

•
•
•

Data Preparation Techniques for Your Database

JOE BALABAN

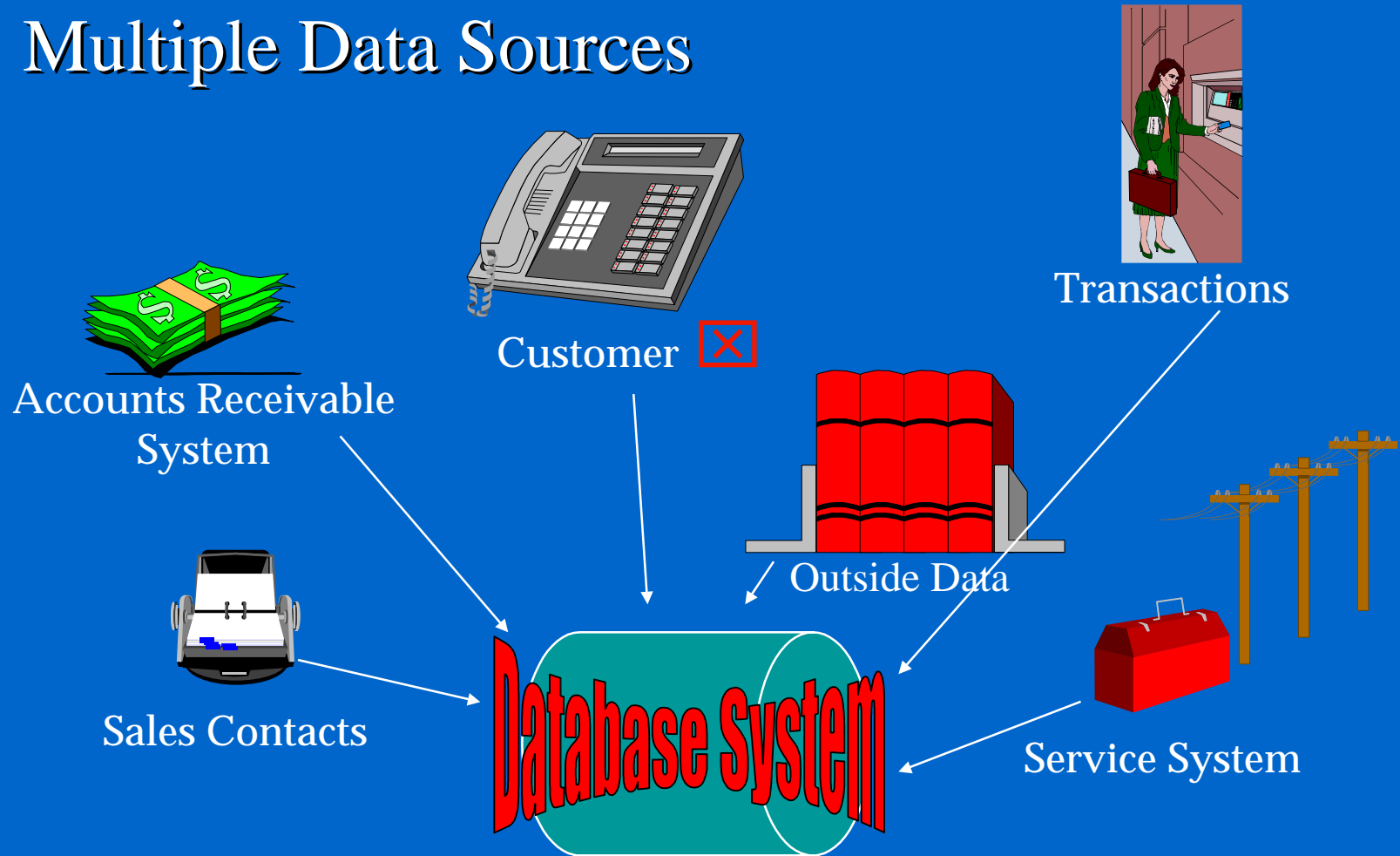


Outline

- Data Sources
- Need For Data Standardization
- Difference in Data
- Data Quality Flow
- Techniques For File Preparation
- Required Data Elements
- Decision Points

