Don’t copy Production Data into Test

DBAs Behind Bars

Session Code G07 May 4, 2011 9:45 – 10:45am

Robert Andresen
Principal Consultant
IBM Certified Database Administrator,
DB2 9 for z/OS
IBM Certified SOA Associate
ITIL Foundation Certified
Robert.Andresen@ca.com

Objectives

- Understand why using production data in test is dangerous
- Learn several strategies for creating valid test data that is not a security risk
- Understand various coding schemes for credit cards, drivers licenses, social security numbers, etc.
- Understand that applications in production and test may not require 100% data consistancy
  - Name, sex & birthdate vs. D/L #
  - Valid credit card numbers
- Understand your potential culpability in the even of unauthorized access or use of production data

Agenda

- Background
- Why companies copy prod to test
- What are the liabilities?
  - Damage to reputation
  - Legal
- Corporate privacy and data protection policy
- Strategies to build safe test databases

- Note: This presentation has corrections I found after creating the IDUG master copy

I’m not a lawyer

- Nor do I play on TV
- Thanks to the invaluable assistance of the CA Technologies Privacy Officer for her invaluable direction and information she provided to help me put this presentation together.
- The errors and sarcasm are all mine

DBA Retirement Home?

Identity Theft

- Identity Theft Resource Center 2010 Data Breach Stats
  - Banking/Credit/Financial # of Breaches: 54 # of Records: 4,853,708
  - Business # of Breaches: 279 # of Records: 6,626,435
  - Educational # of Breaches: 65 # of Records: 1,598,266
  - Government/Military # of Breaches: 104 # of Records: 1,214,773
  - Medical/Healthcare # of Breaches: 160 # of Records: 1,874,360
  - Total breaches: **662**, Records exposed: **16,167,542**

- http://www.idtheftcenter.org/artman2/publish/lib_survey/ITRC_2008_Breach_List.shtml#

Identity Theft Resource Center

- **What is a breach?** A breach is defined as an event in which an individual name plus Social Security Number (SSN), driver's license number, medical record or a financial record/credit/debit card is **potentially** put at risk – either in electronic or paper format.
- Security breaches can be broken down into a number of categories. What they all have in common is that they usually contain personal identifying information in a format easily read by thieves, in other words, not encrypted. The ITRC tracks five categories of data loss methods:
  - Data on the Move
  - Accidental Exposure
  - Insider Theft
  - Subcontractors
  - Hacking

Why Companies copy prod to test

- Great for testing-
  - All production conditions are in test
  - All data is valid and consistent
    - Credit Card numbers
    - SSN
    - Drivers License numbers
    - Area code, zip code and address match
  - RI valid
- It's really easy
- No perceived cost

All of us have two points of view:

- As an employee:
  - You want valid test environments
  - Easy to create and maintain
- As a customer:
  - You want your personal data to be secure
- Your company policies affect your customers
- All of us are customers of other companies

There oughta be a law!

- **Distributed by:** McClure Newspaper Syndicate
  First Appeared: 1948
  Creators: Harry Shorten (writer) and Al Fagaly (artist)
- http://ilovecomix.blogspot.com/2008/07/there-oughta-be-law-926-1950s.html

Well, there is, or are many

- Identity theft is a high profile problem
  - Horror stories in the news
  - Experiences of friends and family
  - Almost everyone has had at least a bogus credit card charge
- Legislative bodies
  - US Congress
  - 55 Federal Agencies with rule making power
  - 37 czars, including cyber security
  - 50 state legislatures
  - Local laws

Legislation in the US

- 46 jurisdictions with laws about protecting financial data
- Any breech needs to be reported to:
  - Individual whose data was compromised
  - Attorney General, or
  - State Police, depending on jurisdiction
- Data is supposed to be retained only for the purpose it was gathered and should not be retained longer than needed.
- The law makes no distinction between “production data” vs. “test data”

Applicable Legislation

- Financial Identity Data Protection Laws
  - Local Laws
  - Laws outside the US
  - Gramm-Leach Bliley Act
  - Massachusetts Regulations
- HIPAA (1996 Health Insurance and Accountability Act)

Legislative language

- Reasonable and Prudent
  - Law doesn't require perfection
  - Reasonable safeguards need to be established
- Reckless and Irresponsible
  - Prosecuting attorneys argument
  - There will always be disagreement as to whether the measures taken were sufficient, if a major breach happens

