



Test Data Privatization: A Real-World Case Study

Presented by Craig Laufer and Jeremy
Sloan of the Erie Insurance Group
QUEST 2008



Test Data Privatization

Average cost of a data breach in the US:
Over six million dollars



Test Data Privatization

How many organizations use live data:

- For testing applications? 69%
- For development? 62%
- For production support? 38%
- Only 25% report no production data usage.

"The Insecurity of Test Data: The Unseen Crisis"

Ponemon Institute LLC, December, 2007



Test Data Privatization

What type of live data do you use for development and testing purposes?

- Customer records - 89%
- Consumer Lists - 74%
- Employee records - 43%

"The Insecurity of Test Data: The Unseen Crisis"
Ponemon Institute LLC, December, 2007



Test Data Privatization

“Organizations may think their test data is immune from privacy threats because testing occurs in a non-production environment. However, test environments are less secure because data is exposed to a variety of unauthorized sources.”

Dr. Larry Ponemon, as reported in the SD Times article “With Your Data, Trust No One”



Test Data Privatization

Management wanted to see:

- A holistic approach to test data management
- Practical short-term results:
 - Improve the time spent mining data
 - Protect customer's nonpublic personal information
- Consideration of longer term concerns
 - Security exposure
 - Legislative landscape



Test Data Privatization

Concerns of Testers:

- Finding test data is time consuming
- Existing tools used in an ad hoc manner
- Some testers do not know all tools available
- Testers largely depend on production data and manual data entry



Test Data Privatization

Choosing an application for the POC

- Important legacy system
- Multiple complex interfaces
- Significant downstream processing
- Many freeform fields
- Build challenges into the POC



Test Data Privatization

Demonstrate proof of all privatization methods you may consider, such as:

- Table-based randomization
- Translation to a unique value
- Character-based randomization
- Translation to a hash-value
- Password-based encryption
- Replacement with static text