

# How to Identify Datasets Containing PCI, PII or Other Sensitive Information

*David Wade*

*CIO/EVP*

*Primerica, Inc.*

*david.wade@primerica.com*

*February 5, 2013*

*Session 12943*

# Primerica: Who We Are

- Life Insurance and Investments
- Middle-income families
- Main Street, not Wall Street
- Entrepreneurial business model
- 4.3 million lives insured
- 2 million investment clients
- Publicly traded “PRI”

# Infrastructure Architecture



Applications	Field Registration & Licensing	Field Compensation	Field Technology	Life Administration			
	Image Technology	Securities Administration		National Benefit Life			
	PeopleSoft / Cognos TM1	HR Administrative Systems		Application Infrastructure Support			
Presentation Layer	MS Internet Explorer, Firefox, Safari, Chrome	IBM Websphere Application Server v6.1 & v7.0	IBM Transaction Server v4.2 (CICS) Transaction Gateway v7.2	IBM TSO/E v1.12	MS .NET Framework v4.0	Tomcat v6	
Database Management	IBM DB2 v10 for z/OS	Oracle 11g	IBM DB2 for LUW v9.7	Microsoft SQL Server for Integrated 3 <sup>rd</sup> Party Application	VSAM – QSAM	NAS	MySQL v5.5
Information Security	IBM Security Server v1.12 (RACF)	Native UNIX & CA eTrust v8		Novell v6.5 eDirectory v8.7	MS Active Directory 2003		
Telecom	Diverse, AT&T Multi-link 155/100 Mbps Internet Access	AT&T OC48 SONET Metropolitan Area Network		Single & Multi-Mode Fiber & CAT 5e Wiring	Siemens 9751 PBXs	CISCO 65xx & 72xx Equipment	
Operating Systems Platforms	PR/SM IBM Mainframe Z/OS v1.12	Power VM IBM Unix-AIX v5.3/6.1	VMWare Vsphere 4 Microsoft Windows 2003/2008	SUSE V10 SP3 Edirectory v8.8 and Novell v6.5 Edirectory v8.7	Red Hat v6 Update 2		
Disaster Recovery	Mature Recovery Plans	End-to-End Application Testing		Network Connectivity	IBM Out of Region Recovery Center	Business Recovery Center - Hembree	

# DataSniff



What's in your mainframe?

Finding PCI, PII and other sensitive information in your legacy mainframe



# Mainframe Legacy

- 30+ years of legacy data
  - Production data classified manually
  - End user and Test data not classified
  
- Structured Data
  - VSAM
  - DB2 (includes user-owned tables)
  
- Unstructured Data (Disk and Tape)
  - QSAM / Sequential, GDG
  - PDS, PDSE
  - Packed Decimal
  - DFHSM ML1 / ML2

# Data Loss Prevention Initiative



Effective DLP Programs Utilize a Multitude of People, Process and Technology

## Data Governance

Policies and Standards

Identification

Risk Assessment

Classification

Architecture

Quality

Source: Ernst & Young, LLP, "Data Loss Prevention, Keeping your sensitive data out of the public domain"

# Data Loss Prevention Initiative

## Data Control

### Key Focus Areas for DLP Controls

→ Structured data ←

#### Data in Motion

Perimeter security

Network monitoring

Internet access control

Data collection and exchange

Messaging (Email, IM)

Remote access

#### Data in Use

Privileged user monitoring

Access/Usage monitoring

Data anonymisation

Use of test data

Data redaction

Export/Save control

#### Data at Rest

EndPoint security

Host Resident/Encryption

Mobile device protection

Network/Intranet storage

Physical media control

Disposal and destruction

→ Unstructured data ←

Source: Ernst & Young, LLP, "Data Loss Prevention, Keeping your sensitive data out of the public domain"

Complete your sessions evaluation online at [SHARE.org/SFEval](http://SHARE.org/SFEval)

# Data Loss Prevention Initiative



## Supporting Information Security Processes

Identity/Access Management	Security information/event management	Configuration management	Vulnerability management
Digital rights management	Incident response	Physical security	Training and awareness
Asset management	Data privacy/document protection	Employee screening and vetting	Third-party management and assurance
Business continuity	Disaster recovery	Regulatory compliance management	Change management/SDLC

Source: Ernst & Young, LLP, "Data Loss Prevention, Keeping your sensitive data out of the public domain"





# If You Can't Measure It...

# ...You Can't Manage It!

# The Task at Hand

## Classify all mainframe data files

- Who owns them?
- What do they contain?
- Where is the confidential, PII, PCI or other sensitive information?

