Essentials for Data Masking for Siebel

Agenda

- The Latest on Data Privacy
- Understanding Data Governance
- The Easiest Way to Expose Private Data
- Understanding the Insider Threat
- Considerations for a Privacy Project
- Success Stories

Does This Define Your Privacy Strategy?

The Latest on Data Privacy

- 2007 statistics
  - **$197**
    - Cost to companies per compromised record
  - **$6.3 Million**
    - Average cost per data breach “incident”
  - **40%**
    - % of breaches where the responsibility was with Outsourcers, contractors, consultants and business partners
  - **217 Million**
    - TOTAL number of records containing sensitive personal information involved in security breaches in the U.S. since 2005

* Sources”: Ponemon Institute, Privacy Rights Clearinghouse, 2007

Did You Hear?

- UK gov’t suffered a massive data breach in Nov. 07
  - HMRC (Her Majesty's Revenue & Customs) UK equivalent to IRS
- Lost 2 disks containing personal information on 25 million people (ALMOST ½ of UK population!)
- Information has a criminal value of $3.1 Billion
- No reported criminal activity to date

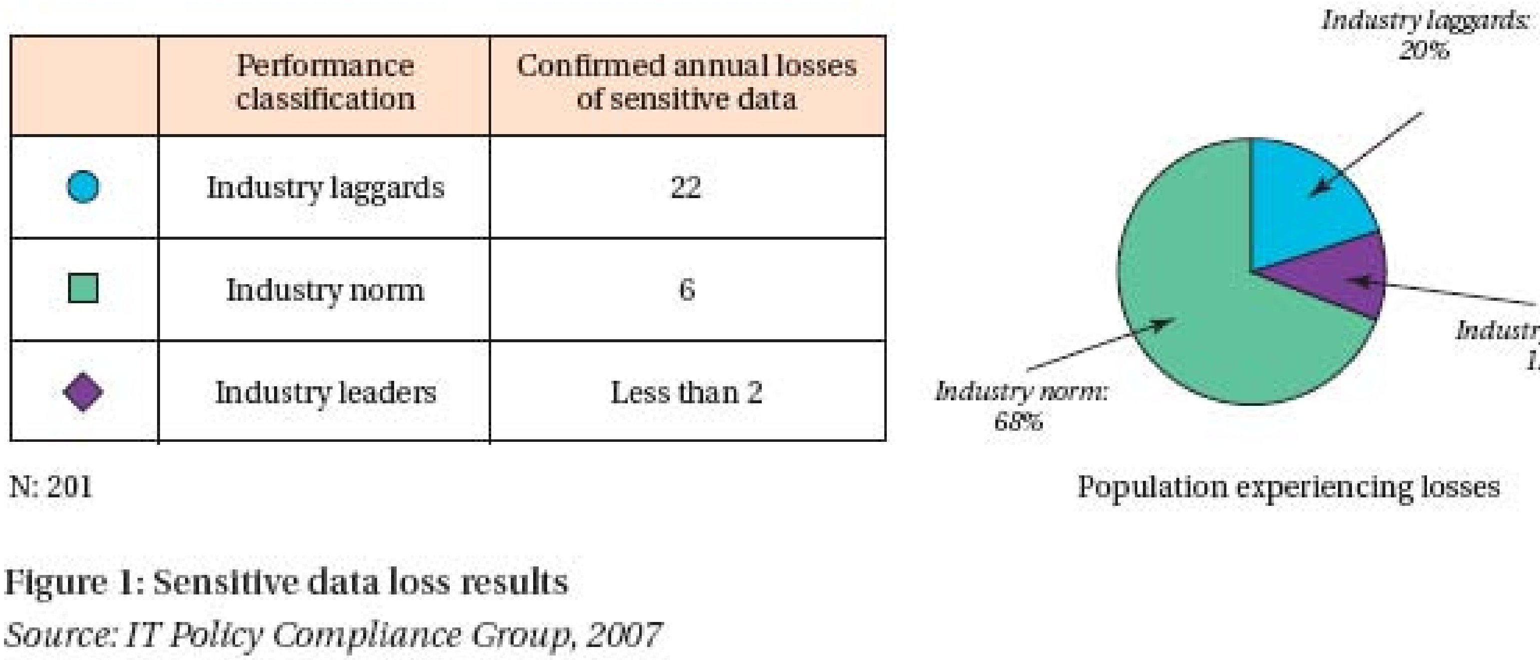
How much is personal data worth?

- Credit Card Number With PIN - $500
- Drivers License - $150
- Birth Certificate - $150
- Social Security Card - $100
- Credit Card Number with Security and Expiration Date - $7-$25
- Paypal account Log-on and Password - $7

Representative asking prices found recently on cybercrime forums.

Source: USA TODAY research 10/06

Where do F1000 Corporations Stand today?



	Performance classification	Confirmed annual losses of sensitive data
●	Industry laggards	22
🟩	Industry norm	6
◆	Industry leaders	Less than 2

