



Digging for the Gold

*Data Mining Initiatives in the U.S. Department
of Housing and Urban Development,
Office of the Inspector General*

Auditor's Forum on Katrina Relief

Baton Rouge, Louisiana

October 10, 2007

Cliff Cole
Computer Audit Specialist
U.S. Dept. Of HUD OIG
Information Systems Audit Division



Agenda

- Definition
- Process
- Data Sources
- History
- Software Tools
- Results – Examples
- Conclusion
- Questions



Data Mining Defined



What is Data Mining?

The process of analyzing data from a variety of perspectives and summarizing it into useful information - information that can be used to maximize audit and investigative resources by:

- ✓ isolating leads
- ✓ identifying problem subsets of the universe
- ✓ decreasing man hours
- ✓ increasing reportable monetary benefits

Data Mining Defined

- Data Mining Includes ...

- ✓ Matching initiatives
- ✓ Filtering data for audit lead/fraud indicators
- ✓ Testing internal controls
- ✓ Quantifying and scoping out relevant universes
- ✓ Using databases as the foundation for a statistical analyses
- ✓ Selecting samples for detailed audit testing

The Data Mining Process

STEP 1: Input

Collect the source data

STEP 2: Analysis

Ask the questions – perform the queries

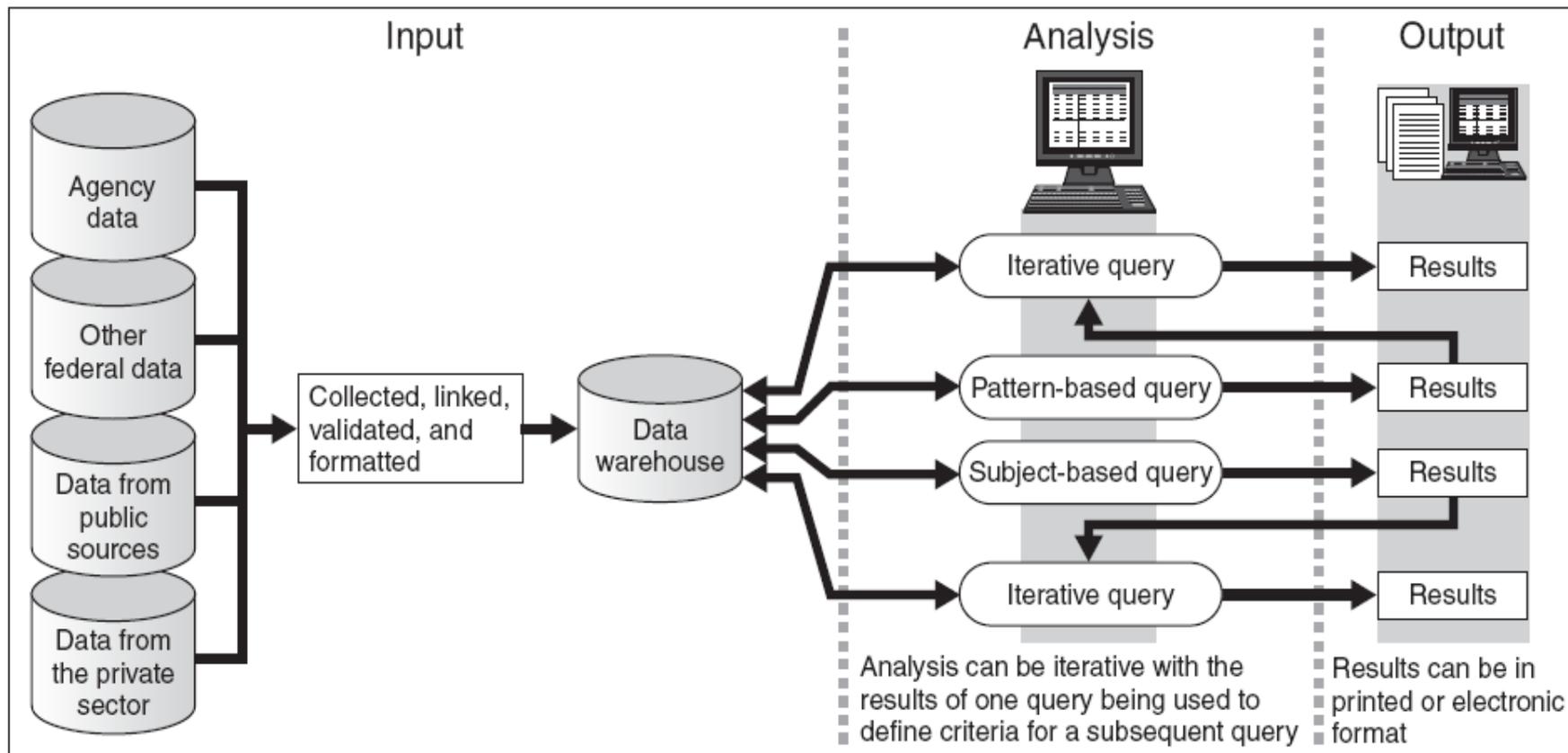
STEP 3: Output

Export the results



The Data Mining Process

The Data Mining Process



Source: GAO, adapted from Vipin Kumar and Mohammed J. Zaki.

HUD Data Sources



Section 8 Rental Assistance

Public Housing tenant-based (PIC)
Multi-Family project-based (TRACS)

Single Family FHA loans & REO

Data warehouse; SAMS;
CHUMS; SFIS

HOME Program and CDBG

IDIS

Data Sources from Other Federal Agencies

OPM – USPS

Federal employee wage data

FBI

Fugitive - Felons

FEMA

Hurricane disaster relief

Other Data Sources

Local Public Housing Agencies

Sec. 8 Rent Rolls
Accounting and general ledger
Inspection reports – HQS data

