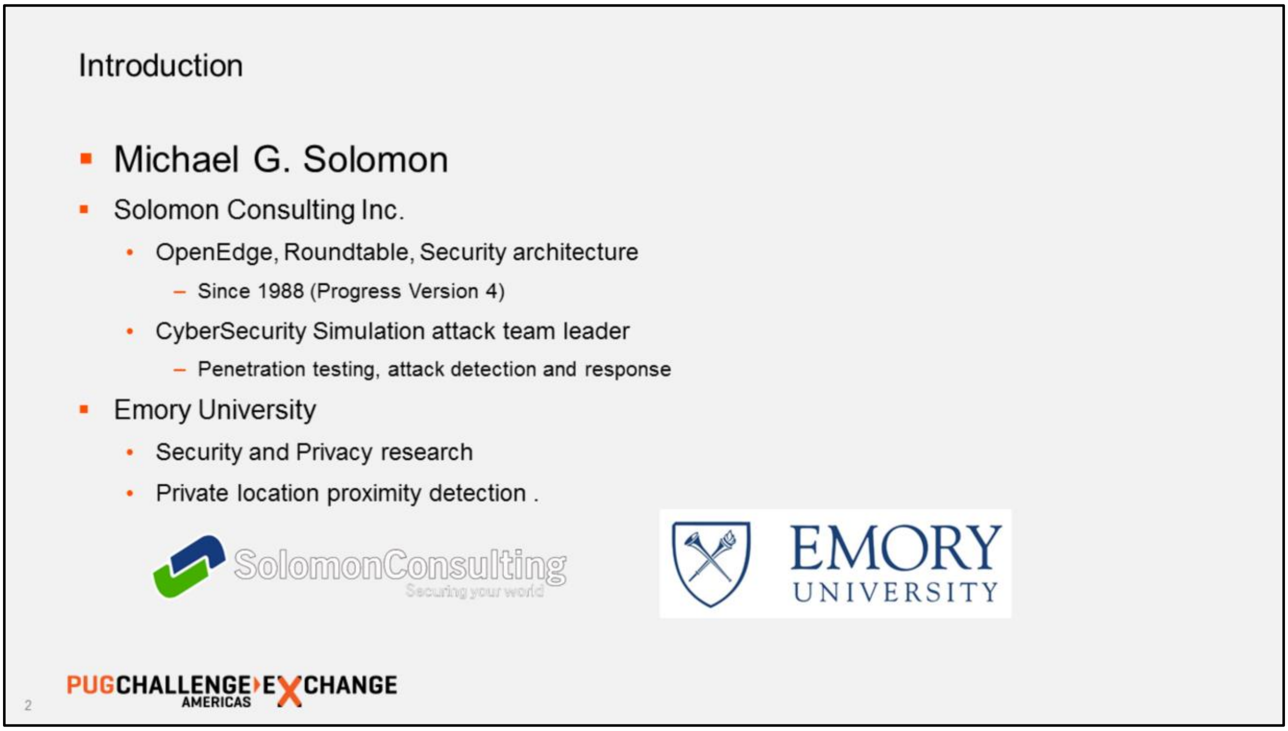
Privacy, Confidentiality, and Security

What's the difference?

Michael Solomon

Solomon Consulting Inc.

PUG CHALLENGE EXCHANGE AMERICAS

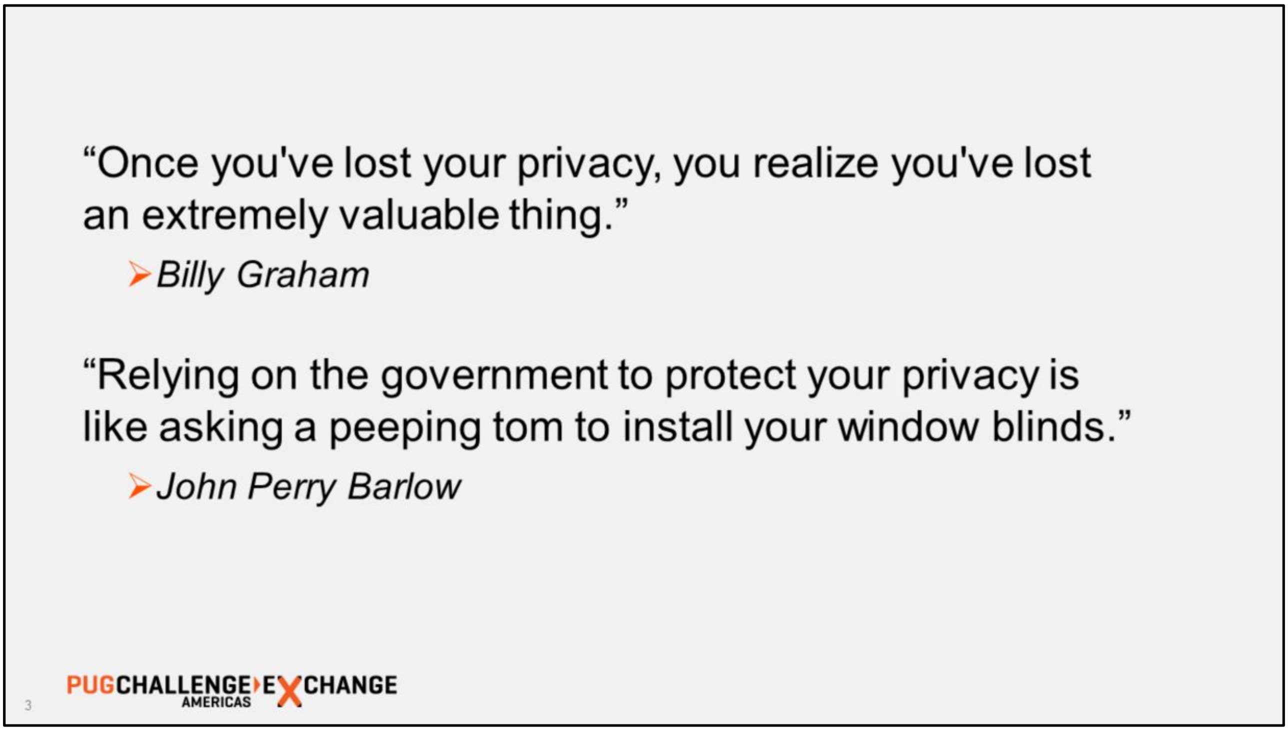


Introduction

- **Michael G. Solomon**
- Solomon Consulting Inc.
  - OpenEdge, Roundtable, Security architecture
    - Since 1988 (Progress Version 4)
  - CyberSecurity Simulation attack team leader
    - Penetration testing, attack detection and response
- Emory University
  - Security and Privacy research
  - Private location proximity detection .

SolomonConsulting
Securing your world

EMORY
UNIVERSITY



“Once you've lost your privacy, you realize you've lost an extremely valuable thing.”

- *Billy Graham*

“Relying on the government to protect your privacy is like asking a peeping tom to install your window blinds.”

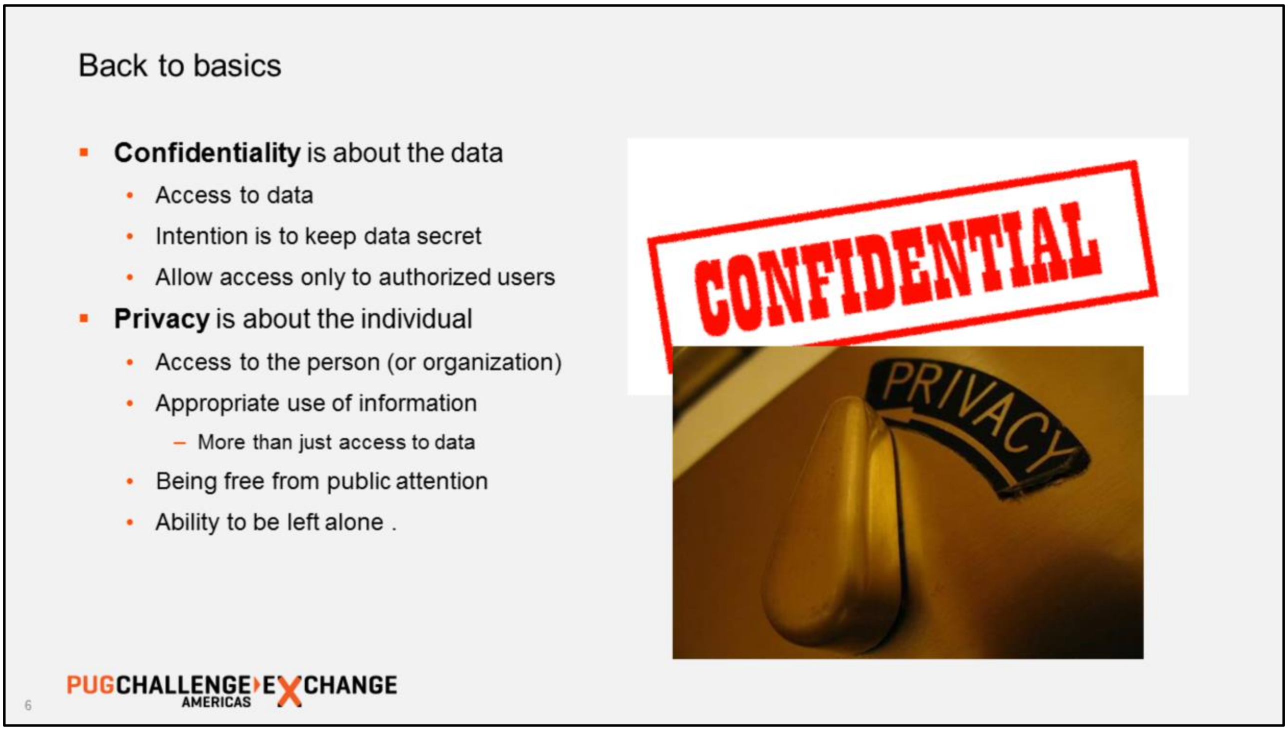
- *John Perry Barlow*

Privacy and Confidentiality

- Common terms in the legal and medical domains
- Often confused
  - Worse yet, interchanged!!
  - Don't be fooled – they are different
- Information Security perspective
  - Subtle differences from commonly accepted meanings
  - It may change your approach to protecting data .