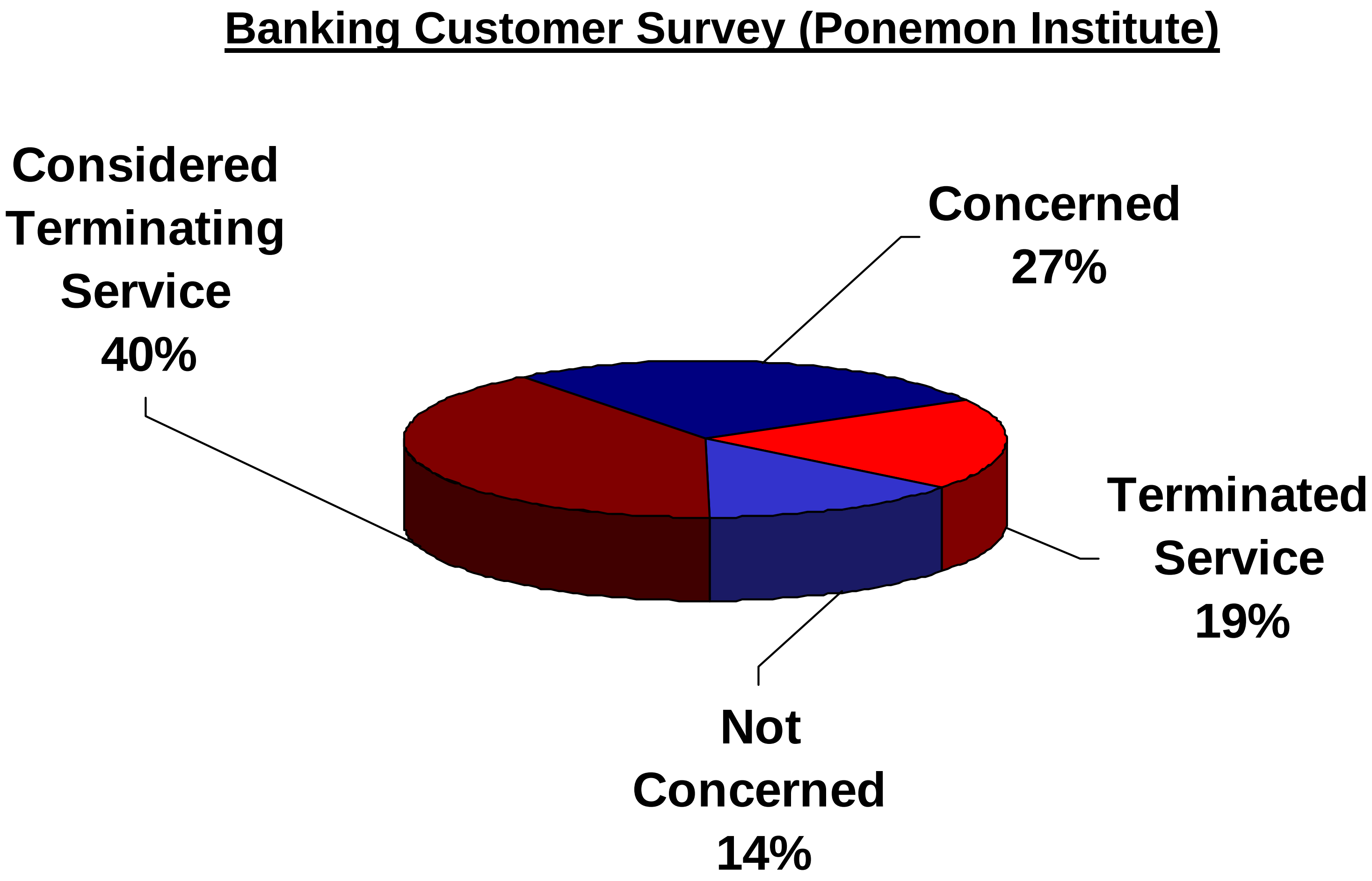
N: 201

Population experiencing losses

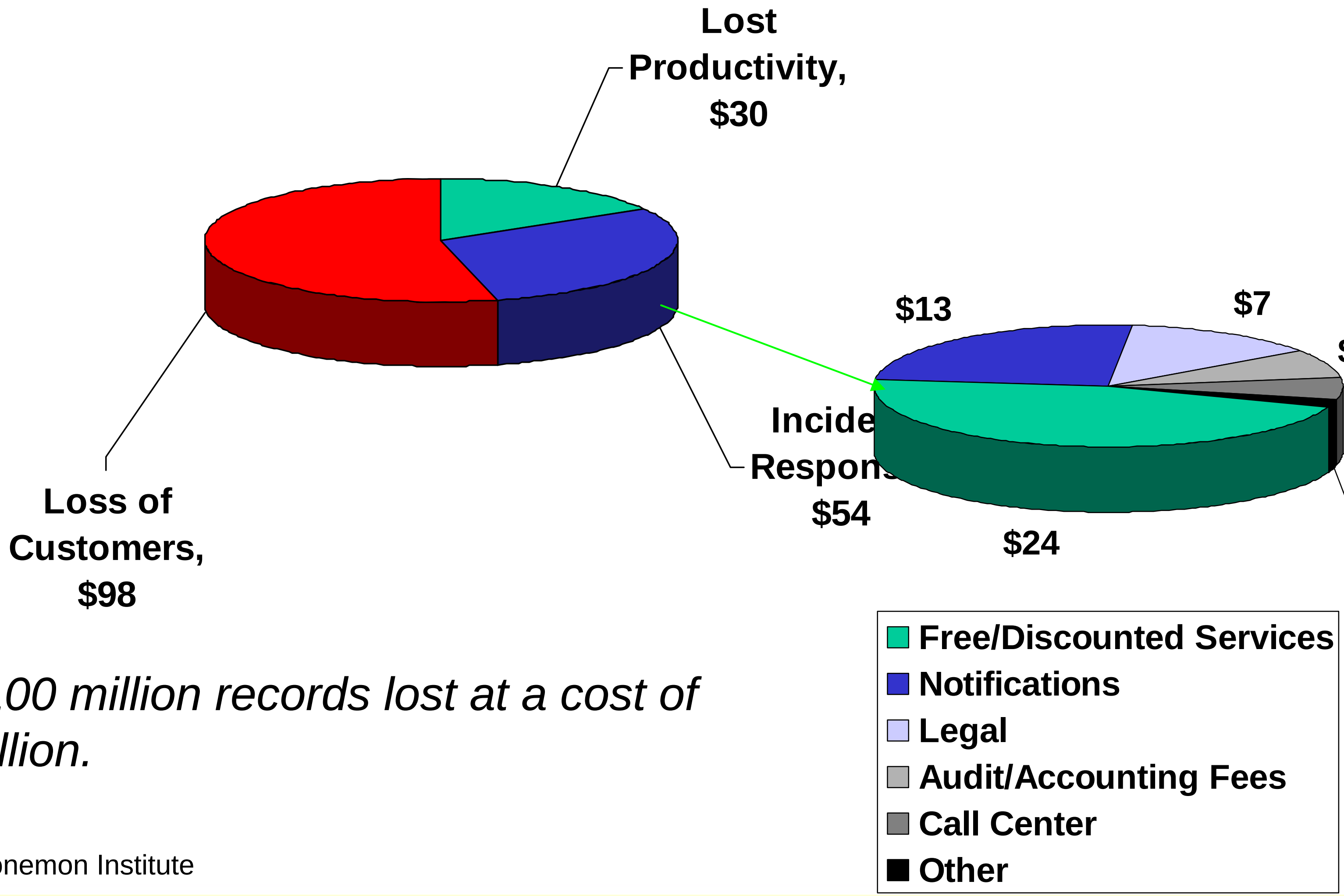
Figure 1: Sensitive data loss results

Source: IT Policy Compliance Group, 2007

Consumer Reaction



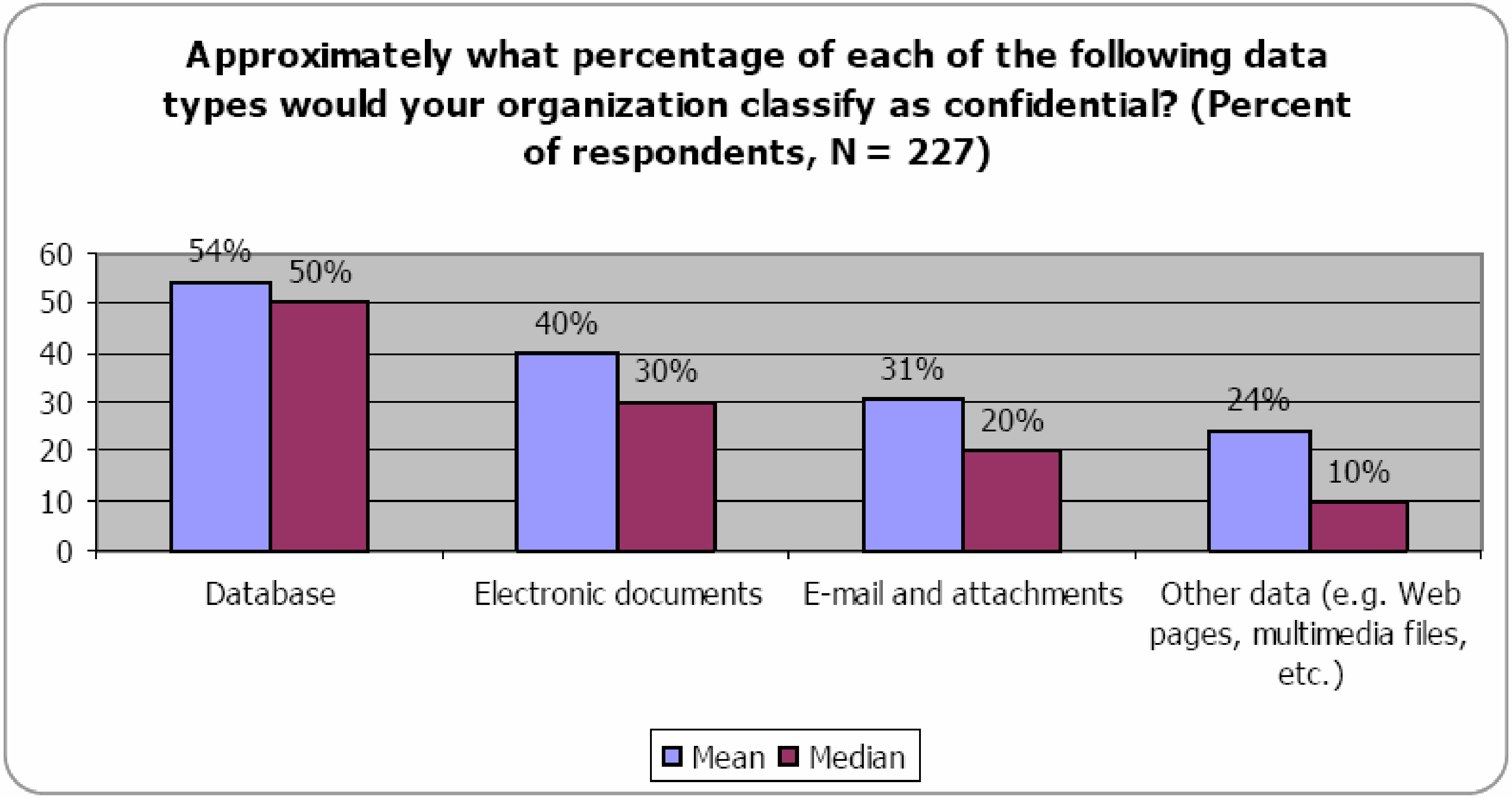
