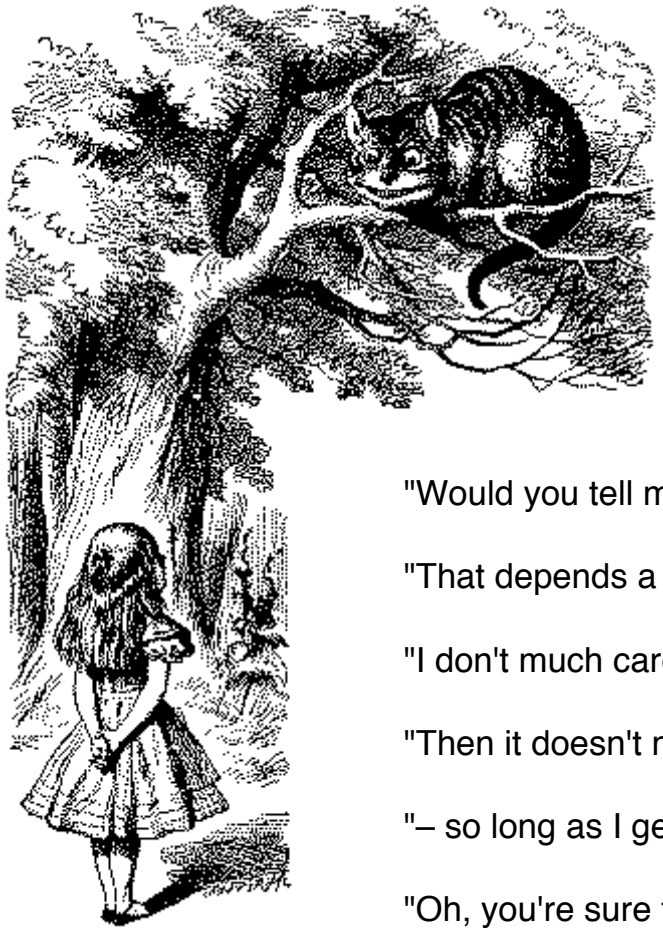
AnyCo+ Example-Cost



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

About Requirements



"Would you tell me, please, which way I ought to go from here?”

"That depends a good deal on where you want to get to," said the Cat.

"I don't much care where –" said Alice.

"Then it doesn't matter which way you go," said the Cat.

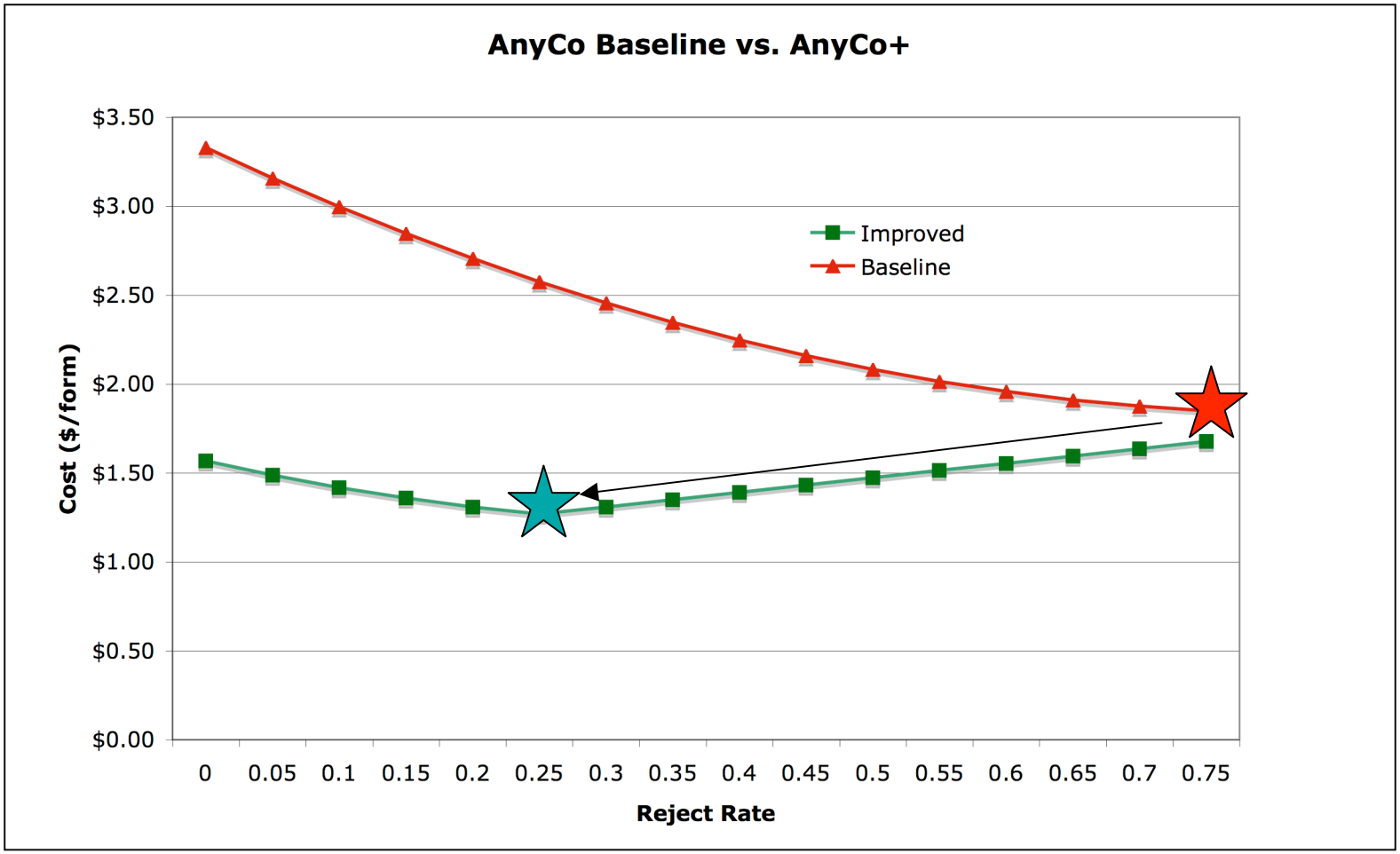
"– so long as I get *somewhere*," Alice added as an explanation.

"Oh, you're sure to do that," said the Cat, "if you only walk long enough."

Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

AnyCo Baseline vs. Improved



Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI

Conclusions from Examples

- Census 2000 was state-of-the-art handprint data capture
- AnyCo is spending about $1.85/form
- If AnyCo improved their data capture system they could potentially reduce their cost over $0.40/form
- Even if it was only $0.10/form...a savings of $0.10/form at 100M forms/year is a savings of $10M/year!

The Bottom Line

Use of robust, quantitative data capture system evaluation, coupled with careful cost analysis can help lead the way to better data capture quality at significantly less cost!

Exact Data accurate. fast. efficient.

ADVANCED IMAGING DOCUMENT ADI