Gramm-Leach Bliley Act

- Financial Modernization Act of 1999
- Protect Consumers personal financial data
- Financial Privacy Rule
  - 8 Federal Agencies + 50 States administer and enforce
  - Financial Institutions + financial products & services to consumers
  - Governs collection and disclose of financial information
- Safeguards Rule
  - Requires design, implementation and maintenance of safeguards to protect customer financial information
- Pretexting Provisions
  - Prohibits obtaining financial information under false pretenses

More GLB

- Privacy Notice
  - Clear, conspicuous and accurate statement of privacy practices
  - What information is collected
  - With whom it is shared
  - How the information is safeguarded
  - Doesn’t apply to public ally recorded information
- Opt-Out Rights
  - Consumer has the right to prevent their data being shared with certain third parties
  - Must explain how, in a reasonable way, to opt-out

Massachusetts Regulations

- **201 CMR 17.00: Standards for the protection of Personal Information of Residents of the Commonwealth**
- Spells out computer system security requirements
  - Pro-active
  - Most other legislation is what to do after the fact
  - Identity theft is a hot topic: expect more laws like this one
- Applies to persons who own or license personal information **about a resident of the Commonwealth of Massachusetts**
  - Person: person, corporation, association, partnership or other legal entity
  - Covers paper and electronic records

17.02 Mass regs definitions

- Personal Information:
  - First name or initial + last name
  - Social Security number
  - Drivers license or state-issued id number
  - Financial account or credit/debit card number, with or without PIN
  - Excludes publicly available information

17.03 Duty to Protect and Standards…

- Comprehensive Information Security Program
- Administrative, technical and physical safeguards appropriate to:
  - Size of company
  - Amount of stored data
  - Need for security and confidentiality
- Security Admins
- Risk assessment
- Ongoing employee training
- Means for detecting and preventing security system failures
- Imposing disciplinary measures for violations
- Preventing terminated employees access
- Oversee overseas providers
- Restrictions on Physical access
- Annual reviews of security measures
- Documenting responsive actions

17.04 Computer System Security Requirements

- Secure user authentication protocols
- Secure access control measures
- Encryption of all transmitted information
- Reasonable monitoring for unauthorized access
- Encryption of all personal information on laptops and portable devices
- Reasonably up-to-date firewalls protection and operating system security patches
- Reasonably up-to-date versions of security software
- Education and training of employees on proper use of computer security and the importance of personal information security

HIPAA

- 1996 Health Insurance and Accountability Act
- Covered entities
  - Health care
  - Business Associates
  - Health Plan Provider
  - Health care Clearing Houses
- Protected Health Information
  - Past, present or future physical or mental health or condition
  - Provision of health care
  - Past, present, or future payment for the provision of health care

HIPAA Penalties

- Civil Penalties
  - $100 to $50,000 or more per violation
  - Calendar year cap: $1,500,000
- Criminal Penalties
  - A person who knowingly obtains or discloses individually identifiable health information in violation of the Privacy Rule may face a criminal penalty of up to $50,000 and up to one-year imprisonment. The criminal penalties increase to $100,000 and up to five years imprisonment if the wrongful conduct involves false pretenses, and to $250,000 and up to 10 years imprisonment if the wrongful conduct involves the intent to sell, transfer, or use identifiable health information for commercial advantage, personal gain or malicious harm. The Department of Justice is responsible for criminal prosecutions under the Privacy Rule.

Recent HIPAA news

http://www.wileyrein.com/publications.cfm?sp=articles&newsletter=4&id=6839

First HIPAA Civil Monetary Penalty Causes Concern

Kirk J. Nahra
March 2011 | *Privacy In Focus*

The big news in the health care privacy world is the imposition by the Health and Human Services (HHS) Office for Civil Rights (OCR) of the first civil monetary penalty for a violation of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. While this isn't the first overall HIPAA penalty, this is the first time that the full HIPAA enforcement process has been used to issue a penalty.

And it's a big one. HHS fined Cignet Health of Prince George's County, Maryland, $4.351 million for its violations of the Privacy Rule.