Typical Situation

Multiple Data Sources

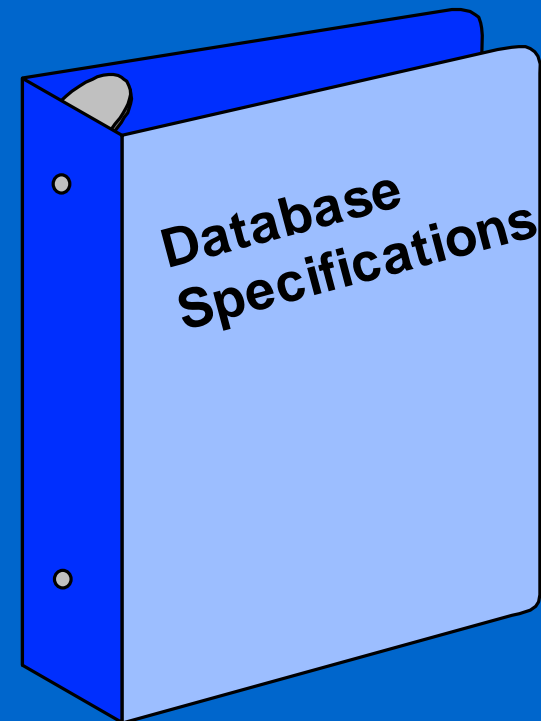


Causes Of Incompatibility

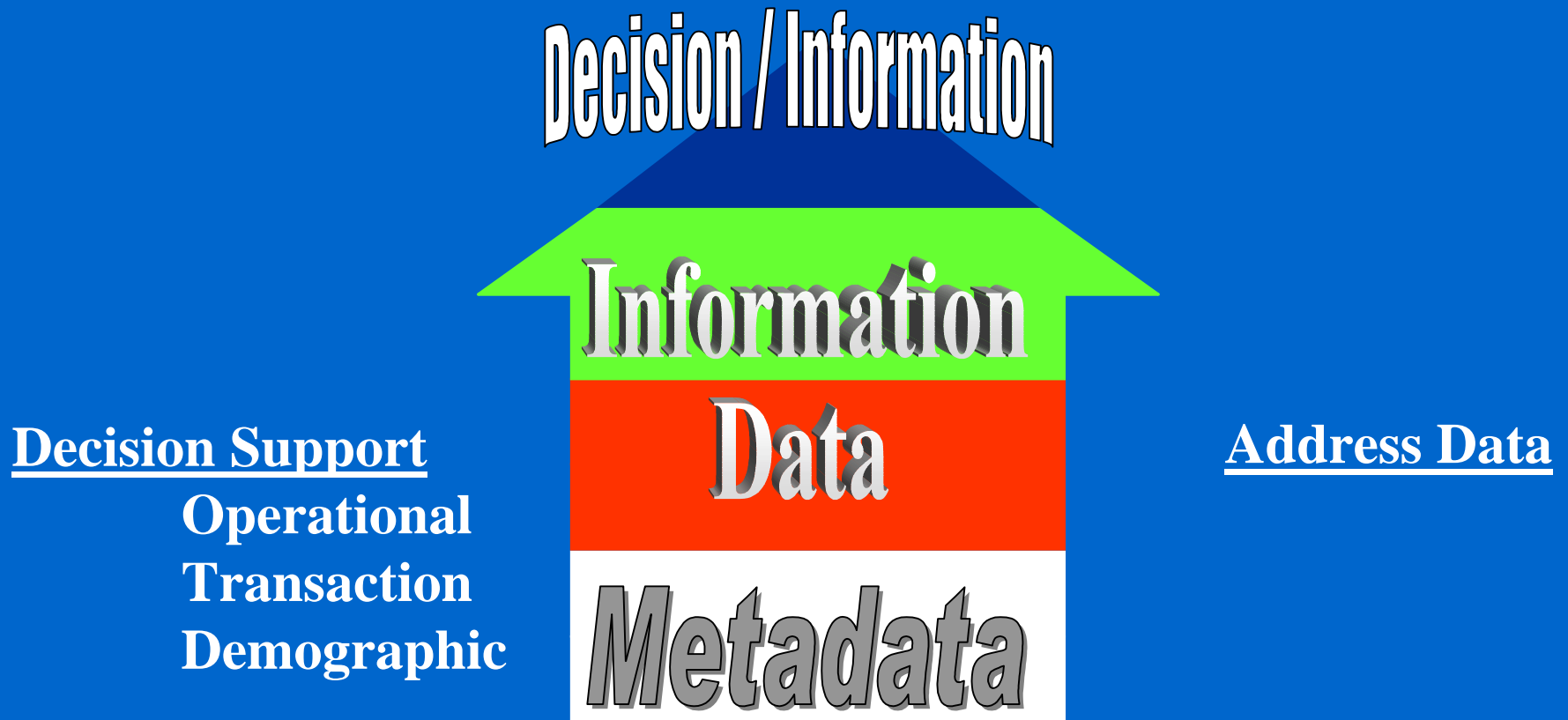
- Non-Machine Readable
- Inconsistent Information
- Name and Address Differences
- Ambiguous Information
- Incomplete Information
- Numbering System Differences
- Multiple Names and Addresses

Create Operational Documents

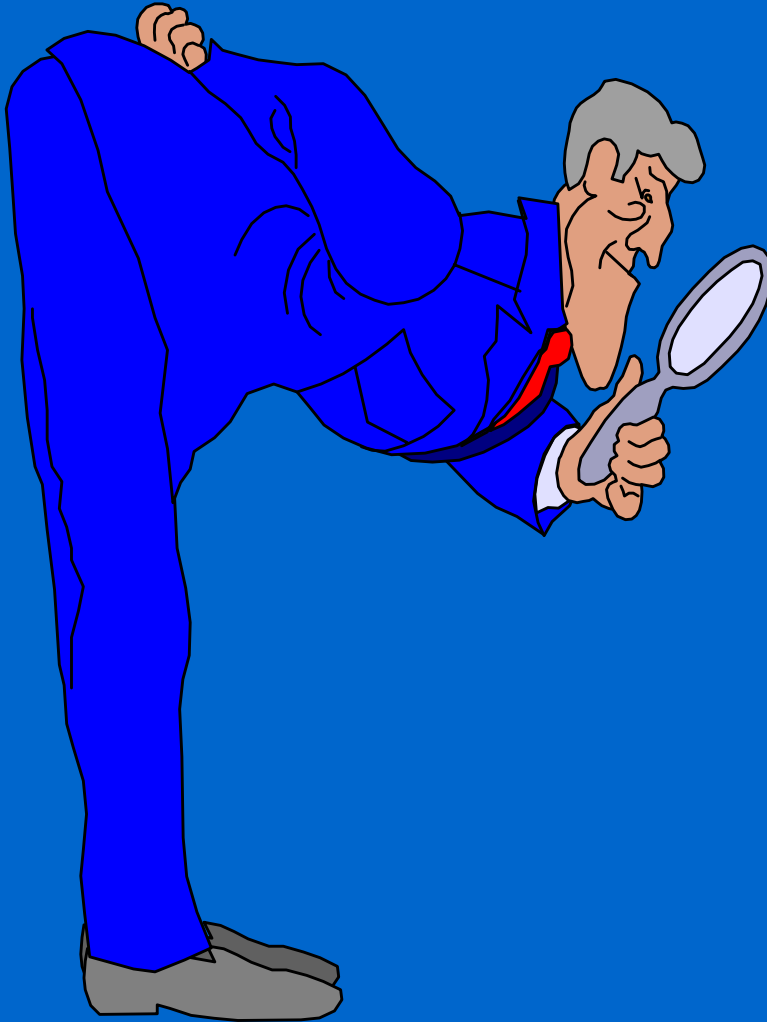
- Must Contain Detailed specs
- Sign-off by all concerned
 - revisions
 - data edits
 - etc. . .



