



## Closing the Data Privacy Gap: How Safe is Your Data?

IBM Software Group  
David L. Alexander  
503.449.8889  
[dalexan@us.ibm.com](mailto:dalexan@us.ibm.com)

## Today's Discussion

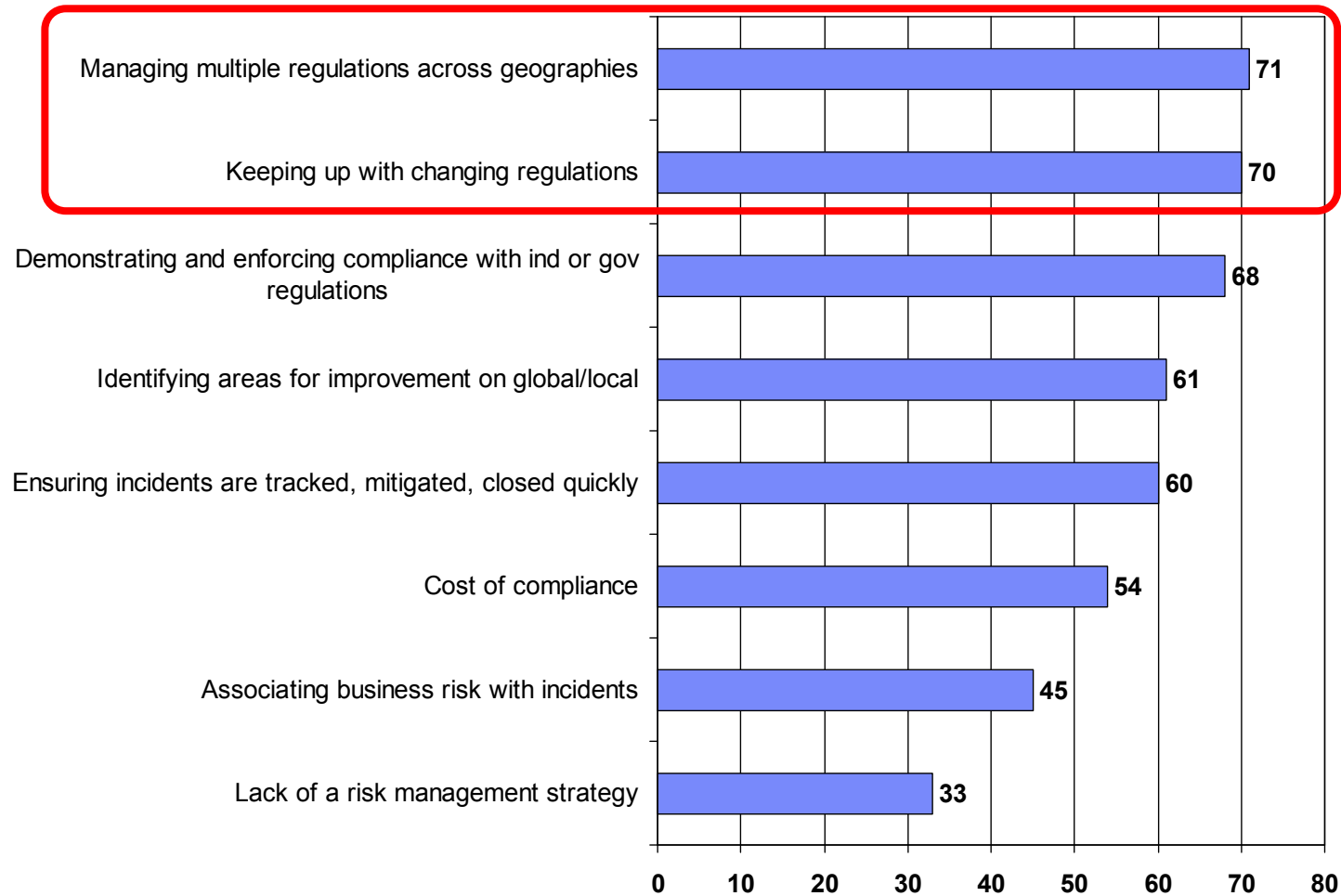
- **The Driving Force Behind IT Risk Initiatives**
- **What's at Stake**
- **Data Privacy – General Principles**
- **Key Elements – Examples**
- **Q&A**