First things first - why do we care?

- One simple word: **Liability**
- The way you handle data can
  - Protect you (and your users) from attacks
  - Protect you from damages (in court)
  - Separate you from competitors
  - Allow your customers to worry about other organizations
- “An ounce or prevention …”
- Good security isn’t cheap
  - But its WAY cheaper than bad security! .

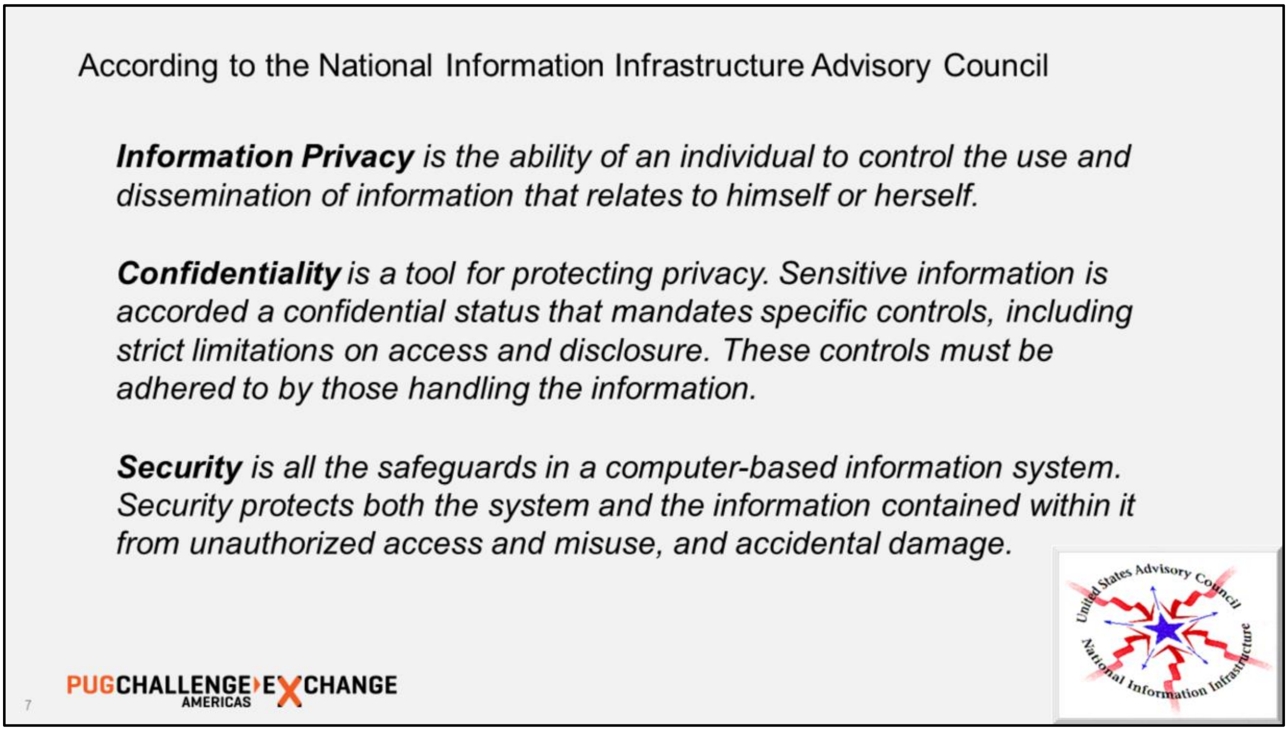


Back to basics

- **Confidentiality** is about the data
  - Access to data
  - Intention is to keep data secret
  - Allow access only to authorized users
- **Privacy** is about the individual
  - Access to the person (or organization)
  - Appropriate use of information
    - More than just access to data
  - Being free from public attention
  - Ability to be left alone .

CONFIDENTIAL
PRIVACY

http://www.differencebetween.com/difference-between-confidentiality-and-vs-privacy/
http://www.research.uky.edu/ori/ORIForms/32-Privacy-vs-Confidentiality.pdf



According to the National Information Infrastructure Advisory Council

Information Privacy *is the ability of an individual to control the use and dissemination of information that relates to himself or herself.*

Confidentiality *is a tool for protecting privacy. Sensitive information is accorded a confidential status that mandates specific controls, including strict limitations on access and disclosure. These controls must be adhered to by those handling the information.*

Security *is all the safeguards in a computer-based information system. Security protects both the system and the information contained within it from unauthorized access and misuse, and accidental damage.*

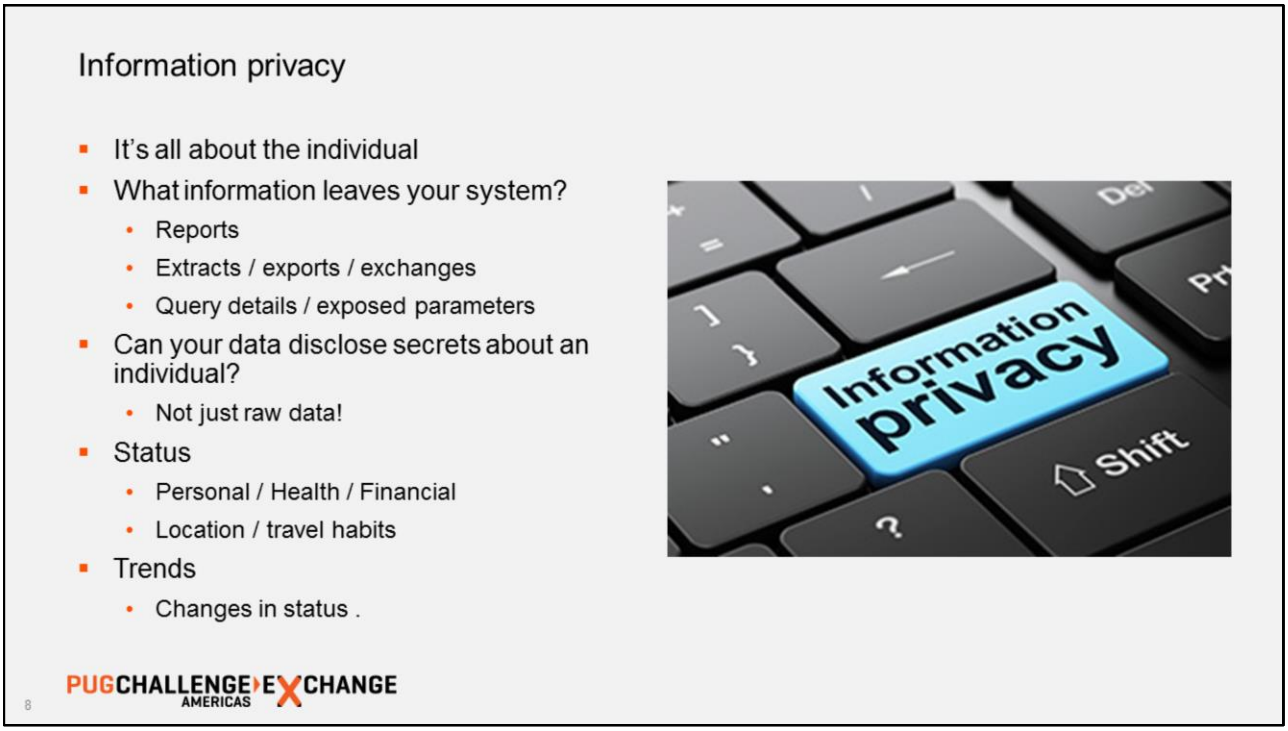
United States Advisory Council
National Information Infrastructure

PUGCHALLENGE AMERICAS EXCHANGE

7

http://www.ntia.doc.gov/legacy/reports/telemed/privacy.htm

The National Information Infrastructure Advisory Council, "Common Ground: Fundamental Principles for the National Information Infrastructure," March 1995.

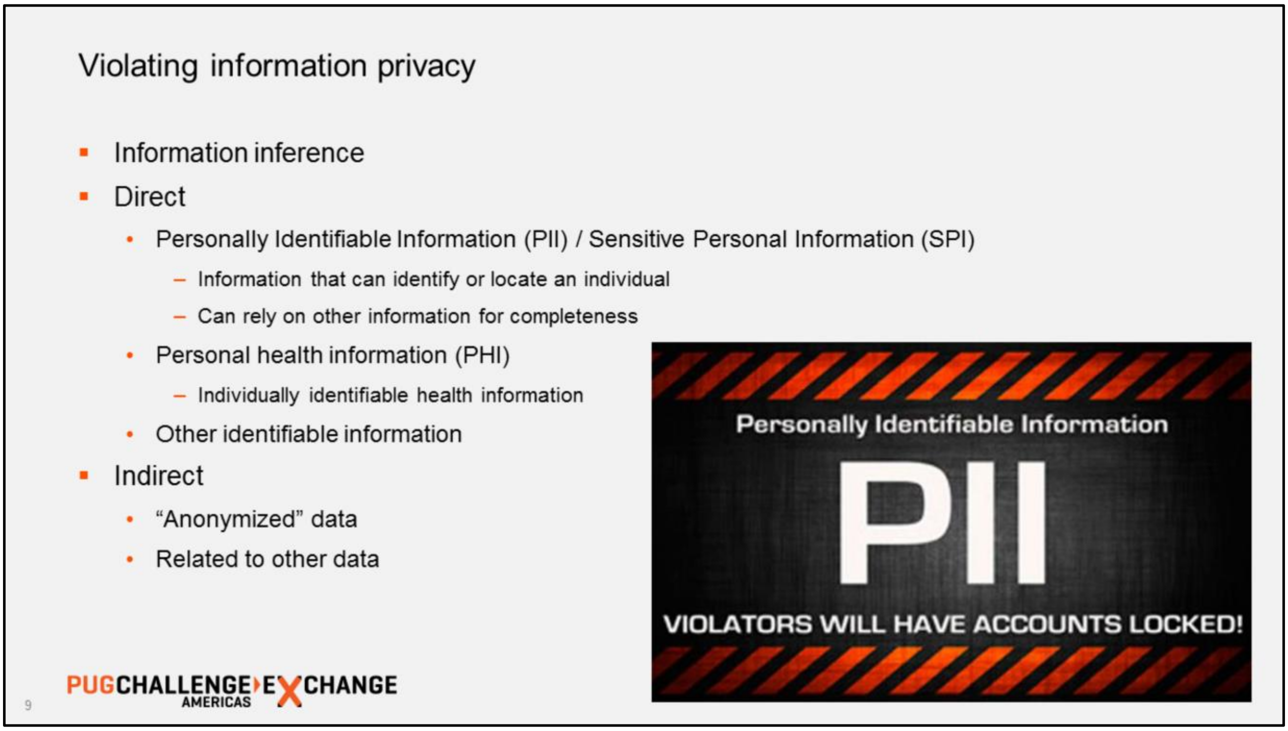


Information privacy

- It’s all about the individual
- What information leaves your system?
  - Reports
  - Extracts / exports / exchanges
  - Query details / exposed parameters
- Can your data disclose secrets about an individual?
  - Not just raw data!
- Status
  - Personal / Health / Financial
  - Location / travel habits
- Trends
  - Changes in status .