Cost to Company per Missing Record: $197



Over 100 million records lost at a cost of $16 Billion.

Source: Ponemon Institute

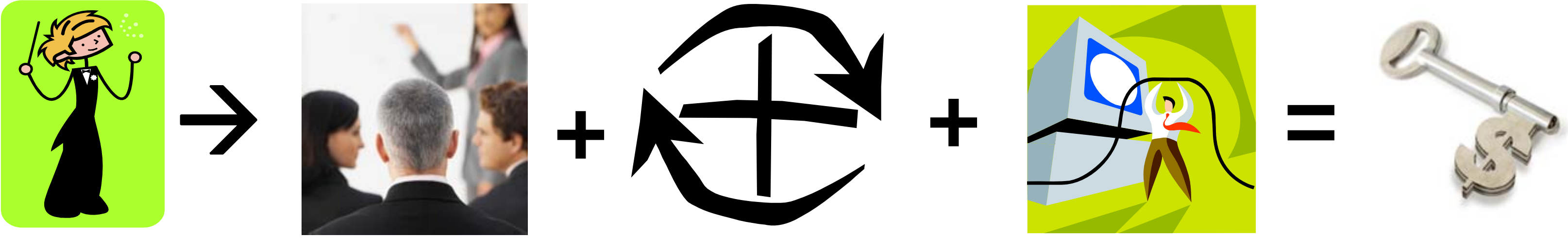
Where is Confidential Data Stored?



[1] ESG Research Report: *Protecting Confidential Data*, March, 2006.

What is Data Governance?