## Does This Define Your Privacy Strategy?



## Challenges Today with Data Privacy

10. What are the THREE biggest challenges your organization is currently facing today regarding data privacy?



Source: AMR Research, Dennis Gaughan – December 2008

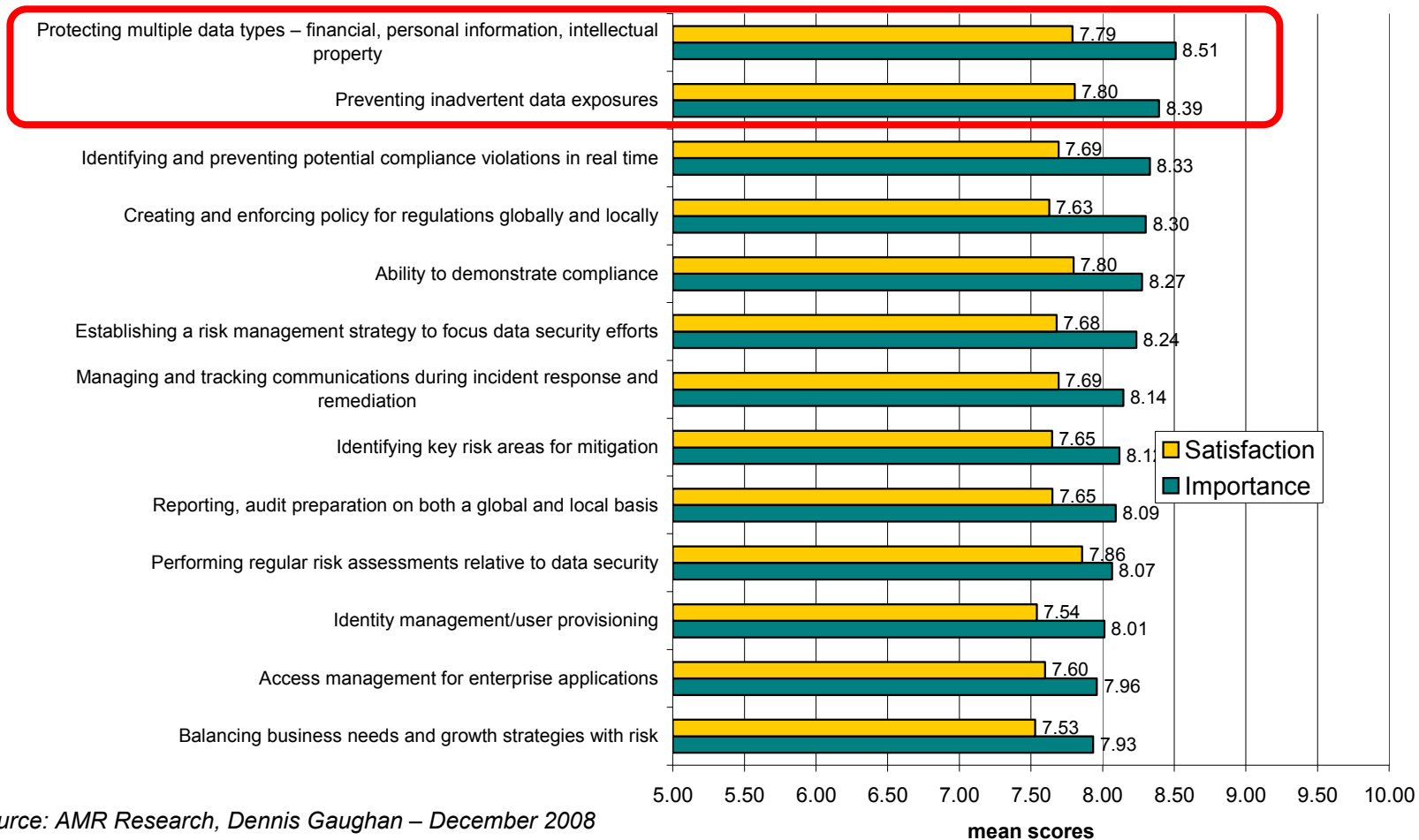
## Common Legislative Themes

- **Government regulations protect consumers**
  - USA: HIPAA, Gramm-Leach-Bliley Act (GLB), California Security Breach Notice Statute
  - Canada: Personal Information Protection and Electronic Documents Act (PIPEDA), Personal Information Protection Act
  - PCI Data Security Standard
  - European Union: Personal Data Protection Directive 1998
  - UK: Data Protection Act of 1998
  - Australia: Privacy Amendment Act of 2000
- **Fines and penalties focus on criminal misconduct**
  - FDIC may levy fines from \$5,000 to \$1,000,000 per day
  - GLB sections 501 & 503 enable criminal penalties

# Data Privacy Gaps

11/12. For each aspect of data privacy listed below, please rate its importance to you, as well as your satisfaction with your company's current level of performance?

**Importance vs. Satisfaction with Existing Processes**



Source: AMR Research, Dennis Gaughan – December 2008

## 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”

## Case in Point...

- **January 2009:** A third-party consulting services firm working on behalf of a large financial institution reported a consultant's computer stolen. The computer had on it the names and Social Security numbers of current and former Financial Advisors and some applicants for employment.