Information privacy
Shift
Del

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

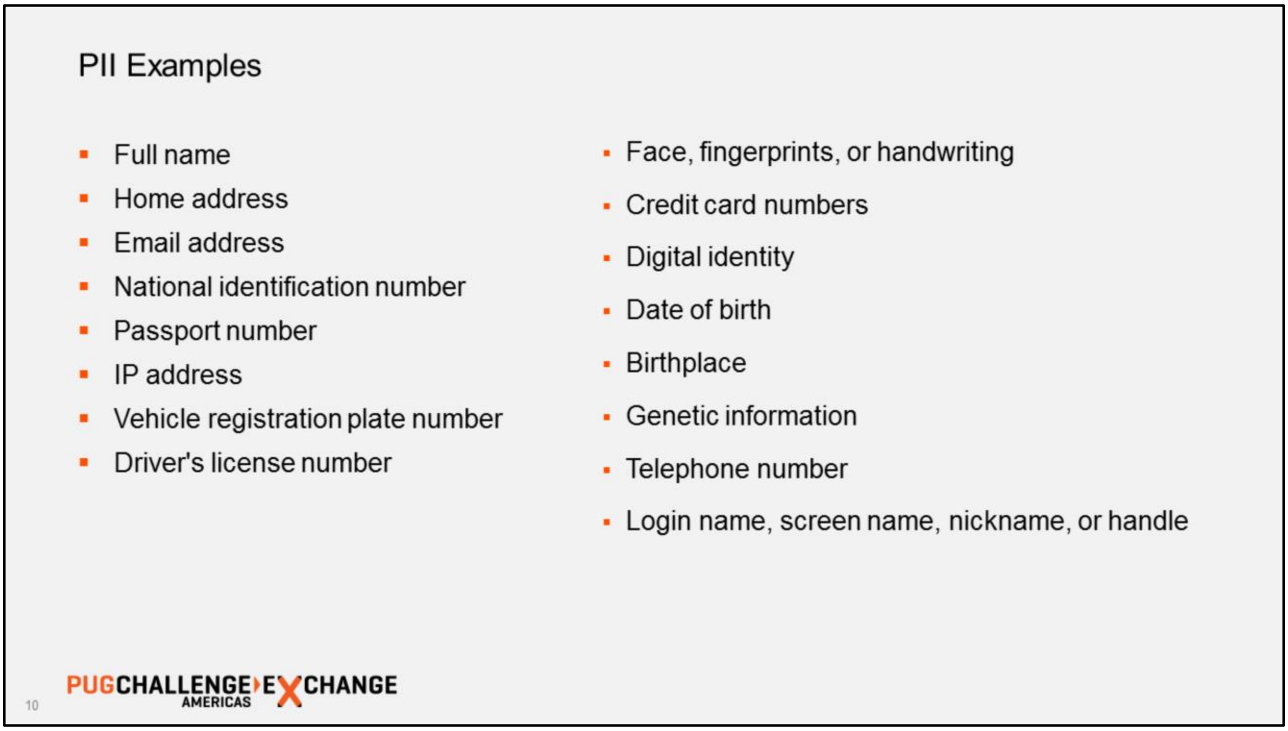


Violating information privacy

- Information inference
- Direct
  - Personally Identifiable Information (PII) / Sensitive Personal Information (SPI)
    - Information that can identify or locate an individual
    - Can rely on other information for completeness
  - Personal health information (PHI)
    - Individually identifiable health information
  - Other identifiable information
- Indirect
  - "Anonymized" data
  - Related to other data

Personally Identifiable Information
PII
VIOLATORS WILL HAVE ACCOUNTS LOCKED!

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



PII Examples

- Full name
- Home address
- Email address
- National identification number
- Passport number
- IP address
- Vehicle registration plate number
- Driver's license number
- Face, fingerprints, or handwriting
- Credit card numbers
- Digital identity
- Date of birth
- Birthplace
- Genetic information
- Telephone number
- Login name, screen name, nickname, or handle

10

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

PHI Examples

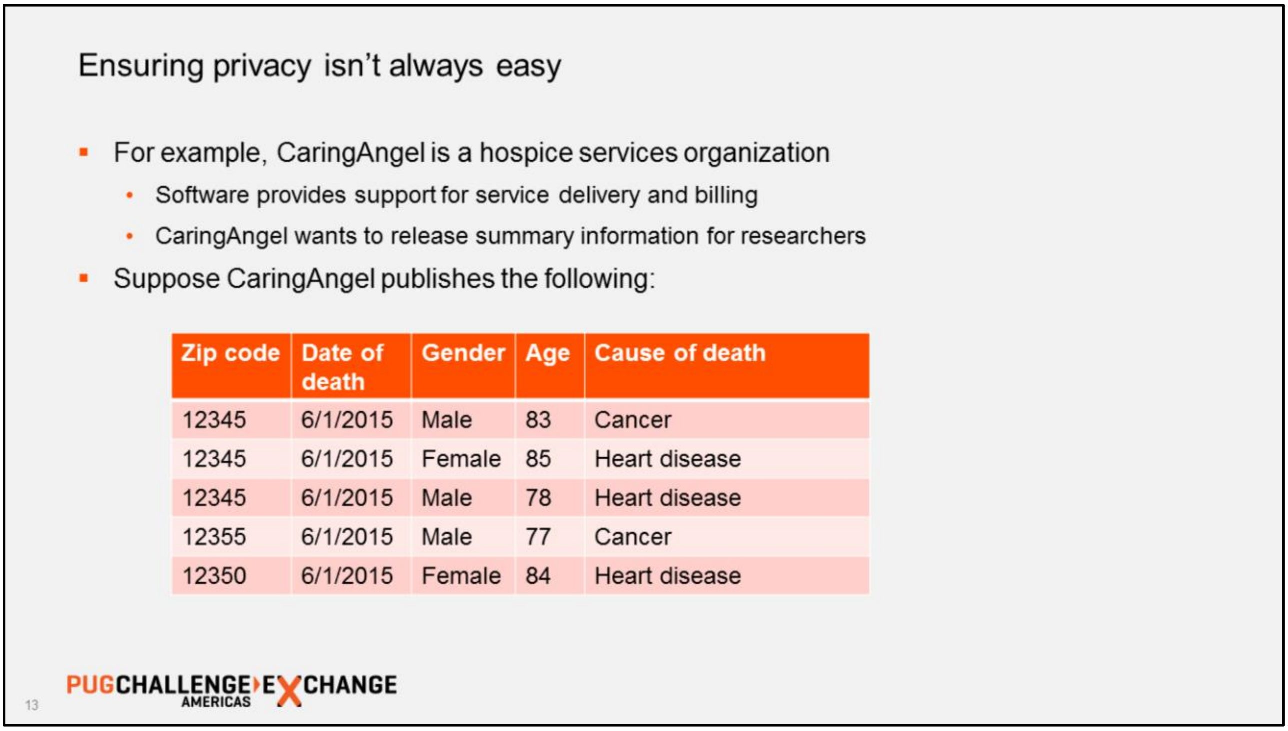
- Name
- All elements of dates except Year
- SSN
- Driver's License Number
- Geographic subdivisions smaller than a State
- URL's and IP's
- Vehicle Identifiers including VIN and License Plates
- Phone numbers

An easy fix?

- Just hide PII, PHI, etc, right?
  - Obfuscate
  - Anonymize
- Unfortunately, no

  For confidentiality, maybe

  Not to protect privacy
- Even “safe” data can violate privacy
- Let’s see how …



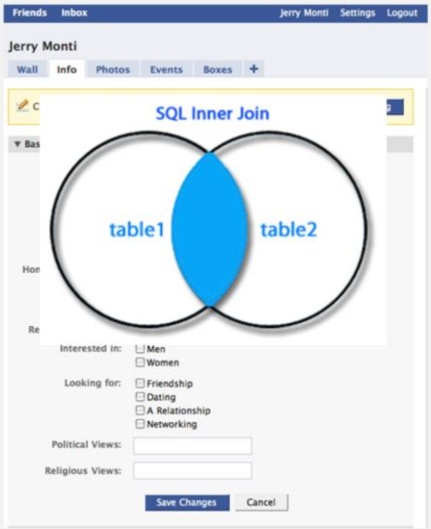
Ensuring privacy isn't always easy

- For example, CaringAngel is a hospice services organization
  - Software provides support for service delivery and billing
  - CaringAngel wants to release summary information for researchers
- Suppose CaringAngel publishes the following:

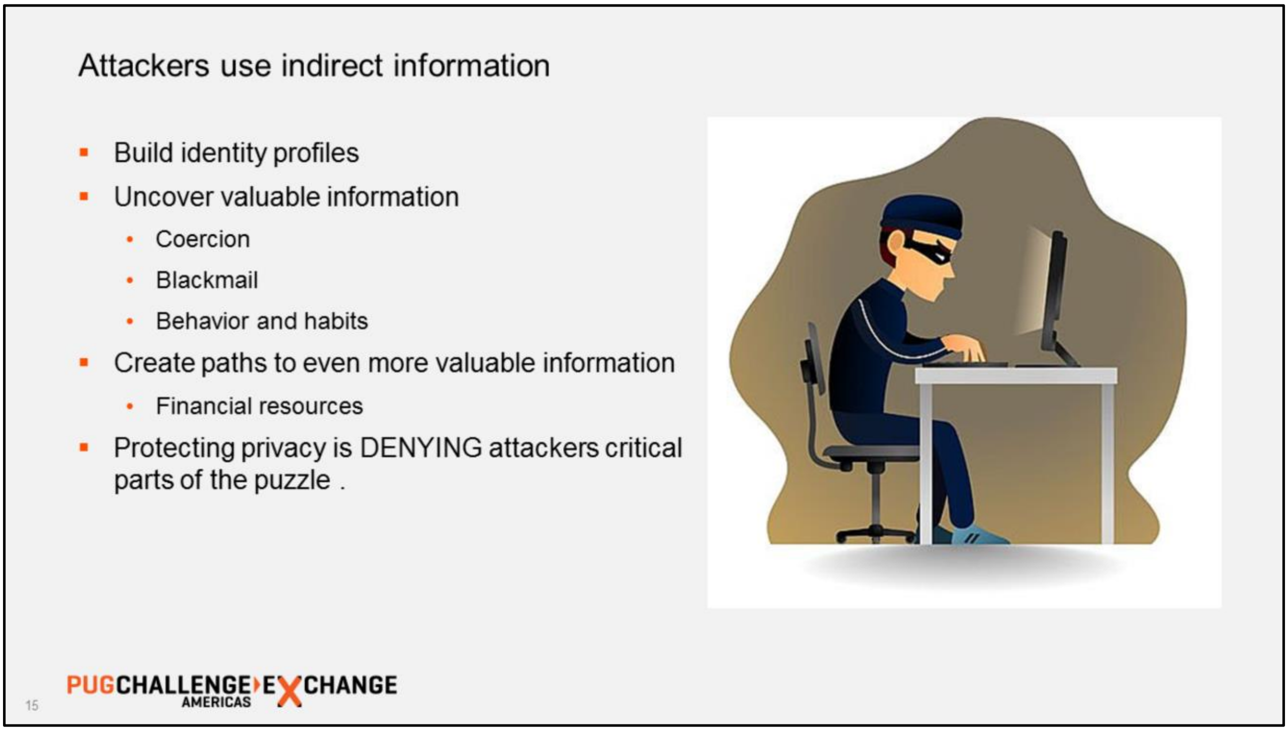
Zip code	Date of death	Gender	Age	Cause of death
12345	6/1/2015	Male	83	Cancer
12345	6/1/2015	Female	85	Heart disease
12345	6/1/2015	Male	78	Heart disease
12355	6/1/2015	Male	77	Cancer
12350	6/1/2015	Female	84	Heart disease

Indirect information

- Cross reference with other information
  - “Other information” is often publicly available
  - Think of a relational JOIN
- Criminal records
- Voter registration records
- Social media sites
  - Gold mines for voluntarily provided personal information
- Many, many others
  - http://www.kn.att.com/wired/fil/pages/liststudentpe3.html
- In the previous slide’s example, how about obituaries? .

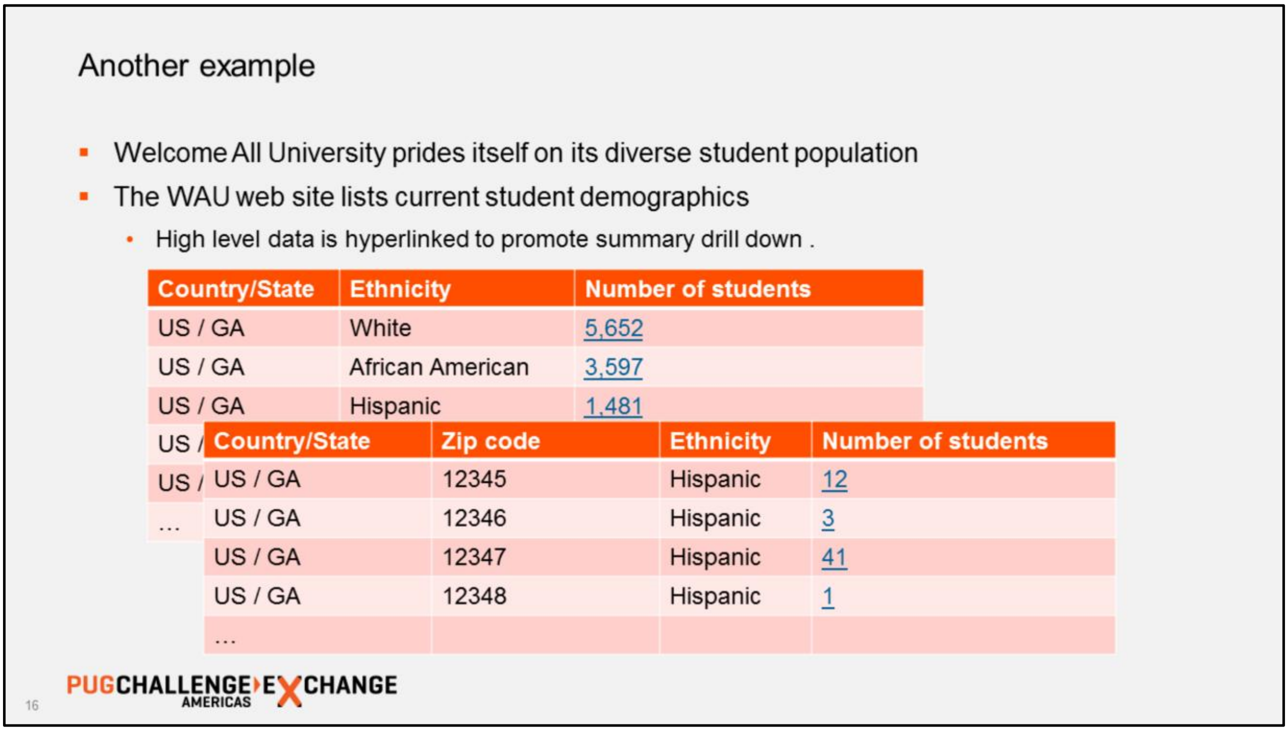


http://www.kn.att.com/wired/fil/pages/liststudentpe3.html



Attackers use indirect information

- Build identity profiles
- Uncover valuable information
  - Coercion
  - Blackmail
  - Behavior and habits
- Create paths to even more valuable information
  - Financial resources
- Protecting privacy is DENYING attackers critical parts of the puzzle .

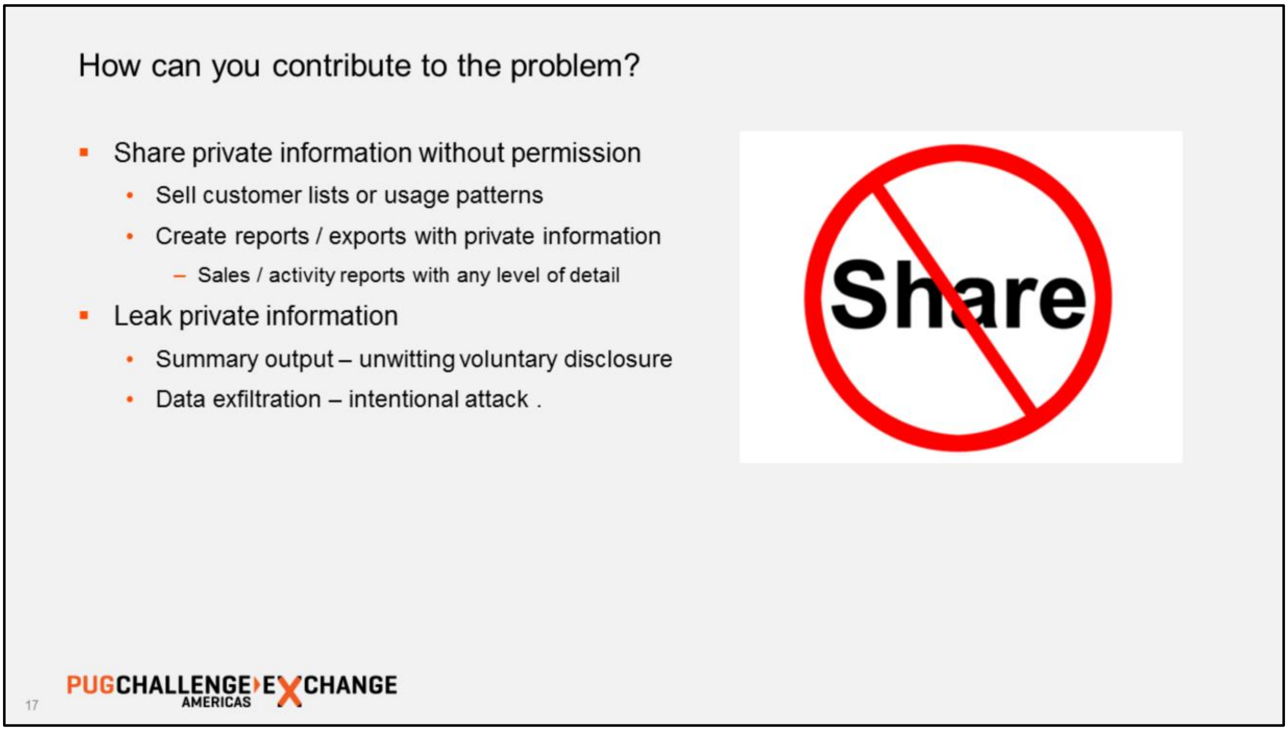


Another example

- Welcome All University prides itself on its diverse student population
- The WAU web site lists current student demographics
  - High level data is hyperlinked to promote summary drill down .