Data Governance is the political process of changing organizational behaviour to enhance and protect data as a strategic enterprise asset



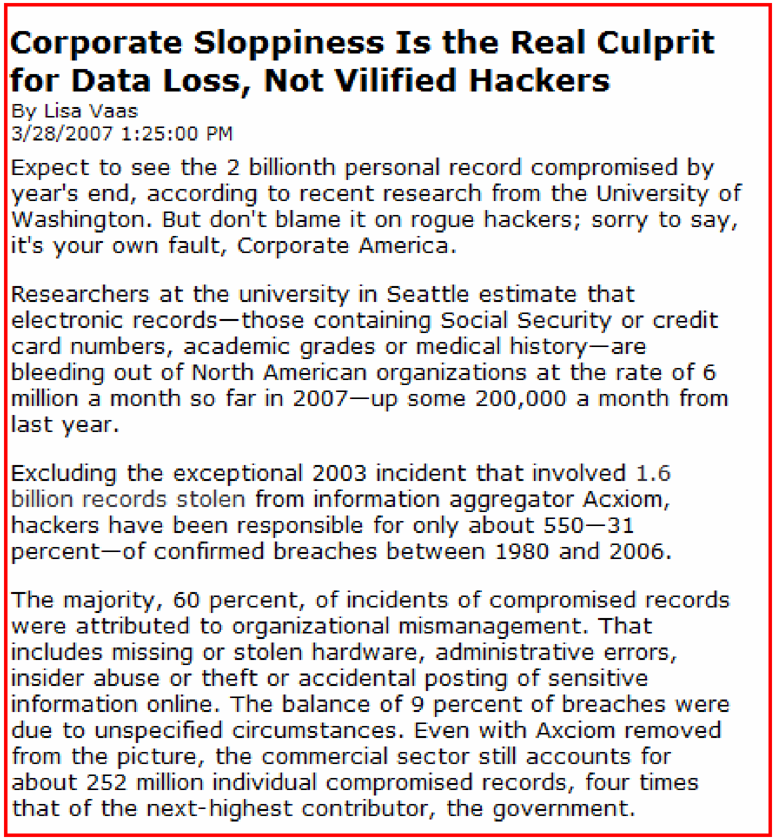
Implementing Data Governance is a fundamental change to the methods & rigor both <u>Business</u> and <u>Information Technology</u> use to define, manage and use of data

The core objectives of a governance program are:

- **Guide information management decision-making**
- **Ensure information is consistently defined and well understood**
- **Increase the use and trust of data as an enterprise asset**
- **Improve consistency of projects across an enterprise**

Without Data Governance…

- **People make mistakes…**
- **Those mistakes more commonly result in losses than hackers…**
- **Those losses effect every aspect of IT and business**
- **But data is still an abstract concept and governance needs technology to be improved…**



Corporate Sloppiness Is the Real Culprit for Data Loss, Not Vilified Hackers

By Lisa Vaas
3/28/2007 1:25:00 PM

Expect to see the 2 billionth personal record compromised by year's end, according to recent research from the University of Washington. But don't blame it on rogue hackers; sorry to say, it's your own fault, Corporate America.

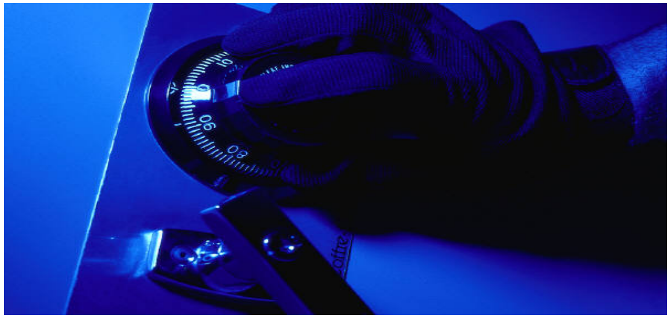
Researchers at the university in Seattle estimate that electronic records—those containing Social Security or credit card numbers, academic grades or medical history—are bleeding out of North American organizations at the rate of 6 million a month so far in 2007—up some 200,000 a month from last year.

Excluding the exceptional 2003 incident that involved 1.6 billion records stolen from information aggregator Acxiom, hackers have been responsible for only about 550—31 percent—of confirmed breaches between 1980 and 2006.

The majority, 60 percent, of incidents of compromised records were attributed to organizational mismanagement. That includes missing or stolen hardware, administrative errors, insider abuse or theft or accidental posting of sensitive information online. The balance of 9 percent of breaches were due to unspecified circumstances. Even with Axciom removed from the picture, the commercial sector still accounts for about 252 million individual compromised records, four times that of the next-highest contributor, the government.

Why the focus on Data Governance?

- **Regulatory Compliance**
  - Consumer privacy
  - Financial Integrity
- **Intellectual Property Theft**
  - Confidential manufacturing processes
  - Financial information
  - Customer lists
  - Digital source code
  - Marketing strategies
  - Research data
- **Economic Espionage**
  - Trade secret



State sues global management consulting company over stolen backup tape. Unencrypted tape contained personal information on 58 taxpayers and nearly 460 state bank accounts.

Over 45 million credit and debit card numbers stolen from large retailer. Estimated costs $1bn over five years (not including lawsuits). $117m costs in 2Q '07 alone.

Who is breaking in and how?

- Hackers take advantage of:
  - Vulnerable network or infrastructure security, poor server or database security standards
- Thieves steal:
  - Physical medium (backup tapes or disk drives), User-ids and passwords
- Employees or Business Partners have authorization to servers, databases and data

•No security strategy is 100% hacker proof

•Most security breaches occur internally

- ***•Accidental opening of firewall***
- ***•Stealing user-id, password from authorized user***
- ***•Have the authority to access the server or the data***

What is Done to Protect Data Today?