Payment Card Industry Data Security Standard (PCI DSS)

- Worldwide information security standard defined by the Payment Card Industry Security Standards Council
  - Visa
  - MasterCard
  - American Express
  - Discover
  - JCB
- Help organizations processing card payments to prevent credit card fraud
  - Increased controls around data and its exposure to compromise
  - Applies to all organizations that hold, process, or exchange cardholder information from any card branded with the logo of one of the card brands

http://en.wikipedia.org/wiki/Payment_Card_Industry_Data_Security_Standard

PCI DSS version 2.0

Control Objectives	PCI DSS Requirements
Build and Maintain a Secure Network	1. Install and maintain a firewall configuration to protect cardholder data
	2. Do not use vendor-supplied defaults for system passwords and other security parameters
Protect Cardholder Data	3. Protect stored cardholder data
	4. Encrypt transmission of cardholder data across open, public networks
Maintain a Vulnerability Management Program	5. Use and regularly update anti-virus software on all systems commonly affected by malware
	6. Develop and maintain secure systems and applications
Implement Strong Access Control Measures	7. Restrict access to cardholder data by business need-to-know
	8. Assign a unique ID to each person with computer access
	9. Restrict physical access to cardholder data
Regularly Monitor and Test Networks	10. Track and monitor all access to network resources and cardholder data
	11. Regularly test security systems and processes
Maintain an Information Security Policy	12. Maintain a policy that addresses information security

http://en.wikipedia.org/wiki/Payment_Card_Industry_Data_Security_Standard

Legislation outside the US

- EU Data Processing Directive
- Frequently more restrictive than US law
- Broader areas to protect
- Forbids release of “sensitive data”
  - Name
  - eMail
  - Telephone
  - Birth date
  - Political Affiliation
- Multi-national companies need to follow the laws of all the countries they do business in

Other Risks besides applicable law

- Loss of Company reputation
  - Financial Industries – 90% of their customers choose them because of trust and reputation
- Loss of Consumer Trust
  - Loss of customers
  - Loss of Revenue
- Civil Lawsuits

Security Breach Hall of Shame

Published on *Privacy Rights Clearinghouse* (http://www.privacyrights.org)
Today's Date: Feb 17, 2011
Source URL (retrieved on *2011-02-17 00:00*): http://www.privacyrights.org/data-breach

Chronology of Data Breaches

Chronology of Data Breaches
Security Breaches 2005-Present

Posted Date: April 20, 2005
Updated Date: February 15, 2011

Security Breach Discussion

- The next three slides describe the worst security breaches of 2008. The website they came from is listed and you can go back and see what they feel the “lessons learned” should be.
- Let’s spend a few minutes and see if we can apply the “reasonable and prudent” safeguards covered in the legislation section to see how you could avoid similar breaches at your company.

Top 10 Security Breaches of 2008

- http://www.bankinfosecurity.com/articles.php?art_id=1120
  - 1. **TJX Case Winds Up, Arrests Made**
    - The August arrest of 11 alleged hackers accused of stealing more than 40 million credit and debit cards brings law enforcement closer to closing what is still the largest hack ever. The U.S. Department of Justice brought charges against 11 alleged hackers from around the globe
  - 2. **Bank of New York Mellon**
    - An unencrypted backup tape with 4.5 million customers of the Bank of New York Mellon went missing on Feb. 27, after it was sent to a storage facility. The missing tape contains social security numbers and bank account information on 4.5 million customers - including several hundred thousand depositors and investors of People's United Bank of Connecticut, which had given Bank of New York Mellon the information so it could offer those consumers an investment opportunity.
  - **3. Hannaford Brothers**
    - In March, the Maine-based Hannaford Brothers grocery store chain announced that 4.2 million customer card transactions had been compromised by the hackers. More than 1800 credit card numbers were immediately used for fraudulent transactions.

Possible causes:

1. Probably insufficient security or security standards or not up to date security patches. All required by legislation. Much hacking is “social engineering” which can be as simple as calling up and asking for a password, pretending to be someone else.

2. Unencrypted tape answers this one in two words. Any time financial data is sent anywhere by any means it must be encrypted

3. Hackers exploit the three areas listed in answer 1.

Top 10 Security Breaches of 2008

- 4. **Countrywide Insider Theft**
  - In August, a former Countrywide Financial Corp. senior financial analyst, Rene Rebollo, was arrested and charged by the FBI for stealing and selling sensitive personal information of an estimated 2 million mortgage loan applicants. How he did it over a two-year period was to download about 20,000 customer profiles each week onto flash drives, working on Sunday nights, when no one else was in the office. Rebollo then took the excel spreadsheets to business center stores to email to buyers.
- 5. **GE Money Backup Tape Goes AWOL**
  - Early in January, Iron Mountain said it could not find a backup tape that belonged to GE Money, containing information on J.C. Penney customers and 100 other retailers. The tape was stored in an Iron Mountain vault, says an Iron Mountain statement issued about the loss, and had been requested by GE Money in October 2007. The tape contained the personal information of about 650,000 J.C. Penney customers and the other 100 retailers. GE Money processes credit cards for those retailers. As a records and archive company that specializes in records management, Iron Mountain was at a loss to explain the tape's whereabouts. Iron Mountain said it was an unfortunate case of a misplaced tape, but asserts that there was no evidence that the information was obtained and used by unauthorized persons. The missing tape also included about 150,000 social security numbers.
- 6. **RSA Report: Half-Million Banking ID's Stolen**
  - In November, security vendor RSA said it found a single Trojan that had taken more than 500,000 online banking accounts credentials, credit cards and other resources. The company's Fraud Action Research Team added that the hacking gang behind the Trojan may have been operating for as long as three years. The compromised data came from hundreds of financial institutions around the world.