Country/State	Ethnicity	Number of students
US / GA	White	5,652
US / GA	African American	3,597
US / GA	Hispanic	1,481
US /		
US /		
…		

Country/State	Zip code	Ethnicity	Number of students
US / GA	12345	Hispanic	12
US / GA	12346	Hispanic	3
US / GA	12347	Hispanic	41
US / GA	12348	Hispanic	1
…			

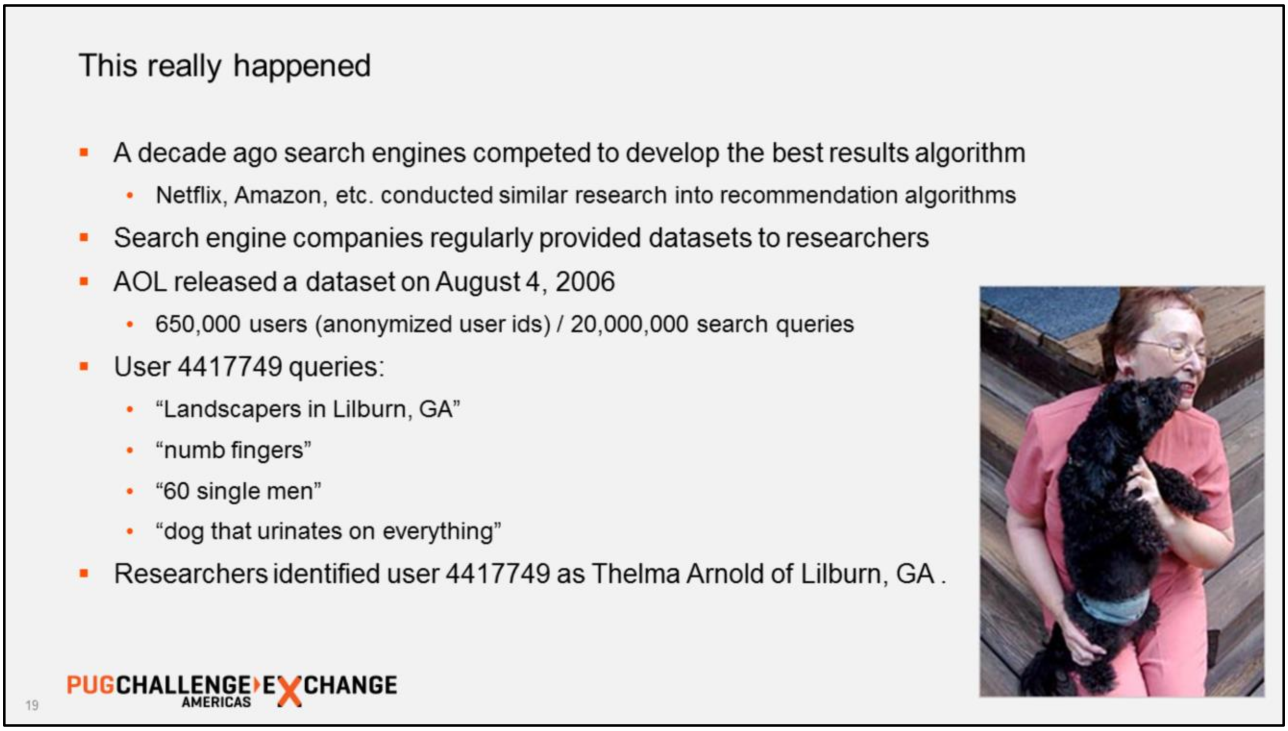


How can you contribute to the problem?

- Share private information without permission
  - Sell customer lists or usage patterns
  - Create reports / exports with private information
    - Sales / activity reports with any level of detail
- Leak private information
  - Summary output – unwitting voluntary disclosure
  - Data exfiltration – intentional attack .

Indirect information example

- Is John trustworthy?
  - Informal relationship question
  - Arrest records are often online – easy to query
  - Public notices of many legal proceedings
  - Don't forget about search engines and social media
  - Inference is not difficult
- The first step is to define "trustworthy"
- Lots of information is already available
- Don't provide the linkage!!
  - Perhaps as simple as name, home state/county, and age?



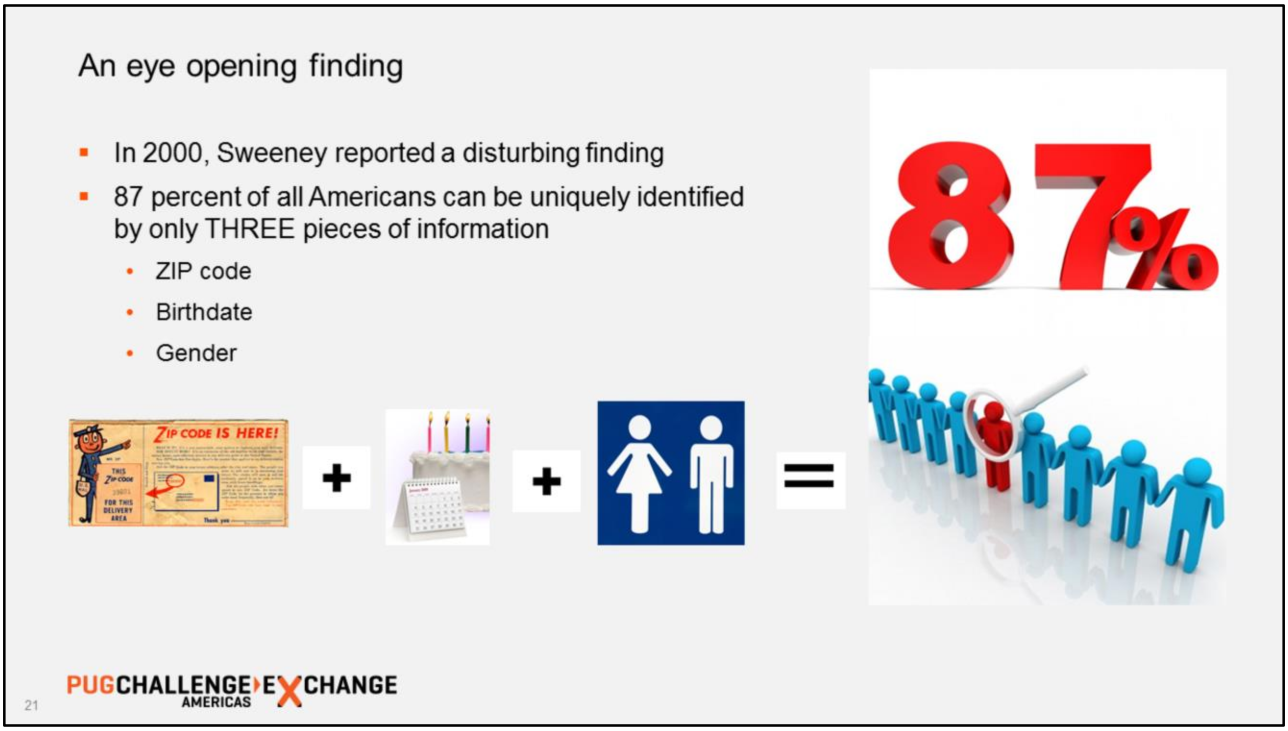
This really happened

- A decade ago search engines competed to develop the best results algorithm
  - Netflix, Amazon, etc. conducted similar research into recommendation algorithms
- Search engine companies regularly provided datasets to researchers
- AOL released a dataset on August 4, 2006
  - 650,000 users (anonymized user ids) / 20,000,000 search queries
- User 4417749 queries:
  - “Landscapers in Lilburn, GA”
  - “numb fingers”
  - “60 single men”
  - “dog that urinates on everything”
- Researchers identified user 4417749 as Thelma Arnold of Lilburn, GA .

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

This happened too

- The Massachusetts Group Insurance Commission
  - Released “anonymized” data of every state employee’s hospital visits
    - Included zip code and birthdate
  - Intention was to promote research
- Latanya Sweeney – MIT CS PhD student
  - Re-identification work
  - Successfully identified current Massachusetts’ Governor William Weld’s medical records
- How?
  - Purchased Cambridge voter rolls for $20
  - 6 in Cambridge w/Weld’s birthdate (3 were women)
  - Only 1 of the remaining 3 lived in Weld’s zip code
    - A MATCH!! .

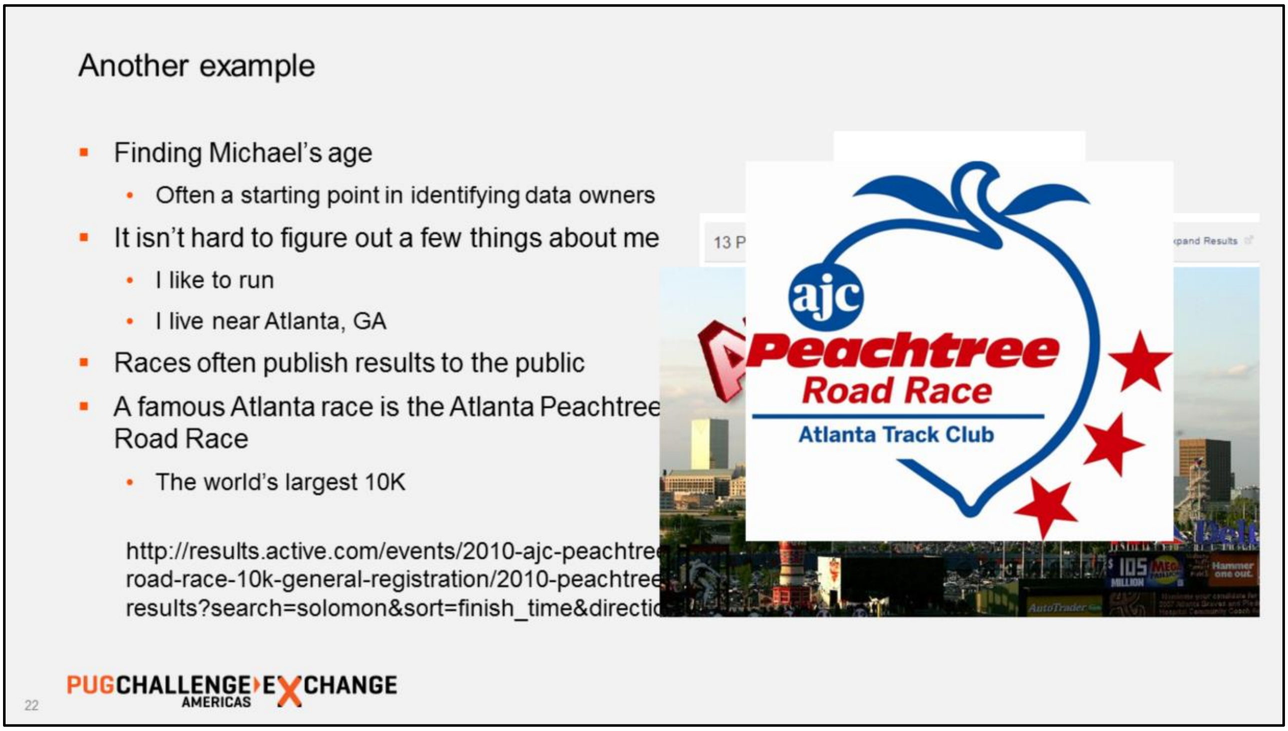


An eye opening finding

- In 2000, Sweeney reported a disturbing finding
- 87 percent of all Americans can be uniquely identified by only THREE pieces of information
  - ZIP code
  - Birthdate
  - Gender