# DataSniff by XBridge Systems



- Data Discovery Tool
  - Runs on the mainframe
  - Automated, native scanning of mainframe data
  - Schedule by Media Type
  
- Analytical Engine
  - Appliance or Windows-based
  
- Browser-based User Interface
  - Schedule scans
  - Manage reporting

# DataSniff Evaluation

- Identify and secure
  - Payment Card Numbers (PCN)
  - Personally Identifiable Information (PII) especially Social Security/Tax ID numbers, home address
- Support all Mainframe data structures
  - DB2 databases
  - VSAM KSDS & ESDS
  - QSAM, GDG files on both disk and tape
  - PDS & PDSE
  - DFHSM ML1 & ML2 on tape
  - Packed decimal data

# Features Evaluated

- Locate SSN/PCN format data in DB2 tables, QSAM/VSAM, & PDS(E)s
- Selection of data using regular expressions
- User control of mainframe resources
- HSM re-migration
- User-written regular expressions
- Results reporting
- Impact on mainframe performance

# Management Console



DLP Scan Status - Microsoft Internet Explorer provided by Primerica Financial Services

http://172.16.33.61/DS/DLP/ScanList.aspx?

File Edit View Favorites Tools Help

Google Search More >>

Sign In

DLP Scan Status

Status	Table Name	Disposition	Read	Analyzed	Hits	Errors	Status Date
✓	AH023.FEB07.PAYROLL	TBD	139	139	101	0	3:34 PM
✓	AH023.JAN10.NAME	HitsBelowThreshold	515	515	8	0	3:31 PM
✓	AH023.APR07.PAYROLL	TBD	160	160	113	0	3:31 PM
✓	AH023.SEP09.PAYROLL	TBD	574	574	338	0	3:31 PM
✓	AH023.JUN08.PAYROLL2	TBD	831	831	505	0	3:31 PM
✓	AH023.NAME0205	TBD	1,044	1,044	19	0	3:31 PM
✓	AH023.ISPF.PROFILE(*.35)	HitsBelowThreshold	973	973	4	0	3:28 PM
✓	.....AH023.ISPF.PROFILE(DSQEEDIT)	HitsBelowThreshold	71	71	2	0	3:28 PM
✓	AH023.APR12.PAYROLL	TBD	671	671	399	0	3:28 PM
✓	AH023.APR12.NAME	TBD	671	671	19	0	3:28 PM

Record #	Run #	Archived	Column Name	Message	Notes
599	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....2936	...
525	1	<input type="checkbox"/>	Dataset record (pos: 111)	PFS SSN found.....4965	...
525	1	<input type="checkbox"/>	Dataset record (pos: 189)	PFS SSN found.....4965	...
524	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....9365	...
522	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....2194	...
500	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....8062	...
463	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....5398	...
462	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....3188	...
454	1	<input type="checkbox"/>	Dataset record (pos: 189)	PFS SSN found.....1068	...
384	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....9462	...
318	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....8887	...
272	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....6109	...
267	1	<input type="checkbox"/>	Dataset record (pos: 0)	PFS SSN found.....1800	...

Local intranet 100%

# Management Console



DLP Scan Status - Microsoft Internet Explorer provided by Primerica Financial Services