Possible causes:

- 4. Security systems allow access to be granted by day and time of day. Physical access to the building should require key card access, which should provide an audit of Rene being inside the building a lot of week ends. Probably a security procedures lapse.

- 5. Its not so much a problem if an encrypted tape goes AWOL, now is it?

- 6. Three years of security updates and spyware/virus scans didn't find a trojan? What about network monitoring to find unaccounted for network activity? Half a million bank accounts might have taken some bandwidth they should have noticed.

Top 10 Security Breaches of 2008

- 7. **Compass Bank Hard Drive Stolen, 1 Million Accounts Taken**
  - At the sentencing of a former bank programmer at Compass Bank in Birmingham, AL. in March, it was revealed that the accused had stolen a hard drive with 1 million customer records and used it to commit debit-card fraud. James Kevin Real is now serving a 42-month sentence and was ordered to pay back the more than $32,000 that he and an accomplice withdrew from Compass Bank customer accounts. The bank claimed that the customer records contained limited information, but Real was able to create 250 counterfeit debit cards. He used 45 of them to access and withdraw cash before being arrested.
  - At the time of Real's sentencing, Alabama was one of 11 states that didn't require companies to automatically notify customers of data breaches.
- **8. Ski Resort Okemo Suffers Hannaford-Like Data Breach**
  - In an attack similar to what hit Hannaford Brothers in March, the Okemo Ski Resort in Vermont said in April it had been hit by hackers that installed malicious software to capture credit card data as it was being processed at the resort. Law enforcement officials at the time said they were investigating as many as 50 other similar incidents in the Northeast.
- 9. **Retailer Montgomery Ward**
  - Six months after a breach happened at the parent company of the Montgomery Ward website, the company Direct Marketing Services finally began notifying customers that their credit card information was stolen in the hack. At least 51,000 records were stolen out of a database in December, 2007. Direct Marketing said it had promptly contacted its payment processor and Visa and MasterCard, and it also notified the U.S. Secret Service.
- 10. **More Than $5 Million Taken By ATM Capers**
  - The Automatic Teller Machine capers are hitting everywhere. In June, two men were charged with making hundreds of withdrawals from New York City ATMs, grabbing $750,000 in the process, using stolen information from a previous computer intrusion into a Citibank server that processes ATM withdrawals. One of the same accused also allegedly took $5 million in withdrawals from iWire prepaid MasterCard accounts.

Possible causes:

- 7. Why did he have access to a hard drive with financial data? I bet this programmer said he couldn't test unless he had access to real data.
- 8. Companies need to pursue Security updates, security procedures and vigorous security auditing to detect and prevent hacking.
- 9. Another hacking incident, similar suggestions.
- 10. I didn't include fake ATMs in this presentation as it isn't so much about them, but you should pay close attention to any ATM you stick your card into. Frequently the crooks tape an extra reader in front of the normal slot. You may just avoid any ATMs not in secured bank locations.

Data Breach Costs

- Ponemon Institute data;
  - Inappropriately access data: $182/record
  - 31 Companies
    - $226k-$22M
- Cardsystems
  - Processed Credit Card Transactions
  - 40 Million records stolen
  - Out of Business

http://www.ponemon.org/index.php

Written Information Security Plan

- Lists Safeguards to protect Financial Identity Data
  - Administrative
  - Technical
  - Physical
- Handling of Financial Identity Data
  - Creation/Receipt
  - Storage
  - Access, Sharing and Disclosure
  - Transmission
  - Disposal

Written Information Security Plan, cont.

- IT Policies and Procedures
  - Electronic Access
  - Network Security
  - Encryption
- Privacy Security Awareness
  - All Employees
  - Consultants, Third Party service providers
- Security Breach or Unauthorized Access
  - Report suspicious or unauthorized use immediately
  - Open ServiceDesk issue
  - Notify Privacy Officer

What does this mean to me?