ZIP CODE IS HERE!
THIS ZIP CODE FOR THIS DELIVERY AREA
+
+
=
87%
PUGCHALLENGE EXCHANGE AMERICAS
21

http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/

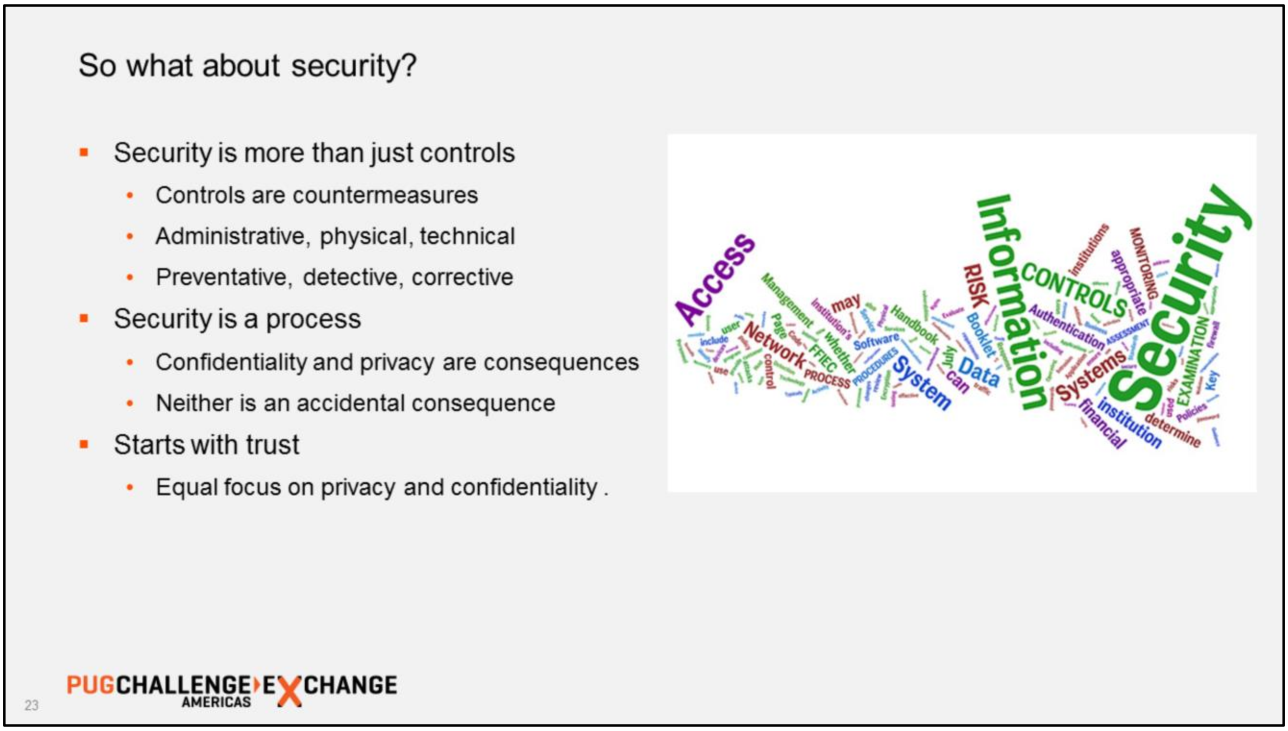


Another example

- Finding Michael's age
  - Often a starting point in identifying data owners
- It isn't hard to figure out a few things about me
  - I like to run
  - I live near Atlanta, GA
- Races often publish results to the public
- A famous Atlanta race is the Atlanta Peachtree Road Race
  - The world's largest 10K

http://results.active.com/events/2010-ajc-peachtree road-race-10k-general-registration/2010-peachtree results?search=solomon&sort=finish_time&directio

http://results.active.com/events/2010-ajc-peachtree-road-race-10k-general-registration/2010-peachtree-results?search=solomon&sort=finish_time&direction=asc



So what about security?

- Security is more than just controls
  - Controls are countermeasures
  - Administrative, physical, technical
  - Preventative, detective, corrective
- Security is a process
  - Confidentiality and privacy are consequences
  - Neither is an accidental consequence
- Starts with trust
  - Equal focus on privacy and confidentiality .

Security
Information
Access
CONTROLS
RISK
Systems
System
Data
Network
Authentication
MONITORING
appropriate
EXAMINATION
institution
financial
determine
Management
FFIEC
Software
Handbook
Booklet
PROCESS
PROCEDURES
may
whether
control
can
Key
Policies

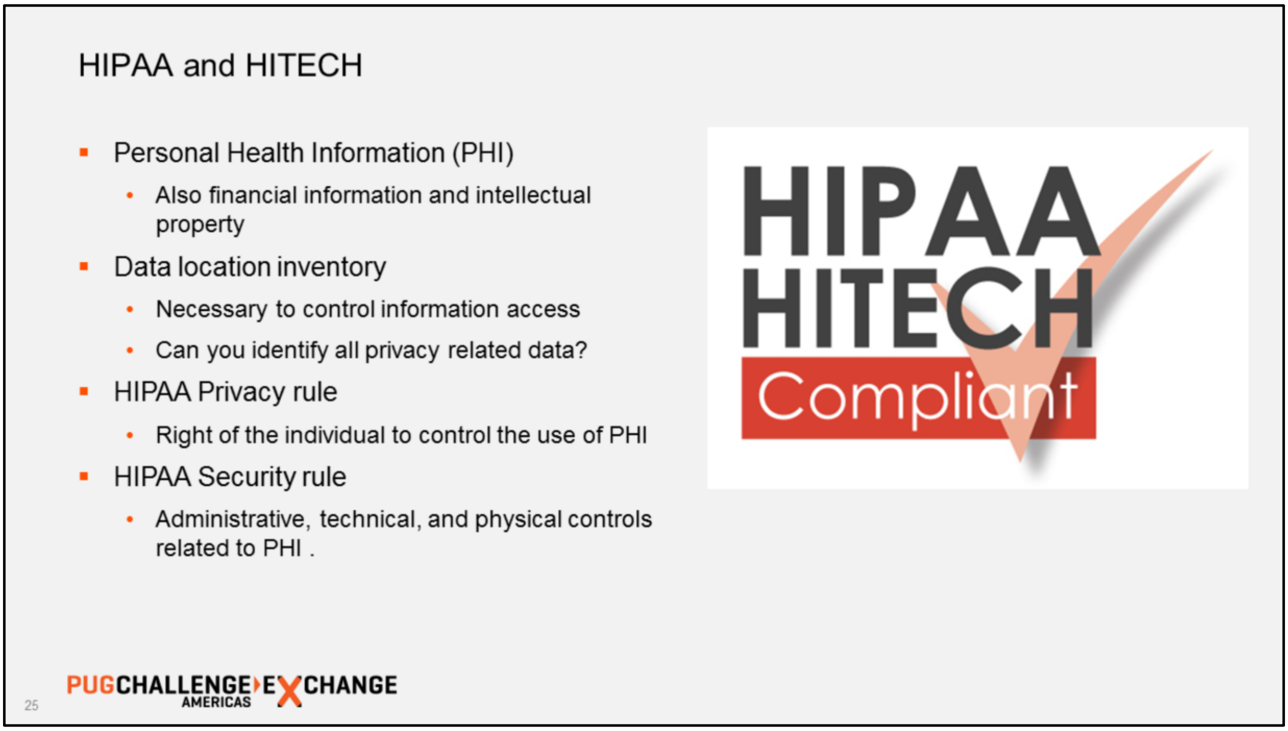
http://www.informationshield.com/papers/Privacy%20and%20Security%20-%20Herold.pdf
http://blog.propay.com/index.php/2010/09/15/what-is-the-difference-between-security-and-privacy/
http://www.washingtonpost.com/blogs/federal-eye/wp/2015/05/26/hackers-stole-personal-information-from-104000-taxpayers-irs-says/

Privacy laws

- National laws – specific to each country
- United States laws
  - Health Insurance Portability and Accountability Act (HIPAA)
  - Financial services Modernization Act (GLB)
  - Final Rule on Privacy of Consumer Financial Information
  - Fair Credit Reporting Act (FCRA)
  - Fair Debt Collections Practices Act (FDCPA)
- More proposed laws affecting privacy .

PRIVACY
LAW

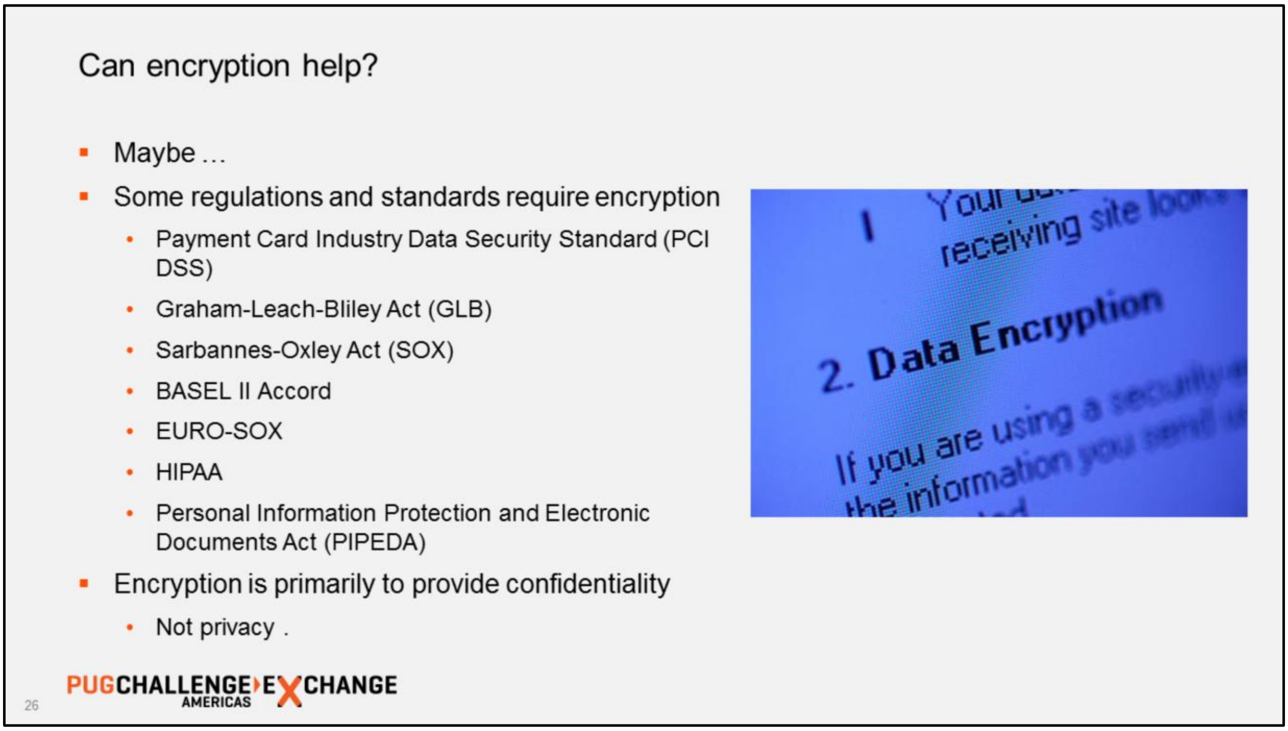
http://en.wikipedia.org/wiki/Privacy_law
http://en.wikipedia.org/wiki/Privacy_laws_of_the_United_States