- Production “Lockdown”
  - Physical entry access controls
  - Network, application and database-level security
  - Multi-factor authentication schemes (tokens, biometrics)
- Unique challenges in Development and Test
  - Replication of production safeguards not sufficient
  - Need “realistic” data to test accurately

The Easiest Way to Expose Private Data … Internally with the Test Environment

- 70% of data breaches occur internally (Gartner)
- Test environments use personally identifiable data
- Standard Non-Disclosure Agreements may not deter a disgruntled employee
- What about test data stored on laptops?
- What about test data sent to outsourced/overseas consultants?
- How about Healthcare/Marketing Analysis of data?
- Payment Card Data Security Industry Reg. 6.3.4 states, **“Production data (real credit card numbers) cannot be used for testing or development”**

*** The Solution is Data De-Identification ***

The Latest Research on Test Data Usage

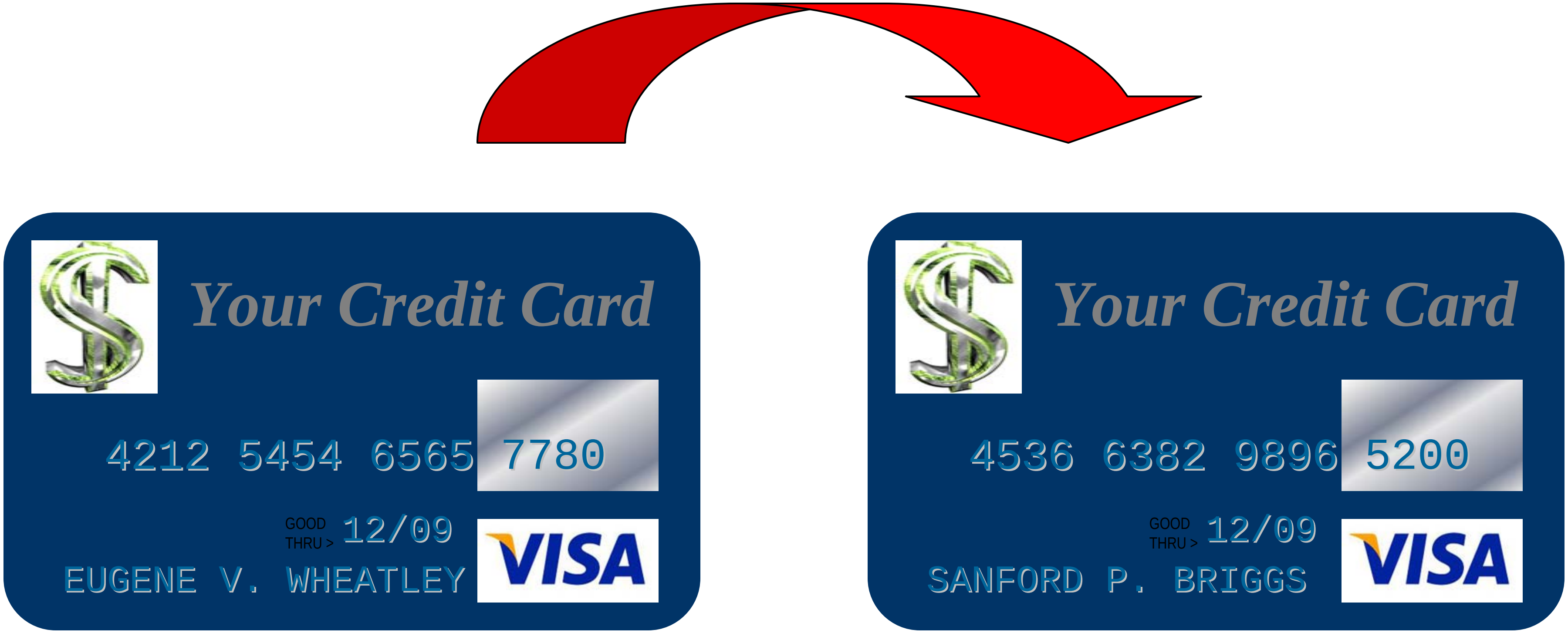
- Overall application testing/development
  - 62% of companies surveyed use actual customer data instead of disguised data to test applications during the development process
  - 50% of respondents have no way of knowing if the data used in testing had been compromised.
- Outsourcing
  - 52% of respondents outsourced application testing
  - 49% shared live data!!!
- Responsibility
  - 26% of respondents said they did not know who was responsible for securing test data

Source: The Ponemon Institute. The Insecurity of Test Data: The Unseen Crisis

What is Data De-Identification?

- AKA data masking, depersonalization, desensitization, obfuscation or data scrubbing
- Technology that helps conceal real data
- Scrambles data to create new, legible data
- Retains the data's properties, such as its width, type, and format
- Common data masking algorithms include random, substring, concatenation, date aging
- Used in Non-Production environments as a Best Practice to protect sensitive data

Masking is transparent to the outside world



Card Holder and Card Number have been masked

Failure Story – A Real Life Insider Threat

- 28 yr. old Software Development Consultant
- Employed by a large Insurance Company in Michigan
- Needed to pay off Gambling debts
- Decided to sell Social Security Numbers and other identity information pilfered from company databases on 110,000 Customers
- Attempted to sell data via the Internet
  - Names/Addresses/SS#s/birth dates
  - 36,000 people for $25,000
- Flew to Nashville to make the deal with…..
- The United States Secret Service (Ooops)

Results:

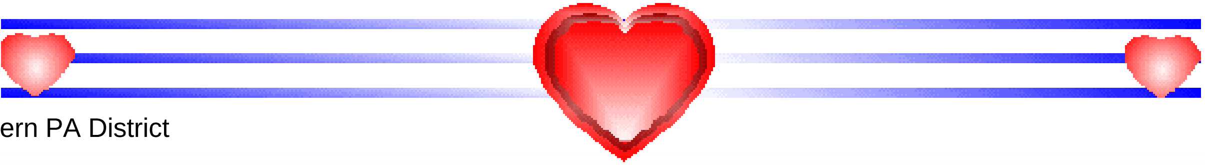
- **Sentenced to 5 Years in Jail**
- **Order to pay company $520,000**

The Top 3 Reasons Why Insiders Steal Data

1. Greed

2. Revenge

3. Love



Source: US Attorney General’s Office, Eastern PA District

How is Risk of Exposure being Mitigated?