- I see what the law is,
- I see the bad things that can happen if my company's data is breached,
- Is that my problem as a "worker bee"?
  - I don't have control over management decisions regarding security
  - They want production data copied into test because there are no extra costs

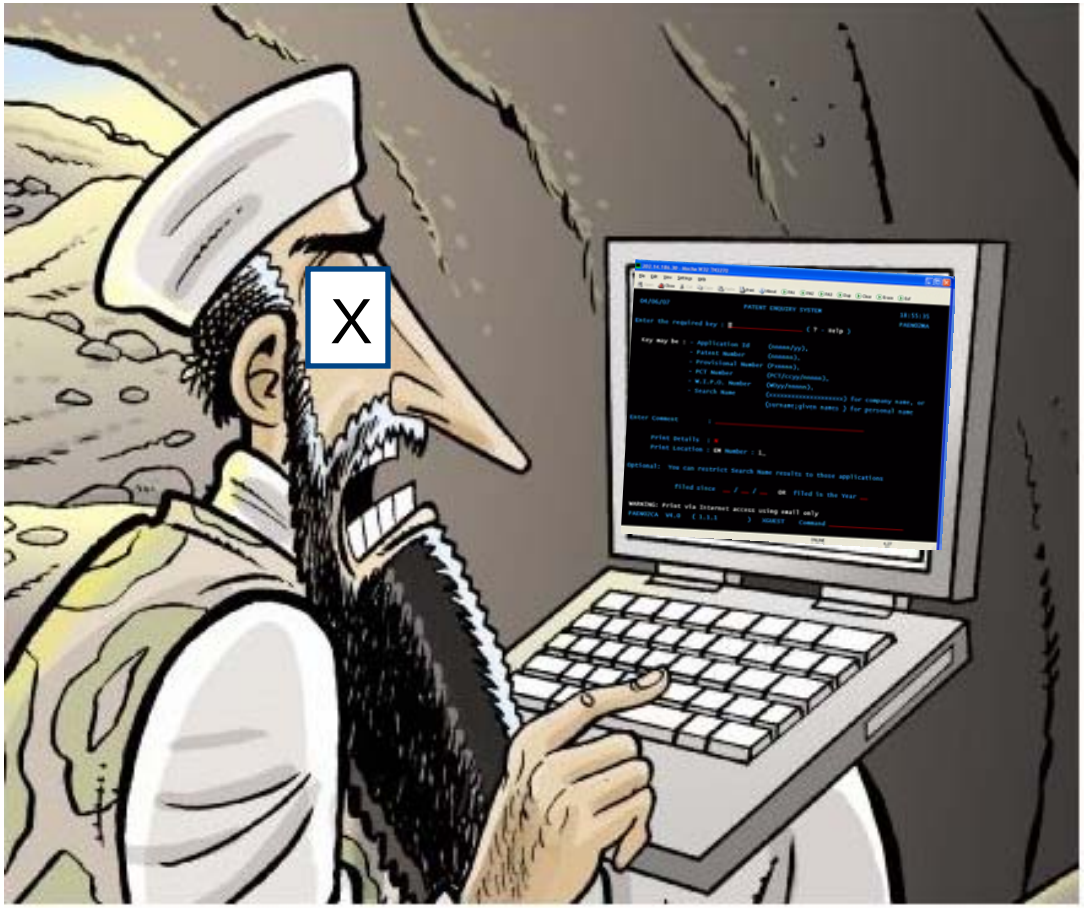
Risks

- Will I go to jail if I copy Production data to test and there is a security breech?
- Probably not
  - Company is responsible – Officers of the Corporation
  - Not much case law against individuals outside of HIPAA violations
  - You might get fired – sacrificial lamb
- Dangers – remember: this is a highly visible problem many people worry about, or have experienced
  - High profile breech +
  - Aggressive DA that wants to run for higher office
    - Dewey/Giuliani: Aggressively prosecuted organized crime
    - Daily: Aggressively prosecuted capital murder cases in Chicago, many overturned afterwards
    - Spitzer: Aggressively prosecuted perceived Wall Street transgressions, now a talk show host for some reason.

So its OK to copy prod data to test?

- Do you have a legitimate reason?
- Laws do not differentiate between production and test
- Production security is usually much more thorough than test systems
- All required safeguards would then need to be implemented on test systems
  - Security
  - Encryption
  - Firewalls
  - Breech Detection/Notification
  - Oversee overseas contractors

Overseeing overseas contractors?