HIPAA and HITECH

- Personal Health Information (PHI)
  - Also financial information and intellectual property
- Data location inventory
  - Necessary to control information access
  - Can you identify all privacy related data?
- HIPAA Privacy rule
  - Right of the individual to control the use of PHI
- HIPAA Security rule
  - Administrative, technical, and physical controls related to PHI .

http://www.privacy.wv.gov/tips/Pages/HIPAAPrivacyHIPAASecurity.aspx

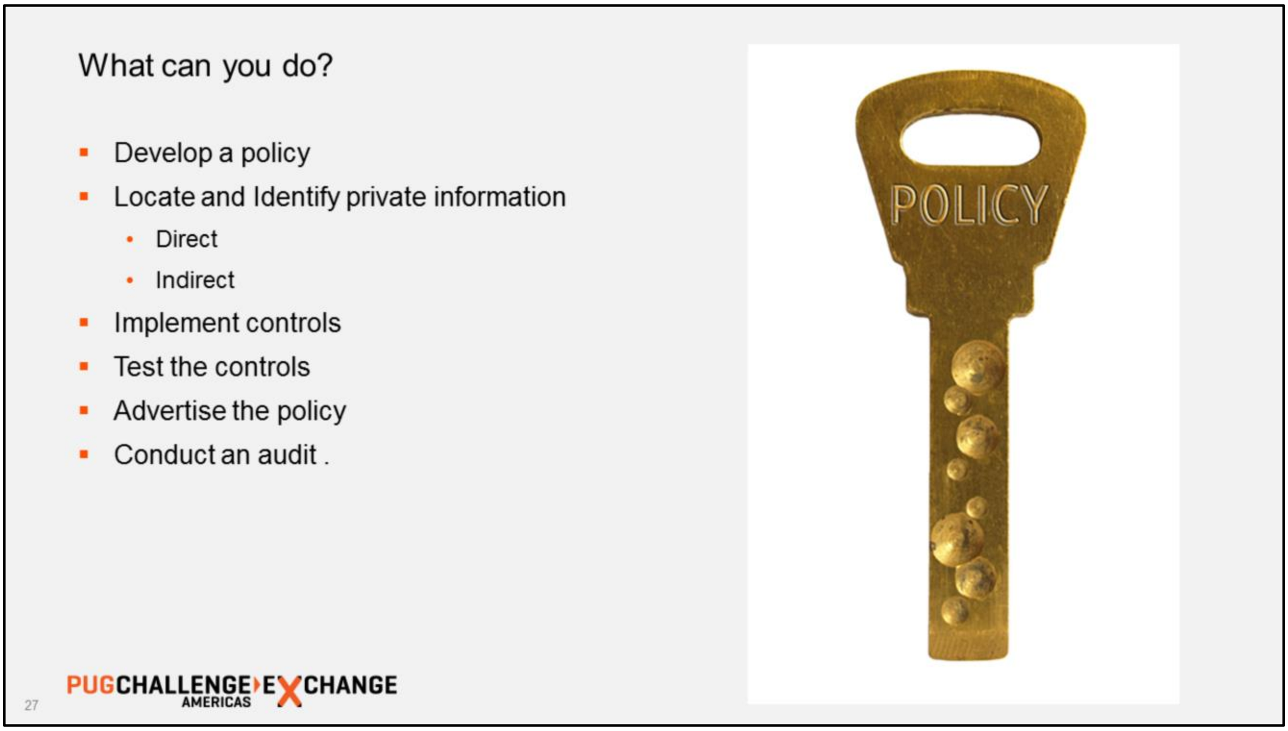


Can encryption help?

- Maybe …
- Some regulations and standards require encryption
  - Payment Card Industry Data Security Standard (PCI DSS)
  - Graham-Leach-Bliley Act (GLB)
  - Sarbannes-Oxley Act (SOX)
  - BASEL II Accord
  - EURO-SOX
  - HIPAA
  - Personal Information Protection and Electronic Documents Act (PIPEDA)
- Encryption is primarily to provide confidentiality
  - Not privacy .

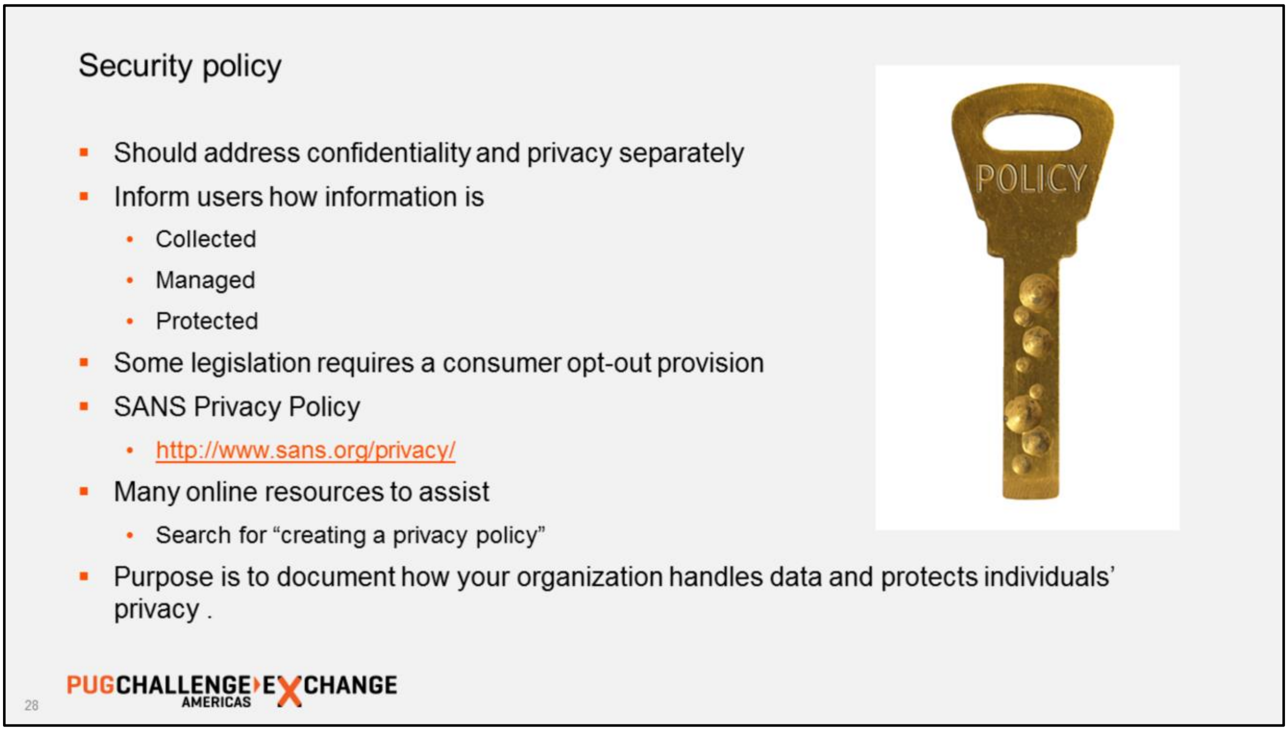
2. Data Encryption
receiving site looks
If you are using a security
the information you send

http://www.sans.org/reading-room/whitepapers/analyst/regulations-standards-encryption-applies-34675



What can you do?

- Develop a policy
- Locate and Identify private information
  - Direct
  - Indirect
- Implement controls
- Test the controls
- Advertise the policy
- Conduct an audit .