- **October 2008:** An IT contractor for a large American oil company used the personal data of four company workers as part of an unemployment insurance claims scam. Employees of a third-party contractor misused information stored in a corporate database. Misused data included names, dates of birth and Social Security numbers.
- **June 2008:** An American university recently notified more than 11,000 current and former students that their Name, Address & SSN have been posted online. Former student employees in the university IT department had posted the information there for remote testing purposes.
- **April 2008:** The theft of a laptop in New York contained the data on over 170,000 people who had used the services of a blood bank. The blood bank states the data had been sent to a US software development company based in NY as part of software upgrade testing for the bank's systems.

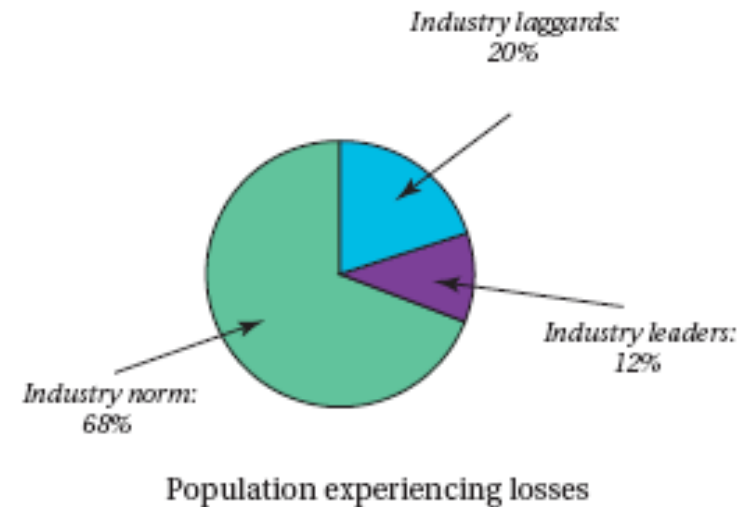
\* Source: Privacy Rights Clearinghouse, 2009



## 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



**Figure 1: Sensitive data loss results**

*Source: IT Policy Compliance Group, 2007*

## The Latest on Data Privacy

- **\$202**
  - Cost to companies per compromised record
- **\$6.6 Million**
  - Average cost per data breach “incident”
- **40%**
  - % of breaches where the responsibility was with Outsourcers, contractors, consultants and business partners

\* Sources: Ponemon Institute, Privacy Rights Clearinghouse, 2008

## The Latest Research on Test Data Usage

- **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.
- **52% of respondents outsourced application testing**
  - 49% shared live data!
- **26% of respondents said they did not know who was responsible for securing test data**