Mortgage Lenders

Insured loan data

Grantees

Draw downs and disbursements
Funding allocations and timeliness

History – Key Events

1996

Data mining efforts formally organized with creation of the nationwide CAATS network

2001

Computer Audit Specialist group established to provide full-time national data mining and CAATS support

Established Computer Data Matching Agreements

2002

FBI - Implemented Fugitive-Felon Initiative

2004

NCMEC - Implemented Operation HUDKid

2005 - 06

USPS – OPM - Implemented Operation FedRent

2006

FEMA – Hurricane disaster relief

History – Key Events

CAS Group Mission

The primary role of the CAS group is to provide timely CAATS related assistance and training in support of mission critical Audit and Investigations operations. Although this support is generally focused on obtaining data (both HUD internal and external entity) and converting it to a usable format consistent with the data analyses software tools available to the OIG, other types of support and services provided include:

- ✓ developing analytical and matching routines,
- ✓ creating unique, job specific software applications,
- ✓ generating statistical sampling plans,
- ✓ providing audit guidance,
- ✓ answering CAATS-related questions,
- ✓ testing hardware and software,
- ✓ working on special CAATS-related projects, and
- ✓ providing on-the-job CAATS and software training

Software - Tools of Choice

Getting the Data - ODBC

ACL
MS Access

Data Conversion

ACL Monarch
MS Excel PDF2TXT

Data Analyses

ACL
MS Excel
MS Access



Software – Tools of Choice

- ACL - the Workhorse of the Agency
 - ✓ Used to analyze large amounts of data (50 million + records)
 - ✓ Never faulted or choked when processing analytical/matching routines on large datasets
 - ✓ Robust built in functions
 - ✓ ODBC interface allows for user friendly data access and downloads
 - ✓ Unparalleled technical support