Data Characteristics



Metadata



Data About Data

Source

Structure

Format

Update

Logical Relationships

Data Model

Consumer

- ☺ Household
- ☺ Individual
- ☺ Transaction
- ☺ Service

Business

- ☺ Enterprise
 - ☺ Multiple Names
 - ☺ Sites
- ☺ Individuals
 - ☺ Contact
 - ☺ Influence
- ☺ Transactions
- ☺ Service

Data

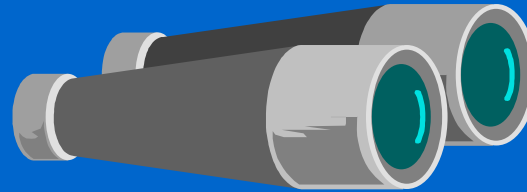
GOAL

- ✓ Consistent
- ✓ Complete
- ✓ Correct
- ✓ Best Information Available
- ✓ Multiple Sources

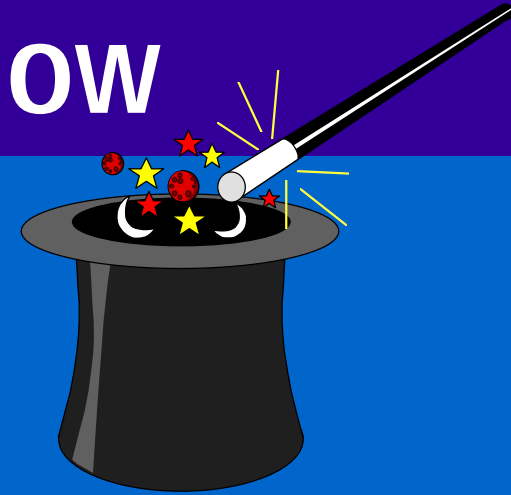
Data Quality Flow



1. Sources



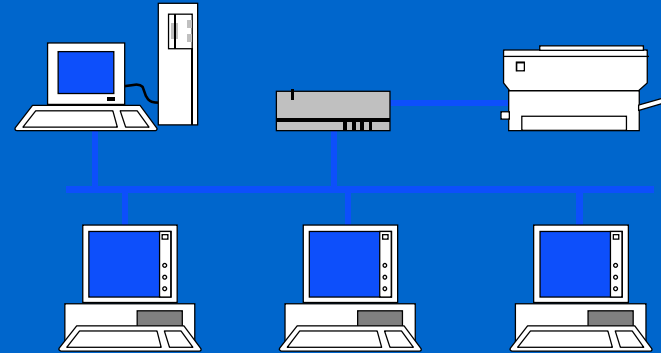
2. Parse, Correct
Standardize



3. Match



4. Consolidate



5. Warehouse/Mart

Building The Database

- Factors Effecting Data Quality
 - ? Business Requirements
 - ? How & Who
 - ? Availability of Data
 - ? Reliability of Data
 - ? Maintainability of Data

Required Data Elements



Postal Requirements
Requirements

Pub. # 28
Information

Business

Operational

Transaction Information
Contact History

Order History

GeoDemographics

Response Data

Address Information

Preparing Data

Preprocessing steps

Identify Sources and Owners

Infrastructure Bias

Front End Edits & Validation

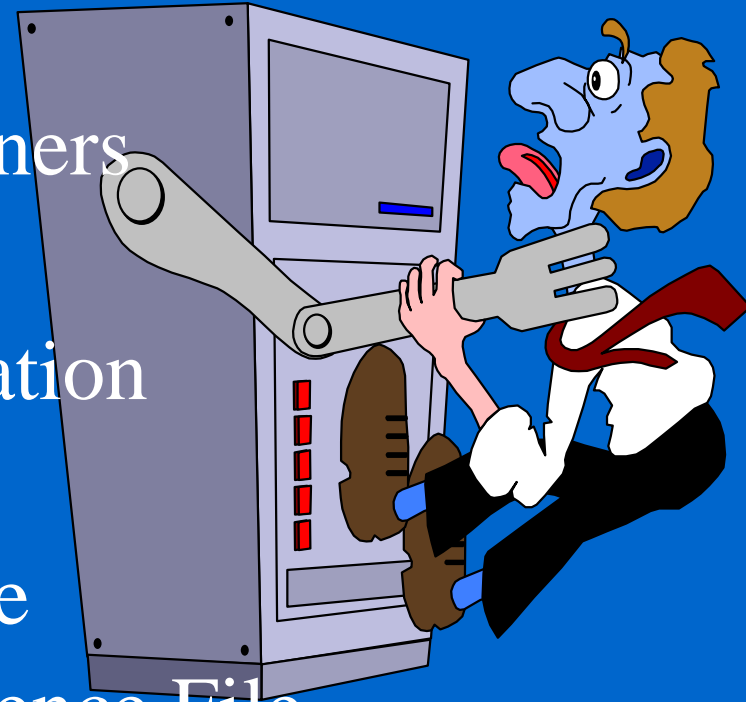
Conversion / De-Code

Standardization / Hygiene

Identify Priority or Reference File

Develop Common Numbering System

Update Source Systems



Inter-Application Communication

Data Bridges Within a Program

- Convert a date from one format to another
- Used after reading a record
- Used before writing a record
- Almost universally applicable
- Simple programming changes required when input or output file date format changes

Data Bridges Between Programs

- Duplicates data
- Requires additional processing time
- May not be applicable to all data
- Easy to remove when date data is compatible

