

Today's Discussion



Data Privacy – What is the Risk?

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



Challenges Today with Data Privacy

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



Source: AMR Research, Dennis Gaughan – December 2008



The Cost of a Data Breach

- \$202
 - ▶ Cost to companies per compromised record
- \$6.6 Million
 - ▶ Average cost per data breach “incident”
- 34% of customers lost
 - ▶ Customers ceasing business with a company after a single privacy breach
- 45% of customers lost
 - ▶ Customers ceasing business when personal information is breached twice

* Sources: Ponemon Institute, Privacy Rights Clearinghouse, 2008



What is Done to Protect Data Today?



- Production Systems – Top of mind
 - ▶ Physical entry access controls
 - ▶ Network, application and database-level security
 - ▶ Multi-factor authentication schemes (tokens, biometrics)
 - ▶ Encryption



- Non-Production Systems – Unique Challenges
 - ▶ Replication of production safeguards not sufficient
 - ▶ Need “realistic” data to test accurately to ensure changes to production are stable



Common Legislative Themes

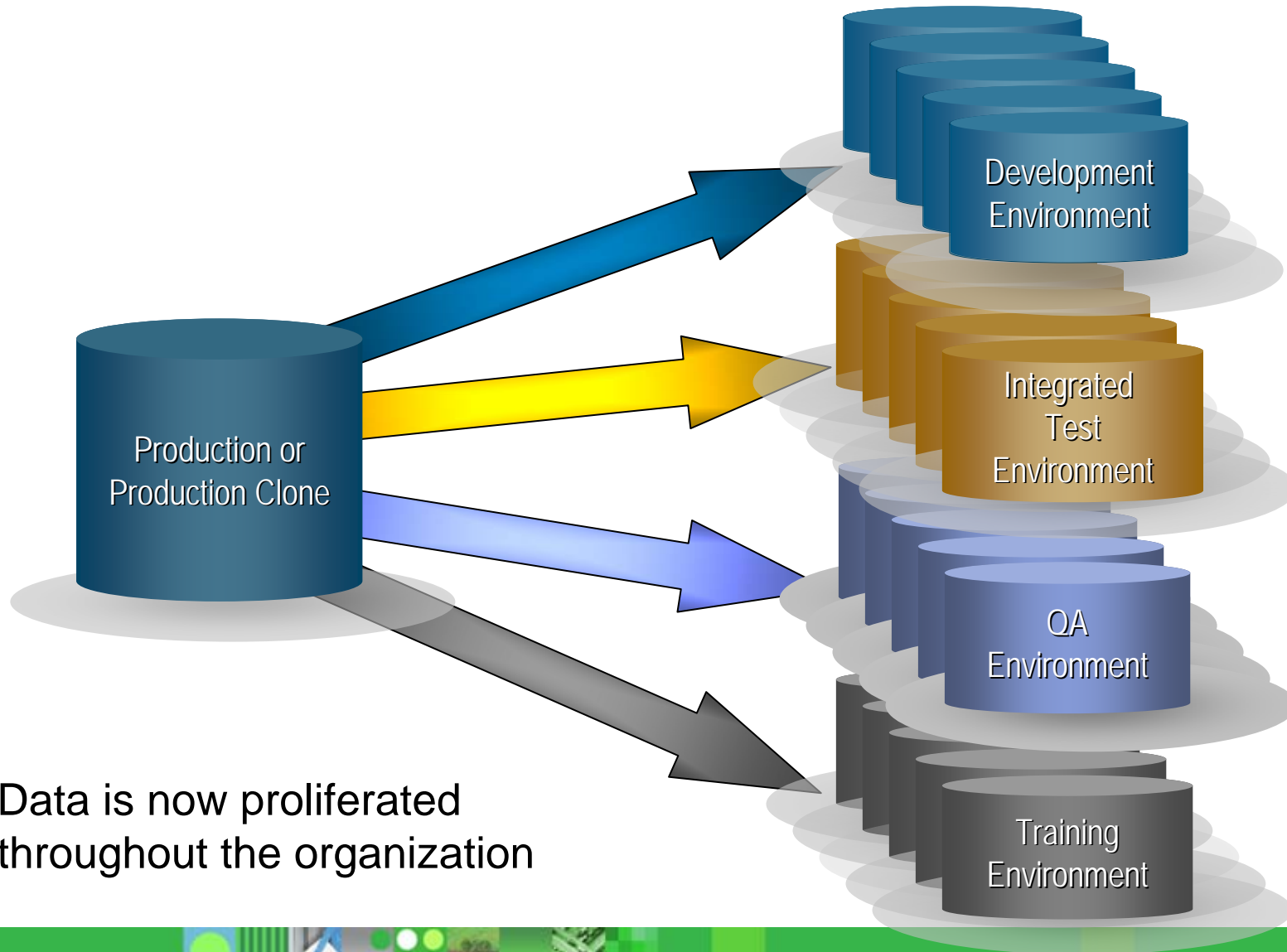
- Government regulations protect consumers
 - ▶ USA: HIPAA, Gramm-Leach-Bliley Act (GLB), California Security Breach Notice Statute
 - ▶ Canada: Personal Information Protection and Electronic Documents Act (PIPEDA), Personal Information Protection Act
 - ▶ PCI Data Security Standard
 - ▶ European Union: 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