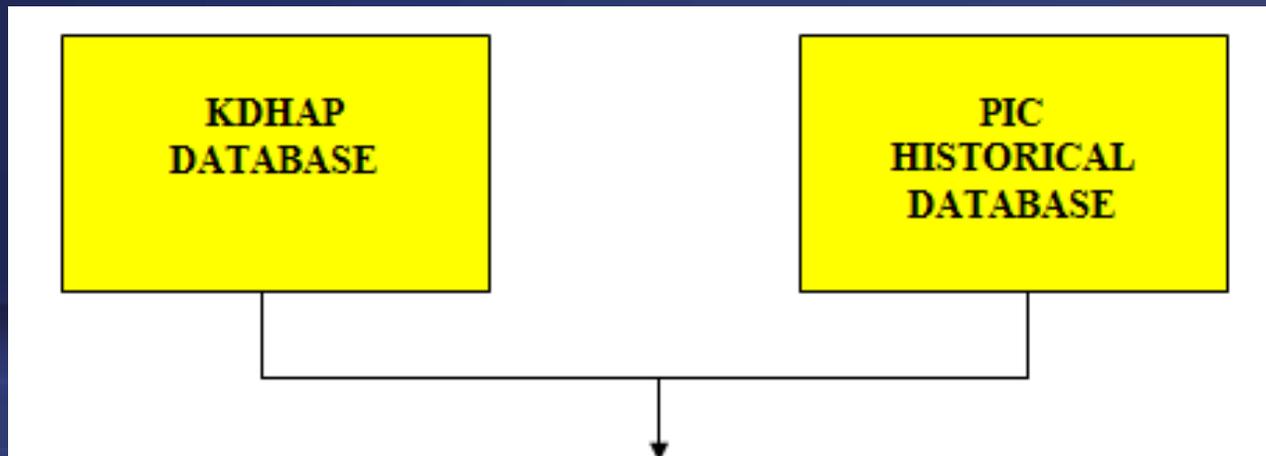
Software – Tools of Choice

- ACL – Complex Analytic Example

1. Data

16,000 records

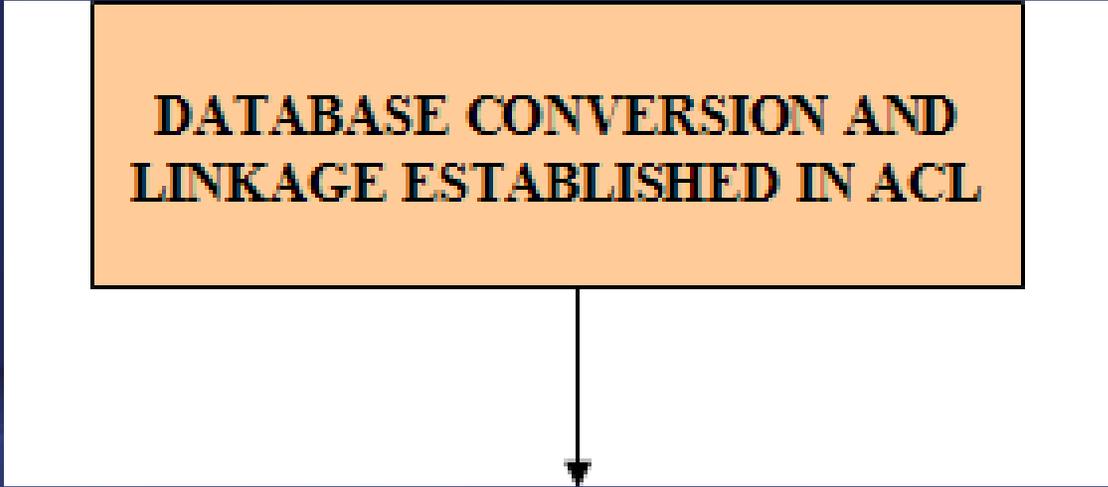
30 + million records



Software – Tools of Choice

- ACL – Complex Analytic Example

2. ACL Project Setup



```
graph TD; A[DATABASE CONVERSION AND LINKAGE ESTABLISHED IN ACL] --> B[ ]
```

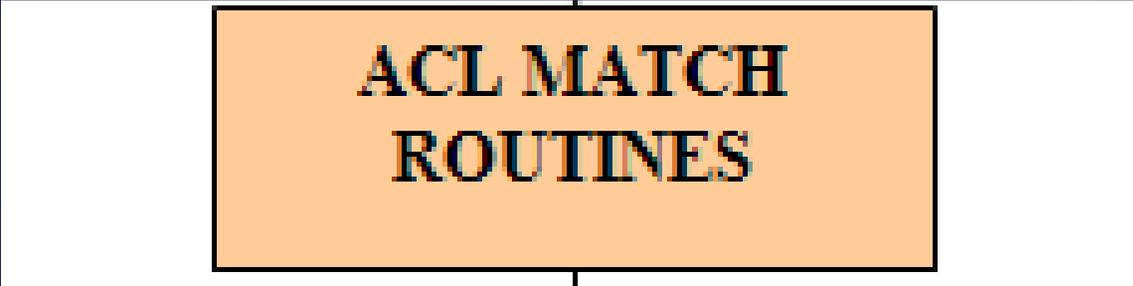
**DATABASE CONVERSION AND
LINKAGE ESTABLISHED IN ACL**

The diagram consists of a central orange rectangular box with a black border containing the text 'DATABASE CONVERSION AND LINKAGE ESTABLISHED IN ACL'. A vertical black line with a downward-pointing arrowhead extends from the bottom center of this box to the bottom center of a larger white rectangular area that serves as the background for the diagram.