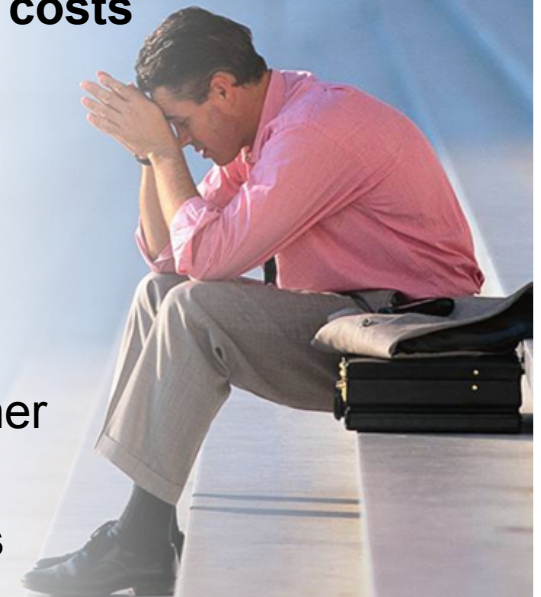
Source: The Ponemon Institute. The Insecurity of Test Data: The Unseen Crisis

## What's at Stake?

- Fines and penalties
- Loss of customer loyalty
- Loss of revenue
- Share price erosion
- Negative publicity
- “Brand equity” damage
- Damage to company reputation
- Increased operations costs

## Compelling Events

- **Globalization and Single Global Instance**
  - International Legislative Requirements
  - Off-shoring, Outsourcing. Off-shore developers in other countries accessing production data
- **Third-Party Access - outsourcers for Payroll, Benefits**
  - Mergers, Acquisitions, Divestiture
- **Introduction of new HR technology or module**



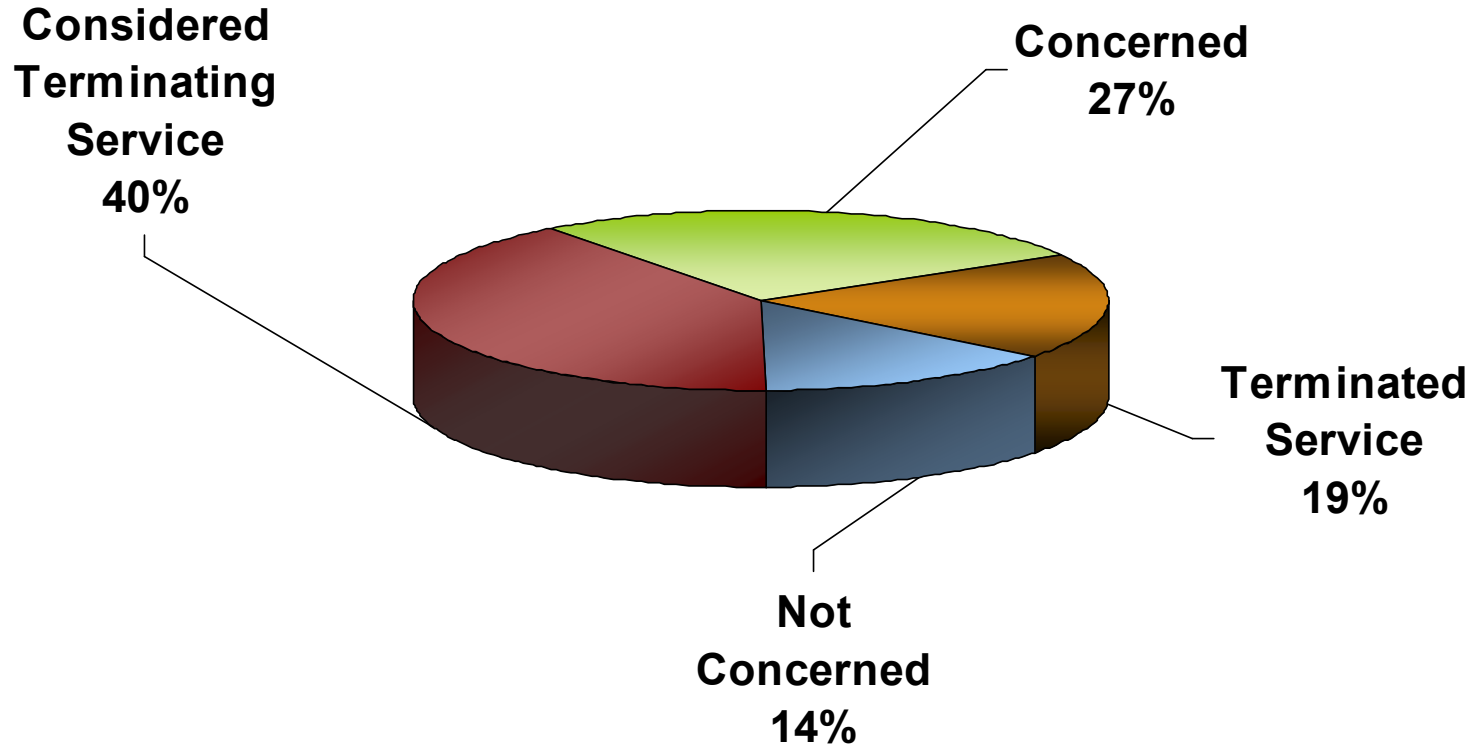
## Best Practices for Securing Information