http://172.16.33.61/DS/DLP/ScanList.aspx

File Edit View Favorites Tools Help

Google Search More >>

DLP Scan Status

Table or Dataset Name	Table Name	Disposition	Read	Analyzed	Hits	Errors	Status Date
LPAR1 DB2A - AT078	AT078 DB2 TABLES	LPAR1 DB2A - AT078	2,000	0	0	0	Randy Gross 5/6/2012
LPAR1 DB2A - AT086	AT086 DB2 TABLES	LPAR1 DB2A - AT086	2,000	0	0	0	Randy Gross 5/6/2012
LPAR1 DB2A - AT088	AT088 DB2 TABLES	LPAR1 DB2A - AT088	2,000	0	0	0	Randy Gross 5/6/2012
LPAR1 DB2A - AT097	AT097 DB2 TABLES	LPAR1 DB2A - AT097	2,000	0	0	0	Randy Gross 5/6/2012
LPAR1 DB2A - AT650	AT650 DB2 TABLES	LPAR1 DB2A - AT650	2,000	379	379	138	51 Randy Gross 5/6/2012
LPAR1 DB2A - AT659	AT659 DB2 TABLES	LPAR1 DB2A - AT659	2,000	5,517	5517	0	15 Randy Gross 5/7/2012
LPAR1 DB2A - AT812	AT812 DB2 TABLES	LPAR1 DB2A - AT812	2,000	995	995	28	305 Randy Gross 5/7/2012
LPAR1 DB2A - AT994	AT994 DB2 TABLES	LPAR1 DB2A - AT994	2,000	206	206	31	34 Randy Gross 5/7/2012
LPAR1 DB2A - AH012	AH012 DB2 TABLES	LPAR1 DB2A - AH012	1,000	43	43	8	0 Randy Gross 5/6/2012
LPAR1 POC DATASETS	LPAR1 POC DATASETS	LPAR1 POC QSAMVSAM	2,000	59	59	47	12 Randy Gross 5/6/2012
LPAR1 POC TABLES	LPAR1 POC TABLES	LPAR1 DB2A - DB2PROD	1,000	5	5	3	0 Randy Gross 5/6/2012
LPAR1 TSO AH023	TSO datasets for AH023	VSAM/QSAM in LPAR1	10,000	365	365	207	15 Randy Gross 5/9/2012
LPAR7 - POC Prod Datasets	AWCPROD.APECS.PHISTORY.G1...	LPAR 7 VSAM/QSAM Datasets	5,000	9	9	7	0 Randy Gross 2:47 PM

  

Status	Table Name	Disposition	Read	Analyzed	Hits	Errors	Status Date
✖	AWCPROD.APECS.PHISTORY.G1625V00	TBD	5,000	5,000	24,585	0	2:47 PM
✖	AWCPROD.APECS.HISTORY	TBD	5,000	5,000	23,712	0	2:47 PM
✖	FANQPROD.AUM.FAF.ACTIVITY.G0001V00	TBD	5,000	5,000	14,073	0	2:47 PM
✖	FANQPROD.AUM.FAF.MASTER.G0001V00	TBD	5,000	5,000	12,142	0	2:47 PM
✖	AWCPROD.IRS.VS.YEINDEX	TBD	5,000	5,000	4,847	0	2:47 PM
✖	FVF2.NBL.CLIM.INSURED.FILE	TBD	5,000	5,000	4,216	0	2:47 PM
✖	FANQPROD.Q8000TRV.COMMFILE.G0002V00	TBD	35	35	144	0	2:47 PM
✔	FVF2.NBL.CMS.MAINT.COMSTRAN	NoHits	1	1	0	0	2:47 PM
✔	AWCPROD.IRS.VS.YEMAST99	NoHits	1	1	0	0	2:47 PM

  

LPAR7 DB2A DB2PROD	All DB2PROD tables in DB2...	LPAR7 DB2A DB2PROD Schema	250	1,444	1444	0	1,444	Randy Gross 2:04 PM
LPAR7 QSAMVSAM Scan of Migrated Dataset	LPAR7 Demo Migrated Datas...	LPAR 7 VSAM/QSAM Datasets	100	2	2	2	0	Randy Gross 2:04 PM
LPAR7 TSO	TSO Datasets in LPAR7	LPAR 7 VSAM/QSAM Datasets	100	102,109	4247	749	1,026	Randy Gross 2:04 PM
LPAR7 VSAM/QSAM Scan	POC test datasets - TSO...	LPAR 7 VSAM/QSAM Datasets	250	253	253	42	125	Randy Gross 2:04 PM

Page size: 50 29 items in 1 pages

Scan List is current as of 5/11/2012 2:48:18 PM

v:4.2.54.12229 © 2010 - 2012 Xbridge Systems, Inc. All rights reserved. Contact Us

Local Intranet 100%

# Management Console



DLP Scan Status - Microsoft Internet Explorer provided by Primerica Financial Services

http://172.16.33.61/DS/DLP/ScanList.aspx

File Edit View Favorites Tools Help

Google Search More >>

Sign In

DLP Scan Status

Table or Dataset Name	Logs
LPAR1 DB2A - AT994	AT994 DB2 TABLES
LPAR1 DB2A -- AH012	AH012 DB2 TABLES
LPAR1 POC DATASETS	LPAR1 POC DATASETS
LPAR1 POC TABLES	LPAR1 POC TABLES
LPAR1 TSO AH023	TSO datasets for AH023
LPAR7 - POC Prod Datasets	AWCPROD.APECS.PHISTORY.G1...