- No laptops allowed in the building
- Development and test devices
  - Do not have USB
  - No write devices (CD, DVD, etc.)
- Employees sign documents
- Off-shore development does not do the testing
- The use of live data is ‘kept quiet’

Encryption is not Enough

- DBMS encryption protects DBMS theft and hackers
- Data decryption occurs as data is retrieved from the DBMS
- Application testing displays data
  - Web screens under development
  - Reports
  - Date entry/update client/server devices
- If data can be seen it can be copied
  - Download
  - Screen captures
  - Simple picture of a screen

Strategic Issues for Implementing Data Privacy

ALIGN

Data Masking Considerations

- Establish a project leader/project group
- Determine what you need to mask
- Understand Application and Business Requirements
- Top Level Masking Components
- Project Methodology

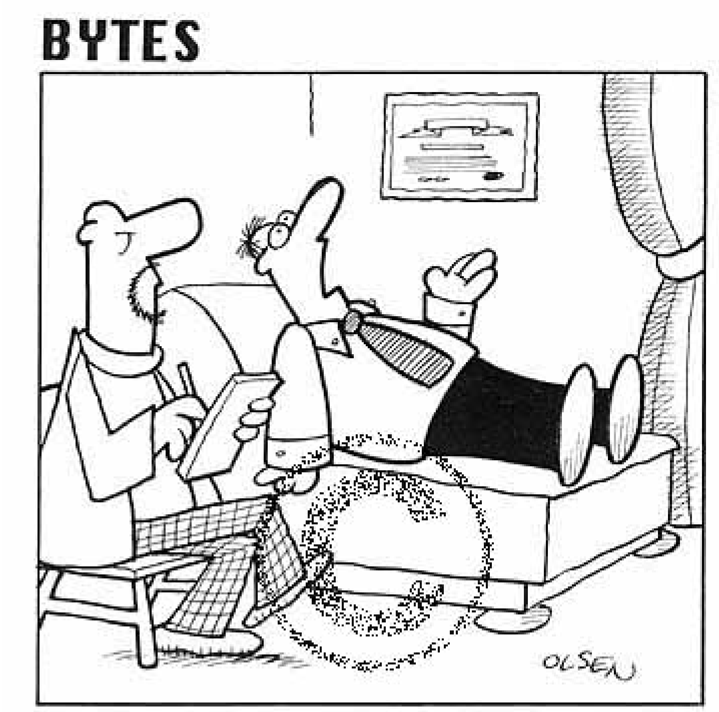
Data Masking Consideration – Step 1

- **Establish a Project Leader/Group**
  - Many questions to be answered/decisions to be made
  - Project Focus
  - Inter-Departmental Cooperation
  - Use for additional Privacy Projects

Data Masking Consideration – Step 2

- **Determine what you need to mask**
  - Customer Information
  - Employee Information
  - Company Trade Secrets
  - Other

Data Masking Consideration – Step 3



"My computer doesn't understand me !"

- **Understand Application and Business Requirements**
  - Where do applications exist?
  - What is the purpose of the application(s)?
  - How close does replacement data need to match the original data?
  - How much data needs to be masked?

Data Masking Consideration – Step 4 Masking Components (Top Level)

- Masking is not simple!
  - Many DBMS
  - Legacy Files
  - Multiple platforms
- Needs to fit within existing processes
- Not a point solution – consider the enterprise
- Not a one time process

Component A - Consistency

- Masking is a repeatable process
- Subsystems need to match originating
- The same mask needs to be applied across the enterprise
  - Predictable changes
  - Random change will not work
- Change all ‘Jane’ to ‘Mary’ again and again

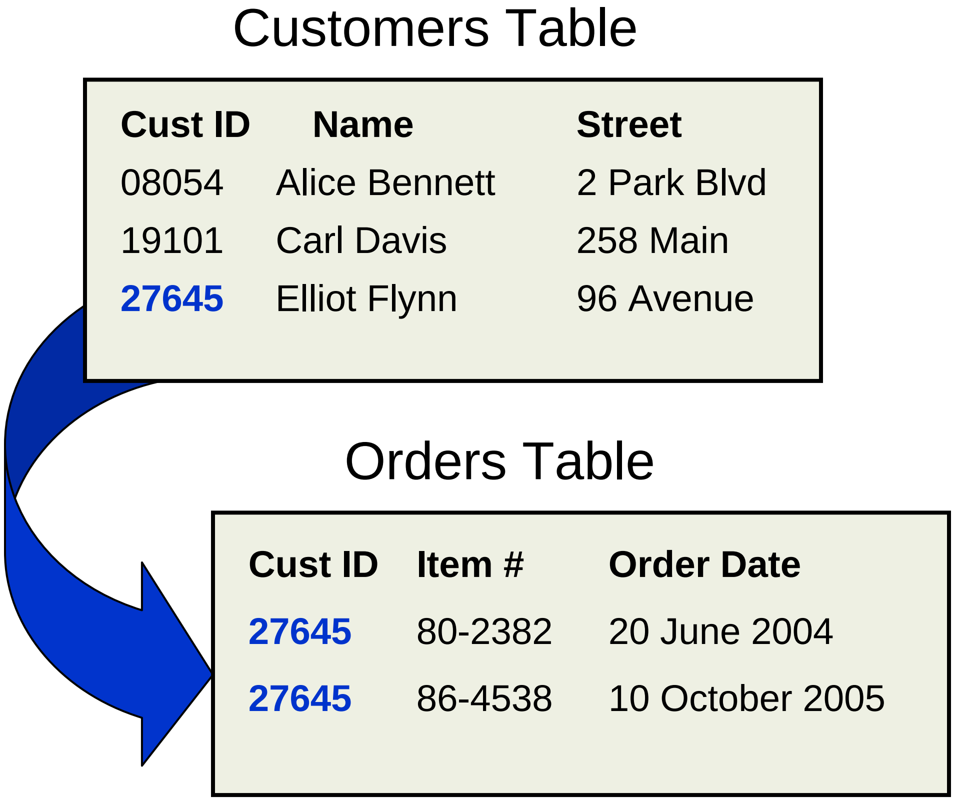
Example: First and Last Name

- Direct Response Marketing, Inc. is testing its order fulfillment system
- To fictionalize customer names, use the a random lookup function to pull first and last names randomly from the Customer Information table:
  - "Gerard Depardieu" becomes "Ronald Smith"
  - "Lucille Ball" becomes "Elena Wu"