- Assess risk to determine where to start
- Address the root of the problem- start with discovery
- Utilize cross-functional teams to design risk management solutions
- Promote a culture of content protection and awareness
- Build a plan to operationalize at the end
- Expand coverage



# Consumer Reaction

Banking Customer Survey (Ponemon Institute)



## What is Done to Protect Data Today?

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

## 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
  - Data entry/update client/server devices
- **If data can be seen it can be copied**
  - Download
  - Screen captures
  - Simple picture of a screen





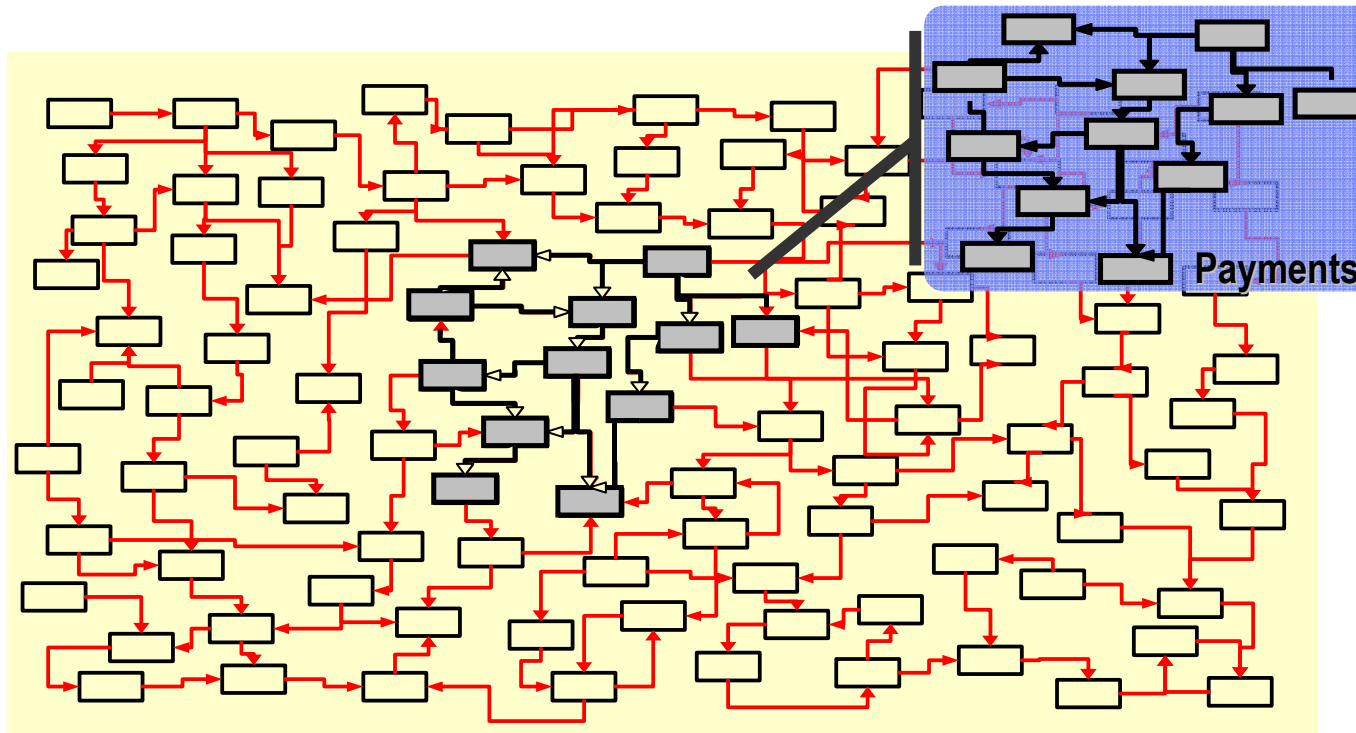
## The “Testing Paradox”