Table or Dataset Name	Disposition	Read	Analyzed	Hits	Errors	Status Date
AWCPROD.APECS.PHISTORY.G1625V00	TBD	5,000	5,000	24,585	0	2:47 PM
AWCPROD.APECS.HISTORY	TBD	5,000	5,000	23,712	0	2:47 PM
FANGPROD.AUM.FAF.ACTIVITY.G0001V00	TBD	5,000	5,000	14,073	0	2:47 PM
FANGPROD.AUM.FAF.MASTER.G0001V00	TBD	5,000	5,000	12,142	0	2:47 PM
AWCPROD.IRS.VS.YEINDEX	TBD	5,000	5,000	4,847	0	2:47 PM
FVF2.NBL.CLM.INSURED.FILE	TBD	5,000	5,000	4,216	0	2:47 PM

Table of Dataset Name

Status

Table Name

Disposition

Read

Analyzed

Hits

Errors

Status Date

AWCPROD.APECS.PHISTORY.G1625V00

TBD

5,000

5,000

24,585

0

2:47 PM

Logs

Hits

Errors

Created	Run #	Archived	Log Entry
2:47 PM	1	<input type="checkbox"/>	Completed (subject to limit) analysis of AWCPROD.APECS.PHISTORY.G1625V00. Analyzed 5000 records in 34.390625 sec. Found 4999 records with potential policy violations.
2:47 PM	1	<input type="checkbox"/>	Reached HitDetailsLimit and all remaining hit details will not be recorded. Hits will continue to be counted.
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@751 size:21:11-49 col:(452/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@692 size:11:11-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@664 size:11:11-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@443 size:11:11-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@419 size:17:17-17 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@117 size:9:7-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@107 size:11:11-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@101 size:11:11-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@95 size:11:11-11 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@89 size:13:13-13 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@63 size:20:11-21 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Searched EBCDIC data and found possible Packed Decimal: PD@48 size:11:11-13 col:(499/500.99%).
2:47 PM	1	<input type="checkbox"/>	Initializing Analysis of AWCPROD.APECS.PHISTORY.G1625V00 for scan "LPAR7 - POC Prod Datasets", on host "LPAR 7 VSAM/QSAM Datasets", run number: 1.
2:47 PM	1	<input type="checkbox"/>	Analysis will use policies dated 5/11/2012 10:13:51 AM.

Done

Local intranet 100%





# Successes

- ✓ PCN/SSN data located with minimum of “false positives”
- ✓ User-defined selection filters
- ✓ Scanned all required data structures: DB2, VSAM, Sequential / QSAM, PDS/PDSE, HSM migrated data sets
- ✓ Scanned packed decimal structures
- ✓ Controlled use of scarce mainframe resources (tape, disk)
- ✓ HSM migrated files automatically recalled and re-migrated to original migration level (ML1, ML2)
- ✓ Scanned large numbers of data stores with no performance impact

# Issues

## Resolved during evaluation

- ✓ Dynamic allocation of HSM migration control dataset caused contention with HSM CDS reorgs.
- ✓ Excessive runtimes for PDS/PDSE scans. Fix has been developed.

# Performance

- Quick data analysis; select group of datastores

	<b>Test LPAR</b>	<b>Prod LPAR</b>
Total Files Scanned	34,960	225,195
By File Type:		
PDS/E	669	1,804
QSAM	5,049	176,852
VSAM	28,447	37,112
DB2 Tables	795	9,428
By Media Type:		
DASD (non-migrated)	32,169	67,812
Tape	115	41,905
Migrated (ML1/ML2)	2,676	115,478

- Minimal mainframe resource usage
- User-controlled mainframe resources:
  - Tape
  - HSM recalls

# Custom Regular Expressions

- Identify SSNs with specific state codes
- Addresses and zip+4's in the form "xxxxx-xxxx".
- Example of SSN regular expression

```
(?:[0-6]\d{2}|7[0-6]\d|77[0-2])(?: {0,2}(?(-)(?:-\ {0,2})))?(?:\d{2})(?: {0,2}(?(-)(?:-\ {0,2})))?(?:\d{4})
```

```
((253|263|260)[0-9]{6})|((253|260|263)-[0-9]{2}-[0-9]{4})
```

# Conclusion

- ✓ Data identification and classification
- ✓ Mainframe data at rest
- ✓ Easy to use
- ✓ Quick and efficient
- ✓ Little or no impact to the operational environment

*DataSniff is a valuable part of Primerica's comprehensive DLP program.*



*XBridge Systems, Inc.*  
*Mountain View, CA*  
[www.xbridgesystems.com](http://www.xbridgesystems.com)



*Primerica, Inc.*  
*Duluth, GA*  
[www.primerica.com](http://www.primerica.com)