Parse Input Data

Input

Parsed Record

Joe G Balaban, VP
Anchor Computr
1900 New Highway
Farmdale NY 11735

Computr

Highway

Field Name

1st name Joe

Middle Initial G

Last Name Balaban

Title VP

Co. Name Anchor

Range 1900

Street New

City Farmdale

State NY

Standardize Elements

Parsed Record

Standardized Data

1st name

Joe

Joseph

Middle Initial

G

G

Last Name

Balaban

Balaban

Title

VP

Vice President

Co. Name

Anchor **Computr**

Anchor

Computer

Range

1900

1900

Street

New Highway

New

Highway

City

Farmdale

Farmingdale

State

NY

NY

Validation Test

Number of Name or ZIP Error Rejects \pm

Records with Valid Last Names and/or Company Names \pm

Of Valid City Names Found

Records with Valid First Names and/or Company Names

Records with Valid Street Types and/or PO Boxes

Records with 0 (Zero) Numerics in Street Address

Records with 1 Numerics in Street Address

Records with 3 Numerics in Street Address

Records with Numerics in Care 1 Field

Records with No Alpha Characters in Street Address

FIRST-CLASS

- AUTOMATION

7/1/97

- *UPDATE ADDRESSES*

- EVERY SIX MONTHS*

- *CASS CERTIFICATION*

Address Hygiene

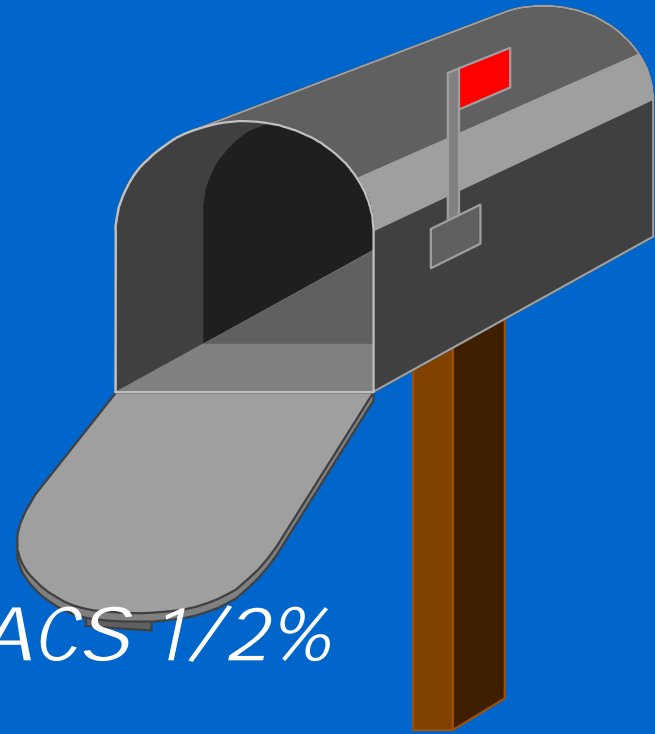
- BENEFITS

- \$ Postage Discounts
- \$ Improved Delivery
- \$ Higher Response Rates
- \$ Increased Revenue
- \$ Increased Cash Flow

- TYPICAL RESULTS

NCOA 3% - DSF 4% - LACS 1 1/2%

TOTAL 7 1/2%



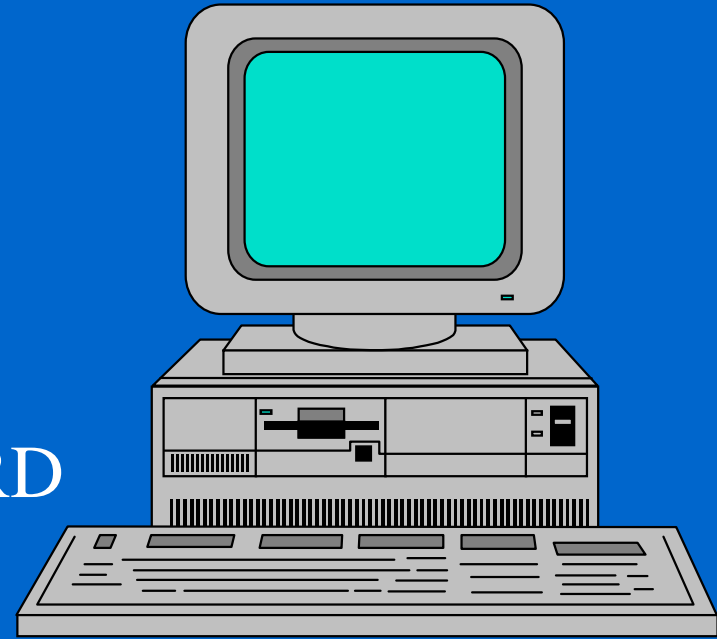
Address Hygiene Products

- POSTAL BULLETIN
- NCOA, NCOA/NIXIE
- DSF
- LACS
- AEC
- Ancillary Endorsements - ACR/ACS
- *FASTforward*
- MLOCs
- PROPRIETARY PROGRAMS



Postal Bulletin

- MONTHLY CHANGES
 - CITY NAME
 - POSTAL UNIT
 - ZIP CODE
 - ON USPS BULLETIN BOARD



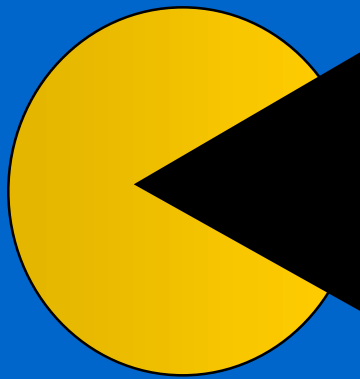
Post Office Changes

<i>Old/ New</i>	<i>Finance No.</i>	<i>ZIP</i>	<i>State</i>	<i>P.O.Name</i>	<i>County/ Parish</i>	<i>Station Branch/unit</i>	<i>Unit Type</i>	<i>Effective Date</i>	<i>Comments</i>
OLD	053018	91740	CA	Glendora	Los Angeles	Main	Post Office	7-1-93	Establish a NEW ZIP Code 91740. Use Glendora, CA 91741 as last line of address.
NEW	053018	91741	CA	Glendora	Los Angeles	Main	Post Office	7-1-93	
OLD	206924	41370	KY	Sebastian	Breathitt	Main	Post Office	2-13-93	Post office and ZIP Code discontinued. Establish a place name. Use Sebastian, KY 41314 as last line of address.
NEW	200864	41314	KY	Booneville	Owsley	Sebastian	Post Office	2-13-93	