Software – Tools of Choice

- ACL – Complex Analytic Example

3. ACL Functions



**ACL MATCH
ROUTINES**

Software – Tools of Choice

- ACL – Complex Analytic Example

4. ACL Functions

↓
SSN
IN KDHAP
NOT IN PIC

KDHAP – PIC SSN MATCH,
ACTION CODE 6 (EOP)
PRESENT IN PIC, KDHAP
LEASE IN EFFECT, NO POST
DISASTER PIC ACTIVITY

KDHAP – PIC SSN MATCH,
LAST PRE-DISASTER PIC
ACTION WAS CODE 6, KDHAP
LEASE INE EFFECT, NO POST
DISASTER PIC ACTIVITY

Software – Tools of Choice

- ACL – Complex Analytic Example

5. Match Routine Logic

1. FIND MATCHING SSN
2. ANALYZE PIC DATA, ID ACTION CODE 6
3. ANALYZE PIC DATA, ID POST DISASTER
ACTIVITY
4. ANALYZE KDHP DATA, ID LEASE IN
EFFECT
5. IF (2) , (4), (5) = YES & (3) = NO, ID
BENEFICIARY AS POTENTIALLY

Stage 1

(1 THRU 5 AND)

6. ANALYZE PIC DATA, ID LAST PRE-
DISASTER ACTION AS CODE 6
7. IF (5) = TRUE & (6) = YES, ID
BENEFICIARY AS HIGHLY LIKELY
INELIGIBLE

Stage 2

Software – Tools of Choice

- ACL – Complex Analytic Example

6. Results

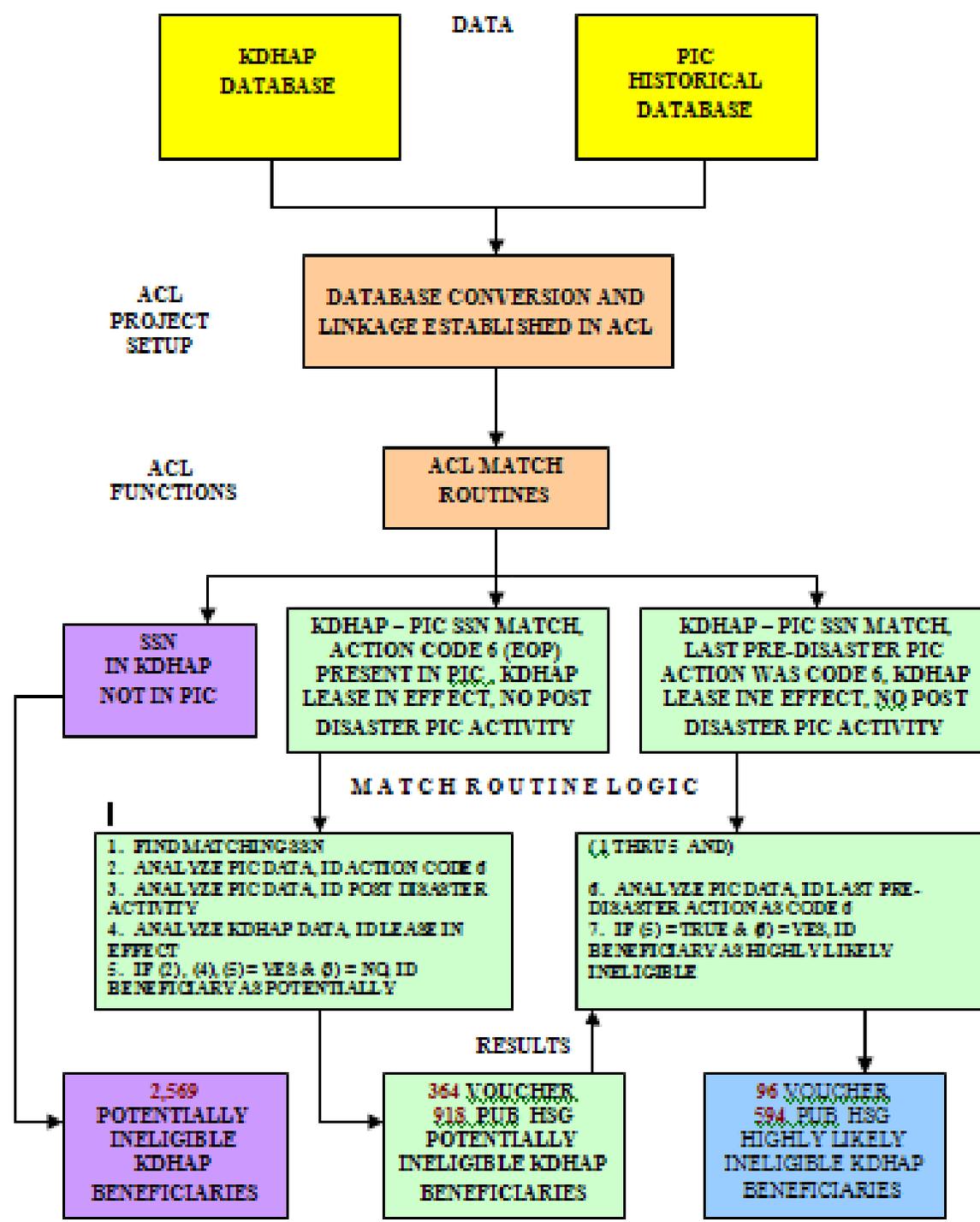
2,569
**POTENTIALLY
INELIGIBLE
KDHAP
BENEFICIARIES**