Example: Bank Account Numbers

- First Financial Bank’s account numbers are formatted “123-4567” with the first three digits representing the type of account (checking, savings, or money market) and the last four digits representing the customer identification number
- To mask account numbers for testing, use the *actual first three digits*, plus a *sequential four-digit number*
- The result is a fictionalized account number with a valid format:
  - “001-9898” becomes “001-1000”
  - “001-4570” becomes “001-1001”

Propagating Masked Data



Customers Table

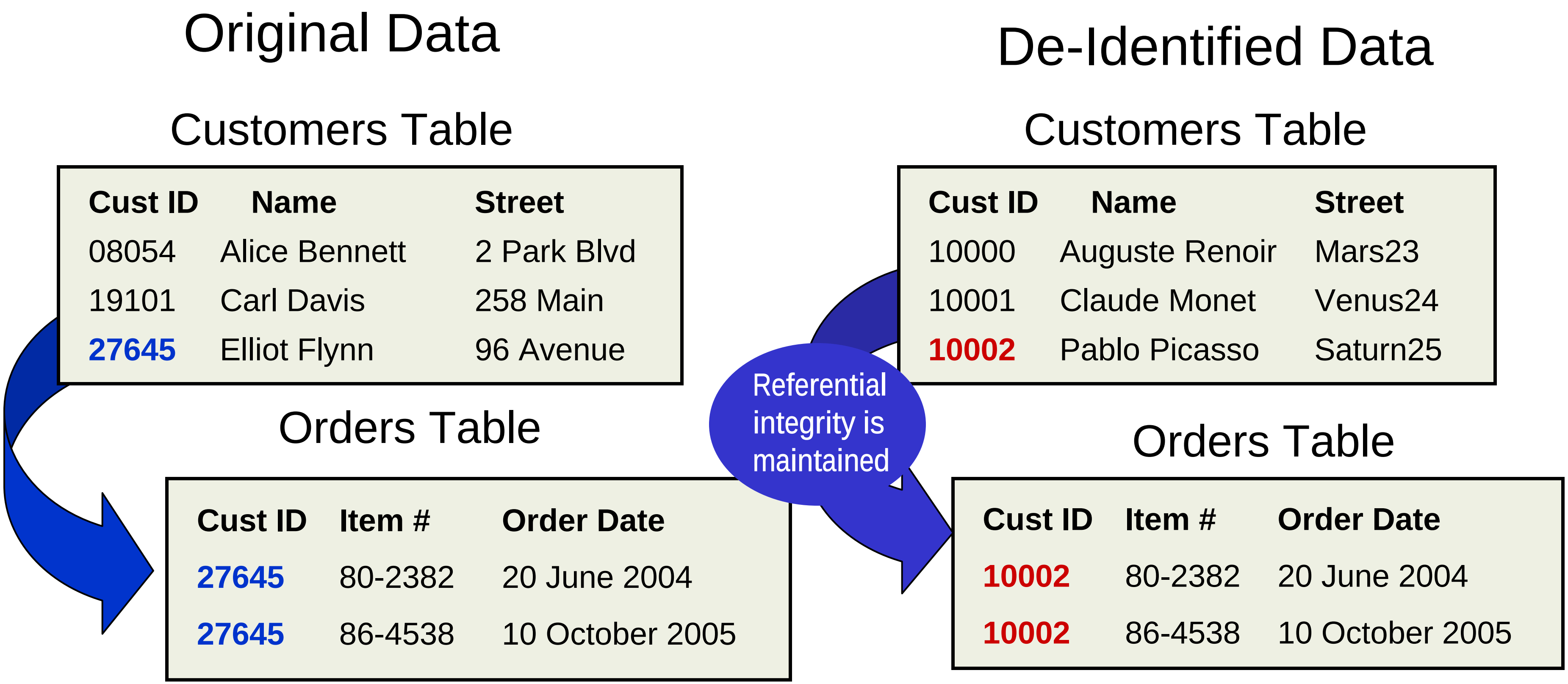
Cust ID	Name	Street
08054	Alice Bennett	2 Park Blvd
19101	Carl Davis	258 Main
27645	Elliot Flynn	96 Avenue

Orders Table

Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005

- Key propagation
  - –Propagate values in the primary key to all related tables
  - –Necessary to maintain referential integrity

Masking with Key Propagation



Original Data

Customers Table

Cust ID	Name	Street
08054	Alice Bennett	2 Park Blvd
19101	Carl Davis	258 Main
27645	Elliot Flynn	96 Avenue

Orders Table

Cust ID	Item #	Order Date
27645	80-2382	20 June 2004
27645	86-4538	10 October 2005