Production vs. Test Security

- Production
  - Tightly locked down
  - Access on a “need to know/update” basis
  - Few authorized users
  - Authorized users often in a few secured locations
  - Tracing/Debugging/Printing of data limited
- Test Systems
  - Frequently wide open, at least usually not as stringent rules
  - Adding more security potentially stops applications from being productive
  - Dozens to thousands of employees or contractors have access
  - Access from many to unlimited places
    - Global access
    - Home access
  - Lots of debugging, sensitive data printed or on screens in cubicles

Preventing unauthorized access to test data

- Prevent “shoulder surfing” to protected data
- Safe disposal or printed traces and dumps to prevent “dumpster diving”
- Security procedure training for all who have access to data or facility
- Encrypt all data being transmitted or copied to external devices
- Is the ease of copying production data worth the risk and the cost?

Do you really need Prod data to test with?

- Do you programs care or know the difference:
  - Valid
    - Credit card
    - SSN
    - Consistent data i.e. zip code, area code, state
    - D/L #
    - etc.
- If these fields are treated as character and just passed along, any random value will do

They just need to trust us

- If they didn't trust you, you wouldn't be working there!
- Your auditors need to sign off on your security procedures
- Do you want to be one of the "usual suspects"?
- Frequently people ask for:
  - Root access
  - Non cancel access or
  - More access than what their job requires
- Test systems have so may people with possible access, wouldn't you want to be able to prove you had no responsibility for a breach?

Allegory of the master catalog

- Years ago, XA allowed ESM control of Master Catalog
- Previously we used a password
- Stopped using password, used ACF2
- Master Catalog should only have what's needed to IPL
  - Usually just SYS1 datasets
  - Aliases to user catalogs
  - We had 100's of other datasets cataloged
- Without a password, if SYSPROGs mis-keyed a high level qualifier, datasets were cataloged into master catalog
- Solution:
  - ESM to protect master catalog for everybody
  - Password to prevent authorized users from making a mistake
- Moral: Don't give me more access than I need to do my job

Strategies to generate non-sensitive test data

- Buy software to generate test data
  - Data generators
    - RI?
    - Valid encoded values
  - Data Extractors
    - RI
    - Subset
    - Scrambling
- Write code to generate test data
- Either way:
  - Copy and scramble
  - Generate random data

Strategies

- Generate valid test data from scratch
  - Does program logic depend on valid, consistent data?
- Copy production into test, then scramble sensitive fields
  - Scramble account numbers
  - Still may need to enforce consistency

Scrambling vs. Encryption

- Encryption implies it can be decoded
  - Don’t want hackers or unauthorized used to get original production data back
- Want a one way process
  - Still valid for the data type
  - Preserves RI

Sausage Grinder:

You want an irreversible scramble
not an “encryption” process

No way for the original
production data to be recreated.

Can’t to turn the crank backwards
and get the pig back.

How I built my demo database

- Rexx: Quick to write, once you have a first few working examples
- Arrays of common:
  - First names
  - Last names
  - Street names
  - City names
- Array of states
- Randomly generate numbers for:
  - Phone number
  - Zip Code
  - CC numbers

What's not perfect with my approach?

- State/Zip Code/Area Code/Exchange won't be consistent
- First/Last names were picked for a recent year
  - Popular names change over time
  - Demographics change over time
- Credit card, phone numbers and zip codes are not valid

- Do the applications know the difference?
- Do your applications process differently if the name is:
  - Abrigal Adams
  - Anferny Adamsky
  - Angel Acosta
  - Axxywr Azsdfry