How is Data Often Lost or Stolen

- Application breaches
 - ▶ Point-of-Sale
 - ▶ Internal systems
- Data “laying around”
 - ▶ Laptops
 - ▶ Hard drives
 - ▶ Thumb drives
 - ▶ Print outs
- Data exposed in testing and training
 - ▶ Outsourcers
 - ▶ Internal employees



Data in Non-Production – “The Untold Story”



Data is now proliferated throughout the organization

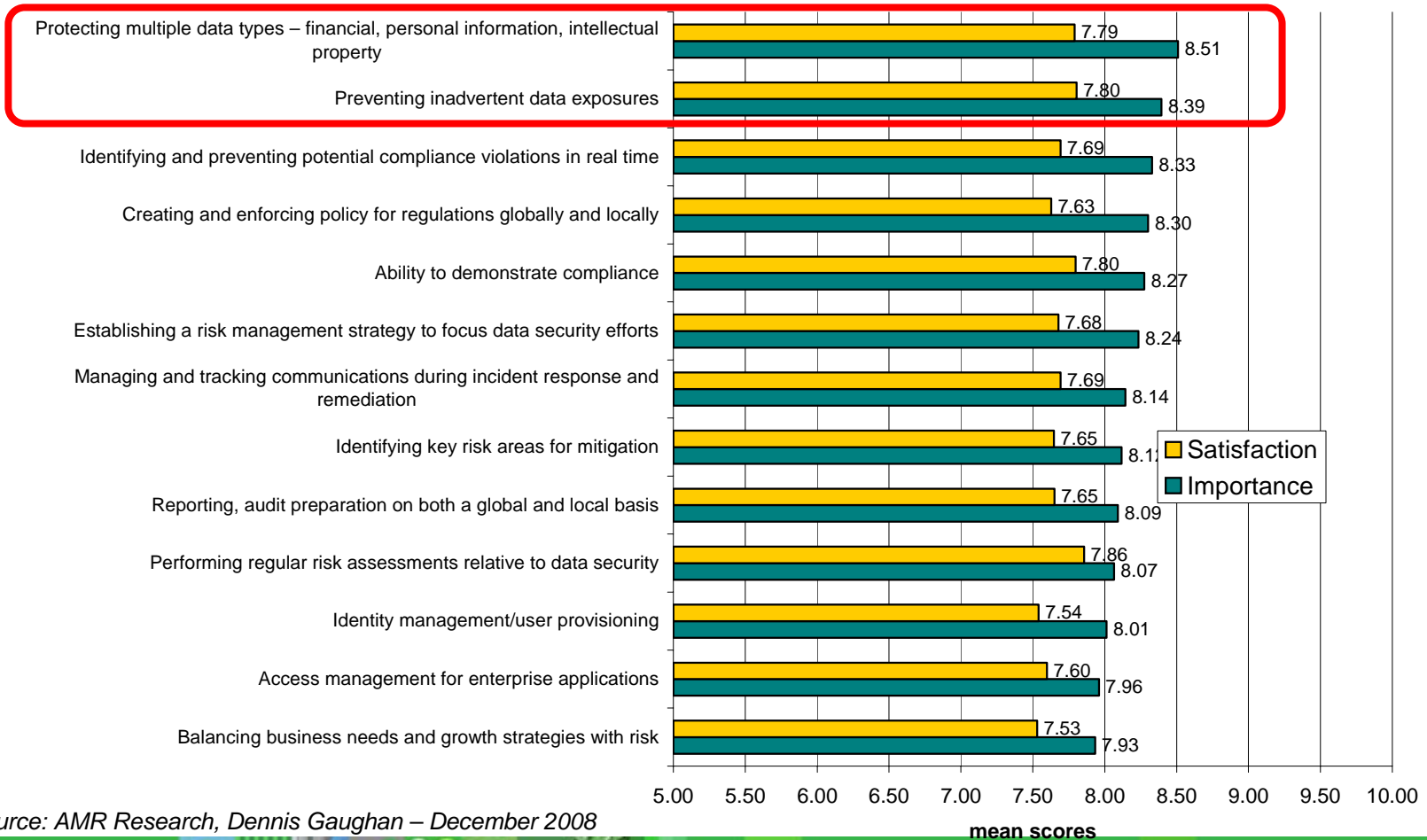




Data Privacy Gaps

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

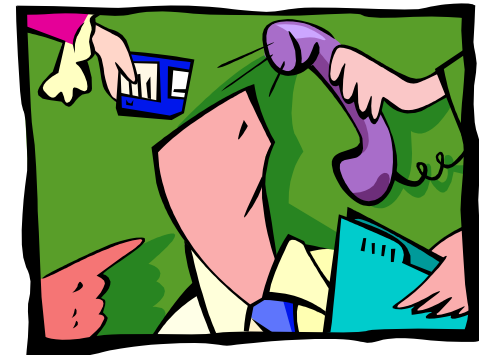


Data Governance

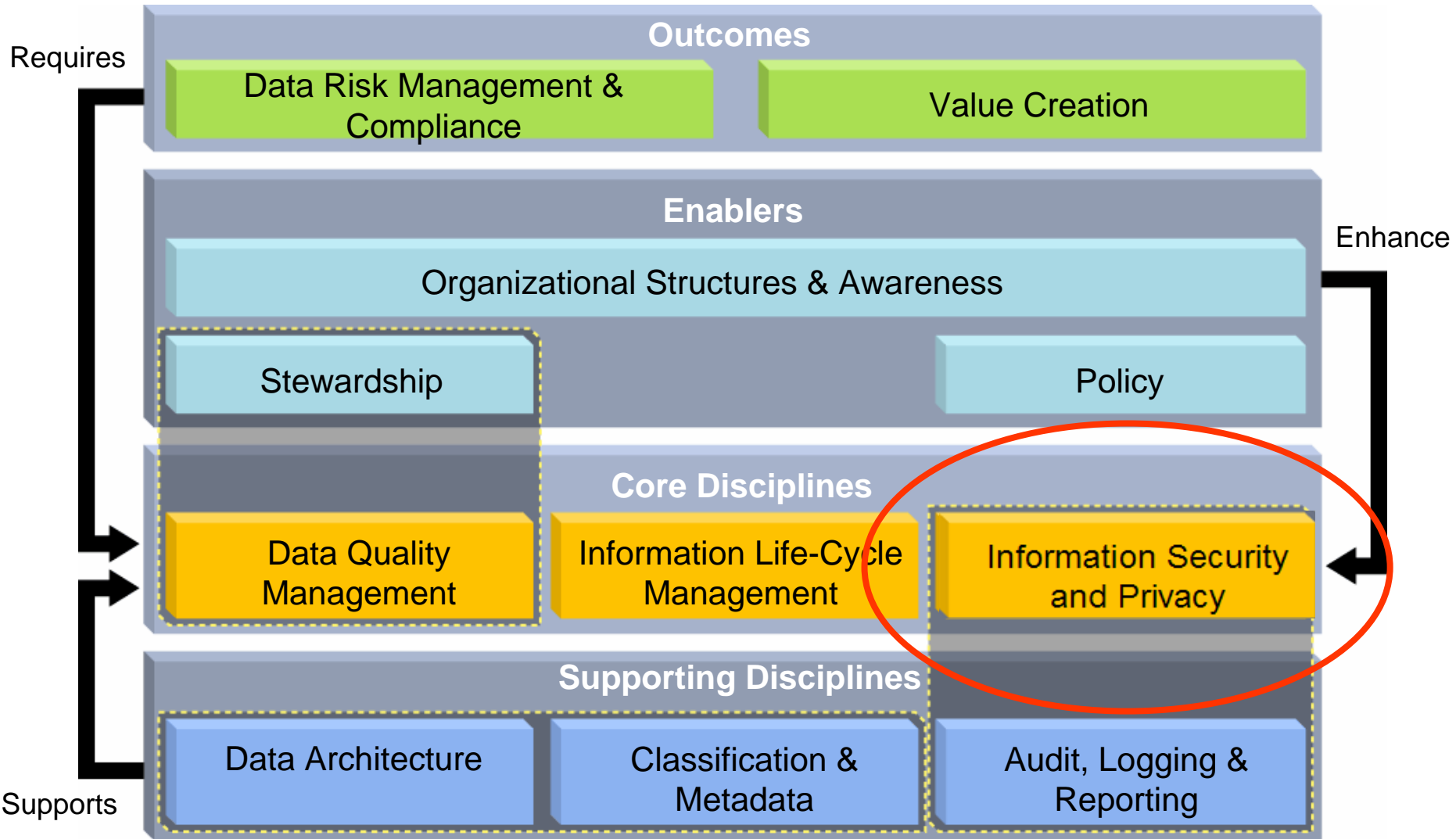
Data governance is the orchestration of people, process and technology to enable an organization to leverage information as an enterprise asset.