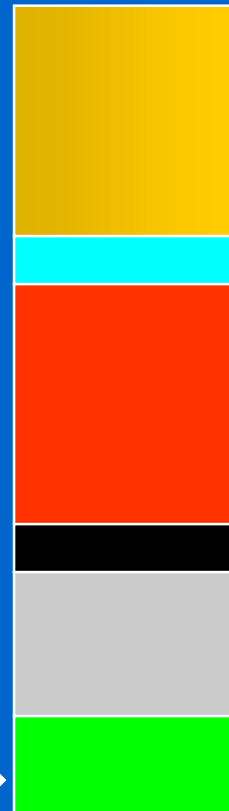
###

MAJOR ADDRESSING DEFICIENCIES

Deliverable
Addresses
66.9%



Errors
33.1%



8.6 % Moved

2.2% Apartment Number

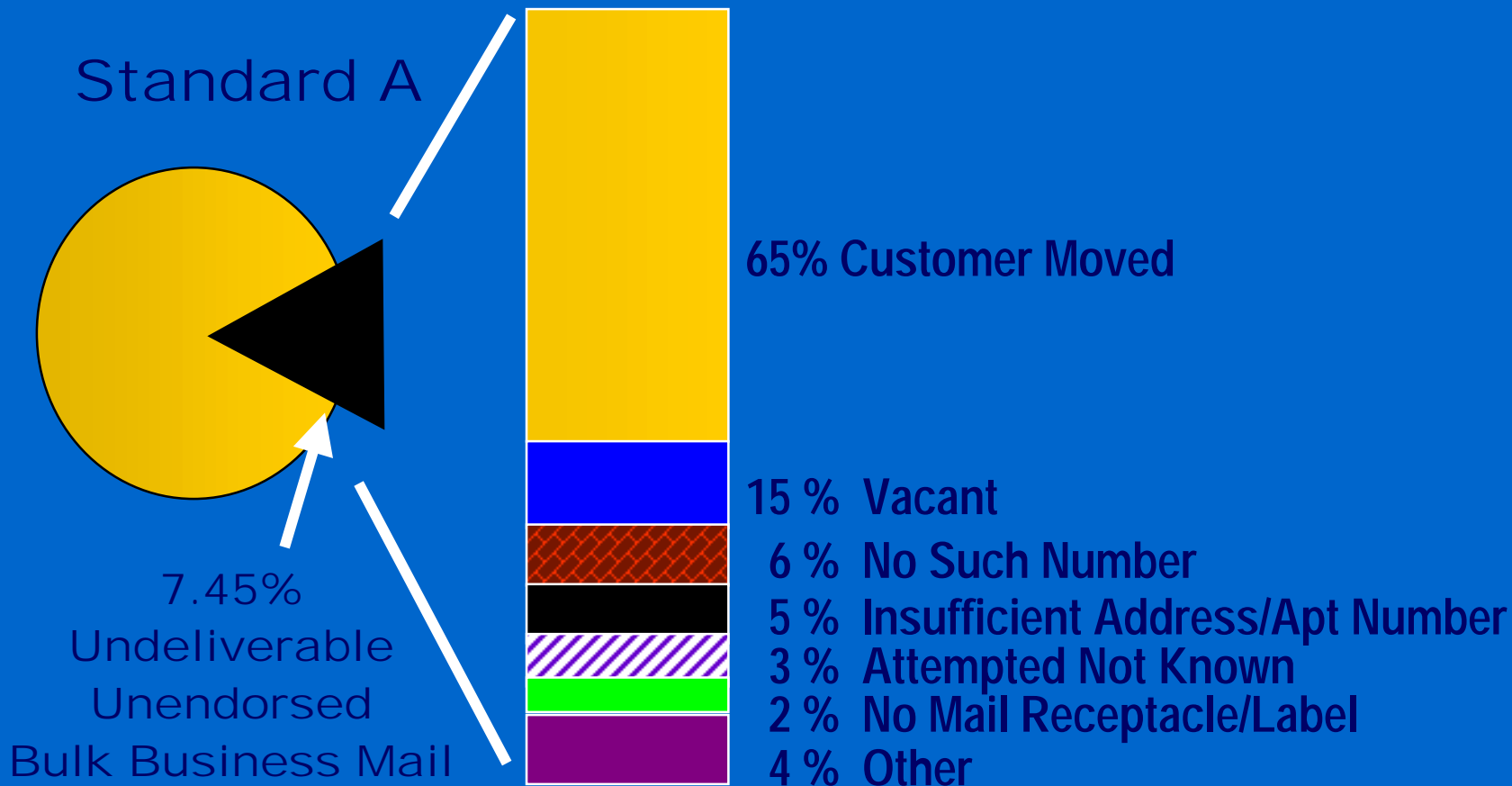
8.9% Directional / Suffix

2.5% Rural Route / Box Number

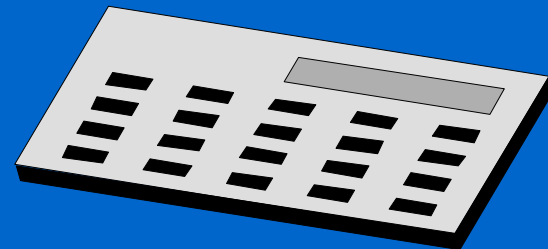
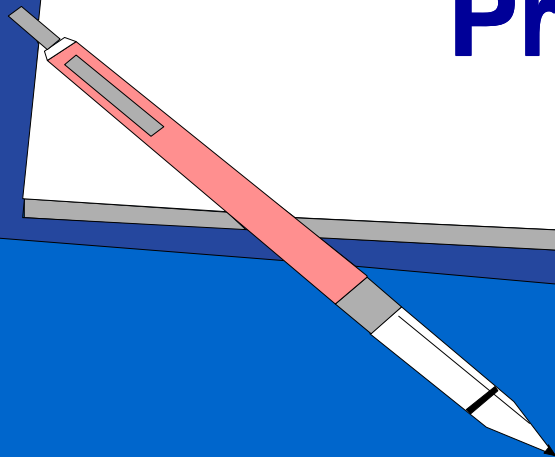
6.2% Street Name / Number

4.7% ZIP Code / City / State

UNDELIVERABLE UNENDORSED BULK BUSINESS MAIL



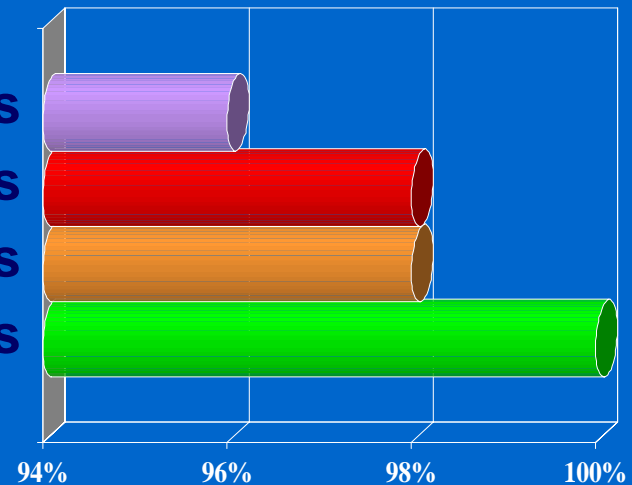
CASS Measures The
Accuracy Of
Address-Matching
Programs