364 VOUCHER
918 PUB HSG
**POTENTIALLY
INELIGIBLE KDHAP
BENEFICIARIES**

Stage 1 Results

96 VOUCHER
594 PUB HSG
**HIGHLY LIKELY
INELIGIBLE KDHAP
BENEFICIARIES**

Stage 2 Results



Software – Tools of Choice

Statistical Sampling

U.S. Army Audit Agency – Statistical Sampling System

<http://www.hqda.army.mil/aaaweb/default.htm>

Defense Contract Audit Agency – EZQuant

<http://www.dcaa.mil/ezquant.htm>

Health & Human Services OIG – RAT-STATS

<http://oig.hhs.gov/organization/OAS/ratstat.html>

State Auditors Office, Texas – Statistical Toolbox

<http://www.sao.state.tx.us/Resources/tools/toolbox.html>

Data Mining In Action – The GOLD

FY 2006 Report Summary
Audits Using Data Mining Techniques



NUMBER OF REPORTS			
	ISSUED	DATA MINING USED	PERCENT
EXTERNAL	138	91	65.9%
INTERNAL	21	9	42.9%
TOTAL	159	100	62.9%

Data Mining In Action – The GOLD

FY 2006 Report Summary Data Mining Results By Region



Region	Total Audits	Audits with Data Mining Use	Percent of Total Audits	Return on Investment
Internal Reports 1 - 10	21	9	43%	\$250,000,000
External Reports				
1	11	6	55%	\$10,000,000
2	14	10	71%	\$23,088,767
3	14	11	79%	\$5,957,060
4	20	13	65%	\$66,154,853
5	20	14	70%	\$34,593,601
6	15	15	100%	\$28,413,436
7/8	18	14	78%	\$11,032,001
9/10	26	8	31%	\$3,350,000
Totals	159	100	63%	\$432,589,718

Data Mining In Action – The GOLD

Other Significant Data Mining Results Katrina Related Audits



- **Duplicate Benefits – LA State Program**
 - ✓ Identified potentially duplicative awards
 - ✓ Mined EGrant and Road Home Loan/Grant databases
- **Duplicate Benefits – Federal KDHAP - DVP**
 - ✓ Identified program participants who appear to be receiving duplicative disaster and non-disaster housing benefits
 - ✓ Mined HUD Disaster and local housing authority Section 8 payment data

Data Mining In Action – The GOLD

Other Significant Data Mining Results
Katrina Related Audits



- **SSN Validation Analyses – MS & LA State Programs**
 - ✓ Identified beneficiary SSN's associated with deceased persons
 - ✓ Beneficiary SSN's associated with red flags

Data Mining In Action – The GOLD

Other Significant Data Mining Results Investigations



- **Fugitive – Felon Initiative**
 - ✓ Over 5,600 arrests
- **Operation FEDRent**
 - ✓ Over 100 cases
- **National Center for Missing and Exploited Children**
 - ✓ Assisted in solving 5 cases

Conclusion

The HUD OIG has a long history of successful use of data mining techniques and continues to be at the forefront of progressive data analyses in the Federal Inspector General community.

Recent Journal of Public Inquiry Article Highlights the HUD OIG Data Mining Journey at:

Fall/Winter 2006/07

<http://www.ignet.gov/randp/jpi1.html>

Questions?



Contact Information

Cliff Cole

Computer Audit Specialist

U.S. Dept. of Housing and Urban Development

Office of the Inspector General

Email: ccole@hudoig.gov

(202) 441-1605