Data Governance safeguards information, keeps auditors and regulators satisfied, uses improved data quality to retain customers and constituents and drive new opportunities

- **Data Governance calls for “stewardship” of sensitive data**
 - ▶ **Enforced by regulations**
 - ▶ **Expected by customers**
 - ▶ **Demanded by executives**



Data Governance Council Maturity Model



What Must be Done? – Data De-identification

- Definition: Removing, masking or transforming elements that could be used to identify an individual
- Also known as: data masking, depersonalization, desensitization, obfuscation, data scrubbing
- Technology that helps conceal real data
 - ▶ Name, address, telephone, SSN / National Identity number, credit card #...
- 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
- Masked data must be appropriate to the context
 - ▶ Within permissible range of values
 - ▶ Application-aware





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 Information
 - ▶ Account Number/Policy Number/Loan Number
 - ▶ Credit Card Number, Bank Account Number, Routing Code
 - ▶ Financial information
 - ▶ Intellectual Property
- Vendor/Customer Details
 - ▶ Name, Address, Phone #, Fax #, Email Address
 - ▶ Vendor ID #s, Vendor SSN #

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...”*
- *“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...”*

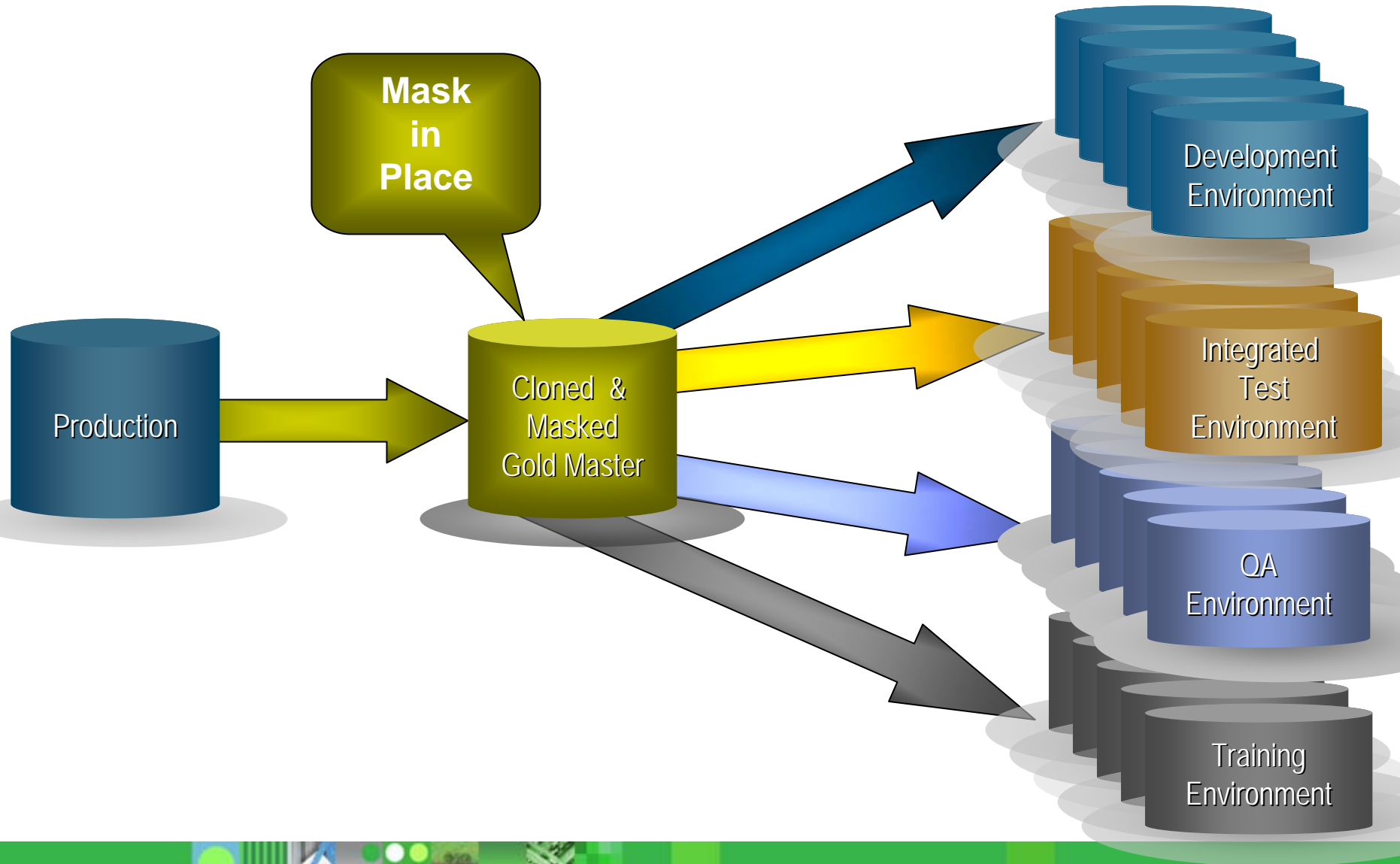


Components of a Privacy Project

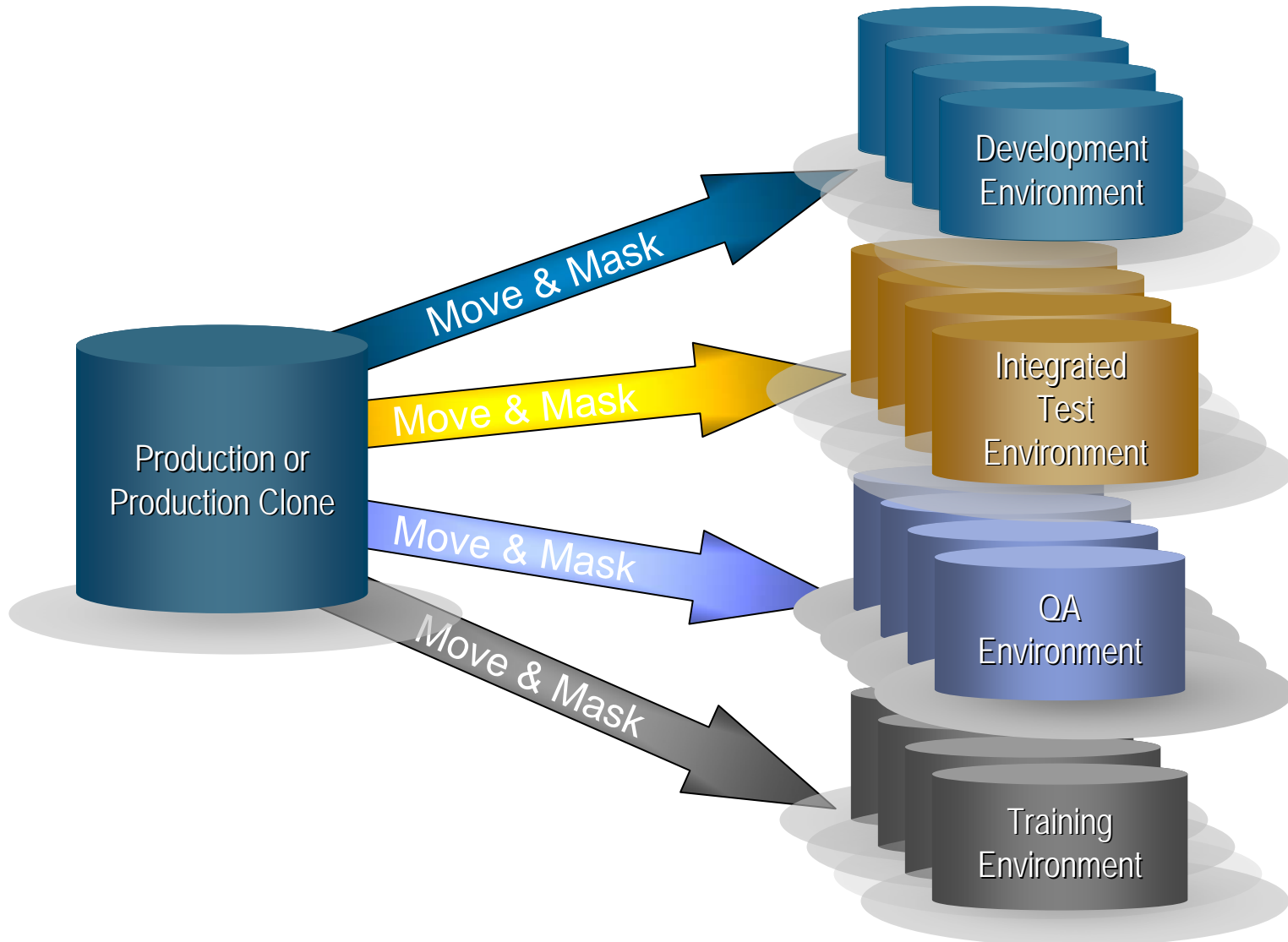
- Understand Application and Business Requirements
 - ▶ Where do applications exist?
 - ▶ What is the purpose of the applications?
 - ▶ How closely does replacement data need to match the original data?
 - ▶ How much data needs to be masked?
- Determine what you need to mask
 - ▶ Sensitive Employee Information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