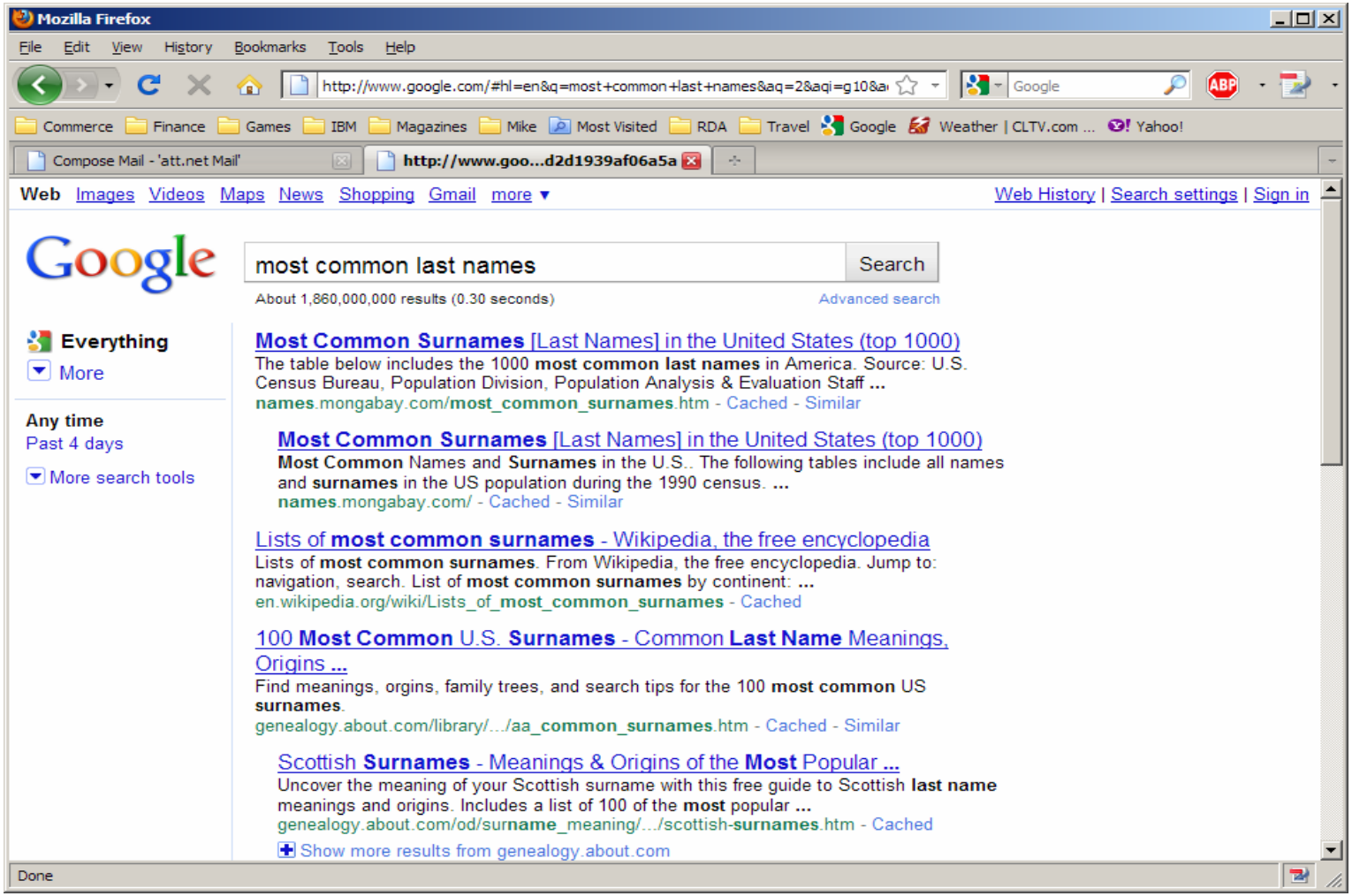
Name Changes for demographics

- US developed DB2 column definitions usually have fields for first and last name
- Sometimes middle name
- As we become more global:
  - Typical US name fields are too small for some nationalities
  - Latin American names include the family name of the father and the mother along with a middle name: Miguel Jose Gutierrez Rodriguez
- These are more likely database design changes rather that issues of creating test databases from production data

Creating safe test databases

- Won't require all the legal safeguards
- No financial/HIPAA data can be compromised
- Create new data –or- Copy prod and update
- Need to understand how the applications process the data
  - Does the zip code have to agree with the area code and exchange
  - Are these numbers treated as unrelated numeric data?
- Are you just concerned that binds in test will pick the same access path as when bound in production?
  - Copy prod stats to your test objects