Referential integrity is maintained

De-Identified Data

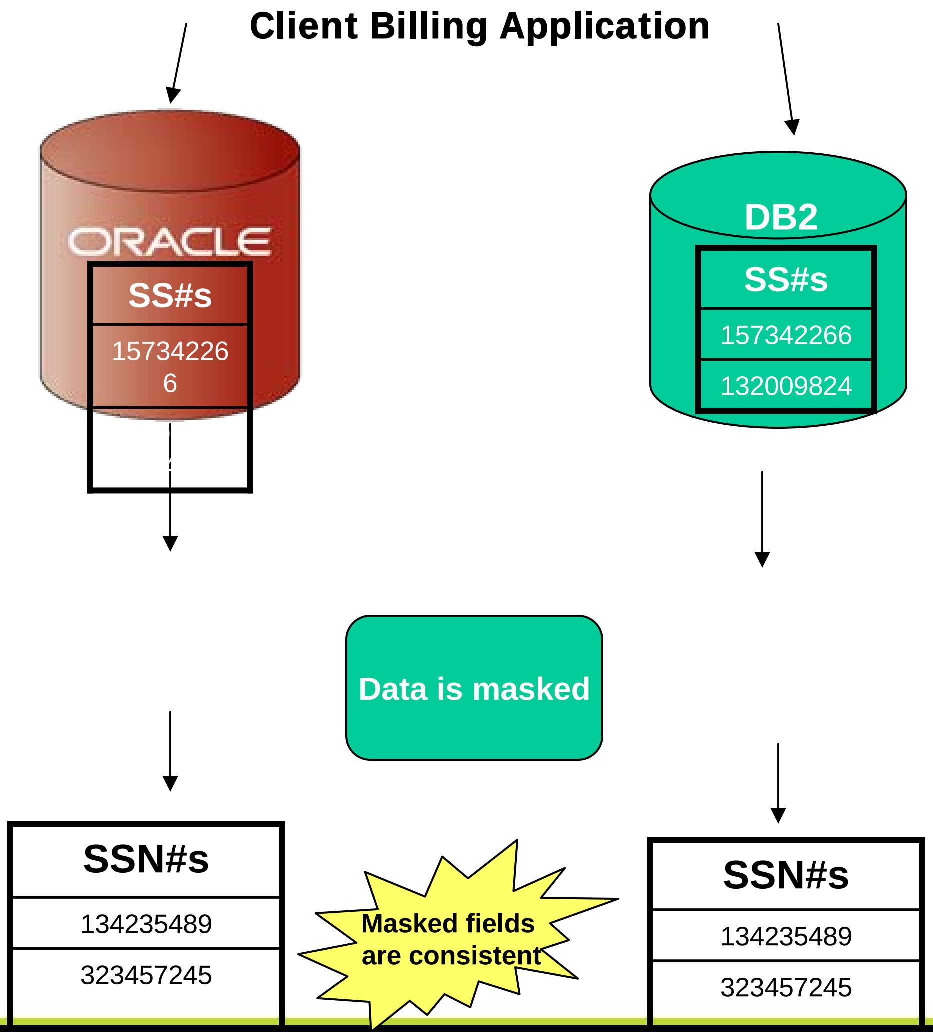
Customers Table

Cust ID	Name	Street
10000	Auguste Renoir	Mars23
10001	Claude Monet	Venus24
10002	Pablo Picasso	Saturn25

Orders Table

Cust ID	Item #	Order Date
10002	80-2382	20 June 2004
10002	86-4538	10 October 2005

Component B - Context



Client Billing Application

ORACLE

SS#s
15734226 6

DB2

SS#s
157342266
132009824

Data is masked

SSN#s
134235489
323457245

Masked fields are consistent

SSN#s
134235489
323457245

- A single mask will affect ‘downstream’ systems
- Column/field values must still pass edits
  - SSN
  - Phone numbers
  - E-mail ID
- Zip code must match
  - Address
  - Phone area code
- Age must match birth date

Component C - Flexibility

- Laws being interpreted
- New regulations being considered
- Change is the only certainty
- ERPs being merged
- Masking routines will change, frequently
- Quick changes will be needed

Data Masking Consideration – Step 5 Project Methodology

- Determine Base Directives
- Compile Data Sources List
- Design Transformation Strategy
- Develop Transformation Process
- Implement Testing Strategy

.

The Market Need

- Corporations have a duty to protect confidential customer information and have gained an understanding that vulnerabilities exist both in the Production and Test Environments
- Companies have begun implementing basic privacy functionality but are requiring more specific and application aware masking capabilities that can be applied across applications

  - *IT organizations require that development databases provide realistic and valid test data (yet not identifiable) after it is masked. This includes: Valid social security #'s, credit card #'s, etc.*

  - *Enterprises require the option to mask data consistently across several different applications, databases, and platforms*

Success with Data Masking

– "Today we don't care if we lose a laptop"

- Large Midwest Financial Company

– "The cost of a data breach is exponentially more expensive than the cost of masking data"

- **Large East Coast Insurer**

Success: Data Privacy

About the Client:
$300 Billion Retailer
Largest Company in the World
Largest Informix installation in the world

- **Application:**
  - Multiple interrelated retail transaction processing applications
- **Challenges:**
  - Comply with Payment Card Industry (PCI) regulations that required credit card data to be masked in the testing environment
  - Implement a strategy where Personally Identifiable Information (PII) is de-identified when being utilized in the application development process
  - Obtain a masking solution that could mask data across the enterprise in both Mainframe and Open Systems environments
- **Solution:**
  - IBM Optim Data Privacy Solution™

- **Client Value:**
  - Satisfied PCI requirements by giving this retailer the capability to mask credit data with fictitious data
  - Masked other PII, such as customer first and last names, to ensure that “real data” cannot be extracted from the development environment
  - Adapted an enterprise focus for protecting privacy by deploying a consistent data masking methodology across applications, databases and operating environments

Success: Data Privacy

About the Client:
$35 Billion Financial Services Company

- **Application:**
  - Custom Banking Applications
- **Challenges:**
  - Complying with a regulatory agency mandate to address increased risk for fraud, related to customer information in the CIS application development and testing environments
  - Implementing a privacy protection strategy in time to support major year-end testing runs and quarterly enterprise application testing activities
  - Expanding data privacy protection to include the mainframe and open systems development and testing environments
- **Solution:**
  - IBM Optim Data Privacy Solution™

- **Client Value:**
  - Satisfied the regulatory agency mandate to prevent fraud and avoided penalties by de-identifying customer financial information in the CIS application development and testing environments
  - In less than 4 months, implemented consistent methods for de-identifying or “scrubbing” personal financial information in time for the next application releases
  - Adapted an enterprise focus for protecting privacy by deploying a consistent data masking methodology across applications, databases and operating environments

How does Data De-Identification Protect Privacy?

- Comprehensive enterprise data masking provides the fundamental components of test data management and enables organizations to ***de-identify, mask and transform*** sensitive data across the enterprise
- Companies can apply a range of transformation techniques to substitute customer data with ***contextually-accurate but fictionalized data*** to produce ***accurate test results***
- By masking personally-identifying information, comprehensive enterprise data masking protects the ***privacy and security*** of confidential customer data, and ***supports compliance*** with local, state, national, international and industry-based privacy regulations

Concluding Thought #1

"It costs much less to protect sensitive data than it does to replace lost customers and incur damage to the image of the organization and its brand—an irreplaceable asset in most cases."

IT Compliance Group Benchmark Study 2/07

Concluding Thought #2

“We're not going to solve this by making data hard to steal. The way we're going to solve it is by making the data hard to use.”

Bruce Schneier, author of "Beyond Fear: Thinking Sensibly About Security in an Uncertain World"

Thank You