Data De-identification with Cloning – Mask-in-Place



Data De-identification without Cloning – Move & Mask



Intelligent Data Masking

A comprehensive set of data masking techniques to transform or de-identify data, including:

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

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

De-Identified Data

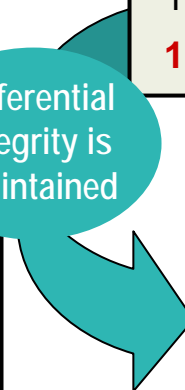
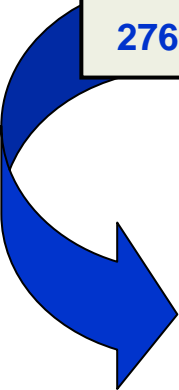
Customers Table

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

Referential integrity is maintained



Example - Intelligent Masking

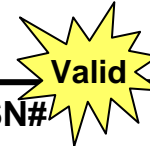
Production Database

F. Name	L. Name	Credit Card#	SSN#
John	Denver	5298774132478855	254-77-6644
Vanessa	Jones	4324115574123654	154-74-7788

**Data before
Masking**

Test Database

F. Name	L. Name	Credit Card#	SSN#
John	Denver	5326458711224956	164-95-3792
Vanessa	Jones	4972584612457744	294-09-1849



**Data after
Masking...
Masked with
Valid CC#
and SS#**

How are these numbers valid?

For Social Security Numbers	For Credit Card Numbers
A Social Security Number (SSN) consists of nine digits. The first three digits is called the "area number." The central, two-digit field is called the "group number." The final four-digit field is called the "serial number." All numbers must fit the latest available criteria for each section.	Most credit card numbers are encoded with a "Check Digit." A check digit is a digit added to a number (either at the end or the beginning) that validates the authenticity of the number. A simple algorithm is applied to the other digits of the number which yields the check digit.



Street Address/City/State/Zip Code Lookup Table

Total Assets	Customers	Street	City	State	Zip Code
\$534,674,233	54,999	12 Buttercup Ln	Cleveland	OH	44101
\$8,777,733,811	105,333	6767 Rte 1 S	Princeton	NJ	08540

1) Client is a Bank who wishes to mask its assets by location

Address
Lookup
Table

288 Elm St	Milwaukee	WI	53201
12 Rodeo Dr	Los Angeles	CA	90001
3526 Diamond Rd	Seattle	WA	98101
12 Street Road	Las Vegas	NV	89101
2 Applegarth Ln	Brunswick	ME	04011