*Live systems are best tested with live data, but exposing live data in a test environment creates a risk of loss, and this is not the purpose for which the data was provided...*

*“Testing Times for HR Systems and EU Data Protection Law” – Gartner, June 2008*

## Complete Business Objects Are Critical for Information-centric Projects



- **Represents application data record – payment, invoice, customer**
  - Referentially-intact subset of data across related tables and applications; includes metadata
- **Provides “historical reference snapshot” of business activity**
- **Federated extract support across enterprise data stores**

## The “Testing Paradox”



*Scrambling at a database level to depersonalize data may not be good for testing purposes... [data masking] requires specific application level scrambling competence and tools...*

## The Solution is De-identification

*“Testing Times for HR Systems and EU Data Protection Law” – Gartner, June 2008*

## What is Data Masking?

- **Also known as: data de-identification, depersonalization, desensitization, obfuscation, 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**

# Components of a Privacy Project

- 1. Understand Application and Business Requirements**
  - Where do applications exist?
  - What is the purpose of the applications?
  - How closely does replacement data need to match the original data?
  - How much data needs to be masked?
- 2. Determine what you need to mask**
  - Sensitive Employee Information like
    - Personal data
    - Bank Details
    - Payroll information
- 3. Choose an enterprise strength data masking solution that**
  - Extends to the existing ERP processes
  - Offers intelligent making routines
  - Easy to use and implement

## Data Privacy – Typical Attributes


- **Employee Personal Information**
  - Name, Address, Phone Number, Email Address, Comp Rate, DOB
- **Employee Beneficiary Information**
  - Name, Address, Phone Number, DOB etc.
- **Identification Numbers**
  - National ID (SSN, SIN, Codice Fiscale etc.)
  - Drivers License, Passport Number, Health Identification Number
- **Business Numbers**
  - Account Number/Policy Number/Loan Number
  - Credit Card Number, Bank Account Number, Routing Code
- **Vendor/Customer Details**
  - Name, Address, Phone #, Fax #, Email Address
  - Vendor ID #s, Vendor SSN #

# Data Masking Techniques

- String literal values
- Character substrings
- Random or sequential numbers
- Arithmetic expressions
- Concatenated expressions
- Date aging
- Lookup values
- Intelligence

## Example 1

Patient Information					
Patient No.	112233	SSN	123-45-6789		
Name	Amanda Winters				
Address	40 Bayberry Drive				
City	Elgin	State	IL	Zip	60123



Patient Information					
Patient No.	123456	SSN	333-22-4444		
Name	Erica Schafer				
Address	12 Murray Court				
City	Austin	State	TX	Zip	78704