CODING ACCURACY SUPPORT SYSTEM CERTIFICATION

**To Be CASS-Certified By The USPS,
Software Must Correctly Code Test
Records To The National ZIP+4 File
With The Following Accuracy:**

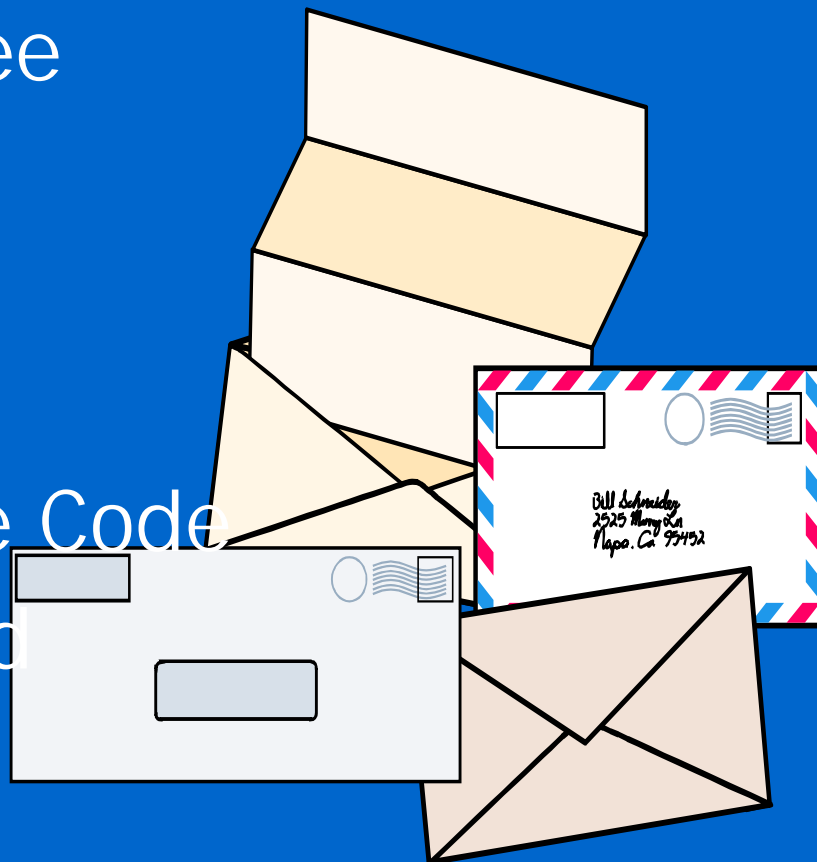
96% For ZIP+4 Codes
98% For 5-Digit ZIP Codes
98% For Carrier Route Codes
100% For Delivery Point Codes



NCOA, NCOA/NIXIE, DSF, LACS

common processing

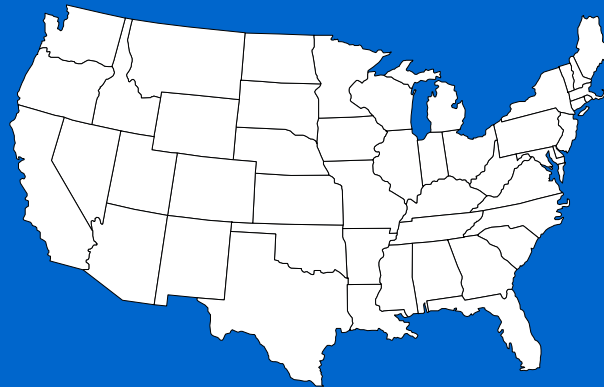
- ☐ USPS Licensee
- ☐ Parse
- ☐ Standardize
- ☐ ZIP Correct
- ☐ Carrier Route Code
- ☐ ZIP+4 Append
- ☐ BAR Code



DELIVERY SEQUENCE FILE ...

DSF

Contains Each Of The 135+ Million Delivery Points
Serviced By The USPS (Except General Delivery)



Offered To Mailing Industry
By Non-Exclusive Licensees

DELIVERY SEQUENCE FILE ...

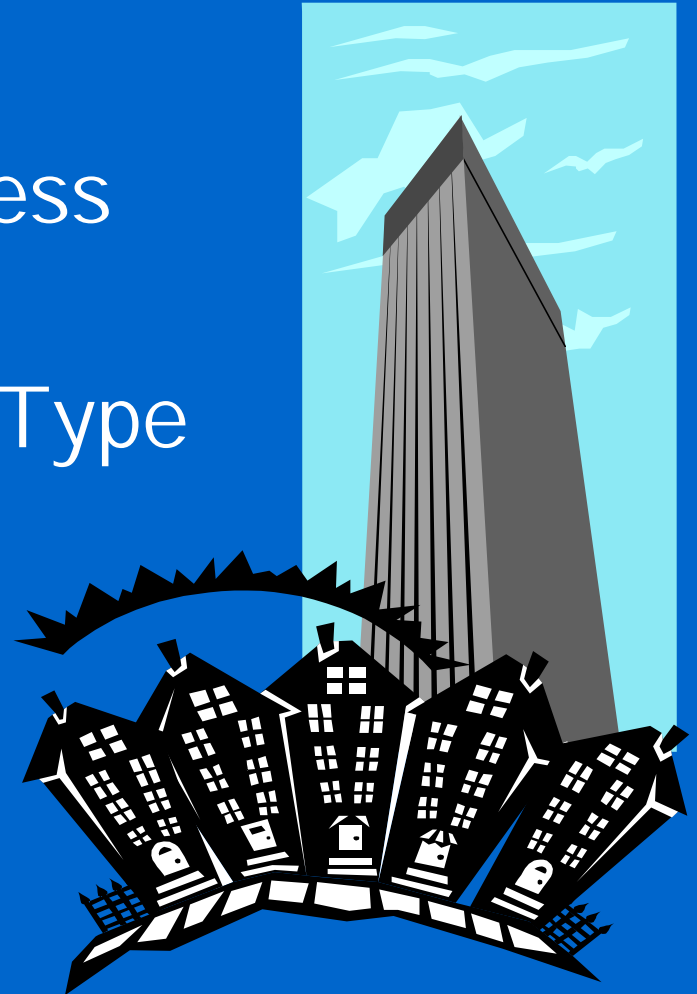
PROVIDES MAILERS WITH
THE FOLLOWING:

- **Address Validation**
- **ZIP+4 Coding**
- **Address Standardization**
- **Carrier Route Coding**
- **Carrier Delivery Sequence**



DSF

- 🏠 136,000,000
- 🏠 Address Only
- 🏠 Delivery Point Address
- 🏠 Vanity
- 🏠 Dwelling & Address Type
- 🏠 Add - Ons'
- 🏠 Walk Sequence
- 🏠 90 Day Vacancies



DELIVERY SEQUENCE FILE ...



- Detects Potentially Undeliverable Addresses For Reasons Other Than Change-Of-Address
- Saves Costs By Improving Address List Quality
- Indicates Delivery-Type Code For Business Or Residence
- Provides Seasonal Delivery Information

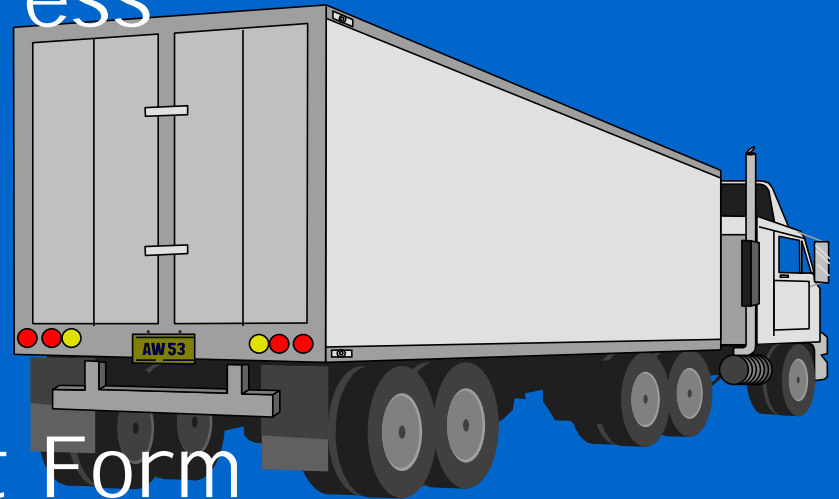
NATIONAL CHANGE OF ADDRESS



- Helps You To Maintain Your Address File
- Provides New Address Information When Your Customers Move
- Processing Services Provided Through Licensees

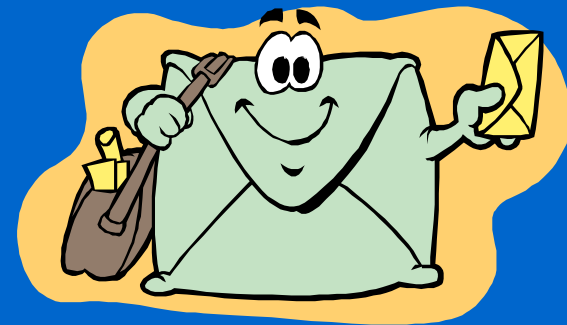
NCOA, NCOA/NIXIE

- 🏠 120,000,000
- 🏠 Name & Address
- 🏠 36 Months / 1 Week
- 🏠 Forwarding Address
- 🏠 NIXIE Option
- 🏠 Move Type
- 🏠 6% Business
- 🏠 Acknowledgment Form



LOCATABLE ADDRESS CONVERSION SYSTEM

- Obtain Updates Of Your Addresses Changed Due To 911 Or Other Reasons
- Old Addresses May No Longer Be In USPS Data Files. Consequently They Can No Longer Be Barcoded
- When An Old Address Can Be Updated To A New Address, You Can Then Barcode The New Address
- Processing Services Provided Through Licensees



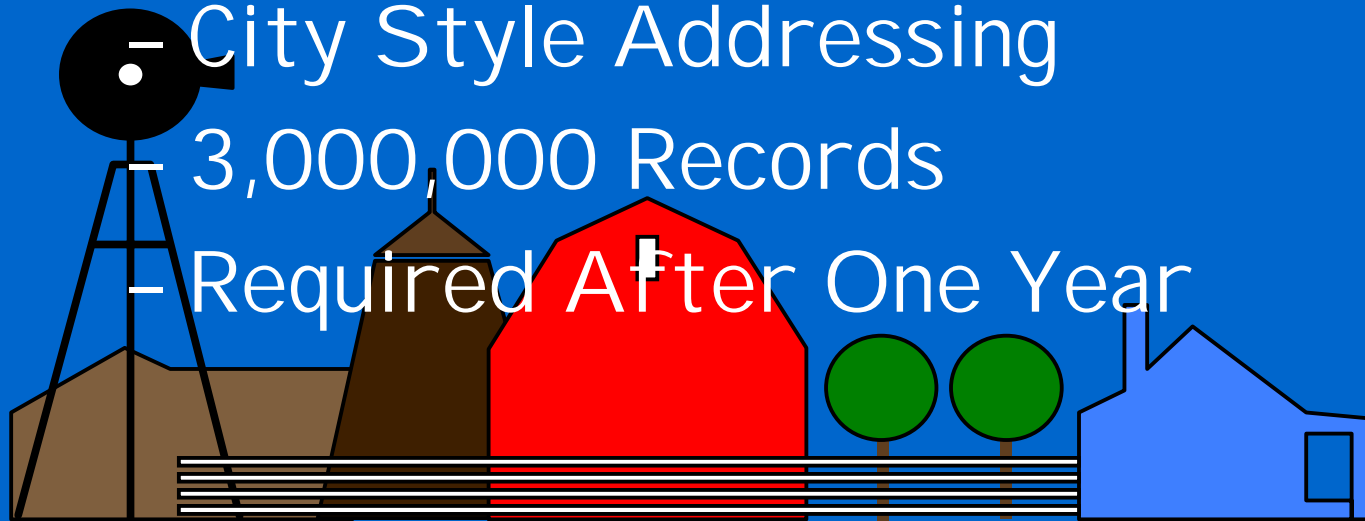
LACS

- IMPROVES DELIVERY

- City Style Addressing

- 3,000,000 Records

- Required After One Year



- Improves Matching

ADDRESS ELEMENT CORRECTION (AEC)