2) Street
Address/City/State/Zip
Codes for masking

New Table with Masked Data

Total Assets	Customers	Street	City	State	Zip Code
\$534,674,233	54,999	3526 Diamond Rd	Seattle	WA	98101
\$8,777,733,811	105,333	21 Street Rd	Las Vegas	NV	89101

3) Entire address
row can be masked



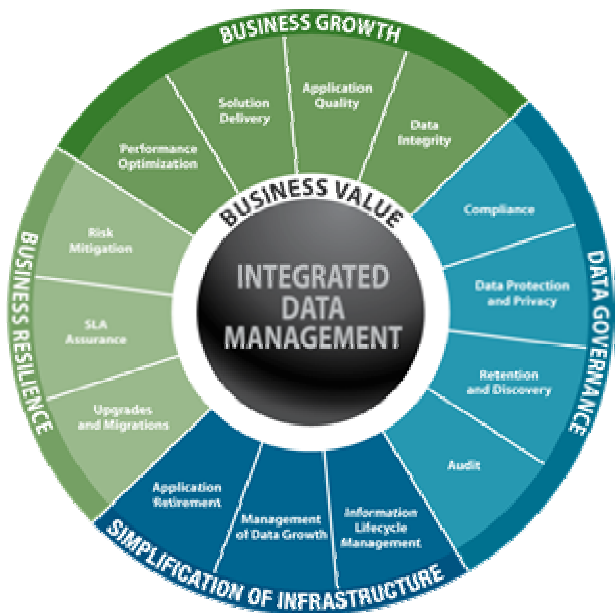
Using Custom Masking Exits

- Apply complex data transformation algorithms and populate the resulting value to the destination column
- Selectively include or exclude rows and apply logic to the masking process
- Valuable where the desired transformation is beyond the scope of supplied Column Map functions
- Example: Generate a value for CUST_ID based on customer location, average account balance, and volume of transaction activity



Introducing Integrated Data Management from IBM Optim™

An integrated, modular environment to manage enterprise application data, and optimize data-driven applications, from requirements to retirement



- **Grow the business, without growing costs**
 - ▶ Develop and deploy business critical applications faster
 - ▶ Mitigate compliance risks with model-driven data governance
 - ▶ Prevent runaway infrastructure spending
 - ▶ Improve performance – of work teams, databases, applications, and business units

IBM Optim Data Privacy Solution

- Mask in place
 - ▶ Apply masking to existing databases as they exist
- Move and mask
 - ▶ Move data into a test environment and mask as it is moved
- Basic masking - character strings, arithmetic functions, random numbers
- Advanced masking - Privacy Models
 - ▶ National Identifiers
 - ▶ Credit Card Numbers
 - ▶ Lookup values - to ensure contextual accuracy (i.e. city, country, postal code)
 - ▶ Replacement data (names, addresses)
- Application-aware masking capabilities, ensuring masked data is realistic but fictional
 - ▶ Custom and Packaged Applications, SAP, Siebel CRM, PeopleSoft Enterprise, Oracle e-Business Suite, JD Edwards EnterpriseOne





Enterprise Environments

IBM Integrated Data Management

Database Design, Development & Administration, Data Growth, Data Privacy, Test Data Management, Application Upgrades & Retirements, Data Retention & E-Discovery

Enterprise Environments



Windows XP/2000 Solaris HP/UX Linux AIX OS/390 Linux z/OS i-series

NAS SAN ATA CAS Optical Tape



Case Study: Global Financial Business Solutions Provider Creates Privatized Test Data Environment with IBM Optim Solutions

- Application:
 - ▶ Over 200 Applications
 - ▶ Mix of CICS and Distributed
- Challenges:
 - ▶ Responsiveness
 - ▶ Stale data, Multiple Environments, Excessive Data
 - ▶ Inappropriate & Invalid Test Cases
 - ▶ Development Teams unable to create test data
 - ▶ Industry and corporate data privatization policies
- Solution:
 - ▶ IBM Optim Test Data Management Solution
 - ▶ IBM Optim Data Privacy Option
- Results:
 - ▶ Creates a secure environment for processing customer information
 - ▶ Delivers an improved and requested solution to our customers
 - ▶ Mitigates the risks of regulatory and legal impact
 - Fines up to \$200,000 per incident; \$90 per lost account record
 - ▶ Reflects adherence to Banks' Information Security policies
 - ▶ Provided savings in processing by reducing volume of Test Data – CPU, DASD
 - ▶ Allows better management, control of test data cases



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...



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



Thank
You