Pick common, valid data



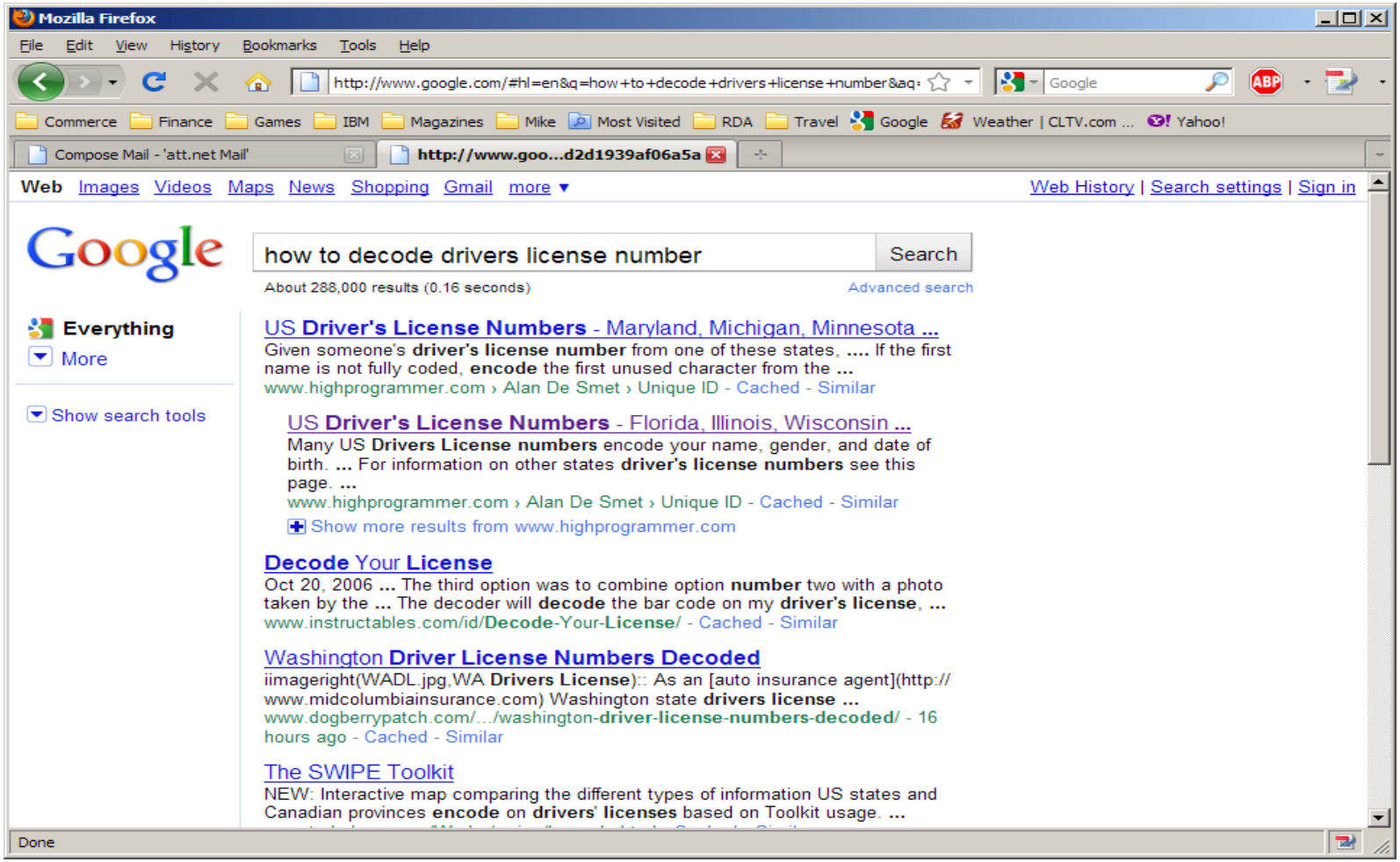
Consistent Data

- Area, Exchange code by City, State
- Zip code by City, State
- Drivers License matches state
- Multiple name fields:
  - Do Initials match generated name
  - Formal vs. familiar name matching
    - Michael – Mike
    - Robert – Bob
    - What about names from other countries and cultures?

ANSI Standard X4.13-1983

- System used by most national credit-card systems.
- The first digit identifies the type of card:
  - 3 is T&E cards
  - 4 is Visa Card
  - 5 is MasterCard
  - 6 is Discover Card.
- The structure of the card number varies by credit card system:
  - American Express - Digits three and four are type and currency, digits five through 11 are the account number, digits 12 through 14 are the card number within the account and digit 15 is a check digit.
  - Visa - Digits two through six are the bank number, digits seven through 12 or seven through 15 are the account number and digit 13 or 16 is a check digit.
  - MasterCard - Digits two and three, two through four, two through five or two through six are the bank number (depending on whether digit two is a 1, 2, 3 or other).
- The digits after the bank number up through digit 15 are the account number, and digit 16 is a check digit.

Drivers Licenses



For example: Illinois Drivers License

- http://www.highprogrammer.com/alan/numbers/dl_us_shared.html
- SSSF-FFYY-DDDS

Segment	Example/Description	
SSSS	F255	Soundex code
FFF	921	First name, middle initial
YY	50	Year of birth
DDD	094	Day and month of birth
NN	03	Overflow (not all states use this)

DBAs behind bars, as it should be

Questions?

- Session Code G07 May 4, 2011 9:45 – 10:45am
- Robert Andresen
- Robert.Andresen@ca.com