Data is masked with contextually correct data to preserve integrity of test data

# Data Masking Techniques

- String literal values
- Character substrings
- Random or sequential numbers
- Arithmetic expressions
- Concatenated expressions
- Date aging
- Lookup values
- Intelligence

## Example 2

Personal Info Table		
PersNbr	FirstName	LastName
08054	Alice	Bennett
19101	Carl	Davis
27645	Elliot	Flynn
	⋮	

Event Table		
PersNbr	FstNEvtOwn	LstNEvtOwn
27645	Elliot	Flynn
27645	Elliot	Flynn



Personal Info Table		
PersNbr	FirstName	LastName
10000	Jeanne	Renoir
10001	Claude	Monet
10002	Pablo	Picasso
	⋮	

Referential integrity is maintained with key propagation

Event Table		
PersNbr	FstNEvtOwn	LstNEvtOwn
10002	Pablo	Picasso
10002	Pablo	Picasso



# Client Success: Data Privacy

## About the Client

### Leading Global Household Goods Manufacturer

Annual Revenue  
**\$11.7 Billion**

Application  
**SAP® Human Capital Management (HCM)**

Solution  
**A Data Privacy Solution for SAP**

- **Challenge:**
  - This leading household goods manufacturer needed to consolidate multiple worldwide instances of the SAP Human Capital Management application. As they created their testing environment, the client wanted to “de-identify” their SAP HCM data so that developers were not using confidential employee HR data in their test environments.
- **Solution:**
  - Using IBM Optim Data Privacy Solution for SAP software, the client plans to reduce the visibility of confidential data in non-production environments. Optim offered proven capabilities for performing complex data masking routines, while preserving the integrity of the SAP HCM data for development and testing purposes
- **Benefits:**
  - Reduced time to manually code the data scrambling routines.
  - Implemented data masking solution, as part of overall support data governance strategy
  - Protected confidential employee information within the testing and development environments, ensuring privacy of HR and payroll information
  - Deployed data masking solution quickly and efficiently, using both out-of-box definitions as well as custom de-identification routines

# Client Success: Data Privacy

## About the Client

**Large Global Retailer;  
Largest Informix  
installation**

Annual Revenue  
**\$300 Billion**

Application  
**Multiple Interrelated  
Retail Processing  
Applications**

Solution  
**A Data Privacy  
Solution**

- **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

- **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

## Concluding Thought



*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...<sup>1</sup>*

<sup>1</sup> Bruce Schneier, Author - "Beyond Fear: Thinking Sensibly About Security in an Uncertain World"

# Questions?

## Disclaimers

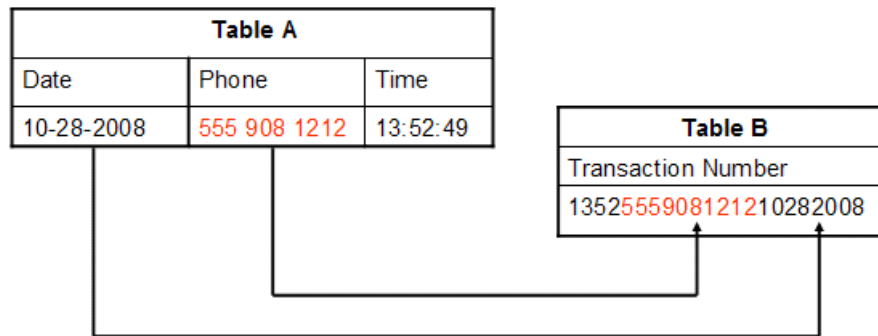
**IBM customers are responsible for ensuring their own compliance with legal requirements. It is the customer's sole responsibility to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws.**

**IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.**

**The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information provided, it is provided "as is" without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.**

## Using InfoSphere Discovery to identify confidential data

- **Some instances of sensitive data are easy to recognize, but others are hidden**
  - Compounded with other data elements in a row
  - Broken apart and spread into multiple columns
  - Buried within comment or text fields



**Hidden instances of private data represent a potential compliance risk**