FROM THE
UNITED STATES
POSTAL SERVICE

ADDRESS ELEMENT CORRECTION (AEC)

A QUALITY PROCESS DEVELOPED BY THE USPS WITH INDUSTRY SUPPORT

Focuses On Addresses
Unable To Be Matched
To A Postal ZIP+4 Code
Using CASS-Certified
Address-Matching Software

Focuses On The 33%
Of All Addresses On Mail
Containing Deficiencies
That Might Cause The Mail
To Be Delayed Or Not Delivered



Mail Stops

Identifies Physical Location in Building or Complex

For Example:

MSK10, RM 7A-377, Bldg 123,
M/S 123, STECS-DA-ID

Common In:

Government Agencies

Large Companies



Mail Stops

Work Place Marketing Mail Preference
Edmund

Scientific

PURPOSE:

To reduce waste by eliminating unwanted mail and improving delivery without deleting customers and prospects who want to receive the Edmund Scientific Catalog.

METHODOLOGY:

Ask the mail room to help reduce waste and appeal to their interest in the environment.

-
-
-

Match Your File To A National Database

Confirm Name and
Address Data and
Identify Characteristics
Shared by Your Best
Customers Enabling You
to Target Promotions to
Your Best Prospects

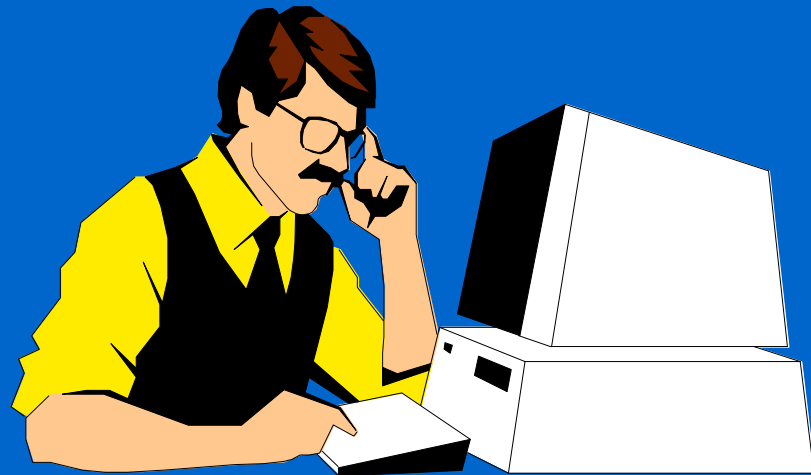
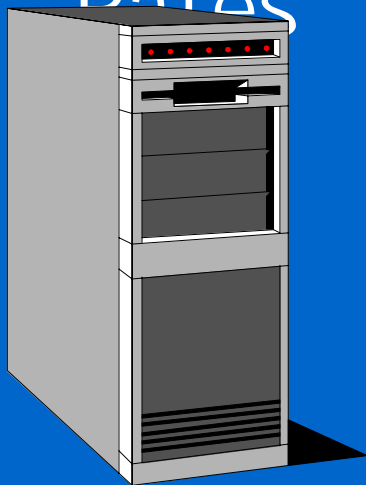


The Enhancement Process

5-15 Days

45-70% + Matches

Clean Files to Get Highest Match Rates



Additional Sources

Department of Commerce

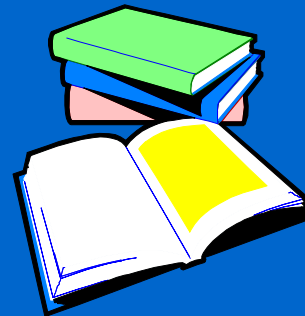
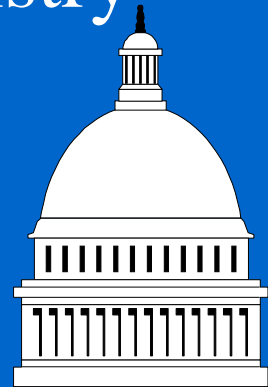
Real Estate Appraisers Registry
County Business Patterns

Chamber of Commerce

Member Listings

Trade Publications

Specialty Titles
Directories



List Enhancements

	Compiled Qual	
	<u>Lists</u>	<u>Card</u>
10 Million Names	X	V
Address Std.	X	V
Age of Executive	X	X
Big Business Executive	X	X
Multiple Executives	X	
Phone Numbers	X	X

•
•
•

List Enhancements (continued)

<u>Card</u>	Compiled	Qual
	<u>Lists</u>	
Hot Listings:		
New Companies	X	
New CEO	X	
New Locations	X	
Year Started		X
Co. Name	Change	X

List Enhancements

(continued)

Compiled	Qual	<u>Lists</u>	<u>Card</u>
Carrier Route Coding		X	X
Census Tract Codes		X	X
Government		X	x
Computer Side Data		X	X
Congressional Distr.		X	
Contact Names		X	X

•
•
•

Questions To Ask:

? *How Large is the Appending File --*

One record per firm?

Under 20 employees?

Under \$1 million in sales?

? *How Current is the Data --*



- ? *SIC-Coding DMA-Approved*
?
- ? *Enhancement Experience ?*
- ? *Security/Confidentiality ?*
- ? *World Wide Coverage ?*



For More Information?

- ✉ **NATIONAL CUSTOMER SUPPORT CENTER**
- **UNITED STATES POSTAL SERVICE**
- **6060 Primacy Pkwy STE 101**
- **MEMPHIS TN 38188-001**
- ◆ **POST OFFICE ACCOUNT REP.**
- ◆ **BUSINESS CENTER**
- ◆ **<http://www.usps.gov/>**
- ◆ **RIBBS 1-800-262-9541**
- ◆ **FEDERAL REGISTER**

THANK YOU



JOE BALABAN
ANCHOR COMPUTER