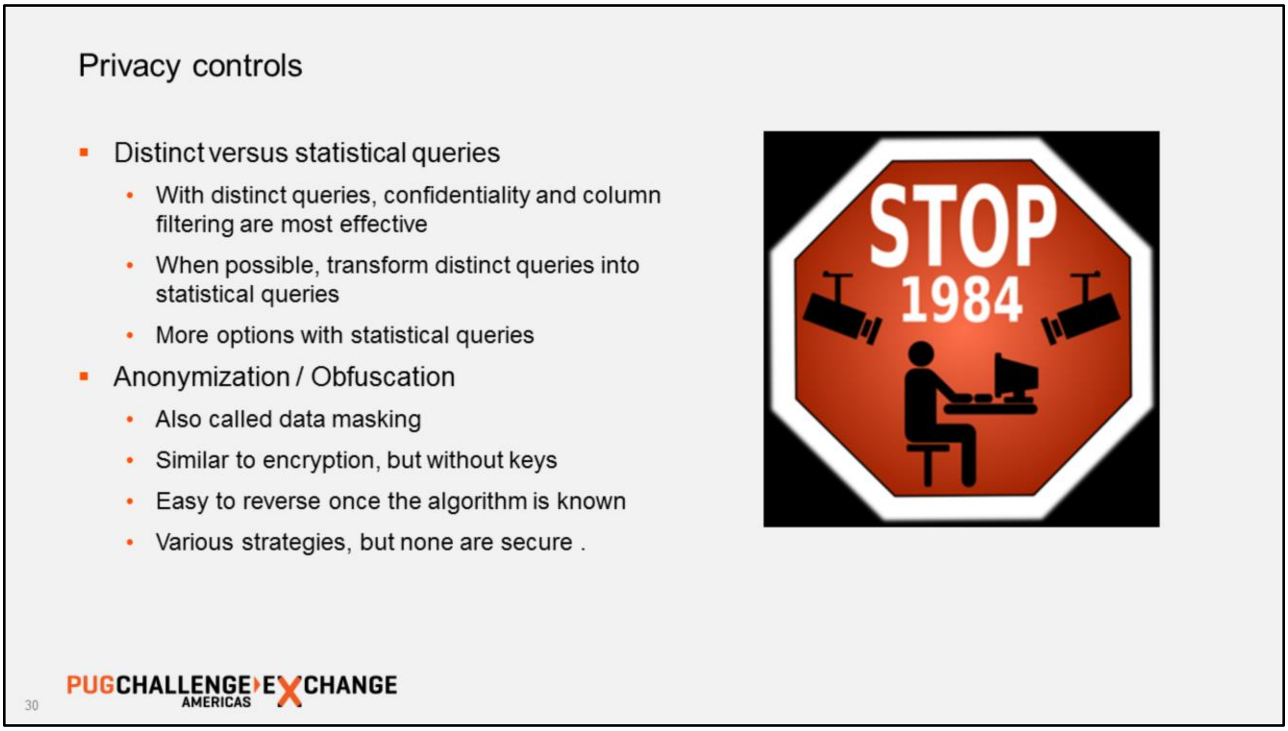
Security policy

- Should address confidentiality and privacy separately
- Inform users how information is
  - Collected
  - Managed
  - Protected
- Some legislation requires a consumer opt-out provision
- SANS Privacy Policy
  - http://www.sans.org/privacy/
- Many online resources to assist
  - Search for “creating a privacy policy”
- Purpose is to document how your organization handles data and protects individuals’ privacy .

http://www.wikihow.com/Create-a-Website-Privacy-Policy

Locate and identify private information

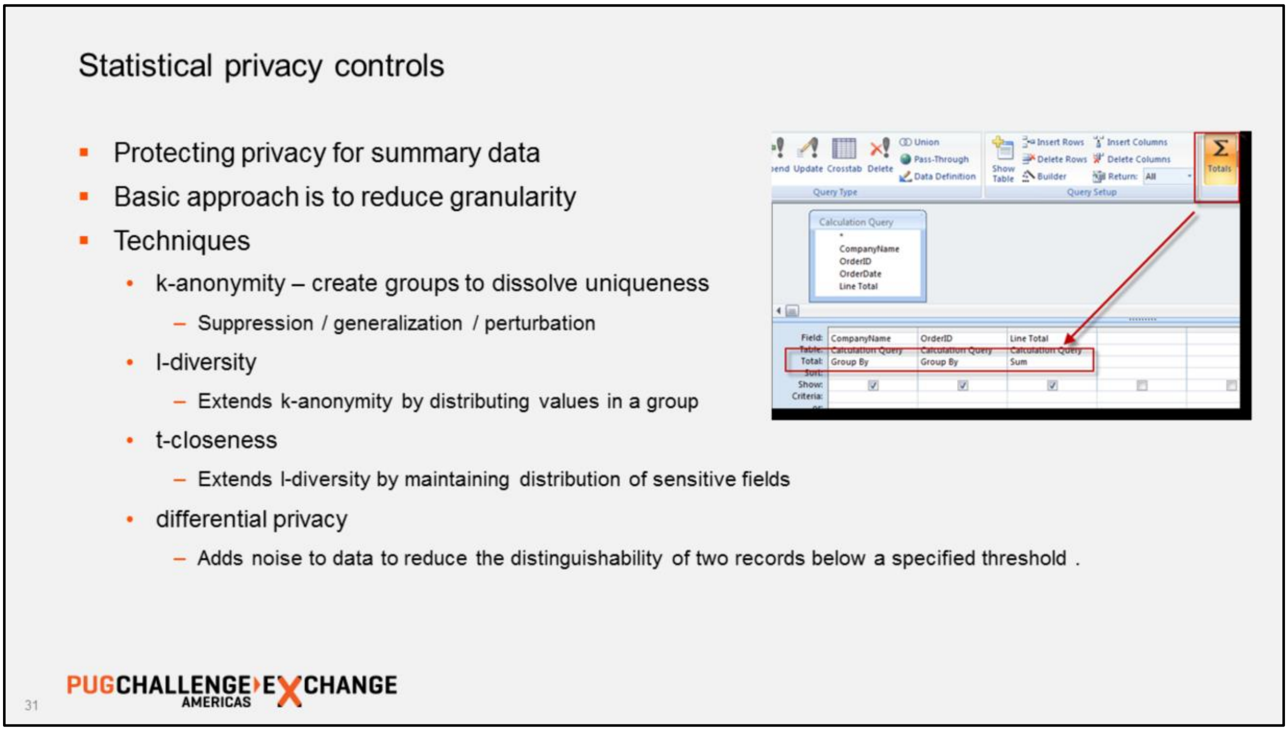
- Direct and indirect
  - PII, PHI, any identifying information
  - Remember Sweeney's findings
- Database
- Central file storage (shares)
- E-mail servers
- Personal devices and removable media
  - Increasingly difficult to enumerate and manage
- Backup media
- End-of-life media .



Privacy controls

- Distinct versus statistical queries
  - With distinct queries, confidentiality and column filtering are most effective
  - When possible, transform distinct queries into statistical queries
  - More options with statistical queries
- Anonymization / Obfuscation
  - Also called data masking
  - Similar to encryption, but without keys
  - Easy to reverse once the algorithm is known
  - Various strategies, but none are secure .

http://www.differencebetween.info/difference-between-obfuscation-and-encryption



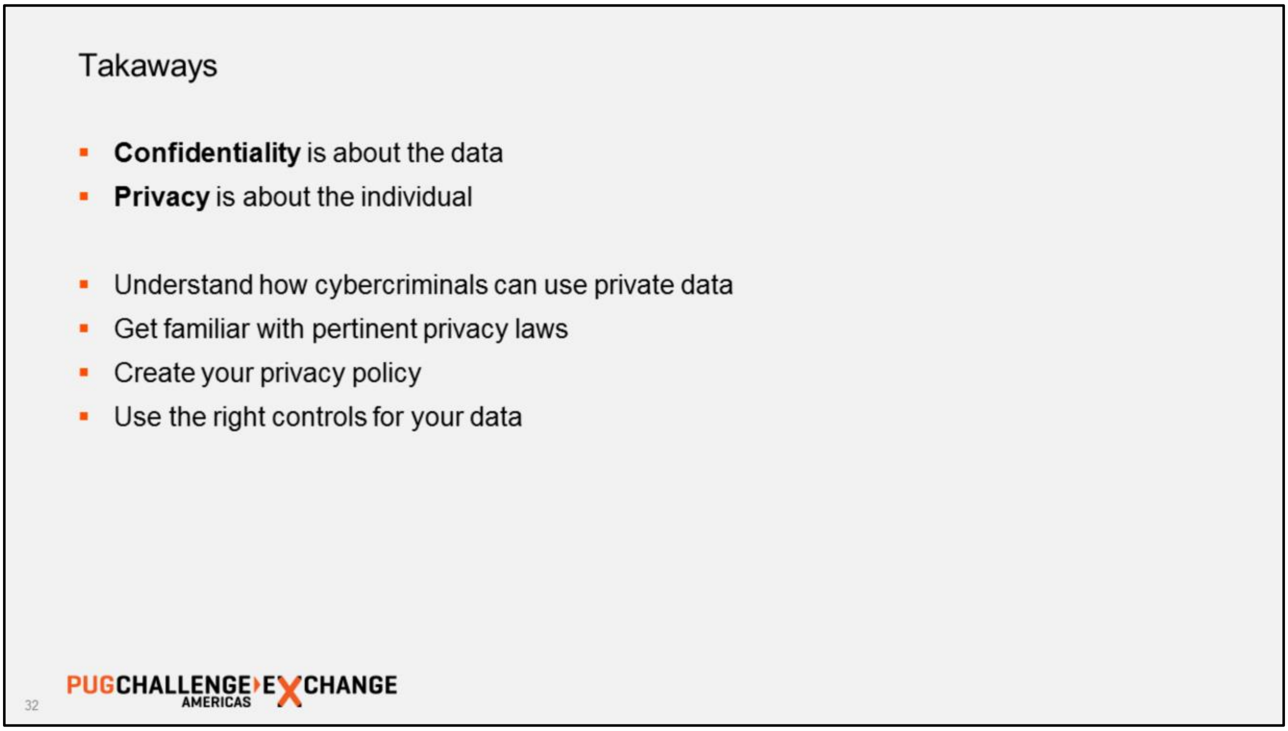
Statistical privacy controls

- Protecting privacy for summary data
- Basic approach is to reduce granularity
- Techniques
  - k-anonymity – create groups to dissolve uniqueness
    - Suppression / generalization / perturbation
  - l-diversity
    - Extends k-anonymity by distributing values in a group
  - t-closeness
    - Extends l-diversity by maintaining distribution of sensitive fields
  - differential privacy
    - Adds noise to data to reduce the distinguishability of two records below a specified threshold .

Field:	CompanyName	OrderID	Line Total		
Table:	Calculation Query	Calculation Query	Calculation Query		
Total:	Group By	Group By	Sum		
Sort:					
Show:	☑	☑	☑	□	□
Criteria:					
or:					

PUGCHALLENGE AMERICAS EXCHANGE

http://en.wikipedia.org/wiki/K-anonymity
http://en.wikipedia.org/wiki/L-diversity
http://en.wikipedia.org/wiki/T-closeness
http://en.wikipedia.org/wiki/Differential_privacy



Takaways

- **Confidentiality** is about the data
- **Privacy** is about the individual

- Understand how cybercriminals can use private data
- Get familiar with pertinent privacy laws
- Create your privacy policy
- Use the right controls for your data

time for questions

THANK
YOU!